eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Konputazio zientziak eta Adimen Artifiziala Saila
Informatika Fakultatea

Estrategias y tecnologías para la colaboración segura entre personas y robots en entornos industriales

Autor:

Iñaki Maurtua Ormaetxea

Directores:

Dr. Basilio Sierra Araujo

Dr. Aitor Ibarguren

Durango, Junio 2017

Iñaki Maurtua

Tesis doctoral: Estrategias y tecnologías para la colaboración segura entre personas y robots en entornos industriales

Agradecimientos

Este trabajo no hubiera sido posible sin el apoyo y colaboración de muchas personas.

En primer lugar agradecer a toda la organización de IK4-TEKNIKER por promover y facilitar que los investigadores llevemos a cabo los programas de doctorado y permitir que los proyectos de generación de conocimiento den soporte a las diferentes tesis propuestas.

Dentro de la organización, una mención especial a todos los miembros de la unidad de Sistemas Autónomos e Inteligentes que con su labor diaria y apoyo en los proyectos han permitido que las contribuciones que se recogen en este trabajo se hayan materializado exitosamente.

Cómo no, agradecer a mis dos directores Basilio Sierra y Aitor Ibarguren su paciencia, comentarios y apoyo técnico en la redacción de este trabajo y a lo largo de estos años de colaboración en proyectos.

Por supuesto, eskerrik asko Jone, Eneko y Antxon que habéis soportado mi ausencia virtual durante este último período (mi cuerpo estaba ahí, pero mi mente probablemente no), dándome ánimos y apoyo en todo momento.

Y también a los ausentes, en especial a ti Ama, que haya al final del siglo pasado ya me animabas a hacer el doctorado, sin saber exactamente lo que significaba pero siguiendo con ese empeño que nos trasladaste a Carlos y a mí para aprender, aprender y aprender. Seguro que te alegrarías al saber que por fin lo presento y tú, Aita, discretamente, te sentirías orgulloso del hijo del carpintero.

Eskerrik asko denoi.

Resumen

La imagen de los robots encerrados entre barreras físicas para proteger a las personas de posibles riesgos está comenzando a cambiar a medida que la legislación establece el marco de diseño de entornos colaborativos y el avance de las tecnologías permite implementar esos diseños.

El cambio de paradigma en el campo de la fabricación, caracterizado por una tendencia a lotes de fabricación más cortos y la personalización masiva, hace inviable en la mayoría de los casos la automatización de los procesos productivos. Es en ese ámbito donde la robótica colaborativa puede jugar un papel relevante en la productividad de nuestras empresas, en particular de las pequeñas y medianas empresas.

Las aplicaciones de robótica colaborativa implican cambios importantes en la forma en la que las personas y los robots interaccionamos. El robot deja de ser una máquina que repite con precisión una tarea programada para convertirse en el ayudante que facilita el trabajo del trabajador, con el que interacciona de forma natural, que es capaz de adaptarse a las condiciones cambiantes del entorno y lleva a cabo sus tareas de forma segura para las personas.

La presente memoria presenta una revisión sobre la actividad de investigación aplicada que se ha realizado en el contexto de diferentes proyectos de investigación en las tres áreas mencionadas.

En el campo de la interacción se presenta la aplicación de tecnologías semánticas que permiten combinar de forma eficiente y robusta diferentes canales de comunicación, así como el desarrollo de sistemas que permiten identificar diferentes categorías de gestos, en particular el gesto de apuntar.

Se presenta la labor de diseño de estrategias y arquitecturas de seguridad desarrolladas en varias proyectos y dos aproximaciones diferentes para la implementación del método de colaboración 'Speed and separation monitoring' (SSM). Se presentan igualmente diferentes experimentos llevados a cabo con potenciales usuarios de la robótica colaborativa al objeto de validar los diseños y estudiar la actitud de las personas y el nivel de confianza ante esta nueva forma de entender la robótica.

Por último se describe la contribución en el campo de la percepción, basada en la visión artificial, para conseguir que los robots sean capaces de caracterizar y adaptarse al entorno.

Laburpena

Inguruko pertsonen segurtasuna bermatzea ahalbidetzeko barrera artean hertsirik dauden roboten irudia aldatzen hasi da legediak ezarritako diseinu marko kolaboratiboak eta diseinu hauek bermatzen dituzten teknologiek aurrera egin duten heinean.

Fabrikazio munduak bizi duen paradigma aldaketak, hau da, pertsonalizazio masiboa eta fabrikazio lote laburragoak ezaugarritzat dituen paradigma aldaketak, ezinezkoa egiten du, kasu gehienetan behintzat, produkzio prozesuen guztizko automatizazioa. Esparru horretan, hain zuzen, joka lezake berebiziko papera robotika kolaboratiboak gure enpresen produktibitatean, batez ere enpresa ertain eta txikietan.

Robotika kolaboratiboaren aplikazioek pertsonen eta roboten artean ematen den interakzioan aldaketak suposatzen ditu. Robotak aurrez programatutako eginkizuna zehaztasunez errepikatzen duen makina izateari utzi eta langileari lana errazten dion laguntzailea bihurtzen da, interakzioa era naturalean gauzatzeko gai dena, inguruko baldintza aldagarrietara egokitzeko gai dena eta bere eginbeharrak inguruko pertsonekiko segurtasuna bermatuz betetzen dituena.

Ondoren aurkeztuko den memoriak aurrez aipatu diren hiru esparruen inguruan aurkeztu eta egin diren proiektu eta ikerketen berri ematea du helburu.

Interakzioaren arloan teknologia semantikoen aplikazioak aurkezten dira, zeinak komunikazio kanal ezberdinen arteko konbinazio eraginkor eta sendoak bermatzen dituzten. Era berean, keinu mota diferenteak identifikatzea ahalbidetzen dituzten sistemen garapena azalduko da, bereziki seinalatzeko erabiltzen den keinuaren ingurukoak.

Proiektu desberdinetan aurkeztutako segurtasunaren inguruko estrategia eta arkitekturen diseinu lana aurkeztuko da, baita '*Speed and separation monitoring*' (SSM) kolaborazio metodoaren inplementaziorako bi hurbilketa desberdin ere. Bestalde, robotika ulertzeko forma berri honen inguruan pertsonek daukaten jarrera eta konfiantza maila zein den eta diseinu hauek balioztatzeko helburuarekin, robotika kolaboratiboaren erabiltzaile potentzialekin egindako esperimentazio lanak ere aurkeztuko dira.

Bukatzeko, pertzepzioaren esparruan egindako ekarpena azalduko da, ikusmen artifizialean oinarritutakoa, robotak ingurua identifikatzeko eta egoera horretara egokitzeko gai izan daitezen.

Índice

Índice de Figuras

Glosario

HRC	Human-Robot Collaboration
AMI	Ambient Intelligence
UCD	User Centered Design
COBOT	Robot colaborativo
SSM	Speed and Separation Monitoring
ROS	Robot Operating System
HCI	Human Computer Interaction
ROI	Región de Interés

Sección I Descripción del trabajo

1. INTRODUCCIÓN

Comienzo la redacción de este documento la semana en la que me acaban de notificar la aceptación definitiva de los dos artículos que completan el conjunto de las publicaciones que se incluyen en este trabajo, '*Natural Multimodal Communication for Human-Robot Collaboration*' y '*Human Robot collaboration in Industrial applications: safety, interaction and trust*'. Los dos artículos recogen las tres líneas de trabajo fundamentales de mi actividad investigadora: la interacción persona-máquina, la seguridad y la robótica colaborativa como marco en el que se integran las anteriores.

Esta larga carrera de 30 años, desarrollada íntegramente como investigador en IK4-TEKNIKER, comenzó con el desarrollo de varias células de fabricación flexible (FMS) en las que aparecieron los primeros robots en mi vida y son el punto de partida de la evolución que está viviendo la robótica:

- Los robots eran elementos 'flexiblemente inflexibles' que por un lado permitían realizar la carga y descarga de piezas diferentes en máquina (*flexible*) pero necesitaban que el entorno fuera estructurado, con una mínima incertidumbre en el posicionamiento, tanto de las piezas como de las máquinas o utillajes con los que interacciona (*inflexible*).
- Los robots eran elementos seguros, sí, pero a costa de encerrarlos en jaulas a las que los operarios sólo podían acceder a través de puertas cuya apertura era detectada por los sensores de enclavamiento correspondientes que hacían caer los motores de los robots.
- La interacción con el robot se limitaba a la utilización de un 'teach-pendant' durante la fase de programación (grabación de puntos y ajuste de programas), inicialización y mantenimiento del sistema. La interacción suponía navegar a través de los menús del dispositivo, usando el teclado o, en el mejor de los casos, el 'joystick' incorporado al dispositivo.

Al final de la década de los noventa, apareció un concepto que iba a cambiar el rumbo de mi trabajo de forma definitiva: la 'Inteligencia Ambiental' (AMI), un entorno capaz de reconocer y responder a la presencia de las personas de una forma constante, no intrusiva y a menudo invisible. Ese concepto se recogía en el documento [1] en el que se identificaban algunas de las características de AMI: (1) facilitar el contacto con las personas, (2) ayudar a mejorar las condiciones de trabajo, (3) inspirar confianza y (4) fácil de usar y controlable por personas 'ordinarias'.

IK4-TEKNIIKER realizó un esfuerzo en el desarrollo de las tecnologías que posibilitaban esa visión de futuro en el campo de la interacción y la facilidad de uso, enfocándolo fundamentalmente al entorno de trabajo. Tras varios proyectos de investigación básica llevados a cabo a nivel nacional (AMIn, AMIGUNE, CONTEXTO) y fundamentalmente el proyecto wearIT@work, en los que desarrollamos conceptos interesantes de interacción basados en gestos, aplicación de la metodología de 'User Centered Design' (UCD), tecnología de agentes y computación llevable, decidí aplicar dichos conceptos en el campo de la robótica para contribuir a crear entornos robóticos que permitieran una interacción natural y segura con las personas y abordar aplicaciones en entornos no estructurados en los que realmente se explotara la flexibilidad de los robots.

El esfuerzo de estos años se ha materializado en una exitosa participación en proyectos europeos en el campo de la robótica, unas veces como colaboradores y aportando nuestro conocimiento (X-ACT [2], ROBOPARTNER [3], SMERobotics, ECHORD [4], ECHORD++ [5], AUTOWARE [6], MANUWORK [7]) y EUROC [8]), otras liderando, además, los consorcios internacionales (ROBOFOOT [9], MAINBOT [10], FOURBYTHREE [11] y CRO-INSPECT [12]).

Los objetivos de esta actividad investigadora han ido encaminados a crear entornos robóticos inteligentes caracterizados por:

Objetivo 1: Interacción
Permitir una interacción natural entre la persona y el robot.

La fusión mediante tecnologías semánticas de la información obtenida a través de diferentes canales de interacción, voz, texto y gestos, permite que las personas interaccionemos con los robots de forma natural y robusta en entornos que presentan condiciones de contorno adversas (ruido, iluminación no controlada).

La **aportación** fundamental ha sido:

- Utilización de tecnología semántica para la interacción natural sensible al contexto.
- Reconocimiento de gestos mediante la utilización de sensores llevables y sistemas de visión.

Esas aportaciones han sido validadas en diferentes entornos: interpretación del lenguaje de signos para personas sordas, monitorización postural como elemento de la comunicación no verbal y nuevas formas de interacción con los medios productivos en entornos de fabricación.

Objetivo 2: Seguridad
Diseñar e implementar estrategias de seguridad que garanticen una colaboración segura entre la persona y el robot.

La eliminación de las barreras físicas utilizadas tradicionalmente como salvaguarda de la seguridad en la utilización de robots es uno de los objetivos fundamentales para conseguir un entorno colaborativo.

Para ello es necesario diseñar una estrategia que permita cumplir con los requisitos definidos en las directivas y normas de seguridad vigentes, desarrollar las tecnologías necesarias para que la implementación de dichas estrategias no suponga una merma en la eficiencia de los robots y conseguir un nivel de confianza adecuado en las personas que deben trabajar junto a los robots sin barreras.

Las **aportaciones** fundamentales han sido:

- Diseño de estrategias de control que permiten la colaboración entre personas y robots en ausencia de barreras físicas de protección.
- Desarrollo de la tecnología necesaria para la implementación eficiente del modo de colaboración 'Speed and separation monitoring' definido en las normas ISO-10218-1:2011 [13], ISO-2018-2:2011 [14] y la especificación técnica ISO/TS 15066:2016 [15]. Esos desarrollos han culminado con la solicitud de la patente 'METODO, SISTEMA Y PROGRAMA INFORMATICO DE DETECCION DE PROXIMIDAD' en la Oficina Española de Patentes y

Marcas.

Los resultados obtenidos han sido validados mediante diferentes experimentos con trabajadores y usuarios ocasionales de robots en ferias y entornos de fabricación.

Objetivo 3: Flexibilidad
Posibilitar que los robots puedan adaptar su comportamiento a las condiciones cambiantes y entornos no completamente estructurados.

La tecnología posibilitadora fundamental es la aplicación de la visión para caracterizar dicho entorno y la generación y control de trayectorias de forma dinámica, 'visual servoing'.

Las **aportaciones** fundamentales han sido:

- Diseño de una estrategia de control que facilite la integración de diferentes tecnologías y de respuesta a diseños de robot modulares.
- Desarrollo de algoritmos de percepción y control para la identificación de la pose de objetos.

Los resultados han sido implementados y validados en diferentes proyectos posibilitando el desarrollo de aplicaciones robóticas en entornos no estructurados.

Esta memoria se organiza en dos Secciones fundamentales: una primera que comprende los capítulos 2, 3 y 4, en los que se proporciona una visión general de la problemática abordada, los proyectos en los que se ha abordado dicha problemática y un resumen de las aportaciones más relevantes, finalizando con el capítulo 5 en el que se recogen las conclusiones y propuestas de investigación futura. En la Sección II se recogen las publicaciones y patentes que refrendan dichas contribuciones.

1.1. IK4- TEKNIKER

IK4-TEKNIKER es un centro tecnológico, constituido jurídicamente como Fundación sin ánimo de lucro, que pertenece a la Alianza Tecnológica IK4. Fundado en el año 1981, cuenta en la actualidad con más de 270 personas en plantilla y sus ingresos totales rondan los 24 M€.

Tecnológicamente se define como especializado en "Manufacturing" y su actividad se puede enmarcar en 4 grandes **Líneas de generación de conocimiento**: fabricación avanzada, ingeniería de producto, ingeniería de superficies y tecnologías de la información y las comunicaciones.

Este conjunto de actividades se orientan a proporcionar **soluciones tecnológicas** a las empresas. Dichas soluciones se engloban en 7 grandes paquetes: sistemas mecatrónicos, mantenimiento industrial, automatización y robótica industrial, superficies multifuncionales, medición e inspección, dispositivos sensores e innovación e inteligencia competitiva.

Para terminar, y dado el carácter horizontal y tecnológico del planteamiento de su oferta, IK4-TEKNIKER sirve a un amplio grupo de **sectores** de actividad: máquina-herramienta y fabricación, energías renovables, aeronáutica y espacio, industria de la ciencia, biomedicina, automoción, infraestructuras y e-salud y tecnología social.

IK4-Tekniker tiene un estrecho vínculo con las asociaciones industriales locales y es miembro de varias plataformas como EFFRA, euRobotics o HISPAROB, que cuentan entre sus objetivos la promoción de la robótica y la excelencia en la fabricación.

1.2. Unidad de Sistemas Autónomos e Inteligente

El objetivo de la Unidad de Sistemas Autónomos e Inteligentes (SAI) es el desarrollo de sistemas con una elevada inteligencia y autonomía en diferentes ámbitos. Para lograrlo investiga en las siguientes líneas:

- Robótica. En el marco de los proyectos que se describen en el capítulo 2 se desarrollan tecnologías para dotar de inteligencia a los robots y mejorar su navegación, generación de trayectorias y la capacidad de manipulación.
- Interacción persona-máquina. Incorporando el reconocimiento de gestos (basado en sensores inerciales y visión) y la utilización de tecnologías semánticas fundamentalmente en aplicaciones robóticas.
- Visión artificial. Dando soporte a las líneas anteriores y para aplicaciones de control de calidad.

Desde la Unidad se lidera la Línea de Especialización en Robótica de IK4-TEKNIKER, definiendo y coordinando las actividades de otros grupos de investigación en este campo.

2. DESCRIPCIÓN GENERAL DE LOS PROYECTOS DE INVESTIGACIÓN

La actividad investigadora se ha desarrollado en el contexto de proyectos de investigación europeos que se resumen en este capítulo. A lo largo del documento y en las propias publicaciones que se incluyen en este trabajo se hace referencia a las contribuciones más importantes ligadas a cada uno de los proyectos.

2.1. Proyectos Europeos coordinados

Se presentan en esta sección aquellos proyectos en los que el autor de este trabajo desarrolla o ha desarrollado la labor de coordinador.

2.1.1. ROBOFOOT

ROBOFOOT (FP7-2010-NMP-ICT-FoF): Smart robotics for high added value footwear industry [9]

El objetivo del proyecto fue la introducción de la robótica en el sector del calzado, caracterizado por un uso extensivo de mano de obra. Se abordaron diferentes operaciones que exigían la aplicación de tecnologías de control de fuerza y visión.

El trabajo desarrollado por TEKNIKER se centró en la aplicación de la visión artificial para la identificación de la posición del calzado en la manovía y como fuente de información para la identificación de defectos superficiales originados en distintas fases del proceso productivo.

En la *Figura 1* se muestra uno de los procesos implementados, consistente en el empaquetado de los zapatos para el que se desarrollaron algoritmos de identificación de la posición del zapato y estrategias de manipulación, así como el desarrollo de una pinza multifunción

Figura 1: Robot manipulando zapatos en la tarea de empaquetado

EL objetivo industrial fue contribuir a que la producción del calzado vuelva a Europa y el proyecto fue nominado por la Comisión Europea como uno de los mejores proyectos en robótica[1].

2.1.2. MAINBOT

Mainbot (FP7-NMP): Mobile robots for inspection and maintenance activities in extensive industrial plants [10].

En el proyecto Mainbot, se desarrollaron un conjunto de tecnologías para la implementación de robots móviles y autónomos capaces de llevar a cabo tareas de inspección en plantas extensas. El proyecto se centró en tres aspectos fundamentales:

- Autonomía. Desarrollando las capacidades ligadas a la navegación autónoma del robot en un entorno no estructurado o semi-estructurado.
- Percepción para inspección y navegación. El objetivo fue dotar a las plataformas de las capacidades de percepción necesarias para la medición, detección y reconocimiento de fallos en la instalación, integrando la información térmica y visión con las capacidades de manipulación y movilidad de la plataforma.
- Desarrollo mecatrónico de plataformas capaces de moverse en grandes espacios o de trepar estructuras verticales como las torres de concentración solar.

Se utilizaron dos tipos de plataformas móviles: una para superficie (la mostrada en la *Figura 2*) y una segunda para movimientos en vertical.

[1] http://europa.eu/rapid/press-release_MEMO-13-11047_en.htm

Figura 2: Plataforma manipuladora móvil utilizada en el proyecto MAINBOT

2.1.3. FOURBYTHREE

FOURBYTHREE (H2020-FoF-2014): Highly customizable robotic solutions for effective and safe human robot collaboration in manufacturing applications [11].

Proyecto para el desarrollo de robots colaborativos mediante un concepto modular que permita a un usuario final o integrador disponer de todos los componentes hardware y software necesarios para el desarrollo de aquella configuración del robot que mejor se adecúe a sus necesidades. En la *Figura 3* se muestra un ejemplo de robot fabricado con los componentes desarrollados en el proyecto.

Figura 3: Uno de los prototipos de robot desarrollado a partir de los módulos desarrollados en FourByThree

IK4-TEKNIKER, además de coordinar el proyecto, se centra en el diseño y desarrollo de la arquitectura de control, herramientas de simulación y programación y el desarrollo de los componentes de control basado en visión.

2.2. Otros Proyectos Europeos

2.2.1. X-ACT

X-ACT (FP7-2012-NMP-ICT-FoF): Expert cooperative robots for highly skilled operations for the factory of the future [2].

El proyecto parte de la utilización de robots bimanipuladores (ver *Figura 4*) para el desarrollo de aplicaciones colaborativas.

IK4-TEKNIKER se ha centrado en el desarrollo de la arquitectura de seguridad y las tecnologías necesarias para el escenario de desensamblado de pequeño electrodoméstico.

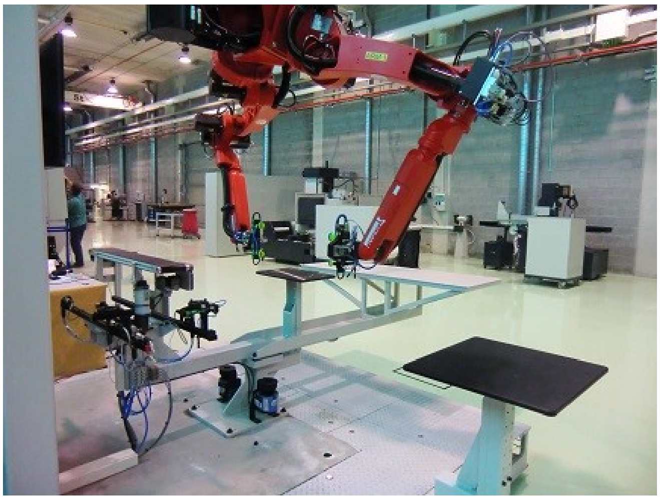


Figura 4: Layout utilizado en los proyectos X-ACT y SMERobotics.

2.2.2. SMERobotics

SMERobotics (FP7-ICT-2011-7): The European Robotics Initiative for Strengthening the Competitiveness of SMEs in Manufacturing by integrating aspects of cognitive systems [16].

IK4-TEKNIKER desarrolló el demostrador FLEXAS: Aeronautic components assembly using flexible dual-arm robotic in close collaboration with human operators.

En este demostrador se siguió avanzando en las desarrollos del proyecto X-ACT, pero centrados en la aplicación de ensamblado de componentes aeronáuticos.

2.2.3. ECHORD

ECHORD (FP7-ICT-2007-3): European Clearing House for Open Robotics Development [4].

IK4-TEKNIKER tomó parte en el experimento EASYPRO: Accurate Manual Guided Robot Programming.

En el experimento se desarrolló un sistema de programación por demostración utilizando un guiado manual y tecnología de visión artificial para caracterizar la trayectoria a seguir por el robot.

2.2.4. ECHORD++

ECHORD ++ (FP7-ICT-2011-9): European Clearing House for Open Robotics Development Plus Plus [5].

IK4-TEKNIKER tomó parte en el experimento DEBUR: Automated robotic system for laser deburring of complex 3D shape parts.

En el experimento se desarrolló una solución robótica para el rebabado de piezas de inyección de aluminio mediante un robot y tecnología láser. La caracterización de la rebaba a eliminar y la identificación de la pieza en el espacio fueron dos de los retos tecnológicos.

2.2.5. ROBO-PARTNER

Robo-Partner (FP7-2013-NMP-ICT-FOF(RTD)): Seamless Human-Robot Cooperation for Intelligent, Flexible and Safe Operations in the Assembly Factories of the Future [3].

Robo-Partner fue un proyecto cuyo objetivo principal fue la generación de las tecnologías necesarias para la integración de robots industriales en entornos de trabajo colaborativos, para que puedan realizar tareas complejas con un elevado grado de autonomía y capacidad de colaboración.

Las tecnologías clave que se desarrollaron en este proyecto fueron:

- La percepción de los robots. Utilizando sensores y sistemas sensoriales capaces de identificar objetos, personas y reconocer el entorno complejo.
- La interacción persona-robot desarrollando tecnologías que permiten interactuar de forma natural incluyendo la voz, gestos, sistemas hápticos y realidad aumentada.
- La seguridad, un aspecto fundamental a la hora de garantizar la colaboración persona robot en un entorno sin barreras, cumpliendo con los requisitos marcados por la normativa.

Como resultado del proyecto se han desarrollado varios prototipos con aplicaciones en diversos ámbitos como la intra-logística (ver la ***Figura 5***), el montaje de electrodomésticos y el montaje en el ámbito de la automoción.

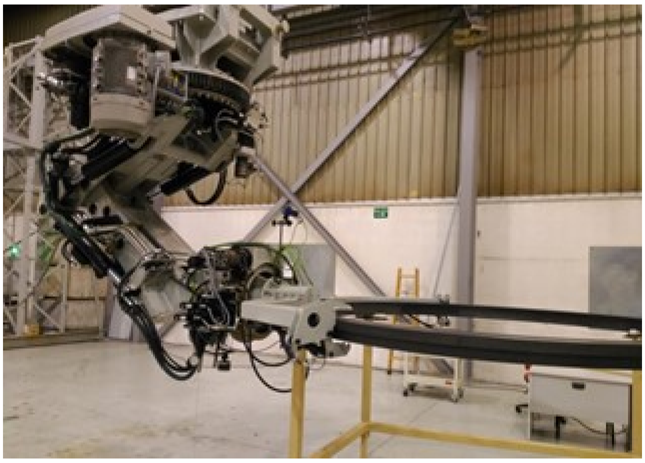


Figura 5: Robot para el manipulado de piezas de grandes dimensiiones

2.2.6. EUROC-1

EUROC (FP7-2013-NMP-ICT-FOF): European Robotics Challenges

IK4-TEKNIKER está participando en el Challenge-1 con el proyecto PIROS: Cooperative, safe and reconfigurable robotic companion for CNC pallets load/unload stations [17].

La labor de IK4-TEKNIKER en este proyecto se centra en la utilización de gestos para interaccionar con el robot y la identificación de la posición de piezas dispuestas aleatoriamente en un contenedor (`bin picking') tal y como se muestra en la ***Figura 6***, así como la generación de las trayectorias libres de colisiones del robot.

Figura 6: Piezas a identificar y extraer del contenedor por el robot

2.2.7. EUROC-2

EUROC (FP7-2013-NMP-ICT-FOF): European Robotics Challenges

IK4-TEKNIKER está participando en el Challenge-2 con el proyecto FLECOOP: Flexible robotized unitary picking in collaborative environments for order preparation in Distribution Centers [18].

El proyecto FLECOOP pretende mejorar un aspecto crítico de la cadena de suministro de los Centros de Distribución, la fase final de preparación de pedidos.

La idea principal consiste en ofrecer un enfoque híbrido, en el que los robots y las personas comparten el mismo espacio de trabajo, combinando una alta automatización con la flexibilidad humana y la seguridad necesaria. En la *Figura 7* se muestra parte del prototipo desarrollado.

El proyecto aborda dos problemas clásicos, el "bin picking" y el "bin packing". En el primero, objetos de diferente tamaño y forma deben ser identificados y cogidos por un robot (manipulador móvil). En el segundo, los objetos deben ser empaquetados en un número finito de contenedores. En este contexto se abordan los siguientes retos técnicos e industriales:

- Segmentación en tiempo real de objetos en contenedores utilizando visión 2D-3D.
- Monitorización del entorno de trabajo para garantizar un comportamiento seguro del robot.
- Implementación de un flujo de trabajo colaborativo.
- Manipulación móvil para operaciones de "pick" y "place".

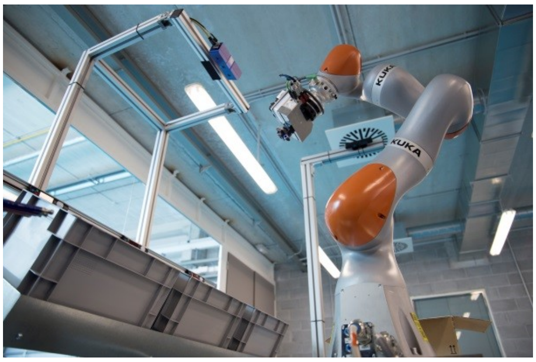


Figura 7: Robot colaborativo en la aplicación de logística propuesta en FLECOOP

2.2.8. CRO-INSPECT

CRO-INSPECT (H2020-CS2-CFP02-2015-01): Collaborative Robotic Solutions for Advanced Inspection of complex Composite Parts [12].

El proyecto CRO-INSPECT tiene como objetivo proporcionar una solución de inspección para piezas composites de aeronáutica, utilizando robots colaborativos que asistan a los operarios de forma eficiente y flexible. La solución se basa en tecnologías robóticas y de inspección mediante ultrasonidos, proporcionando un procedimiento de inspección sistemático y de alta trazabilidad.

Los objetivos técnicos son:

- La introducción de técnicas de inspección con control adaptativo mediante el uso de robots.
- La implementación de funciones robóticas asistivas como el guiado manual, compensación de la carga del sistema de inspección, posicionamiento preciso en puntos de interés.
- La implementación de mecanismos de seguridad de acuerdo a la normativa ISO 10218.
- Desarrollo de nuevas tecnologías de ultrasonidos, como las ondas guiadas.

2.2.9. AUTOWARE

AUTOWARE (H2020-FOF-2016): Wireless autonomous, reliable and resilient production operation architecture for cognitive manufacturing [6]

IK4-TEKNIKER implementa el escenario donde se validan las tecnologías para el desarrollo de los sistemas ciberfísicos. IK4-TEKNIKER dispone de un escenario en el taller donde las tecnologías relacionadas con Industria 4.0 se pueden validar. Concretamente se trata de una celda con un bi-manipulador robótico para montaje colaborativo que se integra con una plataforma móvil que realiza actividades de logística de apoyo.

Este escenario permitirá aproximar las tecnologías relacionadas con los sistemas ciber-físicos (comunicaciones, computación en la nube, etc.) y la robótica a la industria, mostrando el potencial de estas tecnologías en la mejora de los procesos. El escenario de IK4-TEKNIKER corresponde a la robótica flexible, colaborativa y autónoma.

2.2.10. A4BLUE

A4BLUE (H2020-FOF-2016): Adaptive Automation in Assembly For BLUE collar workers satisfaction in Evolvable context [19].

A4BLUE propone el desarrollo y evaluación de una nueva generación de espacios de trabajo sostenibles y adaptables que permitan responder a los requisitos de procesos de fabricación donde se introducen mecanismos de automatización. Se trata de ofrecer herramientas que permitan adaptar los sistemas automatizados a los trabajadores mediante interfaces personalizados y funciones de asistencia basadas en la información contextual.

2.2.11. MANUWORK

MANUWORK (H2020-FOF-2016): Balancing Human and Automation Levels for the Manufacturing Workplaces of the Future [20]

Bajo el mismo programa de investigación que A4BLUE, en MANUWORK IK4-TEKNIKER es responsable de definir las guías de diseño de entornos robóticos colaborativos y desarrollar el escenario en el que un robot colaborativo ayude a desarrollar su actividad laboral y aumentar el grado de satisfacción en el puesto de trabajo a personas con discapacidad intelectual.

2.3. Otros proyectos

En el pasado se ha contribuido en otros proyectos tanto a nivel nacional como internacional en la línea de investigación de la Inteligencia Ambiental (AmI).

El más relevante fue wearIT@work (FP6-2003-IST-2): Empowering the mobile worker by wearable computing [21]. En este proyecto, en el que participaban 36 empresas, universidades y centros de investigación, IK4-TEKNIKER fue responsable de la definición e implementación de la metodología *'User center Design'* en el proyecto. Además desarrollamos el escenario de Producción junto a la empresa Skoda, en una de cuyas plantas se validó una solución basada en tecnología llevable para reducir el tiempo de acceso de los trabajadores a un nuevo puesto en la línea y facilitar las labores de inspección al final de línea. En la ***Figura 8*** se muestran algunas de las tecnologías testeadas, tales como las gafas de realidad aumentada, reconocimiento de voz o sensores para captar señales mioeléctricas.

Figura 8: Testeando tecnología llevable para reducir el tiempo de aprendizaje de una tarea

En los proyectos AMIn, AMIGUNE, AMIROB y CONTEXTO, desarrollados a nivel local, IK4-TEKNIKER exploró el concepto de Inteligencia Ambiental y su aplicación en el entorno de fabricación, centrándose en el desarrollo de agentes software, la caracterización del contexto y facilitar la interacción persona máquina.

Junto con la empresa ADUR, se desarrolló un proyecto de investigación para explorar el uso de trajes sensorizados para identificar y corregir la postura de las personas durante la comunicación no verbal.

Finalmente en ROBAUCO (CONSORCIADO 2007-2009), NOA (KUTXA 2006-2009) y KTBOT (KUTXA 2010-2013) se trabajó en el desarrollo de aplicaciones de robótica móvil, desarrollando tecnología para mejorar la navegación y autonomía de las plataformas y conseguir una interacción natural entre robot y persona.

3. Objeto de la Investigación

Como ya se ha explicado en la introducción, los tres hilos conductores de la investigación han sido la ***Interacción*** persona-máquina, la ***Percepción*** como elemento necesario para dotar de ***Flexibilidad*** a los robots y la ***Seguridad*** en la colaboración persona robot.

La popularización del término Robótica Colaborativa, con sus distintas interpretaciones, algunas de ellas lejos de la realidad actual y probablemente no realizable a medio plazo, ha dado lugar a numerosos estudios sobre el impacto en la economía y el empleo [22], [23], [24]. La interpretación periodística siembra el temor en la opinión pública con titulares como *'Los robots destruirán puestos de trabajo y bajarán sueldos, según un nuevo estudio'* [25] o surgen polémicas como la necesidad de que los robots paguen impuestos por los trabajos que eliminan [26].

El debate está abierto y no es objeto de este trabajo entrar en esos aspectos, por otro lado apasionantes, sino ahondar en los retos tecnológicos que impactan en el despliegue de los robots en los entornos productivos, en particular la interacción, la flexibilidad y la seguridad.

En el cuadro de la ***Figura 4*** se resumen las características fundamentales de los COBOT o robot colaborativos:

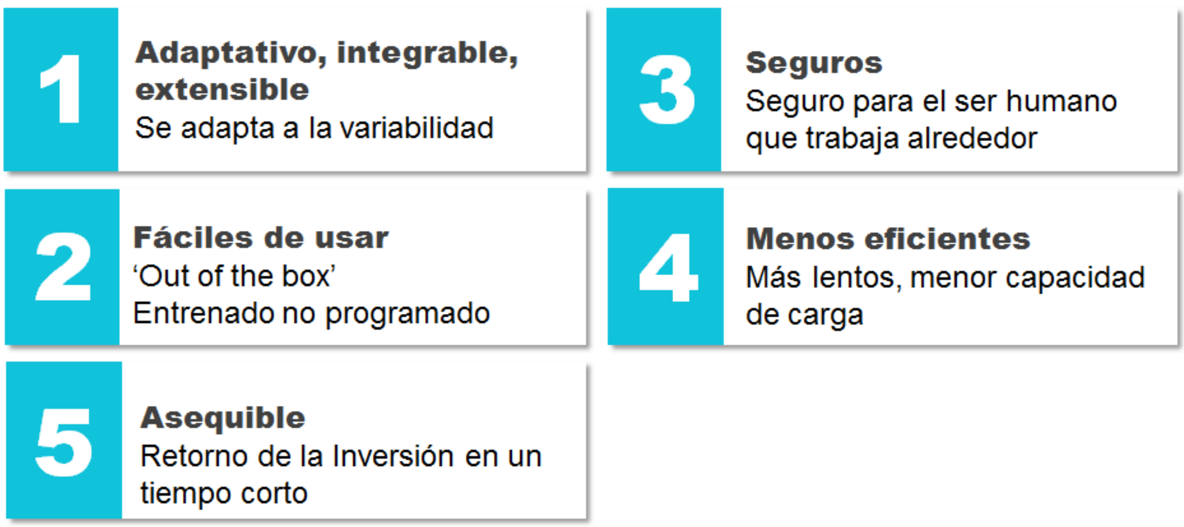


Figura 9:: Características de un robot colaborativo

El nicho de aplicación de los robots colaborativos está definido por las características de los procesos productivos representados en la ***Figura 10***: la cantidad de referencias a gestionar, el volumen a producir de cada una de esas referencias, la necesidad de toma de decisiones para la ejecución de la tarea y el nivel de estructuración del entorno tal y como se representa en la ***Figura 10***, en la que la robótica avanzada engloba tanto la robótica colaborativa como aquella robótica que comporte algunos características singulares como la capacidad de adaptarse al entorno de una forma inteligente (flexibilidad).

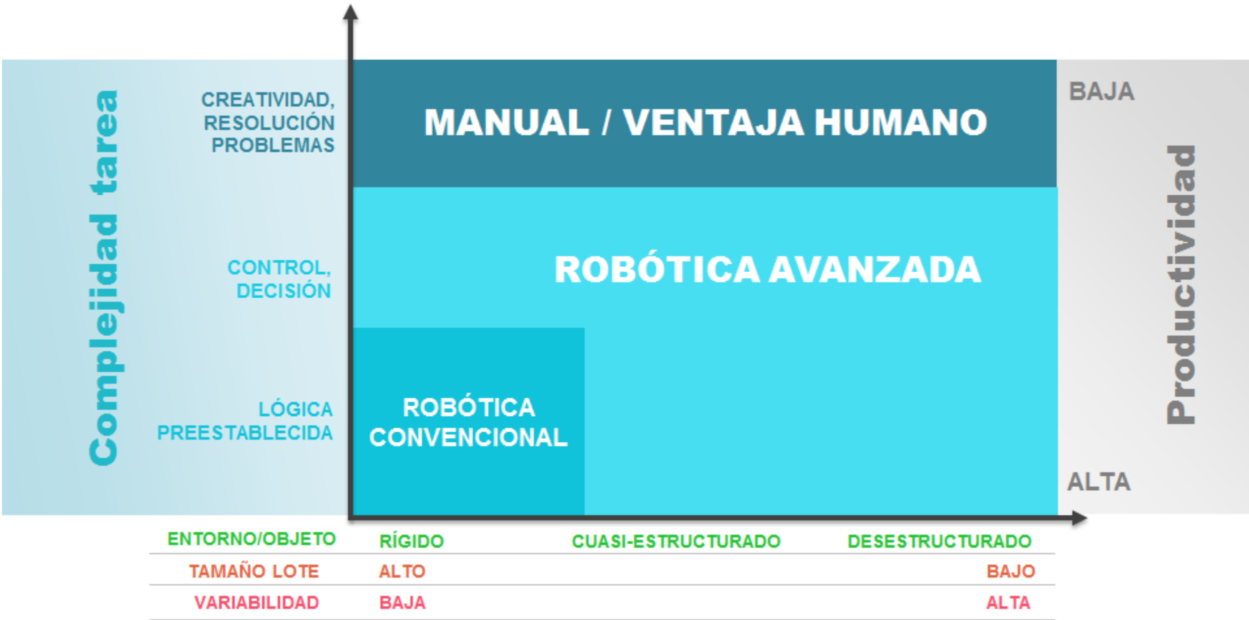


Figura 10: Nicho de aplicación de la robótica avanzada

3.1. Interacción persona-robot

En la actualidad se pueden observar dos líneas de actividad en el campo de la interacción persona computador (HCI) con el objetivo de superar el paradigma WIMP (Windows, Icons, Menus, Pointer) [27]. Por un lado, con los avances en las tecnologías lingüísticas y explotando las capacidades del aprendizaje automático, los interfaces basados en lenguaje natural están alcanzando, gracias a las tecnologías semánticas [28], niveles de madurez suficientes para liberar al usuario de la necesidad de usar mecanismos WIMP y posibilitar la multitarea.

Por otro lado, se observa un interés creciente en la creación de interfaces de usuario físicos (tales como los interfaces 'tangibles' [29], computadores deformables [30], [31], y la 'fisicalización' de la información [32]) que explotan las propiedades físicas del propio interface. De esta forma las interfaces se convierten en objetos en el espacio 3D del usuario, jugando un papel similar a los robots en el campo de la interacción persona-robot, HRC.

Si consideramos la interacción desde el punto de vista de la robótica industrial, tradicionalmente no ha existido un espacio común para la interacción física, ya que ambos actores, robot y persona, estaban separados por motivos de seguridad [33]. La interacción entre ambos se limitaba a la fase de programación en la que el interface seguía el paradigma WIMP. Con los desarrollos en las tecnologías de percepción, es posible implementar entornos en los que las personas y los robots convivan en el mismo espacio físico y la 'socialización' se convierte en un elemento importante tanto para la seguridad como para la productividad.

Uno de los retos en el campo de la interacción persona-robot es dotar a la persona de la percepción del entorno (*'situational awareness'*), es decir, del estado del robot, cómo interpreta éste el contexto colaborativo e información que anticipe las acciones futuras [34]. En una relación de colaboración el robot debe mantener informada a la persona al objeto de reducir su necesidad de mantener un estado de vigilancia ante situaciones inesperadas y de riesgo.

En un entorno de colaboración persona-robot existen diferentes factores que afectan a la experiencia de usuario, *'User experience (UX)'*, tales como la eficiencia y efectividad de la colaboración, la seguridad y la percepción de seguridad, el flujo de la interacción, ergonomía y aspectos estéticos.

Movimientos imprevistos están en la raíz de la mayoría de los accidentes con robots [35], [36]. Incluso en aquellos robots diseñados como intrínsecamente seguros, la experiencia enseña que los usuarios con menor experiencia tienden a una sobre-vigilancia sobre las acciones del robot, lo que supone una pérdida de productividad en el operario por exceso de vigilancia o supervisión frente al objetivo de liberarlo de actividades y permitir un trabajo colaborativo [37]. En la práctica existe un grado de incertidumbre tanto para el robot como para la persona: para la persona existe la incertidumbre de que el robot vaya a ejecutar las acciones previstas, mientras que en el robot existe un grado de incertidumbre relacionado con la ambigüedad de las instrucciones del usuario y la incertidumbre ligada a los propios algoritmos utilizados en la toma de decisiones [38], [39].

Ese efecto negativo sobre la experiencia de usuario debido a la incertidumbre puede verse reducido si el interface mantiene al usuario informado de lo que sucede en el sistema (robot y entorno) y le permite cancelar o modificar las acciones, proporcionando por lo tanto un elevado grado de control al usuario [40].

3.1.1. Reconocimiento de gestos

El reconocimiento de gestos ha sido objeto de interés para muchos investigadores, en el afán de trasladar este medio de interacción natural entre personas al campo de la interacción persona-máquina. Algunos de esos desarrollos comienzan a materializarse en aplicaciones reales como el sistema de reconocimiento de gestos incluido en algunos vehículos de nueva generación.

Existen dos aproximaciones diferentes para la identificación de gestos en función de la tecnología sensorial en la que se basan:

- Utilización de sensores portables, 'vestibles' o llevables. Proporcionan información sobre la posición y movimientos de diferentes partes del cuerpo humano. Su ventaja fundamental es la ausencia de oclusiones que sufren los sistemas basados en visión. Sin embargo exigen que la persona porte los dispositivos que deben ser alimentados eléctricamente y necesitan comunicarse con un dispositivo exterior donde se realiza el procesamiento de la información.

  Los sensores utilizados incluyen acelerómetros y giróscopos como los utilizados en los productos comerciales desarrollados por XSENS [41] o sensores EMG ('*electromiografía*') incluidos en el producto MyO [42].

  [43] presentan un guante sensorizado para capturar los movimientos de la mano aplicado a la teleoperación de robots. Una aproximación en línea con ésta pero aplicada en el reconocimiento de signos es propuesto en [44] y [45].

- Utilización de un sistema de visión. Es la tecnología que mayor interés despierta actualmente, en particular desde la aparición en el mercado de cámaras RGB-D de bajo costo que proporcionan información 3D del entorno.

  Una revisión del uso de imágenes de profundidad para el seguimiento de la mano y el reconocimiento de gestos está disponible en [46] donde se analizan 37 publicaciones en ese campo. Por ejemplo en [47] se presenta un trabajo en el que los autores utilizan una cámara Kinect para la identificación dinámica de los gestos realizados con una mano.

3.1.2. Tecnologías semánticas

La capacidad de entender semánticamente el lenguaje hablado es aún limitado y sigue siendo un reto para la interacción con las máquinas [48]. La importancia de mejorar la interacción natural con los robots queda patente en [49] utilizando un robot de servicio desempeñando la labor de guía en un museo como elemento de validación. Pero también existen ejemplos en entornos industriales, como en [50] que presenta un sistema que combina la mirada, botones virtuales y voz.

La combinación de diferentes canales de comunicación es abordada en varios trabajos. Por ejemplo en [51] se utiliza audio y video (expresiones faciales) para reconocer el estado emocional de una persona; en [52] se utilizan un conjunto de tres sensores inerciales y ocho de fuerza para estimar el modelo cinemático y dinámico de los pasos de una persona al andar. En [53] los autores presentan su propuesta de fusión de diferentes canales de comunicación (voz y reconocimiento de gestos basado en Hidden Markov Models) usando SVM.

3.2. Seguridad

La Seguridad es la condición necesaria para la implementación de un entorno colaborativo. Tradicionalmente los robots se han confinado en espacios protegidos por barreras físicas que impedían un posible contacto con las personas. En el mejor de los casos se utilizan medidas de protección como barreras opto-eléctricas o alfombras sensibles a la presión que permiten eliminar dichas barreras físicas pero no permiten la colaboración con las personas.

Afortunadamente en 2011[2] se publican las normas ISO 10218-1:2011 [13] e ISO 10218-2:2011 [14] que establecen el marco regulatorio para la implantación de un entorno colaborativo y posteriormente la especificación técnica ISO/TS 15066:2016 [15] que proporciona guías para su implementación.

Básicamente dicha normativa define 4 modos de colaboración:

- *Safety-rated monitored stop.*

  Si no hay nadie en la zona de trabajo colaborativo el robot puede trabajar autónomamente, pero si una persona entre en dicho espacio el robot debe pararse y garantizar que la parada es segura (con los motores energizados pero con sistema de monitorización certificado que asegura que el robot no se mueve). Un movimiento detectado por dicha monitorización debe dar lugar a una parada en categoría 0.

  En este modo es posible que un operario y el robot entren en contacto, por ejemplo para cargar o descargar piezas de la pinza del robot o labores de verificación.

- *Hand guiding*

  Es posible, bajo ciertas condiciones, mover el robot con la mano, bien actuando sobre el propio robot o mediante un dispositivo de mano. Esas condiciones son:

  - Previo al contacto el robot debe de estar en el modo 'Safety-rated monitored stop'
  - El dispositivo de guiado debe cumplir las especificaciones de seguridad e incluir una seta de emergencia y un mecanismo de habilitación
  - El operario debe tener una buena visibilidad del entorno de trabajo
  - En caso de dejar de actuar el dispositivo de habilitación ('hombre muerto') el robot debe entrar en el modo 'Safety-rated monitored stop'

- *Speed and separation monitoring*

  El robot es capaz de mantener una separación segura con respecto al operario en todo

[2] Una primera versión se publicó en 2006

momento. Esa separación depende de la velocidad del robot y la persona, alcance de la persona y tiempos de reacción y parada, debiendo establecerse durante el análisis de riesgo de acuerdo a [54], que únicamente considera velocidades normales de la persona (no incluye saltos, correr, caídas, etc.).

- Power and force limiting

  En este método se permite el contacto físico entre persona y robot, tanto de forma accidental como voluntaria. El robot debe estar diseñado específicamente para proporcionar esta funcionalidad, bien por un diseño intrínsecamente seguro o por medio de un sistema de control seguro. Tal y como se especifica en [15], se debe evaluar durante el proceso de Análisis de riesgos cuales son los límites de los valores de fuerza admisibles analizando el riesgo de contactos *cuasi-estáticos* y *permanentes* e identificando las partes del cuerpo humano susceptibles de recibir dichos contactos.

Los dos últimos métodos son los más relevantes para conseguir una colaboración real entre persona y robot ya que permiten que el robot esté en movimiento en la proximidad de la persona. Ya en el año 2.003, Ikuta [55] propuso clasificar las estrategias de seguridad en (1) estrategia pre-contacto y (2) estrategia post-contacto, tal y como se refleja en la ***Figura 11***. La primera intenta reducir el riesgo de colisión y de los efectos negativos de la misma antes de que suceda, mientras que la segunda tiene como objetivo reducir el daño una vez que la colisión ha sucedido.

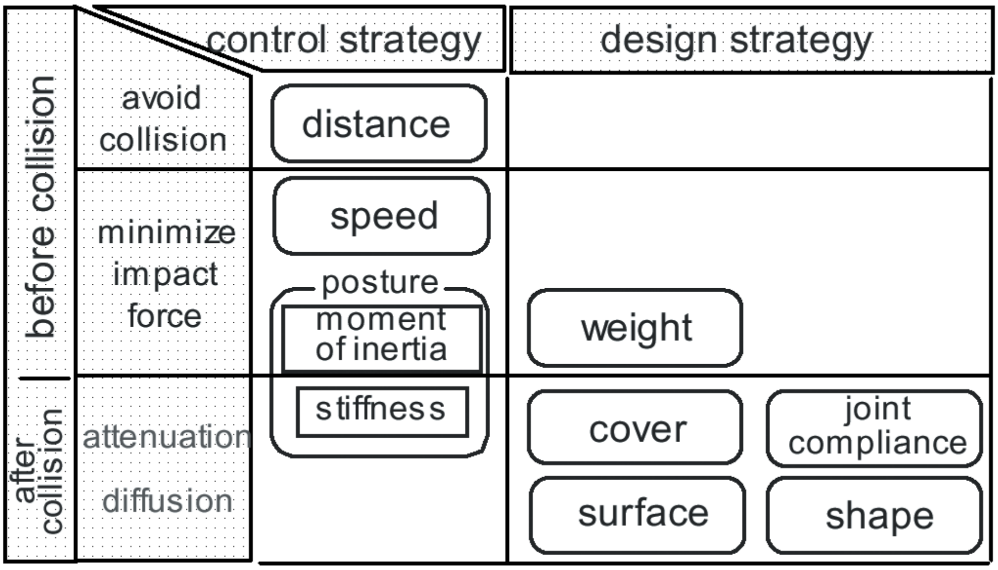


Figura 11: Clasificación de estrategias de seguridad [55]

En [56] se presenta una completa revisión de métricas usadas para medir la probabilidad y grado de severidad de colisiones aplicadas a diferentes sistemas autónomos e inteligentes, incluidos brazos robóticos.

Esos dos métodos han dado lugar a diferentes trabajos de investigación y algunos productos comerciales que se recogen en las revisiones del estado del arte de las publicaciones incluidas como soporte de este trabajo. Se recogen aquí los aspectos más relevantes.

3.2.1. Speed and separation monitoring (SSM)

SSM es el método más habitual de implementación de modos de colaboración segura.

Desde el punto de vista de control, existen dos alternativas de comportamiento: ajuste de la trayectoria del robot alrededor del punto de colisión potencial (evitación de obstáculos activa) y ajuste de la velocidad del robot en su trayectoria a la espera de que la situación de potencial riesgo desaparezca.

La evitación de obstáculos activa, a su vez, puede afrontar dos tipos de problemáticas en función de que los obstáculos sean estáticos o móviles. Los algoritmos más extendidos se basan en el concepto de campos de potencial (*potential fields)* [57]: el algoritmo define un destino objetivo y establece mecanismos de repulsión/expulsión con los objetos presentes en el espacio de trabajo. Si la distancia con respecto al destino se incrementa, la fuerza de atracción aumenta; si por el contrario la distancia con respecto a un objeto se reduce, una fuerza radial lo empuja alejándolo del objeto.

Una alternativa relacionada con los campos de potencial se basa en el concepto de fuerzas virtuales [58] en las que el controlador cuantifica la distancia entre el robot y los objetos como fuerzas simuladas que intervienen en el lazo de control del robot. En estos algoritmos se intenta mantener la adherencia a la trayectoria preestablecida.

En el caso de obstáculos móviles (y las personas pueden asimilarse como tales) el sistema de seguridad debe trazar la posición del obstáculo alrededor del robot. Tanto los métodos basados en campos de potencial como en los basados en fuerzas virtuales han sido extendidos para abordar estas situaciones [59].

Sin embargo el ajuste de velocidad es el mecanismo más común en soluciones aplicadas en entornos reales ya que son más previsibles en los resultados proporcionados desde el punto de seguridad aún a costa de una incertidumbre en el tiempo de operación del brazo robótico, dependiente de que el obstáculo desaparezca de su trayectoria.

Para implementar estos métodos es necesario disponer de sensores externos (ópticos y de proximidad fundamentalmente) que permitan estimar uno o ambos parámetros (velocidad y distancia).

Dejando a un lado las barreras opto-electrónicas tradicionalmente usadas en diferentes aplicaciones robóticas (escáneres laser de seguridad y cortinas ópticas de seguridad fundamentalmente) existe un creciente interés por la utilización de sistemas basados en Vision Artificial para determinar la distancia entre persona y robot.

Consisten en una o varias cámaras dispuestas en el entorno del robot ([60], [61], [62]) que permiten monitorizar la presencia de un ser humano en sus cercanías y disparar una parada o establecer una reducción de velocidad cuando la distancia esté por debajo de un umbral.

La gran mayoría de esos sistemas únicamente miden la distancia con respecto a una coordenada representativa del robot, por ejemplo el tool flange; otros, sin embargo, tienen en cuenta toda la cinemática del robot.

Únicamente un sensor comercial en el mercado basado en un sistema de visión dispone del certificado de seguridad necesario para su uso en un entorno real. Se trata del SafetyEye de la empresa PILZ [63] que utiliza un sistema de visión estéreo que permite definir diferentes volúmenes alrededor del robot. La cámara es capaz de detectar la presencia de cualquier objeto en esos volúmenes, representados en la ***Figura 12***, (no distingue si es una persona o un objeto) y proporcionar dos tipos de señales en función de la configuración del volumen. El sistema es sensible a determinadas condiciones lumínicas y no proporciona información al operario de la distancia relativa al robot, en consecuencia, violaciones involuntarias de las áreas de seguridad merman la eficiencia de los sistemas robóticos que se ven obligados a bajar su velocidad e incluso detener su movimiento ante dichas circunstancias.

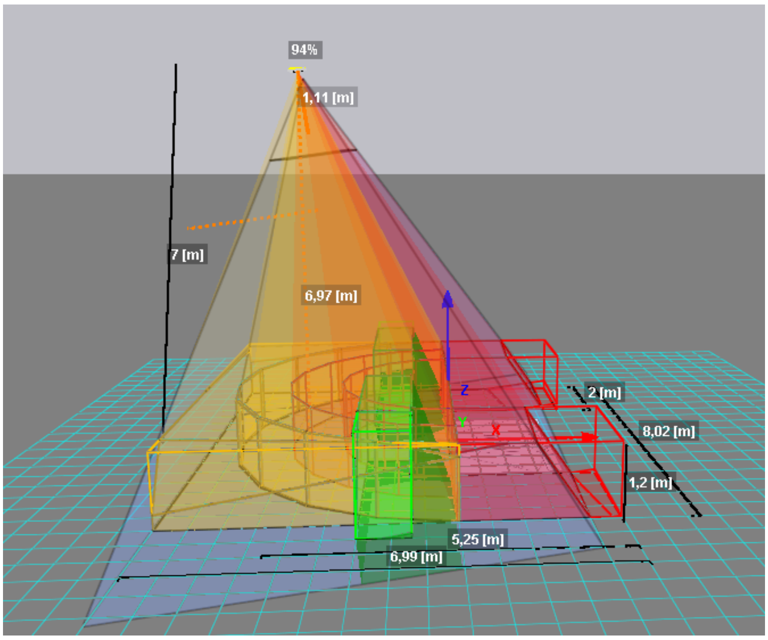


Figura 12: Configuración de zonas en el sistema SafetyEYE. [63]

El Fraunhofer IFF ha desarrollado un ingenioso sistema basado en un sistema de proyección y cámaras 2D que da respuesta a ese problema [64]. El sistema recibe la posición de los joints del robot en cada momento y conocida la cinemática del robot proyecta un patrón a su alrededor. Si una persona interrumpe dicho patrón el sistema de visión capta la circunstancia y se genera una señal de alarma. El funcionamiento se muestra en la *Figura 13*.

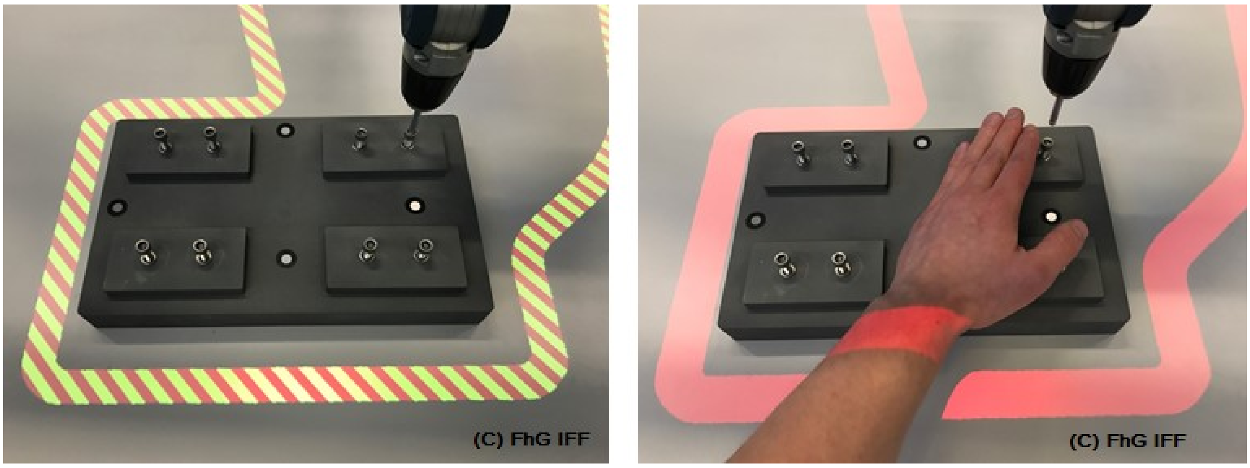


Figura 13: Sistema de proyección para monitorizar un espacio alrededor del robot

Los sensores de proximidad capacitivos permiten la monitorización de la distancia con respecto al robot en un rango de trabajo más corto, apropiado para actividades reales de colaboración, donde otros mecanismos antes descritos no son viables. Dichos sensores se pueden utilizar para detectar la colisión [65] una vez esta sucede, ofreciendo información de la zona en la que se produce y reduciendo el tiempo de respuesta del robot; o bien para detectar la proximidad de la persona con respecto al robot, permitiendo implementar el modo de SSM. Este es el caso del sistema APAS [66], consistente en una piel con sensores capacitivos que permiten detectar la proximidad de una persona generando una señal que el robot puede utilizar para ajustar la velocidad o trayectoria. La piel se fabrica a medida para cada robot.

3.2.2. Force and power limiting

La reducción de daños en caso de colisión se puede realizar mediante un diseño intrínsecamente seguro y mediante una adecuada estrategia de control. Existen en el mercado diferentes robots que implementan estos mecanismos y son los considerados robots intrínsecamente seguros. Este es el caso de KUKA iiwa [67] que incluye unos sensores de par en cada uno de los siete '*joints*' del brazo que le permiten estimar un sobreesfuerzo en caso de contacto con algún elemento exterior; Universal Robot [68] por su parte estima ese sobreesfuerzo a partir de la intensidad consumida por cada uno de los motores, siendo el límite de sensibilidad 5N, sin embargo dependiendo de la herramienta manipulada y la tarea, la fuerza de contacto puede ser difícil de estimar con precisión.

Los robots Baxter y Sawyer desarrollados por Rethink [69] utilizan actuadores del tipo Serial Elastic, que además de permitir medir la fuerza en caso de contacto, absorben parte de la energía de dicho contacto mediante el elemento elástico dispuesto en serie.

Una de las líneas de investigación con mayor impacto en este modo de colaboración es la determinación de los límites que definen un daño '*admisible*'. Para ello se trabaja en el modelado de la dinámica del robot y del cuerpo humano que permita establecer dichos límites y en consecuencia considerarlos en el diseño del robot, así como en la determinación de dichos valores mediante el ensayo con seres humanos [70], [71], [72] o [73]. El resultado de esos y otros estudios han sido recogidos en las tablas de la especificación técnica ISO/TS 15066:2016 [15] que pueden utilizarse como referencia en el análisis de riesgo.

3.2.3. Colaboración persona-robot y percepción de seguridad

Los factores que afectan a la confianza ('*trust*') durante la interacción persona robot han sido estudiados en diferentes trabajos, en particular cuando se desarrolla en aplicaciones militares. En [74] se evalúan y cuantifican diferentes factores y se concluye que los factores ligados al comportamiento del robot son los más relevantes, por delante de los factores ambientales, mientras que los factores ligados a la persona son relegados al tercer puesto.

En [75] y [76] se explica cómo afectan a la confianza el tipo y tamaño del robot, la proximidad y el comportamiento del robot. Según [77] la confianza se ve afectada de forma dinámica por diferentes factores relacionados con el propio robot, el entorno y la naturaleza y características de los miembros del equipo de trabajo. Finalmente en [78, 79] utilizan herramientas de simulación del comportamiento de un ser humano y un robot para medir la confianza durante una jornada laboral en un entorno de fabricación. Utilizan modelos de fatiga muscular para simular la actividad humana y tres modelos de control del robot (control manual por parte del operario, modo autónomo y colaborativo). En ese entorno de simulación el modo colaborativo da lugar a una mayor confianza y una moderada carga de trabajo.

3.3. Flexibilidad / percepción

La capacidad de los robots para realizar múltiples tareas se ha explotado tradicionalmente usando los mecanismos de programación y creando entornos estructurados en los que el robot sólo necesita hacer pequeños ajustes al comportamiento (trayectoria) especificado en el programa.

Sin embargo, la creciente tendencia hacia lotes de producción reducidos e incluso unitarios, el número de variantes en los productos ('*mass-customization*)' y la aparición de entornos colaborativos demanda una mayor capacidad de adaptación que sólo puede conseguirse mediante la introducción de sistemas de percepción y algoritmos de control que exploten la información que proporcionan dichos sistemas.

La visión artificial es la tecnología más relevante para dotar a los robots de niveles de percepción adecuados y es objeto de interés en la identificación de la pose 6DOF de objetos [80]. Las aproximaciones utilizadas por los equipos de investigación incluyen el uso del color y modelos geométricos [81], observándose un creciente interés en el uso de nubes de puntos adquiridas en tiempo real [82].

Numerosos autores han abordado el problema del *'visual servoing'* tal y como se recoge en las publicaciones que acompañan a este trabajo, por lo que aquí sólo citaremos algunos de los más relevantes. [83] propone una sistema basado en estéreo-visión para el guiado de un brazo manipulador que no dispone de información a priori de la distancia relativa con respecto a un objeto. Por su parte Nomura et al. [84] utilizan un filtro de Kalman híbrido para el seguimiento de un objeto desplazándose sobre una cintra transportadora e implementando control visual en 2D y 3D.

Utilizando dos robots manipuladores, [85] proponen un sistema para el seguimiento de objetos implementando filtros de Kalman extendidos para abordar el problema de las oclusiones originadas cuando intervienen ambos brazos. Finalmente en [86] y [87] se propone el uso de filtros de partículas para el seguimiento de objetos basado en la información 3D de sus bordes, permitiendo una detección más robusta que afronta los problemas de ruido en la imagen adquirida.

4. CONTRIBUCIONES

EN esta sección se resumen las contribuciones a los campos de actividad analizados en el capítulo anterior.

4.1. Interacción persona-robot

Las contribuciones en este campo se enmarcan fundamentalmente en la actividad desarrollada en los proyectos FourByThree (2.1.3) y EUROC-PIROS (2.2.6) en los que se plantea conseguir un grado de naturalidad en la interacción mediante el uso y combinación de diferentes mecanismos, como la voz, los gestos y la interacción física.

4.1.1. Tecnologías semánticas

La contribución consiste en un interpretador semántico que es capaz de combinar comandos de voz y gestos realizados por una persona y generar un comando entendible por el robot con un elevado grado de confianza.

Las tecnologías semánticas se utilizan para obtener una fusión coherente de las fuentes de información complementarias que llegan de forma simultánea desde el usuario hacia el robot.

Los seres humanos utilizamos diferentes canales de comunicación para interrelacionarnos entre nosotros., pero la voz y la visión son sin duda los más relevantes.

Mediante el sentido de la vista capturamos información que nos permite interpretar el estado de ánimo y filtrar la información que nos puedan proporcionar otros canales como el verbal (por ejemplo interpretando un gesto de la cara como expresión de duda o asombro en el hablante), pero es además el canal por el que podemos interpretar gestos explícitos por parte de nuestro interlocutor.

En la interacción entre personas combinamos ambos canales, voz y gestos, de una forma natural. La información que proporcionan ambos canales puede ser:

- **Complementaria**. Uno de los canales ofrece información que permite desambiguar la del otro. Un ejemplo puede ser: *El operario dice 'Coge esa pieza' mientras señala con la mano/dedo el objeto a coger*.
    - o Ninguno de los canales proporciona información suficiente para que el robot ejecute una acción
    - o Combinando ambos el robot es capaz de entender el comando
- **Redundante**. Ambos canales ofrecen una misma información (salvo error en su interpretación). Por ejemplo: *El operario dice 'Para' mientras levanta la palma de la mano (gesto*

universalmente aceptado como equivalente a una orden de detenerse)

- Cualquiera de los dos canales ofrece información suficiente para que el robot ejecute la acción
- Si ambos canales ofrecen información contradictoria (por un error en el usuario o por un error en la interpretación de la voz o del gesto) será necesario definir un heurístico que resuelva la contradicción.

En el caso de entornos industriales la información redundante permite desarrollar una interacción natural incluso cuando las condiciones ambientales disminuyen el grado de confianza de la interpretación de cada uno de los canales: la voz en entornos ruidosos o los gestos en condiciones de luminosidad inadecuadas.

Dicha contribución se refleja en la arquitectura de la ***Figura 14***:

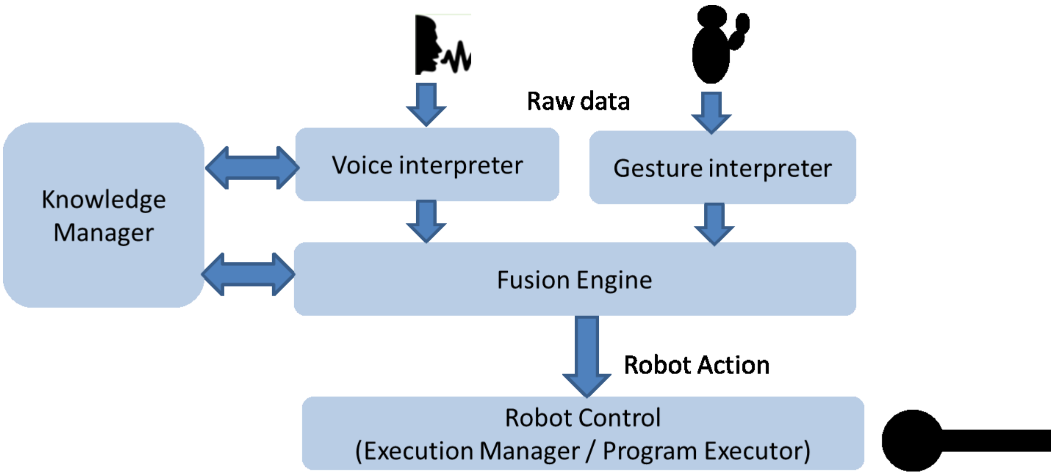


Figura 14: Arquitectura del sistema multimodal basado en tecnología semántica

- Knowledge Manager

  Gestiona las acciones que son ejecutables por el robot. Utiliza una ontología para modelizar el entorno, las capacidades del robot y las relaciones entre los elementos del modelo (entendido como un motor de reglas). El uso de ontologías en la nube permite reutilizar los desarrollos sin necesidad de reprogramar la lógica y compartir el conocimiento entre diferentes robots conectados en la nube.

- Voice Interpreter

  Recibe como input un comando de voz y genera un comando para el robot. Ese proceso se desglosa en tres etapas: (1) transcripción a texto (es posible utilizar un sistema comercial, en nuestro caso la Google Speech API (GoogleApi) [88]); (2) extracción de los elementos clave del texto basado en reglas; (3) correspondencia entre los elementos claves y las acciones admisibles para el robot.

  La extracción de los elementos clave se consigue utilizando técnicas de NLP (Natural Language Processing). Se utilizan reglas basadas en información morfo-sintáctica que se obtiene mediante la herramienta Freeling [89]. Para definir las reglas se aplica el análisis morfosintáctico y de dependencias a una base de ejemplos de instrucciones proporcionado por diferentes personas. Previamente y de forma manual se identifican los patrones morfosintácticos más habituales y se implementan en forma de reglas para la extracción automática de los elementos clave: elementos que indican acciones, objetos y lugares, términos relacionados con gestos (*aquí, allí,* etc.).

La correspondencia entre palabras clave y posibles acciones/elementos/objetos se realiza a partir de la base de conocimiento (*Knowledge Base)* de la que se extraen las tareas posibles mediante el lenguaje de consulta semántica SPARQL [90] y finalmente se lleva a cabo la desambiguación y la comprobación de coherencia utilizando la información destino/elemento mediante inferencia/razonamiento en OWL. El resultado final es un conjunto de posibles comandos y la identificación de la posible necesidad de fusionar con información proveniente del intérprete de gestos.

- Gesture Interpreter

  Identifica el gesto a partir de la nube de puntos proporcionada por un sensor 3D (láser o sistema de visión RGB-D). Ver la sección 4.1.2.

- Fusion Engine

  El motor de fusión se encarga de combinar el resultado del intérprete de voz y el intérprete de gestos para obtener el comando final a enviar al controlador del robot, bien en forma de acción directa (‘parar’ o ‘reanudar’) o ejecución de un programa (se ha diseñado el sistema de forma que acciones complejas como ‘atornillar un tornillo’ se asocien a un programa de robot).

  El proceso de fusión se representa en la ***Figura 15***

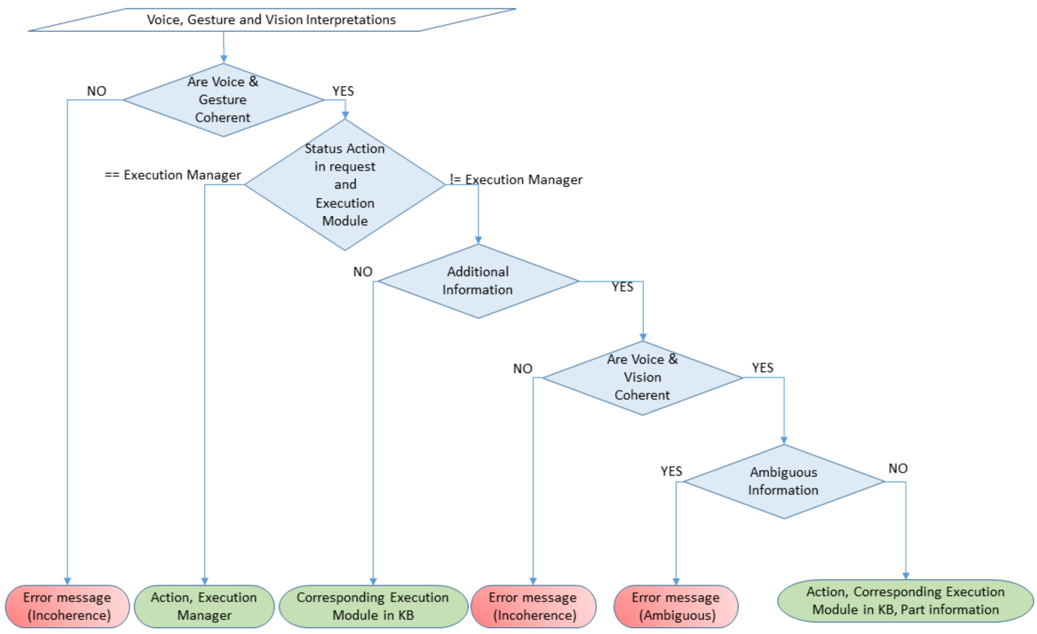


Figura 15: Lógica del Engine Fusion

Detalles de la implementación y el proceso de validación están disponibles en el artículo [91].

4.1.2. Interacción basada en gestos

Desde el punto de vista de interacción los gestos los podemos clasificar en dos categorías:

- Gestos que tienen un significado por sí mismo, fundamentalmente asociados a acciones como *parar* o *reanudar*.
- Gestos que complementan un comando de voz, ligados al concepto de apuntar y que son utilizados para indicar una dirección de movimiento, el objeto sobre el que ejecutar una

acción, etc.

En esta sección se resume la contribución en el campo del reconocimiento del gesto de apuntar, validado en diferentes proyectos (EUROC-PIROS y FourByThree) y descrito en detalle en [91]. Se adelanta también el trabajo en curso para el análisis de gestos denotando acciones con significado propio.

4.1.2.1. Identificación del gesto de apuntar

El gesto de apuntar juega un papel muy importante en la interacción natural ya que complementa al canal de la voz. Efectivamente, expresiones como 'vete allí' o 'coge esa pieza' únicamente son entendibles por el robot si se proporciona información que resuelva la indefinición de los conceptos '*allí*' o '*esa*'. De igual manera que los seres humanos resolvemos esa problemática mediante la ejecución del gesto de apuntar, la contribución aquí descrita tiene como objeto identificar el gesto de apuntar y estimar las coordenadas x, y, z del punto señalado.

En la contribución se parte de la nube de puntos adquirida por un sensor 3D que en la implementación que se describe consiste en una cámara RGB-D. Dicho sensor, referenciado con respecto a la base del robot, se debe colocar de tal forma que su campo de visión abarque la zona a apuntar y la zona del tronco superior de la persona que realiza el gesto. En la ***Figura 16*** se muestra la información proporcionada por la cámara RGB-D como nube de puntos e imagen de profundidad.

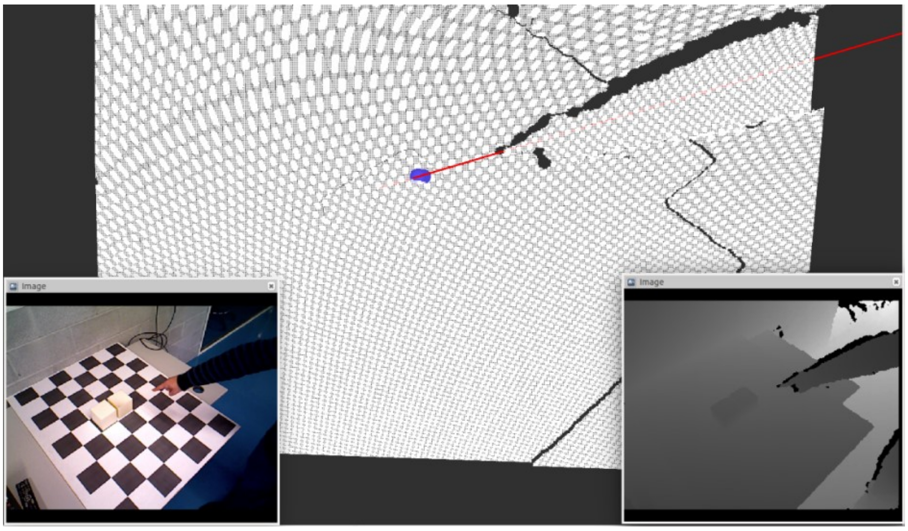


Figura 16: Gesto de apuntar, nube de puntos e imagen de profundidad

A partir de la nube de puntos se deben resolver dos problemas:

- Identificación del gesto de apuntar

  En un entorno colaborativo el trabajador desempeña su labor junto al robot, llevando a cabo diferentes actividades que suponen movimientos de sus manos. Es necesario discriminar dichos movimientos del propio gesto de apuntar. El algoritmo de identificación se dispara por eventos exteriores (una solicitud del robot que desea resolver una tarea para la que desconoce el destino o, más habitualmente, como resultado de un comando de voz que necesita ser completado con la información del gesto, por ejemplo '*coge esa pieza*').

  El algoritmo define dos regiones de interés en forma ortoédrica. La primera de ellas define el espacio de búsqueda del antebrazo del operario, y la segunda el espacio de búsqueda del punto en el espacio al que se está apuntando. La nube de puntos correspondiente al brazo del operario es segmentada y aislada de la nube de puntos original. Teniendo en cuenta la forma natural del antebrazo, asimilable a una forma cilíndrica, se aplica un best-fit de esta forma geométrica básica a dicha nube de puntos. El eje principal de dicho cilindro representa la línea en el espacio que se utilizará en la identificación de la zona apuntada. Para un mejor resultado en este ajuste, no se toma en cuenta la mano del operario.

- Identificación de la zona apuntada

  De forma iterativa se generan regiones de interés de forma cúbicas a lo largo de la línea del eje del brazo comenzando por aquellos puntos de la línea más alejados del brazo a partir del cual se ha estimado la línea, dentro del espacio de trabajo definido. Se identifican los puntos de la nube original que se encuentran contenidos en cada uno de los clúster definidos iterativamente, estableciéndose como criterio de intersección la presencia de N_a puntos de la nube de puntos original en dicho cubo. Se toma como intersección el centroide del cubo.

  El punto de origen de la línea se encuentra contenido en el centro del eje principal del cilindro estimado a partir del antebrazo del usuario. Por lo tanto, se establece una distancia euclídea mínima del punto de intersección respecto a este punto de origen de la línea, dependiendo de la aplicación.

  Una vez seleccionado un punto de intersección, se vuelve a repetir el proceso N_b veces con nuevas nubes de puntos, con lo que se obtienen N_b puntos de intersección. Para que el gesto sea estable y no confundirlo con un movimiento no controlado de la mano se realiza un filtrado espacial: se toma el primer punto de intersección como centroide de un cubo de dimensiones configurables. Si N_c puntos de intersección identificados están contenidos en el cubo se considera un gesto válido y las coordenadas del último punto se devuelven como resultado; en caso contrario se descarta la primera nube de puntos del conjunto tratado, se incorpora una nueva nube de puntos y se repite el proceso.

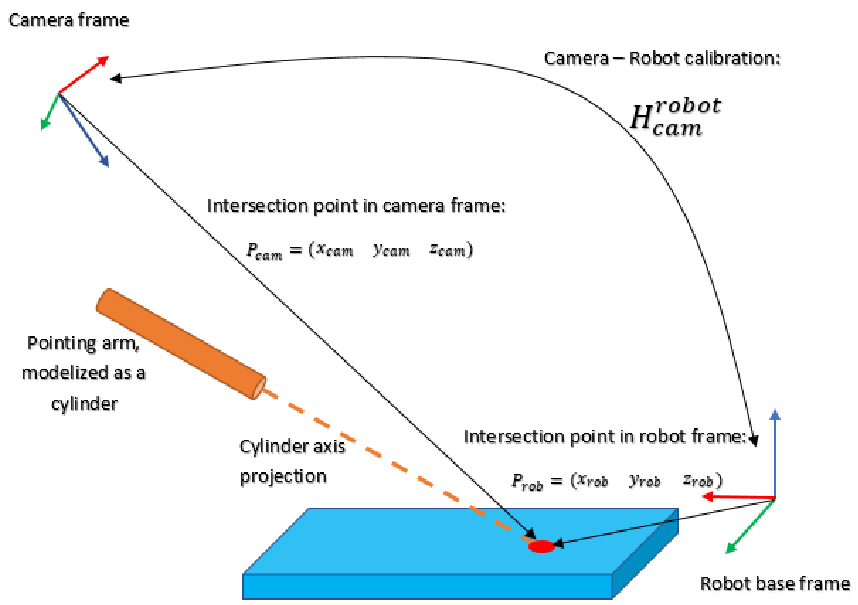


Figura 17: Punto de intersección en el gesto de apuntar

Una segunda versión del algoritmo asume que el punto destino se encuentra sobre una superficie plana (válido en un gran número de casos). Se resuelve matemáticamente la intersección recta - plano, teniendo en cuenta la ecuación de la recta estimada a partir de la modelización del cilindro del antebrazo y la superficie objetivo definida matemáticamente como un plano. Este plano se define a partir de los puntos de la nube dentro de la región de interés.

Esta segunda aproximación consigue un mejor comportamiento en tiempo de cómputo, pero está limitada a aquellos casos en los que la asunción de la existencia de un plano es válida.

El sistema ha sido validado en el contexto del proyecto EUROC-PIROS. Dentro de su fase Freestyle, se planteó un caso de uso ligado con las células de fabricación flexible (FMS). En dichos procesos se utilizan pallets en los que se cargan piezas que deben ser mecanizadas. Con la introducción de robótica colaborativa se plantean varias aplicaciones en las que el gesto de apuntar cobra sentido, por ejemplo el operario indica al robot en qué destino debe colocar una pieza, o sobre cual debe realizar una operación de inspección.

Para la evaluación se utilizó un pallet con 12 posibles posiciones en cada una de las cuales se disponía de un pequeño utillaje de amarre. Las doce posiciones se dispusieron en cuatro filas (etiquetadas del 1 a 4) y tres columnas (etiquetadas como A, B, C). El operario debía señalar una de esas posiciones (*A3*, *B4*, etc.) a instancias del evaluador, el sistema de detección de gestos detectarlo y el robot se debía moverse hasta dicha posición.

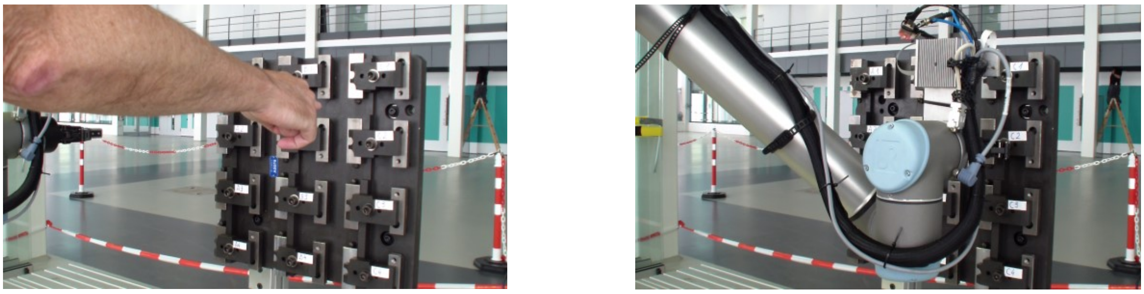


Figura 18: Proceso de evaluación en el contexto del proyecto EUROC-PIROS

En el procedimiento de evaluación el operario (una persona ajena a nuestro equipo que ejercía el papel de evaluador externo) indicaba de forma aleatoria una celda y una persona realizaba el gesto de apuntar a dicha celda. Se realizaron dos rondas y en cada una se apuntó a las doce casillas. Se evaluaron dos parámetros:

- Precisión: identificación correcta del destino. Las 24 celdas fueron detectadas correctamente.
- Tiempo de identificación. Siendo el gesto de apuntar un elemento de interacción se debe tener en cuenta que de acuerdo a [92] tiempos de reacción del sistema por debajo de 0,1 segundos son percibidos como una respuesta inmediata por el usuario de un interface y que hasta 1 segundo se consideran aceptables y no es necesario dar un feedback adicional.

  El resultado obtenido fue de 0,49 segundos con una desviación estándar de 0,11 segundos, por lo que se considera válido para obtener una interacción natural.

4.1.2.2. Detección de gestos

Para la detección de gestos no ligados con el acto de apuntar un destino, se está desarrollando un sistema que a partir de la nube de puntos identifica las coordenadas de las diferentes partes del cuerpo humano (un modelo simplificado) y estima el gesto realizado.

Utilizando esta aproximación, es posible definir múltiples tipos de gestos, generando un vector numérico que los define, siguiendo los siguientes pasos:

- Se toma como origen de referencia las coordenadas 3D del torso del usuario. La utilización del torso como punto de referencia invariante para el cálculo de las distancias relativas de las articulaciones hace este cálculo robusto frente a la variación de posición del operario respecto al origen de coordenadas de la cámara.
- Se calculan las coordenadas relativas de las diferentes articulaciones del esqueleto del usuario respecto a la posición del torso. En concreto para la definición de gestos se utilizan el torso, cuello, cabeza, hombros, codos y manos.

- Con dichas coordenadas se crea un vector que define el gesto.

Es posible definir un gesto teniendo en cuenta la posición global del esqueleto del usuario, en cuyo caso el vector tendrá 24 componentes numéricas (8 puntos del esqueleto * 3 componentes cada uno), o solamente cada una de las partes del cuerpo, izquierda o derecha, en cuyo caso, el vector de definición del gesto tendrá 15 componentes (5 puntos del esqueleto * 3 componentes cada uno).

Para la detección de los gestos se ha utilizado una aproximación basada en Machine Learning. La utilización de esta aproximación viene motivada fundamentalmente por dos razones:

- La variabilidad en la ejecución de los gestos: (1) una misma persona no siempre realiza un gesto exactamente igual y (2) el sistema debe ser válido para su utilización con distintas personas.
- El sistema debe posibilitar la definición de nuevos gestos de forma sencilla y sin modificación de código.

El algoritmo utilizado es KNN (K-Nearest Neighbours), al no requerir un proceso de entrenamiento previo.

El sistema está estructurado en dos módulos:

- Módulo de aprendizaje.

  Para cada uno de los gestos se crea un fichero en el que los diferentes vectores de definición de dicho gesto son almacenados, de esta forma se pueden definir un número N de gestos diferentes.

- Módulo de detección de gestos en tiempo real.

  En la inicialización del sistema, los diferentes ficheros de definición de gestos son cargados en la memoria del sistema creando la base de datos de vectores de definición de gestos.

  En tiempo real, se monitoriza el cuerpo del operario con una cámara RGB-D y se extraen las coordenadas de sus articulaciones respecto al torso, al igual que en la etapa de entrenamiento.

  Finalmente se aplica el algoritmo KNN con este vector y se identifica el gesto.

Al igual que en el caso del gesto de apuntar, se realiza un filtrado temporal para establecer que efectivamente se ha detectado un gesto y no un movimiento casual del operario.

4.1.3. Estudios experimentales sobre diferentes formas de interacción en un entorno de robótica colaborativa

4.1.3.1. Experimentación en condiciones de laboratorio en IK4-TEKNIKER

Como parte de la validación de las estrategias de seguridad y diferentes formas de interacción en un entorno colaborativo implementadas en el proyecto X-ACT se diseñó y llevó a cabo el siguiente experimento (presentado en el workshop 'Safety for Human-Robot Interaction in Industrial Settings', Congreso IROS 2015 celebrado el 2 Octubre 2015 en Hamburgo, Alemania)

Objetivo: analizar la aceptación de las diferentes formas de interacción propuestas (voz, gestos, implícita y botonería)

Participantes: Se reclutaron 17 personas pertenecientes a la plantilla de IK4-TEKNIKER con las siguientes características: (1) la participación fue voluntaria, (2) ninguna formaba parte del grupo de investigación en robótica (3) el 82% tenían experiencia de más de seis años como operarios de máquina, instalación de sistemas de automatización o diseño de sistemas productivos, (4) tres de los participantes formaban parte del comité de Salud y Seguridad Laboral (5) once de los participantes habían conocido accidentes de trabajo en su entorno.

Organización del experimento: Cada participante participó en dos sesiones individuales separadas en el tiempo una semana aproximadamente, con el objeto de evaluar el efecto de recuerdo y la evolución de la confianza en el sistema.

En cada sesión el participante debía llevar a cabo un conjunto de actividades que implicaban la necesidad de trabajar en la cercanía del robot sin barrera física. Al final de cada sesión el participante rellenaba un cuestionario con preguntas cerradas (usando una escala de Lieker de 5) y abiertas.

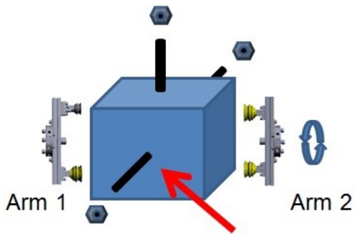
Trabajo colaborativo: Se diseñó una tarea sencilla que simulaba una actividad real en un entorno de colaboración persona-robot.

- El robot bimanipulador movía una caja (ver la ***Figura 19***) describiendo un movimiento circular con el operario fuera de la zona de trabajo del robot.
- Cuando se necesitaba la intervención del operario, el robot posicionaba la caja en una posición colaborativa (con los brazos hacia adelante) y solicitaba la presencia del operario (con un mensaje en un monitor situado en la base del robot).
- Una vez en la zona colaborativa, el operario debía atornillar un tornillo en cada uno de las tres caras accesibles, pero con la condición de que el atornillado se debía ejecutar sobre la cara vertical frontal, por lo que tras cada acción de atornillado debía solicitar al robot que girara la caja para poder proceder con la siguiente cara.
- Una vez finalizada la operación el operario abandonaba la zona de trabajo y el robot reanudaba el movimiento circular.
- La misma secuencia se debía repetir para desatornillar los tres tornillos.
- El operario debía utilizar cuatro formas de interacción:
    - Un pulsador. Cada vez que se pulsaba el robot giraba la caja presentando una nueva cara para atornillar. Con la última pulsación el operario señalaba el fin de la actividad colaborativa (ver la ***Figura 20***).
    - Comandos de voz. Se implementaron dos comandos: '*girar*' y '*fin*'.
    - Gestos. Dos simples gestos: levantando una mano para indicar que el robot debía girar, levantando las dos manos para indicar el fin de la actividad (ver la ***Figura 20***).
    - Interacción implícita. Es decir, el robot identificaba la finalización de las acciones del operario y ejecutaba el movimiento sin necesidad de un comando explícito por parte del operario.

  Dado que el objetivo del experimento no era validar las tecnologías de interacción, se utilizó la técnica de Mago de Oz: uno de los dos técnicos que tomaban parte en el experimento simulaba el comportamiento esperado (la interpretación de la voz, el reconocimiento de gestos o la interacción implícita)

- Una vez el operario fuera de la zona colaborativa, se solicitaba de forma aleatoria que entrara en la zona de trabajo del robot (se habían dispuesto tres objetos sobre la mesa de trabajo en la zona colaborativa; el experimentador solicitaba al trabajador que trajera una de esas piezas). El operario podía proceder de cualquiera de las siguientes formas:
    - Utilizar la llave de bloqueo: con ella se ordenaba la parada del robot y se entraba en la zona colaborativa con la garantía de que el robot estaba parado

- Confiar en que el sistema de monitorización y el método SSM funcionaba correctamente y entrar en la zona de movimientos del robot. Éste se paraba una vez que el operario alcanzaba la zona de seguridad



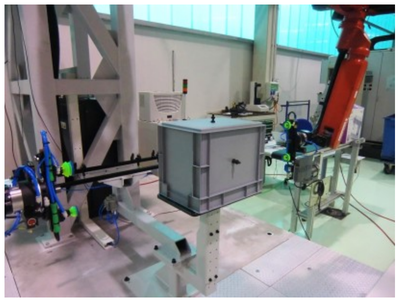


Figura 19: Elementos utilizados para la simulación de una actividad colaborativa

Resultados: Se resumen los datos obtenidos en los cuestionarios y como fruto de la observación durante el experimento.

- Todos los sistemas de seguridad implementados (SafetyEYE, seguimiento de personas mediante los láseres, llave de bloqueo) contribuyen de forma similar a la percepción de seguridad.
- Ninguno de los participantes se sintió inseguro: once completamente seguros y seis seguros. Esto explica la respuesta a la pregunta '*¿Aceptarías trabajar en el futuro en un entorno robótico sin barreras físicas*?: únicamente uno de los participantes no aceptaría trabajar sin barreras físicas.
- La demarcación de la zona colaborativa de forma visible (se pintó una raya en su perímetro) se valoró positivamente por la mayoría de los participantes.
- El uso de la llave de bloqueo para entrar en la zona de trabajo del robot era una forma de evaluar la confianza en la estrategia SSM implementada. Sólo 3 de los participantes respondieron que siempre la utilizarían.
- Hubo unanimidad en la conveniencia de disponer de una Seta de Seguridad en la zona colaborativa: para once debería ser obligatorio, para siete sería conveniente.
- La posición de trabajo con los brazos del robot extendido se consideró positivamente por la mayoría de los participantes.
- La velocidad del robot se consideró muy lenta durante la operación de giro que se llevaba a cabo en presencia del operario.
- Preguntados sobre la posibilidad de que el robot cambiara de trayectoria en lugar de reducir la velocidad en presencia del operario, todos los participantes optaron por esta última alternativa.
- Al final de la segunda sesión el 53% de los participantes afirmaron que su percepción de seguridad había mejorado.

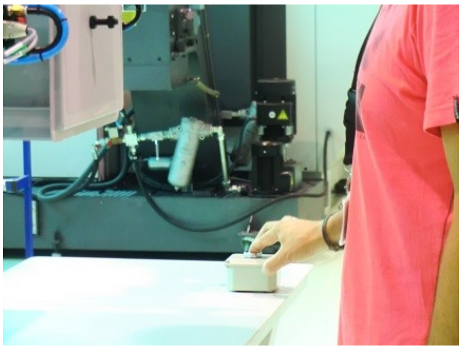
 

Figura 20: Un participante usando el pulsador y gestos para comandar el robot

El resultado del experimento demostraba un grado de confianza muy elevado y una buena disposición para trabajar en entornos colaborativos una vez que las medidas de seguridad adecuadas fueran implementadas y explicadas a los potenciales usuarios. Los resultados han servido de guía para la implementación de la estrategia de seguridad en el proyecto FourByThree.

4.1.3.2. Experimentación en la Bienal de Máquina Herramienta y feria Technishow

En el marco del proyecto FourByThree se ha podido llevar a cabo un conjunto de experimentos en entornos fuera de laboratorio que han permitido que un número significativo de personas ajenas al proyecto y al propio IK4-TEKNIKER hayan tenido la oportunidad de testear los desarrollos y expresar su percepción sobre diferentes aspectos ligados a la robótica colaborativa y en particular sobre los conceptos de seguridad e interacción.

Dicha experimentación se ha realizado en dos ferias industriales en las que los visitantes han sido invitados a participar en el experimento, testear de primera mano las tecnologías y finalmente responder a un cuestionario para expresar sus opiniones. En conjunto 115 personas se ofrecieron voluntarias para participar en la experimentación.

- TECHNISHOW

  Es una feria industrial que se celebra anualmente en Utrecht en la que, en colaboración con un socio del proyecto FourByThree, se llevó a cabo un experimento con un robot Universal Robot y diferentes tecnologías: reconocimiento de gestos, detección de proximidad, guiado manual y detección de colisión. En la *Figura 21* se muestran algunos de los participantes durante el experimento.

- BIEMH

  En la Bienal de Máquina Herramienta se dispuso de un entorno experimental en el stand de IK4-TEKNIKER. En él, los participantes pudieron testear el sistema de detección de proximidad, interacción mediante gestos e interacción física con un robot KUKA iiwa.

Figura 21: Experimento en Technishow (izquierda) y BIEMH (derecha)

Una descripción detallada de ambos experimentos y el análisis de los resultados obtenidos se encuentra disponible en los artículos [91] y [93].

Desde el punto de vista de interacción los resultados más relevantes fueron:

- El sistema de reconocimiento del gesto de apuntar fue positivamente valorado en ambos estudios, tal y como se muestra en la *Figura 22*. Sin embargo es digno de señalar la mejora en la percepción del tiempo de respuesta en el estudio llevado a cabo en la feria BIEMH. La razón, además de pequeñas mejoras en el algoritmo, reside en el hecho de que se incluyera un mecanismo visual que se activaba para avisar al participante de que podía realizar el gesto y se desactivaba al reconocer el gesto.

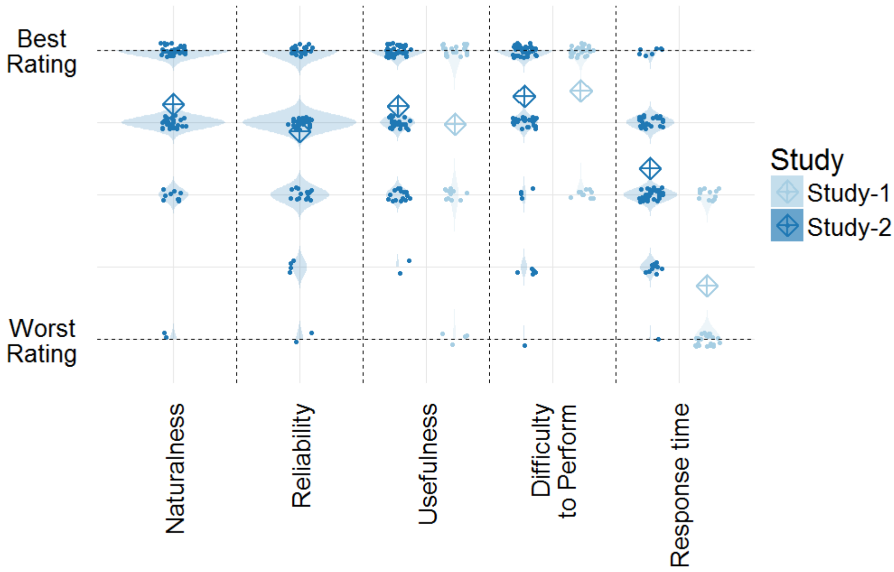


Figura 22: Valoración del reconocimiento del gesto de apuntar

- La fuerza necesaria para mover manualmente el robot (sólo se validó en la feria Technishow con el robot UR10) se consideró en el límite de lo aceptable, aunque el mecanismo de interacción se consideró útil y fácil de utilizar.
- La interacción física (un pequeño golpe en el robot) fue la peor valorada con un 8% de participantes afirmando que no era natural. Dado que es un mecanismo que no necesita una sensorización adicional (se pueden usar la capacidad de medir fuerza, propia de los robots colaborativos) se considera que es una opción que se debe poner a disposición del usuario.

4.2. Seguridad

4.2.1. Seguimiento de personas

En el proyecto X-ACT [2] se ha desarrollado un sistema para la monitorización de personas alrededor del robot basado en la información proporcionada por dos láseres dispuestos a diferentes alturas. La información obtenida es utilizada para modificar la velocidad del robot e implementar el método de colaboración SSM.

El sistema ha sido desarrollado según el esquema de la ***Figura 23***.

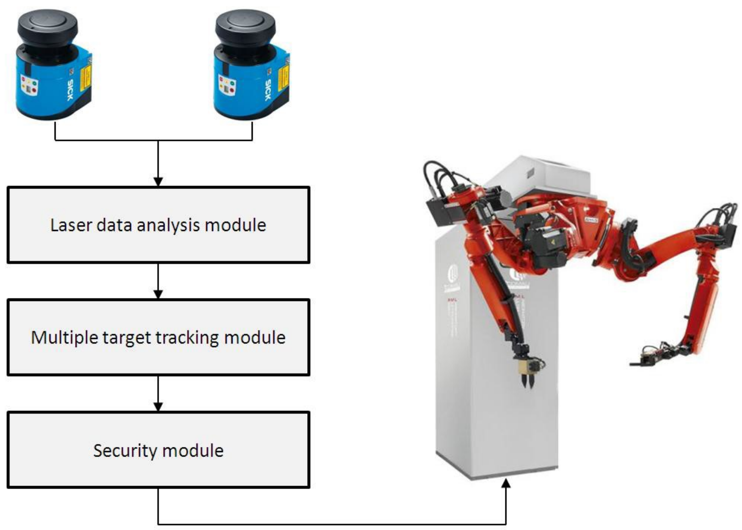


Figura 23: Elementos del sistema de seguimiento de personas

- Módulo de análisis de los datos proporcionados por los sensores Laser

  Este módulo recibe los datos proporcionados por dos escáneres laser y los analiza para identificar la presencia de cualquier obstáculo. Para ello, los datos son comparados con un patrón del entorno previamente registrado en el que se caracteriza cualquier elemento estático (dado que el entorno industrial real es cambiante, ese patrón puede actualizarse en cualquier momento a demanda del usuario). Las coordenadas de aquellos obstáculos detectados son enviadas al *módulo de seguimiento de objetivo*.

- Módulo de seguimiento de múltiples objetivos

  A partir de la información proporcionada por el módulo anterior, implementa un algoritmo de seguimiento de múltiples objetivos basado en los algoritmos *Sequential Monte Carlo filtering* y *Joint Probabilistic Data Association*. Este filtro probabilístico permite integrar múltiples fuentes de información (en este caso los dos sensores laser) en problemas de seguimiento de múltiples objetivos (personas alrededor del robot). La idea principal es la de analizar la probabilidad de que cada observación (detección del módulo previo) pertenezca a un objetivo, obteniendo como resultado la asociación observación-objetivo más verosímil. Esto permite tratar tanto el ruido de los sensores como el problema de las múltiples detecciones de un mismo objetivo de forma sencilla y minuciosa gracias al acercamiento probabilístico del filtro. Como resultado, este segundo módulo obtiene la posición y velocidad de las personas alrededor del robot que son enviadas al Módulo de seguridad.

- Módulo de seguridad

Genera los comandos necesarios para modificar la velocidad del robot a partir de la información de posición y velocidad de cada uno de los obstáculos detectados

La información de posición de cada obstáculo y la velocidad de ajuste del robot son mostradas en un monitor presente en la zona colaborativa, tal y como se muestra en la ***Figura 24*** en la que el punto rojo representa la persona identificada sobre la que se realiza el seguimiento.

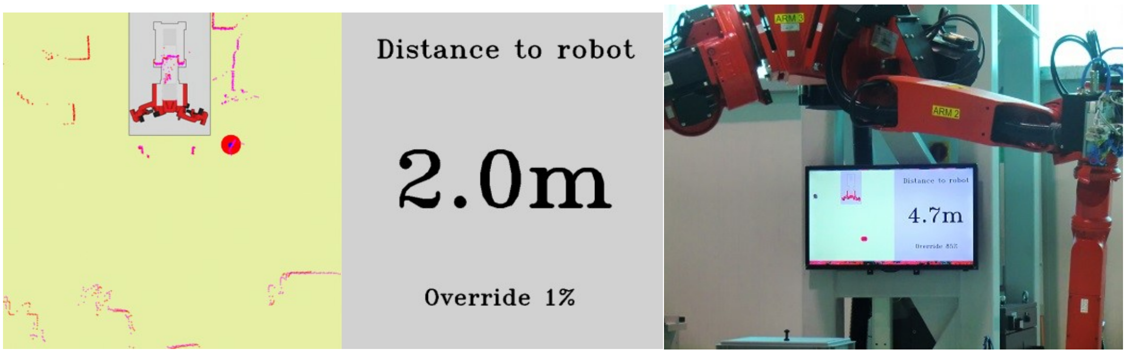


Figura 24: Entorno de visualización del sistema de seguimiento

Detalles de los algoritmos implementados y el proceso de validación llevado a cabo están disponibles en el artículo [94].

4.2.2. Detección de proximidad

El objetivo de esta actividad ha sido el desarrollo de un sistema que permita implementar el método de SSM, monitorizando todo el volumen alrededor del robot y calculando la mínima distancia entre el obstáculo y el robot, considerando toda su geometría. Da respuesta a las limitaciones de otros sistemas como el de proyección antes descrito (sólo trabaja en un plano, el de la mesa, y la mínima distancia es entre la proyección del brazo robot y la línea proyectada) o el SafetyEYE que impide monitorizar la zona de trabajo del robot propiamente dicha (se debe configurar como no monitorizada). Permite, por lo tanto, monitorizar la distancia para rangos de acción cortos, propios de las actividades reales de colaboración en las que se supone que el trabajador y el robot se encuentran trabajando 'codo con codo'.

Los elementos fundamentales del sistema desarrollado se muestran en la ***Figura 25***:

- Utiliza una nube de puntos obtenida mediante uno o varios sensores. En los proyectos ligados a esta contribución se han utilizado sensores Laser 3D (que nos permiten disponer de una nube de puntos densa en un volumen de trabajo grande, requerimiento del proyecto ROBOPARTNER) y cámaras RGB-D (para aplicaciones con robos colaborativos de menor tamaño).
- Zona de detección. Es el volumen en el que se desea llevar a cabo la monitorización de la distancia.
- El modelo geométrico de colisión del robot ('*robot's collision geometry*') definido por un conjunto de cubos que representan un modelo simplificado del robot.
- Información en tiempo real de la posición del robot.
- Modelo Octomap 3D: el modelo de ocupación 3D utilizado para representar obstáculos y el espacio libre en el volumen monitorizado.

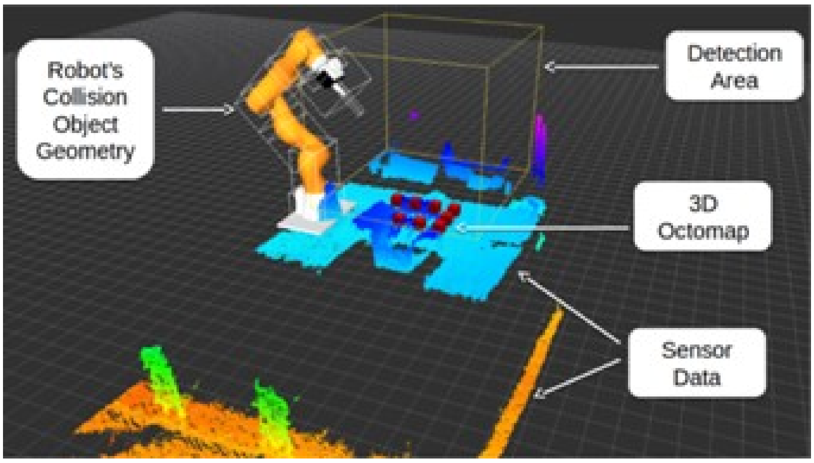


Figura 25: Elementos fundamentales del sistema de detección de proximidad

En la ***Figura 26*** se muestran los componentes fundamentales del sistema:

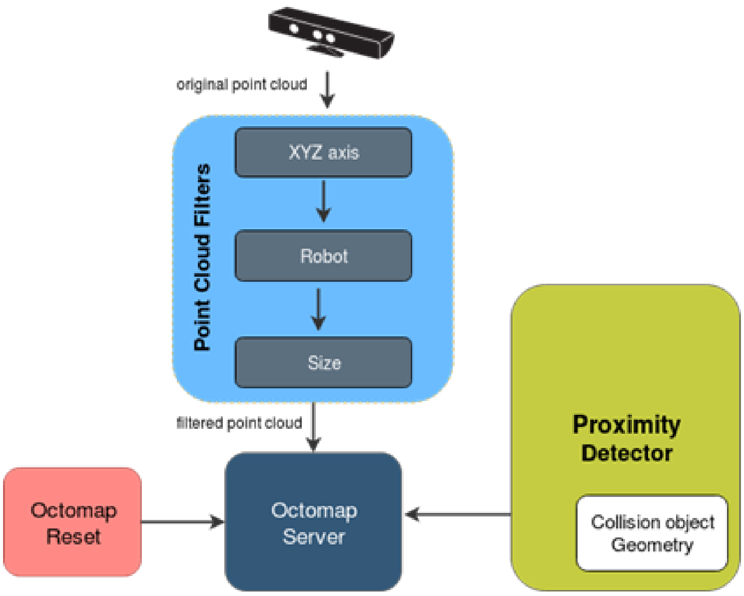


Figura 26: Componentes del sistema de monitorización de proximidad

- Point cloud filters

  Se aplican diferentes filtros a la nube de puntos original proporcionada por el sensor. Cada filtro es implementado como un nodo de ROS que se subscribe a un *topic* para recibir la nube de puntos y publica en otro *topic* la nube una vez filtrada.

  - XYZ Axis. Este filtro elimina aquellos puntos de la nube que se encuentran fuera de las dimensiones del volumen de detección definido por el usuario.
  - Robot: Elimina los puntos de la nube correspondientes al robot, que de otra forma podrían ser considerados obstáculos.
  - Size: Elimina los clúster de puntos por debajo de un tamaño, eliminando de esta forma posibles ruidos en la señal.

- Octomap server

  Este nodo es el encargado de mantener una representación volumétrica del entorno mediante un mapa de rejillas tridimensional. Se basa en Octrees y utiliza estimaciones probabilísticas de ocupación, representando explícitamente tanto el espacio ocupado como el espacio libre y desconocido con un formato binario compacto.

- Octomap reset

  El nodo limpia el mapa 3D con una frecuencia configurable. Utilizando el Octomap server y el Octomap reset es posible mantener actualizado un mapa 3D del volumen a monitorizar.

- Proximity detector

  Es el nodo que computa la distancia entre el robot y posibles obstáculos. Carga el modelo geométrico de colisión del robot y en cada ciclo solicita al Octomap server el modelo más reciente de mapa. Con esa información y haciendo uso de la librería Flexible Collision Library [95], se computa la distancia mínima entre el robot y el obstáculo.

El sistema se ha desarrollado y validado en los proyectos FourByThree [11] y ROBOPARTNER [3], dando lugar a la solicitud de patente P201730325 con el título '**MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD**' (sección 8.1).

4.2.3. Estrategias y Arquitectura de seguridad

La contribución ha consistido en el diseño una estrategia de seguridad alrededor del robot bimanipulador COMAU-RML basado en el método SSM, ya que se trata de un robot industrial que no implementa un control de fuerza. Se ha llevado a cabo en el proyecto X-ACT e incluye los siguientes elementos (ver esquema en *Figura 27*):

- SafetyEye. Es un sensor de seguridad basado en tecnología de estereovisión [63] que permite definir diferentes volúmenes alrededor del robot. Cada uno dichos volúmenes se asocia a una de las dos categorías posibles (aviso y alarma) o la categoría de no monitorado. El sensor SafetyEye, pese a ser un sensor certificado para su uso en aplicaciones de robótica colaborativa, tiene las siguientes limitaciones: (1) por su principio de funcionamiento, debe definirse un volumen que envuelve al robot en el que el sensor no actúa, en consecuencia cualquier persona que en un momento determinado se encuentre en dicha zona no podrá ser monitorizado; (2) únicamente permite conocer la presencia de un elemento intruso en alguno de los volúmenes definidos, pero no su velocidad ni su trayectoria.
- Dos láseres situados en la base de la zona de colaboración. Permiten solventar las dos limitaciones anteriores, monitorizando en dos planos situados a distintas alturas la presencia de una persona, utilizando la aproximación descrita en la sección 4.2.1.

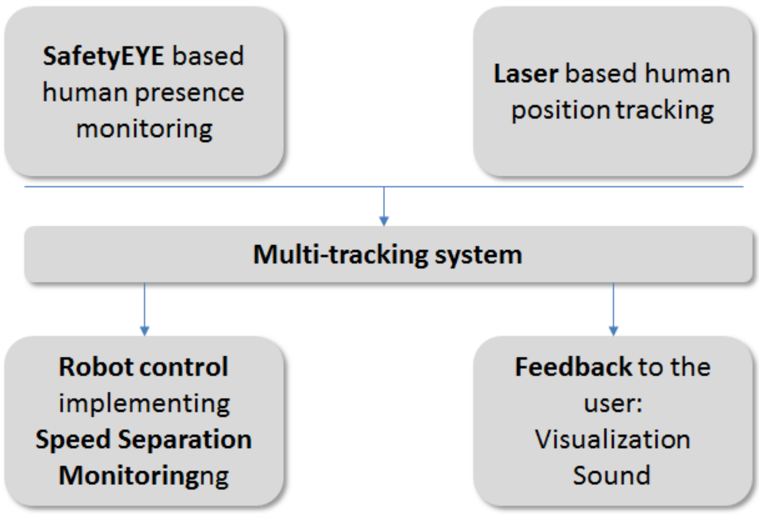


Figura 27: SSM método implementado en X-ACT

Esa implementación del método SSM forma parte de la estrategia global en la que se incluyen además otros elementos:

- Se define el área colaborativa de forma que el robot trabaje en ella con los brazos extendidos (pero alejados de posibles singularidades), asumiendo que un operario se sentirá más seguro sabiendo que el robot no va a mover los brazos hacia adelante (hacia el operario). Esa posición se muestra en la ***Figura 28***.
- Se definen regiones de interferencia en el robot de forma que los movimientos del robot se confinan, bien en esa zona o bien fuera de la misma. En la ***Figura 28*** se ha representado dicho volumen.
- Una llave de bloqueo en la zona exterior no monitorizada, permite acceder a la zona colaborativa de forma segura. Es de uso voluntario.

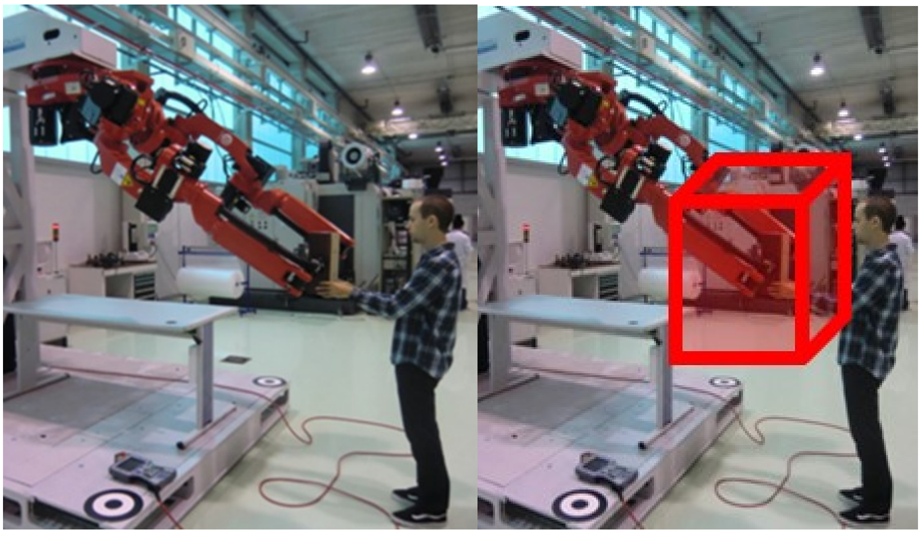


Figura 28: Robot en posición de colaboración y zona de interferencia

La segunda contribución en este ámbito corresponde al diseño de la arquitectura de seguridad en el proyecto FourByThree [11] que como se describe en [93] se basa en cuatro pilares fundamentales:

- Diseño intrínsecamente seguro, eliminando bordes, reduciendo los riesgos de atrapamiento, etc.
- Utilización de actuadores serie-elásticos (*serial elastic*) que permiten medir la fuerza y el par y proporcionan una medida redundante tanto de la posición como el par que es estimado utilizando dos fuentes de información diferentes: medida de la corriente de los motores y la deformación del muelle. Se comparan la señal de los sensores de posición a ambos lados del motor y la transmisión y en caso de divergencia se genera una señal de alarma.
- Los sistemas externos de monitorización del entorno que permiten implementar el modo de colaboración SSM, detectando una posible violación de la zona de seguridad.
    - Se implementa el sistema de proyección descrito en [64] y optimizado en el contexto del proyecto.
    - El sistema de detección de proximidad descrito en la sección 4.2.2.
- Sistema de control dinámico de la rigidez (*variable stiffness*) que permite ajustar los valores de la misma en función de diferentes factores, como la presencia de un operario o la tarea que se está ejecutando.

En la ***Figura 29*** se presenta dicha arquitectura.

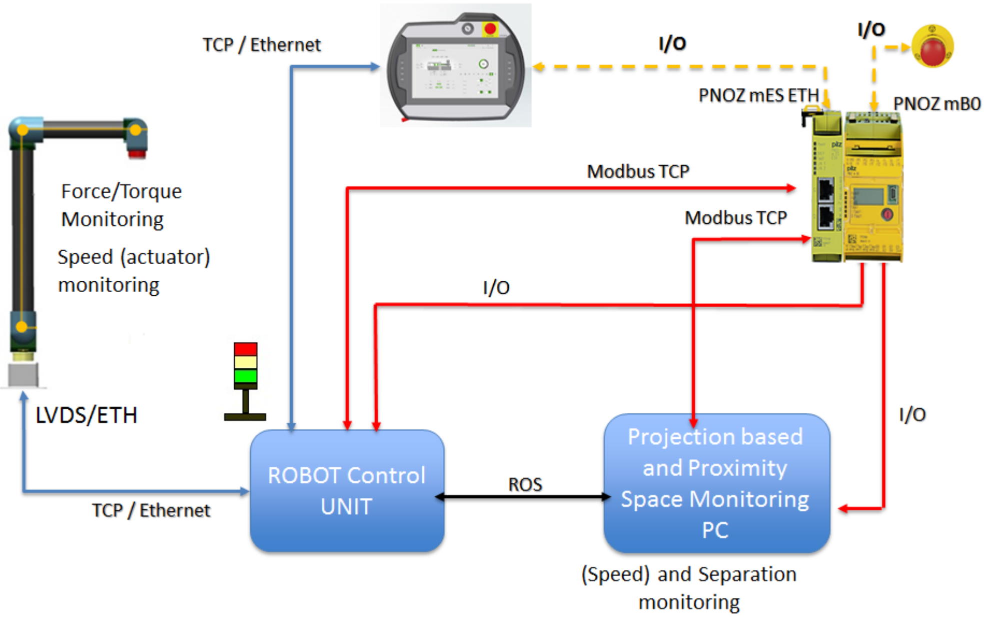


Figura 29: Arquitectura de seguridad implementada en FourByThree

4.2.4. Estudios experimentales sobre Confianza en un entorno de robótica colaborativa

La implementación de mecanismos y arquitecturas de seguridad conforme a la norma es una condición necesaria pero no suficiente para conseguir que los seres humanos acepten trabajar junto a un robot sin ninguna barrera física que los separe.

En el contexto de los proyectos X-ACT [2] y FourByThree [11] hemos desarrollado un conjunto de experimentos encaminados al estudio del nivel de aceptación por parte de los usuarios de las estrategias de seguridad propuestas.

4.2.4.1. Experimentación en condiciones de laboratorio en IK4-TEKNIKER

Utilizando el procedimiento descrito en la sección 4.1.3.1 y de forma simultánea a la experimentación en los aspectos relacionados con la interacción, se planteó un segundo objetivo:

Objetivo: analizar cómo perciben los trabajadores la estrategia de seguridad implementada que permite trabajar junto a un robot sin barreras físicas.

Resultados: Se resumen los datos obtenidos en los cuestionarios y fruto de la observación durante el experimento.

- Los participantes debían identificar su sistema de interacción preferido
  - Pulsador: siete participantes lo seleccionaron en primer lugar y seis en segundo. Ninguno dudó en la forma de usarlo.
  - Voz: cuatro en primera opción y tres como segunda.
  - Gestos: cuatro como opción preferida y cinco como segunda opción.
  - Interacción implícita: fue la peor valorada ya que el 53% de los participantes dudaron de que el robot hubiera interpretado la acción a realizar

  Aunque cinco participantes consideraban interesante la posibilidad de disponer de todos los sistemas simultáneamente; el resto lo consideraba una posible fuente de confusión

- El 53% consideraban interesante disponer de información sobre la acción a ejecutar en la pantalla situada en la zona de trabajo, si bien el 64% afirmaron que siguieron las instrucciones en todo momento. El 83% afirmaron que la información en el monitor ayudaba a realizar la actividad de forma segura.

- El sonido (un 'beep' que cambiaba de frecuencia a medida que la distancia con respecto al robot decrecía) no se consideró intrusiva.
- El 59% de los participantes prefirieron usar un número limitado de comandos de voz en lugar de usar lenguaje natural, uno prefirió lenguaje natural y para el resto (seis) era indiferente
- Ninguno de los participantes se sintió ridículo realizando gestos pero la mayoría (trece) mostraron dudas sobre la posibilidad de confundir los gestos a realizar (en efecto dos dudaron durante la primera sesión y cuatro durante la segunda), en particular si el número de comandos fuera más extenso (hasta 20 comandos). En el mismo sentido, sólo seis participantes consideraron posible 'recordar' más de 20 comandos

El resultado final más significativo fue que el 100% de los participantes consideraron el sistema fácil de usar (35% fácil y el 65% muy fácil). Estos resultados han sido tomados como referencia para implementar los mecanismos de interacción en el proyecto FourByThree.

4.2.4.2. Experimentación en la Bienal de Máquina Herramienta y feria Technishow

Tal y como se describe en la sección 4.1.3.2 se ha podido llevar a cabo una experimentación fuera de laboratorio, en el marco de dos ferias industriales en las que los visitantes han sido invitados a participar en el experimento, testear de primera mano las tecnologías y finalmente responder a un cuestionario para expresar sus opiniones. En conjunto 115 personas se ofrecieron voluntarias para participar en la experimentación.

Una descripción detallada de ambos experimentos y el análisis de los resultados obtenidos se encuentra disponible en el artículo [93].

Desde el punto de vista de seguridad la conclusión más relevante (ver *Figura 30*) es que las personas perciben el entorno colaborativo seguro que les lleva a admitir la posibilidad de trabajar junto a un robot colaborativo sin barreras físicas de separación.

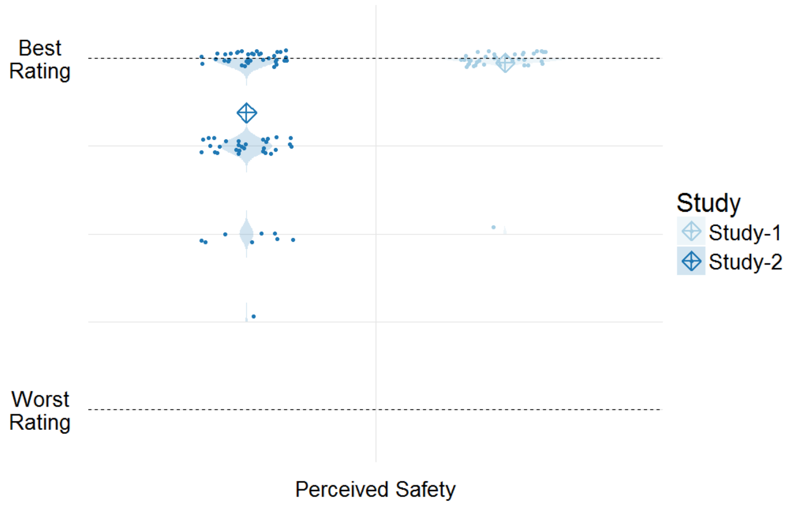


Figura 30: Percepción de seguridad en la experimentación

4.3. Flexibilidad y percepción

La contribución en el campo de la flexibilidad y la percepción se centra en dos aspectos fundamentalmente:

- Aplicación de la visión artificial para el control de movimientos de brazos robóticos
- Arquitecturas de control y tecnologías para la modularidad y flexibilidad

También se incluye una contribución en la inspección de defectos mediante técnicas visuales.

4.3.1. Aplicación de la visión artificial para el control de movimientos de brazos robóticos

Las aplicaciones de robótica colaborativa se caracterizan por la necesidad de que el robot se adapte a las condiciones cambiantes en el entorno, el carácter no necesariamente estructurado del mismo y la incertidumbre en el comportamiento de las personas. La visión artificial ofrece la posibilidad de abordar esa problemática de forma eficiente.

La contribución en esta línea ha permitido desarrollar soluciones para el control y generación de trayectorias a partir de la identificación de la posición de un objeto en el espacio (visual servoing), siendo por lo tanto extensibles a aplicaciones de robótica industrial avanzada.

Los proyectos que han permitido estos desarrollos son ROBOFOOT, EUROC-PIROS, EUROC-RSAII, MAINBOT y FOURBYTHREE.

Las publicaciones relacionadas se incluyen en el apartado de anexos (7.3, 7.4, 7.13, 7.17, 7.20 y 7.22 fundamentalmente).

En [96] y [97] se presenta una solución *look-and-move* de visual servoing para la gestión de la incertidumbre en la posición de zapatos en una línea de producción. Uno de los objetivos del proyecto ROBOFOOT era la introducción de la robótica en el sector de la fabricación del calzado introduciendo las mínimas modificaciones en la línea productiva y conviviendo con procesos manuales (tal y como se ejecutan en la actualidad). Por ello debía hacerse frente a la necesidad de manipular zapatos que venían simplemente apoyados en la cadena de producción (manovía en el argot del sector), tal y como se muestra en la ***Figura 31***.

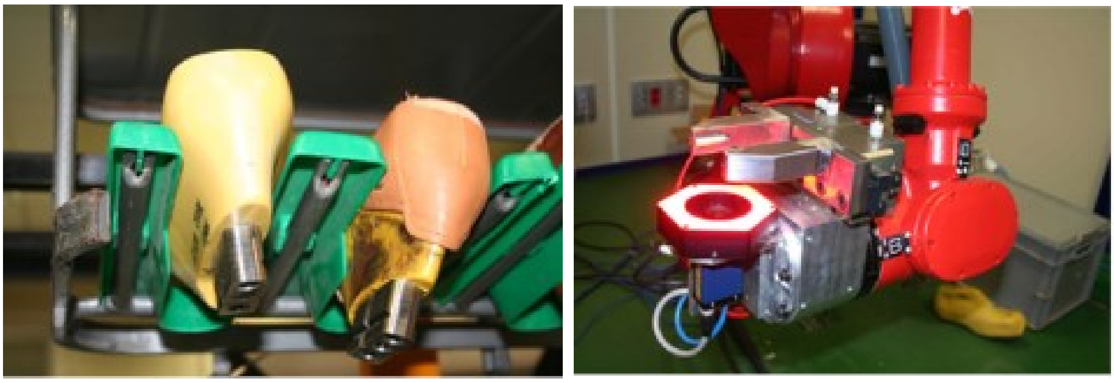


Figura 31: Zapatos apoyados en la manovía (izquierda) y configuración del sistema de visión y pinza (derecha)

Para mejorar la estimación de la posición del zapato, basada en la extracción de características significativas del mismo a partir de la imagen de una cámara embarcada en el robot, se ha usado un filtro de partículas. Este filtro bayesiano permite realizar una estimación precisa de la posición del objetivo (el zapato en este caso) a partir de una secuencia de imágenes. Para ello se obtienen N vectores de características, cada uno de ellos correspondiente a un análisis de imagen. Esos vectores representan una 'observación' que alimenta al filtro de partículas que contiene múltiples hipótesis de la posición del zapato. Una vez que el filtro ha sido alimentado por una serie de observaciones, es posible estimar la posición más probable o verosímil del zapato. El uso de este algoritmo permite el manejo de múltiples hipótesis de forma sencilla, permitiendo evitar los problemas asociados al ruido de sensores. Como resultado, el filtro de partículas proporciona la pose del zapato que será utilizada para generar el movimiento del robot.

En [98] se describe un sistema de tracking de un objetivo a partir de las imágenes obtenidas por un sistema de visión termográfica implementando un algoritmo basado en filtro de partículas. El sistema permite que un robot manipulador desplazándose por un terreno con irregularidades pueda realizar la inspección de 90 km de tubo colector en una planta termo-solar.

En [97] se presenta una aproximación basada en nubes de puntos para la estimación de la posición espacial de un objeto. En esa contribución se describe con detalle el proceso que consiste básicamente en una reducción del ruido en la nube de puntos, un filtrado espacial para reducir el campo de búsqueda, una clusterización de los puntos usando 'K-d tree' [99] y aplicación del algoritmo '*Iterative Closest Point*' para obtener la correspondencia con el modelo CAD de la pieza y finalmente estimar la posición espacial del objeto. La ***Figura 32*** presenta el modelo de CAD original, su transformación en nube de puntos y el resultado del best-fit con la nube real.

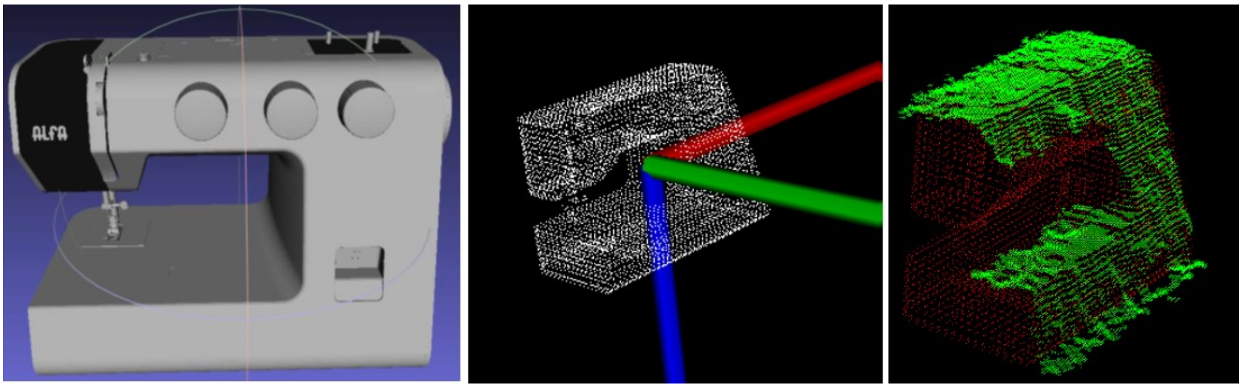


Figura 32: Modelo CAD, nube de puntos extraídos del CAD y best fit con la nube real

Finalmente, en el contexto de EUROC-PIROS se utiliza esta misma aproximación para la identificación de la posición espacial de una puerta de automóvil sobre la que el robot debe insertar un componente. En esta contribución [100] se incluye el proceso de reconstrucción de nubes de puntos parciales como paso previo a la aplicación del algoritmo de estimación de la posición.

4.3.2. Inspección de defectos mediante técnicas visuales

Las publicaciones relacionadas se incluyen en el apartado de anexos (7.2, 7.4, 7.12, 7.18 fundamentalmente).

En MAINBOT se desarrolló un Sistema de identificación de defectos en las instalaciones de una planta termosolar. El trabajo se centró en el análisis de imágenes termográficas para el análisis de la rotura de espejos y la detección de pérdidas de vacío y fugas en los tubos portadores del líquido termoportador. Para la detección de pérdidas de vacío, se desarrolló y validó un método que partiendo de la imagen termográfica, en la que cada pixel proporciona una información de temperatura, se normalizan dichos valores a la escala de grises y a continuación se identifica la línea correspondiente al colector. Sobre esta línea es posible identificar los cambios de temperatura y asociarlos con una pérdida de vacío.

4.3.3. Arquitecturas de control y tecnologías para la modularidad y flexibilidad

En este apartado se resumen las contribuciones en el diseño de una arquitectura de control que permite dotar a los robots de la flexibilidad. El trabajo se ha desarrollado fundamentalmente en el contexto del proyecto FourByThree y en menor medida en Mainbot.

En el proyecto Mainbot, la arquitectura debe dar respuesta a las necesidades de robots de servicio para el mantenimiento de plantas industriales extensas. Dadas las características de la problemática ha sido necesario diseñar dos tipologías de robots:

- Una plataforma móvil sobre ruedas con un brazo manipulador con capacidad de navegar por una planta de aproximadamente 400 hectáreas. El brazo es capaz de manipular varios sensores (reflectómetro y cámara termográfica) para llevar a cabo la inspección de diferentes parámetros, como pérdida de vacío, rotura de espejos y suciedad de los espejos [101].
- Un robot trepador con capacidad de 'escalar' sobre una torre termosolar ejecutando labores de inspección mediante visión y corrientes inducidas [102].

El diseño de la arquitectura [103], desarrollada sobre ROS, ha permitido diseñar misiones de inspección para dos tipologías diferentes de robot, simplificando su desarrollo y la integración con los sistemas de gestión del mantenimiento de la planta.

A partir de esa experiencia, los esfuerzos y la contribución han ido encaminados al desarrollo de una arquitectura de control para robots modulares y aplicaciones de robótica colaborativa en el contexto del proyecto FourByThree tal y como se recoge en diferentes publicaciones. Los resultados de dicha contribución se recogerán en una publicación en preparación y se adelantan en esta sección.

La arquitectura, basada en ROS, consta de los componentes de la ***Figura 33*** que se resumen a continuación.

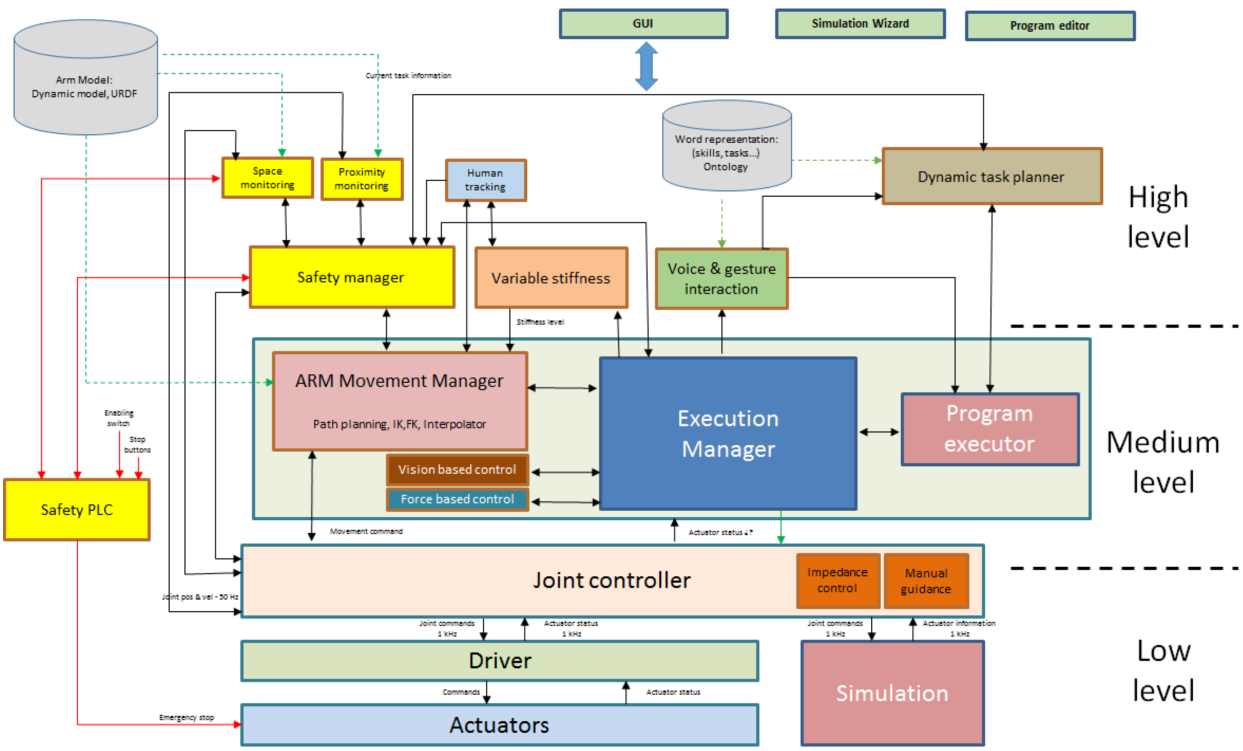


Figura 33: Arquitectura de control para robots modulares en aplicaciones colaborativas

- Low level

Incluye aquellos componentes cercanos al hardware de control del robot y está distribuida en la plataforma electrónica embebida en los propios actuadores y el sistema de control desarrollado en ROS. En este nivel se implementan aquellos servicios y funcionalidades más exigentes en términos de respuesta temporal (el control de impedancia y el guiado manual).

- Medium level

  Es el corazón de la arquitectura e incluye los módulos que permiten llevar a cabo la planificación y control de trayectorias, gestión y despacho de acciones para el brazo robótico y gestión y ejecución de programas.

  Es de destacar que se ha definido un lenguaje de programación, a semejanza de los robots industriales convencionales, con el ánimo de que en un futuro pueda ser adoptado por la comunidad robótica trabajando sobre ROS y ROS Industrial. Frente a los lenguajes de programación propietarios, el lenguaje se ha desarrollado con carácter 'universal' y abierto, pudiendo ser utilizable con otras plataformas robóticas, incluidas las comerciales siempre y cuando se disponga de un driver de ROS adecuado. En la ***Figura 34*** se presenta la gestión de programas en la arquitectura y en la ***Figura 35*** un ejemplo del lenguaje propuesto en su entorno de uso.

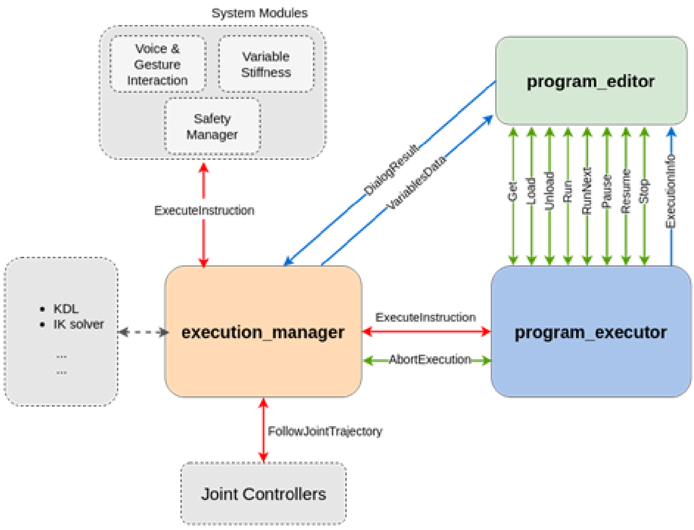


Figura 34: Gestión de programas en la arquitectura

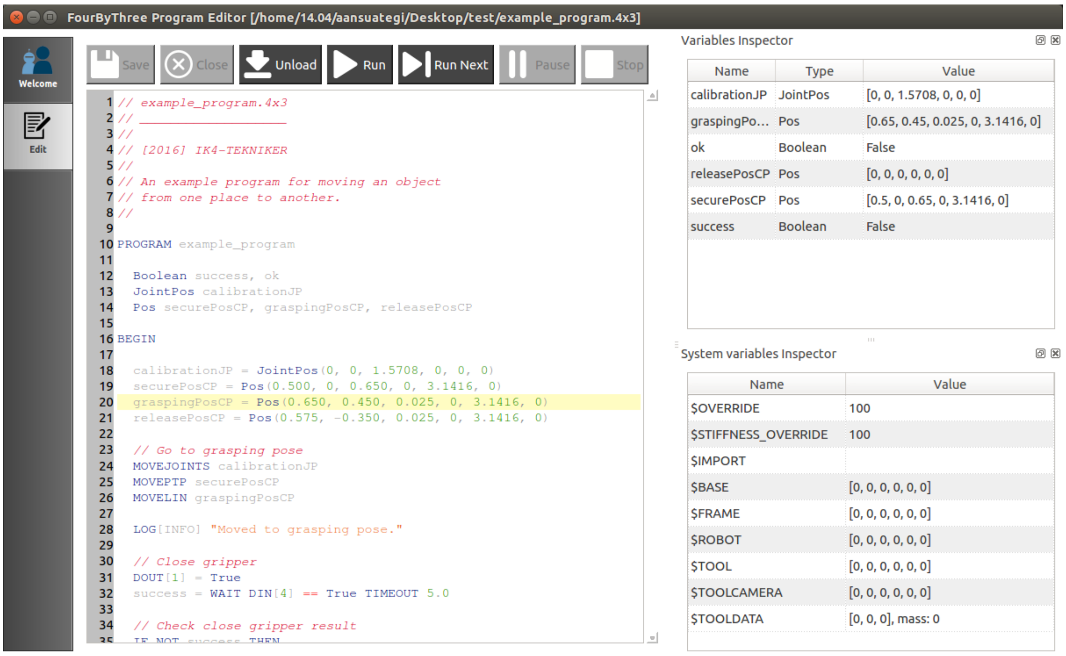


Figura 35: Ejemplo de programa 'universal'

- High level

  Se incluyen a este nivel el resto de componentes que permiten crear un entorno colaborativo, fundamentalmente los relacionados con la interacción, ergonomía, planificación dinámica e interface de usuario.

Cualquier nuevo componente desarrollado por terceros puede incorporarse de forma sencilla en esta arquitectura que, por otro lado, permite el control de diferentes cinemáticas como las definidas en el proyecto FourByThree, donde se han desarrollado robots de 4, 5 y 6 grados de libertad que comparten dicha arquitectura.

5. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se han presentado diferentes contribuciones encaminadas a facilitar el desarrollo de soluciones robóticas colaborativas fáciles de usar, flexibles y seguras.

Fáciles de usar mediante la utilización de tecnologías semánticas que permiten combinar dos mecanismos de **interacción**, los gestos y la voz, para dotar de la robustez necesaria para posibilitar su aplicación en entornos industriales (información redundante) y complementarse para resolver ambigüedades en determinadas ocasiones. La contribución incluye, además, el desarrollo de la tecnología necesaria para el reconocimiento de gestos.

La **seguridad** es una condición necesaria para la creación de entornos colaborativos y exige el desarrollo de tecnologías que posibiliten la implementación de los modos de colaboración definidos por la normativa existente. La contribución se ha centrado en la definición de arquitecturas y estrategias de seguridad, así como en el desarrollo de tecnologías que permiten implementar el modo SSM: el seguimiento de personas y la monitorización de proximidad. Además se ha experimentado con potenciales usuarios de la robótica colaborativa para conocer el grado de aceptación de las diferentes tecnologías desarrolladas, tanto para la seguridad como para la interacción. El resultado de esa experimentación ha servido, también, para adaptar ciertos desarrollos.

Sin embargo, sabemos que conseguir que un robot sea fácil de usar y seguro es condición necesaria pero no suficiente para facilitar el desarrollo de la robótica colaborativa en entornos industriales. Efectivamente, no podemos olvidar que un robot debe realizar el trabajo encomendado de forma eficiente y, para ello, debe saber adaptarse a las condiciones cambiantes del entorno (**flexibilidad**). La contribución aquí descrita se ha centrado en dos aspectos:

- La utilización de la visión artificial para potenciar las capacidades perceptivas de los robots, en particular para posibilitar que puedan identificar la posición de elementos en su entorno con los que tiene que interaccionar, así como para determinar su estado.
- El desarrollo de arquitecturas de control para el desarrollo de aplicaciones colaborativas.

Todos esos desarrollos se han logrado en el ámbito de proyectos de investigación y deben ser refrendados en condiciones de uso real.

- En el campo de la seguridad, la gran barrera es la certificación de los sistemas. El primer paso ha sido la presentación de una patente que permita proteger los desarrollos realizados y se deberá proceder a analizar de estrategia para conseguir su certificación.
- En el campo de la interacción, los desarrollos pueden ser aplicados para diferentes robots comerciales. Su validación dependerá en gran medida del avance real de las aplicaciones colaborativas en el mercado.
- En el caso de las tecnologías ligadas a la flexibilida, el proceso de validación industrial está en curso con varios proyectos de trasferencia industrial fundamentalmente en aplicaciones

de ‘bin-picking’.

La robótica colaborativa sigue ofreciendo retos que deberán ser abordados en el futuro. Las líneas de investigación abiertas van encaminadas a dar respuesta a algunos de esos retos:

- Interacción.

  En la implementación actual la identificación de los patrones morfosintácticos más habituales se realiza manualmente a partir de una base de datos de ejemplos. El objetivo es que dichas reglas puedan extraerse de forma automática aplicando técnicas de aprendizaje automático.

  El reconocimiento de gestos aquí presentado es válido únicamente para gestos estáticos, pero no para aquellos que exigen un movimiento para dotarlos de significado. Esa problemática será abordada en próximos proyectos.

  Por último debemos seguir trabajando en el concepto del robot como interface, es decir posibilitar utilizar ‘la piel’ del robot como elemento de comuncación, tanto de entrada (piel sensible) como de salida (permitiendo utilizarlo como sistema de visualización).

- Seguridad

  Un área de actividad interesante es la creación de metodologías que permitan facilitar el proceso de análisis de riesgos que, siendo obligatorios, dificultan el despliegue de robots colaborativos como elementos de producción flexibles (no podemos olvidar que en el análisis de riesgos se debe especificar la tarea a realizar por el robot, lo que entra en contradicción con la flexibilidad que se les exige).

  El sistema de detección de proximidad exige nuevos desarrollos para conseguir que conviva con condiciones del entorno físico cambiante, es decir que permita adaptar de forma dinámica el mapa 3D del modelo de dicho entorno.

  Un aspecto interesante es el desarrollo de tecnologías no intrusivas que permitan conocer el estado de la persona y sus intenciones, que además de mejorar la estrategia de seguridad, adaptando el comportamiento del robot, puede ser explotado para mejorar la interacción persona-robot.

- Percepción

  El trabajo a corto plazo se centra en mejorar el tratamiento de nubes de puntos que permita explotar las capacidades de los nuevos sensores en el mercado, desarrollando algoritmos con tiempos de respuesta propios de aplicaciones industriales. Igualmente interesante es el desarrollo de tecnologías que permitan realizar medidas de forma robusta y rápida a partir de las nubes de puntos para aplicaciones metrológicas y de control de calidad.

  Finalmente la utilización de las tecnologías de ‘deep learning’ abre un abanico de posibilidades para aplicaciones de detección de defectos.

Pero además es necesario seguir trabajando en los aspectos de divulgación de la tecnología no sólo en el ámbito académico y de la investigación sino contribuyendo a acercarlo al público en general, a las empresas y el mundo laboral.

Sección II Publicaciones y patentes

6. RESUMEN PUBLICACIONES Y PATENTES

En esta sección se presentan las publicaciones, eventos organizados y patentes que dan soporte a este trabajo. Se resaltan en negrita aquellas directamente relacionados con el presente trabajo.

El resto de los trabajos publicados no se incluyen en este documento ya que son, o bien resultados parciales, o resultado de otras actividades realizadas en proyectos que no están directamente relacionadas con los trabajos de investigación presentados en esta memoria.

6.1. Resumen de las publicaciones incluidas en esta memoria

<u>**Revistas**</u>

Iñaki Maurtua, Izaskun Fernández, Alberto Tellaeche, Johan Kildal, Aitor Ibarguren, Basilio Sierra. Natural Multimodal Communication for Human-Robot Collaboration. International Journal of Advanced Robotic Systems. 2017 [91]

Iñaki Maurtua, Aitor Ibarguren, Johan Kildal, Loreto Susperregi, Basilio Sierra. Human Robot collaboration in Industrial applications: safety, interaction and trust. International Journal of Advanced Robotic Systems. 2017 [93].

Aitor Ibarguren, Iñaki Maurtua, Miguel Angel Pérez, Basilio Sierra: Multiple target tracking based on particle filtering for safety in industrial robotic cells. Robotics and Autonomous Systems 72: 105-113 (2015) [94]

Torsten Felsch, Gunnar Strauss, Carmen Perez, José M. Rego, Iñaki Maurtua, Loreto Susperregi, Jorge R. Rodríguez: Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques. Robotics 4(2): 103-119 (2015) [102]

Aitor Ibarguren, José María Martínez-Otzeta, Iñaki Maurtua: Particle Filtering for Industrial 6DOF Visual Servoing. Journal of Intelligent and Robotic Systems 74(3-4): 689-696 (2014) [104]

Aitor Ibarguren, Jorge Molina, Loreto Susperregi, Iñaki Maurtua: Thermal Tracking in Mobile Robots for Leak Inspection Activities. Sensors 13(10): 13560-13574 (2013) [98]

Aitor Ibarguren, Iñaki Maurtua, Basilio Sierra: Layered Architecture for Real-Time Sign Recognition. The Computer Journal. 53(8): 1169-1183 (2010) [45]

Aitor Ibarguren, Iñaki Maurtua, Basilio Sierra: Layered architecture for real time sign recognition: Hand gesture and movement. Engineering Applications of Artificial Intelligence, 23(7): 1216-1228 (2010) [44]

Capítulos en Libros

Iñaki Maurtua. Wearable Technology in Automotive Industry: from Training to Real Production, Human-Computer Interaction, Chapter 4, Inaki Maurtua (Ed.), InTech, (2009) DOI: 10.5772/7742. Available from: https://www.intechopen.com/books/human-computer-interaction/wearable-technology-in-automotive-industry-from-training-to-real-production [105]

Maurtua I., Unceta M., Pérez M.A. (2007) Experimenting Wearable Solutions for Workers' Training in Manufacturing. In: Jacko J.A. (eds) Human-Computer Interaction. HCI Applications and Services. HCI 2007. Lecture Notes in Computer Science, vol 4553 (663-671). Springer, Berlin, Heidelberg [106]

Editor de libros

Human Machine Interaction - Getting Closer. Edited by Maurtua Inaki, ISBN 978-953-307-890-8, 270 pages, Publisher: InTech, Chapters published January 25, 2012 [107]

Human-Computer Interaction, Edited by Inaki Maurtua, ISBN 978-953-307-022-3, 570 pages, Publisher: InTech, Chapters published December 01, 2009 [108]

Congresos

Johan Kildal, Koray Tahiroglu, Juan Carlos Vasquez, Iñaki Maurtua. Studying Human-Robot Collaboration in an Artistic Creative Processes. 2017 Conference on Human-Robot Interaction (HRI2017), ReHRI'17 – International Workshop on reproducible HRI experiments: scientific endeavors, benchmarking and standardization. 2017, Viena [109]

Iñaki Maurtua, Izaskun Fernández, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibarguren: Enhancing safe human-robot collaboration through natural multimodal communication. ETFA 2016: 1-8 [110]

Iñaki Maurtua, Nicola Pedrocchi, Andrea Orlandini, Jose de Gea Fernandez, Christian Vogel, Aaron Geenen, Kaspar Althoefer, Ali Shafti: FourByThree: Imagine humans and robots working hand in hand. ETFA 2016: 1-8 [111]

Iñaki Maurtua, Loreto Susperregi, Ander Ansuategui, Ane Fernández, Aitor Ibarguren, Jorge Molina, Carlos Tubio, Cristobal Villasante, Torsten Felsch, Carmen Pérez, Jorge R. Rodriguez,

and Meftah Ghrissi: Non-destructive inspection in industrial equipment using robotic mobile manipulation, AIP Conference Proceedings 1734, 130013 (2016) [101]

Alberto Tellaeche, Iñaki Maurtua, Aitor Ibarguren: Use of machine vision in collaborative robotics: An industrial case. ETFA 2016: 1-6 [100]

Iñaki Maurtua, Izaskun Fernandez, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibarguren: Interacting with collaborative robots in industrial environments: A semantic approach. ICAPS 2016 Workshop on "Planning, Scheduling and Dependability in Safe Human-Robot Interactions", 2016, Londres. [112]

Revisiting the end user's perspective in collaborative human-robot interaction: Proceedings of the 19th International Conference on CLAWAR 2016. In book: Advances in Cooperative Robotics, pp.196-204 [113]

Alberto Tellaeche, Iñaki Maurtua, Aitor Ibarguren: Human robot interaction in industrial robotics. Examples from research centers to industry. ETFA 2015: 1-6 [114]

Alberto Tellaeche, Iñaki Maurtua: 6DOF pose estimation of objects for robotic manipulation. A review of different options. ETFA 2014: 1-8 [97]

Iñaki Maurtua, Loreto Susperregi, Ane Fernández, Carlos Tubio, Torsten Felsch, Carmen Pérez, Jorge R. Rodriguez, and Meftah Ghrissi: MAINBOT – Mobile Robots for Inspection and Maintenance in Extensive Industrial Plants. Energy Procedia, Volume 49, Pages 1-2532 (2014), Proceedings of the SolarPACES 2013 International Conference, Las Vegas [103]

Iñaki Maurtua, Aitor Ibarguren, Alberto Tellaeche: Robotic solutions for Footwear Industry. ETFA 2012: 1-4, Cracovia [115]

Aitor Ibarguren, José María Martínez-Otzeta, Iñaki Maurtua: Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments. ICINCO, 9th International Conference on Informatics in Control, Automation and Robotics (2) 2012: 161-166. Roma [96]

Iñaki Maurtua, Aitor Ibarguren, Alberto Tellaeche: Robotics for the Benefit of Footwear Industry. ICIRA, 5th International Conference on Intelligent Robotics and Applications (2) 2012: 235-244, Montreal [116]

Alberto Tellaeche, Ramón Arana, Iñaki Maurtua: Accurate Correction of Robot Trajectories Generated by Teaching Using 3D Vision by Laser Triangulation. ICIRA, 5th International Conference on Intelligent Robotics and Applications (3) 2012: 385-394, Montreal [117]

Loreto Susperregi, Izaskun Fernández, Ane Fernandez, Santiago Fernandez, Iñaki Maurtua, Irene Lopez de Vallejo: Interacting with a Robot: A Guide Robot Understanding Natural Language Instructions. UCAmI, 6th International Conference on Ubiquitous Computing and Ambient Intelligence, 2012: 185-192 [118]

Maria Isabel de la Fuente, Javier Echanobe, Inés del Campo, Loreto Susperregui, Iñaki Maurtua: Hardware Implementation of a Neural-Network Recognition Module for Visual Servoing in a Mobile Robot. DEXA 21st International Conference on Database and Expert Systems Applications 2010: 226-232. Bilbao [119]

Maria Isabel de la Fuente, Javier Echanobe, Inés del Campo, Loreto Susperregui, Iñaki Maurtua: Development of an Embedded System for Visual Servoing in an Industrial Scenario. SIES 2010: 192-196, Trento [120]

Iñaki Maurtua, Pierre T. Kirisci, Thomas Stiefmeier, Marco Luca Sbodio, Hendrik Witt. A Wearable Computing Prototype for supporting training activities in Automotive Production. 4th International Forum on Applied Wearable Computing 2007. Tel Aviv [121]

Miguel Ángel Pérez, Loreto Susperregi, Inaki Maurtua, Aitor Ibarguren, and Basilio Sierra. Software Agents for Ambient Intelligence based Manufacturing. In IEEE Workshop on Distributed Intelligent Systems, pages 139–144, Prague, Czech Republic, 2006. [122]

Loreto Susperregi, Inaki Maurtua, Carlos Tubío, Inigo Segovia, Miguel Ángel Pérez, and Basilio Sierra. Context aware agents for Ambient Intelligence in Manufacturing at Tekniker.. AgentLink, 18:28–30, August 2005. [123]

Irene Lopez de Vallejo, Iñaki Maurtua, Miren Unceta. Ambient Intelligence in Manufacturing: Organizational Implications. Workshop on 'Ubiquitous Computing and effects on Social Issues' Portland, Oregon. 2005 [124]

Loreto Susperregi, Iñaki Maurtua, Carlos Tubío, Iñigo Segovia, Miguel Angel Pérez, and Basilio. An agent based ambient intelligence experience in manufacturing. Ambient Intelligence and (Everyday) Life International Workshop, pp. 191–204, (Donostia-San Sebastian, Spain), 2005. [125]

6.2. Organización de Workshops

- **IROS 2015, 2 Octubre 2015, Hamburgo, Alemania**

  Full-Day Workshop: Safety for Human-Robot Interaction in Industrial Settings

  Organizadores: Kaspar Althoefer, Iñaki Maurtua, Hongbin Liu, Helge A Wurdemann, José de Gea Fernández

- **ICAPS2016, International Conference on Automated Planning and Scheduling, 13 Junio**

2016, Londres, UK

SAFEPLAN Workshop on "Planning, Scheduling and Dependability in Safe Human-Robot Interactions"
Organizadores: Ali Shafti, Kaspar Althoefer, Helge A.Wurdemann, Amedeo Cesta, AndreA Orlandini, and Iñaki Maurtua

- ERF16, 2016 European Robotic Forum, 23 Marzo 2016, Ljubljana, Slovenia

  3rd Workshop on Hybrid Production Systems

  Organizadores: Makris Sotiris, George Michalos, Iñaki Maurtua, Ramez Awad

- ETFA16, 22nd IEEE International Conference on Emerging Technologies And Factory Automation, 6 de Septiembre 2016, Berlin, Germany

  Special session: SS03. Safe Human-Robot Collaboration

  Organizadores: Kaspar Althoefer, Amedeo Cesta, Iñaki Maurtua, Andrea Orlandini, Ali Shafti,

- CLAWAR16, 19th International Conference on Climbing and Walking Robots and Support Technologies for Mobile Machines. 12 Septiembre 2016, Londres, UK

  Workshop on Collaborative Robots for Industrial Applications

  Organizadores: Roger Bostelman, Roger Eastman, Ali Shafti, Helge A. Wurdemann, Kaspar Althoefer, Iñaki Maurtua

- ERF17, 2017 European Robotic Forum, 23 Marzo 2016, Edimburgo, UK

  4rd Workshop on Hybrid Production Systems

  Organizadores: Makris Sotiris, George Michalos, Iñaki Maurtua, Ramez Awad

- ETFA17, 22nd IEEE International Conference on Emerging Technologies And Factory Automation 12-15 Septiembre Limassol, Chipre

  Special Session SS07. Safe Human-Robot Collaboration

  Organizadores: Amedeo Cesta, Iñaki Maurtua, Andrea Orlandini, Nicola Pedrocchi, Stefania Pellegrinelli.

6.3. Patentes

MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD

7. PUBLICACIONES

7.1. Natural Multimodal Communication for Human-Robot Collaboration

Iñaki Maurtua, Izaskun Fernández, Alberto Tellaeche, Johan Kildal, Aitor Ibarguren, Basilio Sierra. Natural Multimodal Communication for Human-Robot Collaboration. International Journal of Advanced Robotic Systems. 2017 [91]

Journal Title
XX(X):1–11

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Natural Multimodal Communication for Human-Robot Collaboration

Iñaki Maurtua[1], Izaskun Fernandez[1], Alberto Tellaeche[1], Johan Kildal[1], Aitor Ibarguren[2] and Basilio Sierra[3]

Abstract

This paper presents a semantic approach for multimodal interaction between humans and industrial robots to enhance the dependability and naturalness of the collaboration between them in real industrial settings. The fusion of several interaction mechanisms is particularly relevant in industrial applications in which adverse environmental conditions might affect the performance of vision-based interaction (e.g., poor or changing lighting) or voice-based interaction (e.g., environmental noise). Our approach relies on the recognition of speech and gestures for the processing of requests, dealing with information that can potentially be contradictory or complementary. For disambiguation, it uses semantic technologies that describe the robot characteristics and capabilities as well as the context of the scenario. Although the proposed approach is generic and applicable in different scenarios, this paper explains in detail how it has been implemented in two real industrial cases in which a robot and a worker collaborate in assembly and deburring operations.

Keywords

Safe Human-Robot Collaboration, Collaborative robots, Multimodal Interaction, Natural Communication, Semantic Web Technologies, Reasoning, Fusion

Introduction

In designing the factories of the future, achieving safe and flexible cooperation between robots and human operators is considered as a way of enhancing productivity. The problem of robots performing tasks in collaboration with humans poses the following main challenges: robots must be able to perform tasks in complex, unstructured environments and, at the same time, they must be able to interact naturally with the workers they are collaborating with, while guaranteeing safety at all times.

The present work describes a natural communication approach between person and robot, as developed within the *FourByThree* European project (FBT). This project aims to create a new generation of modular industrial robotic solutions that are suitable for efficient task execution in safe collaboration with humans, at the same time as being easy to use and program. This work will allow system integrators and end users to develop custom robots that best answer their needs. To achieve this, the project provides a set of hardware and software components, ranging from low level control to interaction modules. The outcome from the project will be validated in 4 industrial settings: Investment Casting, Aeronautical sector, Machining and Metallic Part Manufacturing, in which industrially-relevant applications are being implemented for human-robot collaboration: assembly, deburring, welding, riveting and machine tending.

A requirement for natural human-robot collaboration is to endow the robot with the capability to capture, process and understand accurately and robustly human requests. Using voice and gestures in combination is a natural way for humans to communicate with other humans. By analogy, they can be considered equally relevant to achieve natural communication also between humans and robots. In such a multimodal communication scenario, the information coming from the different channels can be complementary or redundant, as shown in these examples:

- Complementary: a worker saying '*Take this*' while pointing at an object.
- Redundant: a worker saying '*Stop!*' while performing a gesture by lifting a hand and showing the palm.

In the first example, the need for different communication channels complementing each other is evident. However, redundancy can also be beneficial (Bannat et al. (2009)) in e.g. industrial scenarios in which noise and variable lighting conditions may reduce the robustness of each channel when considered independently.

In this paper, we present a semantic approach that supports the multimodal interaction between humans and industrial robots, in real industrial settings. Our approach relies on the recognition of verbal commands and gestures for requests processing, first independently, and later fusing

[1]IK4-Tekniker, Spain
[2]Tecnalia, Spain
[3] RSAIT Group, University of the Basque Country, Spain

Corresponding author:
Iñaki Maurtua, IK4-Tekniker Parke Teknologikoa, Iñaki Goenaga, 5 20600, Eibar Spain
Email: imaurtua@tekniker.es

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

both channels, managing information that can potentially be contradictory or complementary. Semantic technologies are used to describe characteristics of collaborative robots as well as scenario contexts, being a key component for request interpretation.

This generic approach can be applied to different industrial scenarios by modifying the information about the environment in which the communication takes place. In this paper, we validate this approach in the scenario described in the section bellow. We continue with a brief review of previous work and the description of the main components used in our approach. After describing the performed test and evaluation, the conclusions and planned future work are presented.

Case Study

As explained before, the FourByThree project includes four industrial scenarios of human-robot collaboration. For an initial validation of our semantic multimodal interaction approach, we selected an application scenario at an Investment Casting process at the Spanish manufacturing company 'ALFA'. This scenario includes two collaborative tasks: a die-assembly task (involving screwing and unscrewing operations) and the deburring of wax pieces as the second task.

- **Assembly task**: the human worker and the robot worked independently (un)screwing bolts on the four different parts of a rectangular die. The job was shared between person and robot by assigning two of the sides to each of them, as decided by the person. The human worker had to specify which two of the four sides of the die the robot had to operate on. Due to the proximity between the worker and the part, performing a pointing gesture was not considered feasible. Thus, the worker specified the side by voice only.
- **Deburring task**: the human and the robot performed sequential tasks on the same workpiece in a synchronized manner. While the person glued parts and removed burrs that were difficult to access, the robot performed the rest of the deburring. In this case, the person provided instructions to the robot about where to deburr using a combination of speech and gesture.

To accommodate language requirements from the end users, we implemented the speech processing for Spanish.

Related work

Over the last two decades, a considerable number of robotic systems have been developed that include Human-Robot Interaction (HRI) capabilities (Fong et al. (2003); Goodrich and Schultz (2007)). Although recent robot platforms integrate advanced human-robot interfaces (incorporating body language, gestures, facial expressions and speech) (Stiefelhagen et al. (2004); Burger et al. (2010)) their capability to understand human speech semantically remains rather limited. Endowing a robot with semantic understanding capabilities is a very challenging task, although an important one. For instance, previous experiences with tour-guide robots (Thrun et al. (1999); Gunhee et al. (2004)) showed the importance of improving natural human-robot interaction in order to ease the acceptance of robots by visitors. In Jinny's HRI system (Gunhee et al. (2004)), voice input was converted to text strings, which were decomposed into several keyword patterns and a specialized algorithm found the most probable response for that input. With this technique, two questions like 'Where is the toilet?' and 'Where can I find the toilet' are interpreted in the same way, since keyword patterns that included 'where' and 'toilet' are extracted from both cases.

Gesture and posture recognition is commonly supported by 3D cameras (stereo vision) and other sensors providing Point Clouds (e.g., LEAP motion) and 2D color cameras (Suarez and Murphy (2012)). Other wearable approaches, based on e.g. accelerometers and gyroscopes are also used for very specific applications (Xse, Myo). Fang et al. (2017) present a data glove which can capture the motion of the arm and hand by inertial and magnetic sensors. The proposed data glove is used to provide the information of the gestures and teleoperate the robotic arm-hand. The use of vision based gesture analysis has the advantage of not needing the user to wear any additional sensor. Gesture information is derived from e.g. depth segmentation and skeleton tracking (Wang et al. (2012)), 3D sensing or from color segmentation (detection of hands, head or tags) Yan et al. (2002) with 2D.

Human-robot natural interactions have also been researched in industrial scenarios. For instance, Bannat et al. (2009) introduced an interaction that consisted of different input channels such as gaze, soft-buttons and voice in an industrial scenario. Although voice constituted the main interaction channel in that use scenario, it was solved by command-word-based recognition.

SHRDLU is an early example of a system that was able to process instructions in natural language and perform manipulations in a virtual environment (Winograd (1971)). Later on, researchers extended SHRDLU's capabilities to real world environments. Those efforts branched out into tackling various sub-problems, including Natural Language Processing (NLP) and Robotics Systems. Notably, MacMahon et al. (2006) and Kollar et al. (2010) developed methods to follow route instructions provided through natural language. Tenorth et al. (2010) developed robotic systems capable of inferring and acting upon implicit commands using knowledge databases. A similar knowledge representation was proposed by Wang and Chen (2011) using semantic representation standards such as the W3C Web Ontology Language (OWL) to describe an indoor environment.

Recent research projects in different fields (Dobrišek et al. (2013), Liu et al. (2017a), Zheng et al. (2014), Liu et al. (2017b)) show how combining different interaction channels outperforms the accuracy and robustness of the interaction with respect to the individual use of each channel. Dobrišek et al. (2013) show how audio and video combination for emotion recognition increase the performance to 77.5%, increasing in 5% with respect to the best of the individual channels, i.e audio. Liu et al. (2017a) uses visual information in combination with information from tactile sensors to capture multiple object properties, such as textures, roughness, spatial features, compliance and friction, bridging the gap for objects that are not visually distinguishable, leading to better recognition results without any relevant drawback in execution time.

Rossi et al. (2013) addressed a multi-channel fusion problem aimed for a robust communication between human and robot. In the case study that they presented, they described a generic and extensible architecture that included gesture and voice recognition, plus a fusion engine to improved the robustness of the interpretation of the human-robot communication. Their implementation was able to recognise 7 different gestures, and it was based on Hidden Markov Models. Voice recognition is frame based, where a simple action or a directed action and its corresponding target object are extracted from the instructions conveyed in speech. The fusion engine is a SVM-based model that delivers the final output considering both input channels, gesture and voice. Their evaluation showed that interaction accuracy increased when combining both inputs (91%) instead of using them individually (56% in the case of gestures and 83% for voice). Furthermore, the average time for processing both channels was similar to the time needed for speech processing.

Our work is based on this extensible architecture approach, combining gesture and speech channels. However, unlike in Rossi et al. (2013), we add semantics both for voice interpretation and for the fusion step. With this, we aim to ensure the coherence within each channel and in the final combination of both, thus avoiding logical contradictions as well as taking advantage of complementary information.

As stated by Liu and Kavakli (2010) in their survey, it is possible to identify three main levels of fusion strategies: signal level fusion (signals of different input devices are merged), feature level (early fusion that concatenates the feature vectors from multiple modalities to obtain a combined feature vector which is then used for the classification task) and the last level that includes several different features and is sometimes named as decision-level fusion, late-fusion or even semantic level fusion where each input channel is treated independently and the output of each of those interpretations integrated sequentially.

According to Atrey et al. (2010), at the feature-level fusion, the features extracted may be extracted from different modalities such as visual, audio, text, motion or metadata. This approach presents the advantages that it makes the use of correlation among different sources of data available and it requires only one learning phase for the combined features. However it also presents the important drawback that it makes the time synchronization between the features very hard to handle. The decision-level fusion, is the most extended one in the field of Human Machine interaction, due to its easy scalability.

Different variants of the late-fusion method have been proposed, being the following three categories the most relevant ones: Rule based, as described in Holzapfel et al. (2010) and Burger et al. (2012); Estimator based, like in Wagner et al. (2011); and Classification based, as described by Kessous et al. (2010) and Wu et al. (2004). In the latter, Support Vector Machines are used as a super-kernel algorithm to fuse the decisions of the previous individual classifiers. In the rule based case, in contrast, feature level fusion and decision level fusion are compared, both based on Bayesian classifiers. In the feature level approach, the Bayesian classifier is used for the fusion itself, while in the decision level a Bayesian classifier is modelled for each channel interpretation and their outputs are later fused based on an estimation strategy. Both fusion strategies outperform significantly the mono-channel systems performances (the best with 67,1%) reaching up to 78,3%.

In Ngiam et al. (2011) deep networks to learn features over multiple modalities are presented, building Deep bimodal representations by modelling the correlations across the learned shallow representations.

Multimodal Interaction Semantic Approach

The approach proposed in this work contributes to create a safe human-robot collaborative environment in which interactions between both actors happen in a natural way, i.e. communication based on voice and gestures in this case. Some examples of this natural interaction could be: human operator asking the robot to complete a certain task; robot asking for clarification when a request is not clear; in some specific scenarios in which human intervention is necessary during an automatized robot task execution, the robot asking the operator to complete some tasks, and once that is done, the operator informing the robot that it can resume its task. This natural communication facilitates the coordination between both actors, enhancing the safe collaboration between robot and human.

To address such a natural communication, we propose a semantic multimodal interpreter that is able to handle voice and gesture-based natural requests from a person, and combine both inputs to generate an understandable and reliable command for industrial robots, facilitating a safe collaboration.

For such a semantic interpretation, we have developed four main modules, as it is shown in Figure 1: a *Knowledge-Manager* module that describes and manages the environment and the actions that are affordable for robots, using semantic representation technologies; a *Voice Interpreter* module that, given a voice request, extracts the key elements on the text and translates them into a robot-understandable representation, combining NLP and semantic technologies; a *Gesture Interpretation* module to resolve pointing gestures and some simple orders like stopping an activity (out of the scope of the work presented in this paper); and a *Fusion Engine* for combining both mechanisms and constructing a complete and reliable robot-commanding mechanism.

These modules are described in detail in the following subsections.

Knowledge Manager

The knowledge manager uses an ontology to model the environment and the robot capabilities, as well as the relationships between the elements in the model, which can be understood as implicit rules that the reasoner exploits to infer new information. Thus, the reasoner can be understood as a rule engine in which human knowledge can be represented as rules or relations. Ontologies are reusable and flexible at adapting to dynamic changes, thus avoiding to have to re-compile the application and its logic whenever a

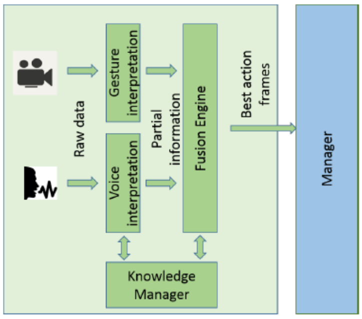


Figure 1. Multimodal semantic approach architecture

change is needed. Being in the cloud makes ontologies even more reusable, since different robots can exploit them, as it was demonstrated in (Di Marco et al. (2013)).

Through ontologies, we model the industrial scenarios in which robots collaborate with humans. The model includes robot behaviors, actions they can accomplish and the objects they can manipulate/handle. It also considers features and descriptors of these objects. We distinguish two main kinds of actions: actions that imply a change on the status of the robot operation, e.g., *start* or *stop*, and actions involving the robot capabilities, e.g., *screw*, *carry* or *deburring*. This is shown in Figure 2.

Using OWL equivalentClass built-in property, *ActionOverPoint* and *ActionOverObject* classes have also been defined. Both are *ModeAction* sub-classes, with the following restrictions: for a start, the action must be related to a point (defined by *hasEffectOnPoint min 1 Point* restriction in the ontology). In addition, the action must be related to an object (defined by *hasEffectOnObject min 1 Object* restriction in the ontology). This way, applying a semantic reasoner able to interpret this OWL statements, it will infer that a *ModeAction* instance like *deburring* with *hasEffectOnObject burr* property also belongs to the *ActionOverObject* class. Same inference effect for *ModeAction* instances with *hasEffectOnPoint point1* property defined, that will be also considered instances of *ActionOverPoint*.

For each individual action or object, we include *tag* property data for listing the most common expression(s) used in natural language to refer to them, including reference to the language used. An automatic semantic extension of those tags exploiting Spanish WordNet (Gonzalez-Agirre et al. (2012)) is done at initialization time. In this way, we obtain different candidate terms referring to a certain concept.

Besides task/programs and objects, the ontology also includes relations between the concepts, as it is shown in Figure 3. These relations are used by the interpreter for disambiguation at run-time. This ability is very useful for text interpretation because sometimes the same expression can be used to refer to different actions. For instance, people can use the expression *remove* to request the robot to *remove a burr*, but also to *remove a screw*, depending on whether the desired action is *deburring* or *unscrewing*, respectively. If the relationships between the actions and the objects over which the actions are performed are known, the text interpretation is more accurate; it will be possible to discern in each case to which of both options the expression *remove* corresponds. Without this kind of knowledge representation, this disambiguation problem is more difficult to solve. These relations are formally modeled in the ontology as it is shown in Figure 3.

For our current implementation, the two contexts described in the Case Study section have been considered. The possible tasks the robot can fulfill in both scenarios have been identified and a knowledge base (KB) created, populating the knowledge manager ontology with instances representing those tasks. The knowledge base also includes the elements that take part in both processes as instances of *Object* and *Point* classes, as well as the relationships they have with respect to the tasks. This knowledge base is published in StarDog 4.0.5 Community version (Stardog) and extended with WordNet as explained before.

In this way, the semantic representation of the scenarios will be available to support the request interpretation process, not only to infer which is the desired action to perform, but also to ensure that all the necessary information is available and coherent in order to be possible for the robot to perform it. Furthermore, the information is coherent as a whole. For instance, when someone ask to the robot via voice to *Remove the bolts from that piece*, pointing at a position where a piece without bolts is placed, the multi-channel interpreter will first use the semantic representation as explained before to select the action that best matches among the feasible ones described in the knowledge base (unscrew), thus taking advantage of the semantic relations between actions, working-pieces and the objects in scope. Once the piece pointed at is recognised through vision, and after fusing channels, the interpreter will (in the case of this example) decline the petition due to incoherence, since performing unscrewing actions on a piece without bolts is impossible. This kind of apparently evident incoherence detection, requires of the semantic representation of the scenarios to also be detectable by machines.

Voice Interpreter

Given as input a human request in which a person indicates the desired action via voice, the purpose of this module is to understand exactly what the person wants the robot to do and, if the information is complete, to generate the corresponding command for the robot. For instance, if a worker says *Remove the burrs from there*, the voice interpreter should interpret that the verb *remove* corresponds to the deburring action and check if it is a feasible action in the current collaborative robot application. If affirmative, it should then generate the necessary information for the robot to execute the request.

For such an interpretation, the module follows three main steps: an initial speech recognition step that deals with the voice-to-text transcription; following, a rule-based step for the extraction of key elements from the transcribed text; and finally, the matching between the key elements and the tasks that is feasible for the robot, based on the KB.

The speech recognition step is based on the Google Speech API (GoogleApi). A recorded audio file including the request of the operator is sent to the Google Speech API, which returns the corresponding text.

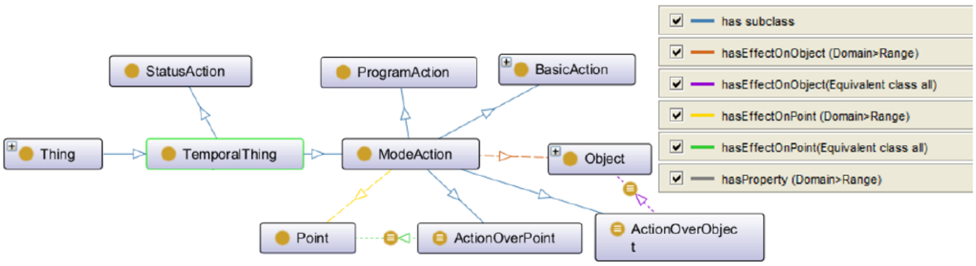


Figure 2. An excerpt of the Knowledge Manager Ontology

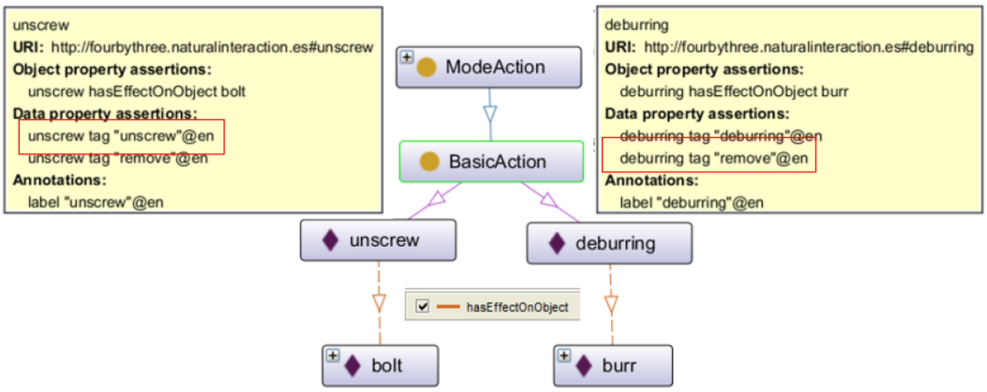


Figure 3. An excerpt of the FourByThree Semantic Interpreter Knowledge Base: Disambiguation

For the second step (extraction of the key elements from the text transcribed in the previous step), Natural Language Processing techniques are used. The main idea is to use syntactic information by means of rules for the extraction of key elements. In our current implementation, the Spanish version of FreeLing (an open source suite of language analysis tools (Padró and Stanilovsky (2012))) is used for this structural analysis.

For the definition of rules, use Freeling for the morphosyntactic analysis and dependency parsing of a set of request examples obtained from different persons. We revise the complete information manually and identify the most frequent morphosyntactic patterns that are relevant for extracting the key elements: elements denoting actions, objects/destinations (target onward) and explicit expressions denoting gestures, such as *there* and *that*. Finally, we implement those patterns as rules. In this way, given a Spanish FreeLing-tagged sentence, it is possible to extract the key elements present in it during execution. In the future, it will be possible to use this process to extend the interpreter to other languages.

Once the key elements are extracted, it is necessary to identify which one of the tasks that the robot is able to perform suit the request best. We undertake this last step by making use of the KB information described above. First, we verify if the identified actions were among the feasible tasks described in the KB, accessing the tag data property of the actions in the KB by using the semantic query language, SPARQL (Harris et al. (2013)). Then, we apply a disambiguation step using the target information, as explained before (taking advantage of the OWL logic inference capabilities). The final output from the voice interpreter consists of frames, one for each potential task candidate, including information denoting gestures, if any exists.

As an example, for the request in Spanish *Quita las rebabas (Remove the burrs)* the system would check, via a Sparql query, which actions in the KB contained *tag* data property with the value *quitar (remove)*. This would return *unscrew* and *deburring* as potential actions. In the following step in which key elements were extracted, *burrs* would be recognized as target element. This element would be checked back with the KB via Sparql in order to see if any of the two candidate tasks was related to it, which would return *deburring* as directly related. As a result, the module would discard *unscrew* as potential action, leaving *deburring* as the only eligible potential task.

The following example illustrates yet another instance in which the disambiguation capability could be exploited. Let us consider these two similar but different requests: *Acércate* (*come here*) and *Acerca la caja* (bring the box closer). In both sentences, the element to look for among *tag* data properties of the tasks represented in the KB would be *acercar* (the infinitive form of the verb), which would match with the tasks *go* and *bring*. As it is shown in Figure 4, both belong to the *ActionOverPoint* class, while the latter also belongs to the

ActionOverObject class. Making use of this information (and although *box* is not explicitly represented in the KB, it would have been identified as the target element in the previous step), the interpreter would be able to infer that *bring* fitted as a task better with this request than *go*.

If the verb did not match any of the tasks in the KB, or if no verb was identified in the request but some target or pointing information had been extracted in the preceding step, the interpreter was able to generate task suggestions, based on the relations and equivalences defined in the KB. For instance, if *tornillo* (*bolt*) was the only information extracted from the voice request, the interpreter would be able to generate *screw* and *unscrew* as potential task candidates. With this information, the robot could potentially request the worker to help with the disambiguation process.

For the sake of illustration, Figure 5 presents the example of asking a robot to deburr a piece starting from a certain point (*Remove burrs starting here*).

Gesture Interpretation

Two kinds of gestures are addressed within the *FourByThree* project: pointing gestures and gestures for simple commands such as stop or start. This paper only deals with pointing gestures that are recognized by means of point-cloud processing. The system is able to handle different pointing gestures that are performed within a certain temporal window, providing the x,y,z coordinates of the target position.

The setup used consists of a collaborative robot and a sensor capable of providing dense point clouds, such as the ASUS Xtion sensor, the Microsoft Kinect sensor or the industrial-grade Ensenso system by IDS. The camera is placed pointing towards the working area of the robot in the region above the human operator. Thanks to this configuration, the point cloud obtained contains information about the working area and part of the human body (in particular, the arm and hand used to perform the gesture). (see Figure 6).

Two cuboid regions of interest (ROIs) are then defined in the point cloud, one for the operator's pointing gesture detection(which essentially includes his/her forearm) and a second ROI where the intersection area (i.e., the target point), has to be identified.

With this setup, two main problems need to be solved for the interaction between the person and the robot to succeed:

Robust estimation of the pointing gesture

The ROI for the pointing gesture detection is initially defined by specifying in the environment a cuboid with respect to the reference frame. In this case, the reference frame is the sensor frame, but it can also be defined using another working frame, provided a tf transformation exists between the frame used and the sensor frame. For robustness, the pointing gesture is defined using the forearm of the human operator. To identify the arm unequivocally, an Euclidean cluster extraction is performed.

Intersection of the pointing gesture with the working area of the robot

The main objective of a pointing gesture is to determine the point on the working area that is being pointed at. To identify this point, the points in the cloud corresponding to the pointing line are selected, from the furthest one all the way to the origin of the line that corresponds to the pointing arm. For each one of the points, a small cuboid is defined, and the ROI of the working area of the robot is filtered with it. If more than N points of the working area are present inside the small centered cuboid defined in the points of the projection line, an intersection has been found. The final intersection point that is published is the closest one to the origin of the projection line. As a threshold, a minimum Euclidean distance value is defined in order to avoid detecting intersections corresponding to the point cloud of the arm that generates the pointing gesture.

To detect gestures within a time frane, we have implemented spatial filtering to distinguish between real stable pointing gestures and natural arm movements. The system monitors the intersection points obtained by the algorithm, and once a valid intersection point is obtained, the spatial filtering monitoring is launched. To detect a stable gesture, N consecutive intersection points must be contained in a defined cube, the centroid of which is the first intersection point obtained. The number of consecutive intersection points and the side of the filtering cube are defined as parameters. A pointing gesture is considered stable and valid if it fulfils the previous condition. If not, the points from the last filtering operation are discarded. Valid points are queued during the time frame, and returned at the end of the acquisition time according to the format described below.

```
{"points": [
    {"x": "x1","y": "y1","z": "z1"},
    ... ,
    {"x": "xN","y": "yN", "z": "zN"}
]}
```

Fusion Engine

The fusion engine aims to merge both the text and the gesture outputs in order to deliver the most accurate request to the Execution Manager, the element in the robot control architecture in charge of controlling the execution of commands. The engine considers different situations, according to the complementary and/or contradictory levels of both information sources.

As a result, the fusion engine will send to the Execution Manager the most coherent and reliable request that is understandable by the robot if no incoherence is found. Otherwise an error is returned.

The decision strategy of the fusion engine is summarized in Figure 7.

Semantic Multimodal Interpreter Testing

This section summarizes the results that we obtained after testing the Semantic Multimodal Interpreter with different types of requests comprising different voice and gesture inputs.

Voice interpreter testing

We designed a simple experiment to test three different features of the voice based interaction:

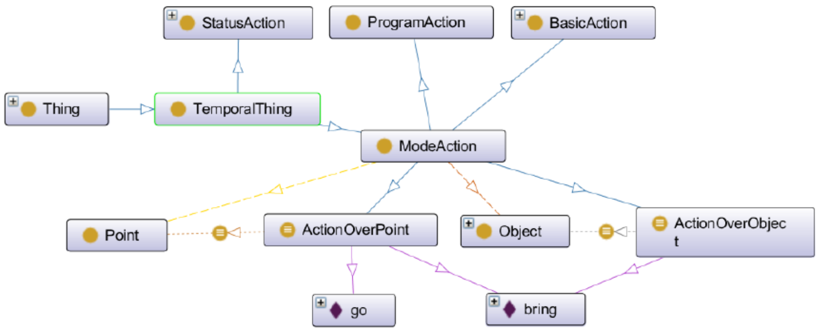


Figure 4. An excerpt of the FourByThree Semantic Interpreter Knowledge Base: *Go* and *Bring*

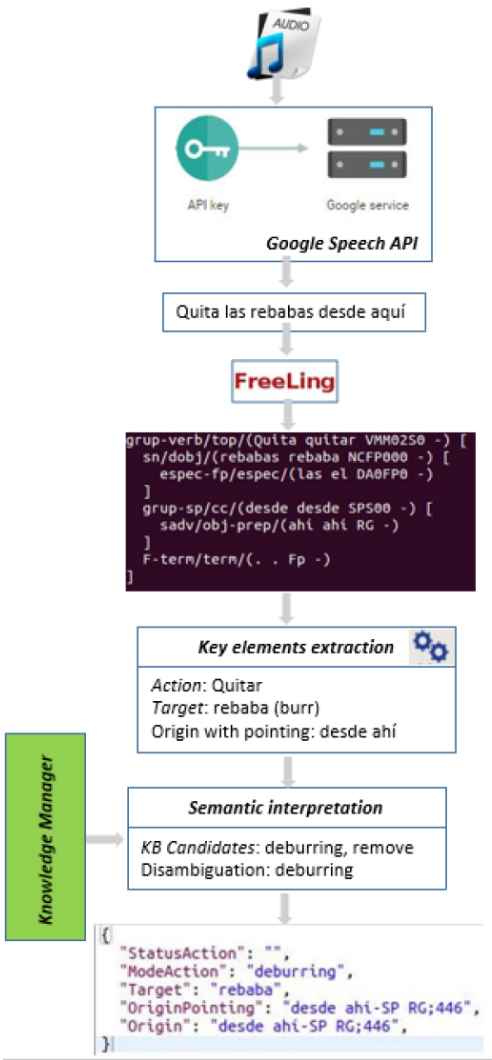


Figure 5. Voice interpreter execution sequence

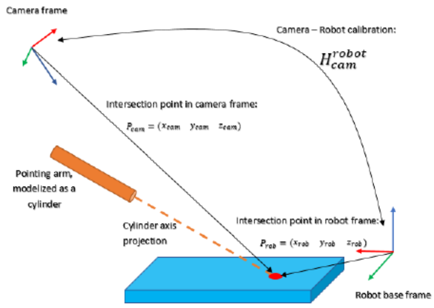


Figure 6. Worker and target frames

1. Time needed to understand a voice command
2. Accuracy of voice request understanding
3. Performance of simple multilingual approach

As the procedure, we fed the system with recorded voice commands (with different lengths and complexities) that corresponded to the Alfa scenario, so that it could run its semantic interpretation.

Time needed to understand a voice command

The first objective of this test was to measure the time needed to transform a voice command into a request understandable by the robot.

The process to interpret a request includes an initialization step and the interpretation itself. The initialization is done simultaneously with the robot initialization, and it involves creating the KB and extending it with WordNet. It takes between 15 and 25 seconds (depending on the number of instances that the KB has to extend, which is closely related to scenario requirements). As the process is done only once during the robot's boot process, it does not affect the user experience.

Table 1 shows the total time required to process the different requests tried out in the experiment, as well as the partial time used by each of the three components of the system, i.e., *Google API* for speech to text transformation, *Freeling* for morphosyntactic analysis and *Understanding* for the semantic interpretation.

Prepared using sagej.cls

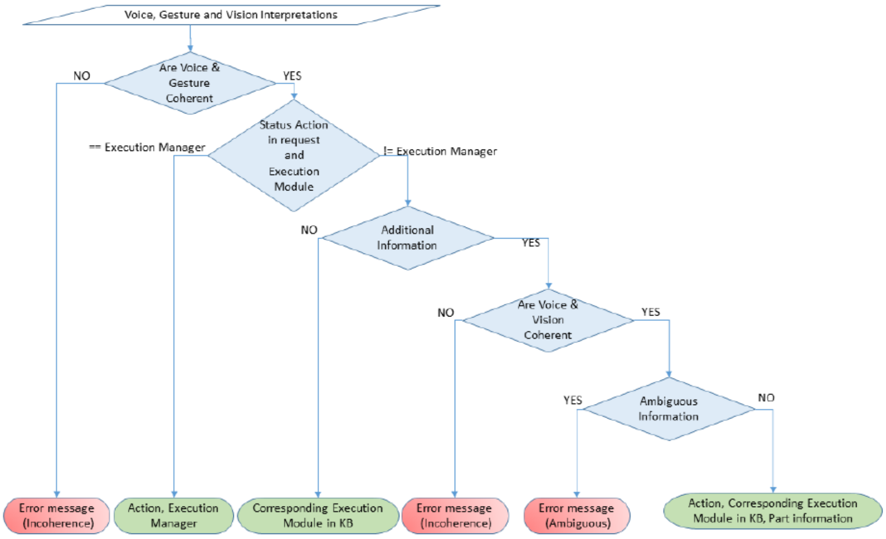


Figure 7. Fusion Engine Decision Strategy

Table 1. Voice interpreter disaggregated times(seconds)

Voice Request	GSA	Freeling	SemInt	Total
Quita ese tornillo (Remove that bolt)	2.16	0.02	0.08	2.27
Quita esa rebaba (Remove that burr)	2.13	.02	0.08	2.25
Empieza a atornillar la pieza (Start to screw the piece)	2.22	0.01	0.12	2.37
Comienza el desatornillado de la pieza redonda de allí (Start unscrewing the round piece that is over there)	2.68	0.02	0.14	2.86
Empieza a apretar los tornillos de la cara superior de esa pieza (Begin to tighten the bolts on the top side of that piece)	2.54	0.03	0.12	2.71
Empieza (Start)	1.85	0.01	0.08	1.96
Detén el desatornillado (Stop unscrewing)	2.12	0.02	0.11	2.31

Time variation is due to the different complexity and length of the speech petition. As it is shown in Table 1, *Google Speech API* is the most time-demanding process, which takes between 1.8 and 2.7 seconds to process each petition, similar to the tests reported in Rossi et al. (2013). In contrast, the sum of *Freeling* and *Understanding* is under 200 miliseconds in most of the cases, depending on the amount of elements to handle and the disambiguation step required.

Voice request understanding performance

The second objective of this test was to evaluate the performance of the text interpreter: given a text request, it was considered that the system had provided a semantically-correct result if the following features were properly identified:

- StatusAction: verbs that modify the status of the operation a robot is doing or has to do, e.g. start, pause, stop.
- ModeAction: actions identifying an operation the robot can do, e.g. screw, debur.
- Target: the element the identified ModeAction applies to.
- Target attribute: any qualifier of the target, such as colour, size, geometry, etc.
- OriginPointing: an expression inside the request denoting the starting point for the ModeAction, or the origin position of the target. It is relevant when a pointing gesture is used in combination with the voice command.
- DestinationPointing: an expression inside the request denoting the destination or end point of the ModeAction. It is relevant when a pointing gesture is used in combination with the voice command.

Although these last two features are not required in the current scenarios, we consider they can be very relevant in other scenarios and have been included in the evaluation.

In our test, 18 out of the 20 requests were correctly interpreted, i.e., the key elements were correctly identified and they created a coherent request in a given context. This final affirmation would not be possible without using semantic technologies that relate possible actions and robot capabilities, actions on specific targets etc. Furthermore, these descriptions lead to an information inference that would not be possible without their semantic representation. Therefore, many requests would not be possible to interpret due to a lack of information, implicit in the request.

The origin of the two errors in the interpretation of the requests are due to mixing up features of the origin and the destination. In the request *Empieza con la pieza que está **en esa esquina*** (Start with the part **on that corner**), instead of origin point, the system returns ***en esa esquina*** (on that corner) as destination point. The same for ***de ahí*** (over there) in the request *Lleva la pieza a la máquina grande **de ahí**.* (Bring the part to the big machine **over there**).

Performance of simple multilingual approach

The objective of this test was to measure the re-usability of the core component, i.e., the text interpreter, when the input was done in a different language. To this aim, Google Translator was used to automatically translate from English into Spanish, and then apply the Spanish semantic interpreter. We used 24 Spanish requests that the Spanish Text Interpreter could already solve correctly. For the experiment, the requests were translated manually into English and the resulting text was introduced in the system. Google Translator was then used to translate those requests from English into Spanish, and the interpeter was then applied on those. We found that 14 out of the 24 initial English requests were properly interpreted, while the remaining 10 were erroneously done. The errors can be classified in 3 groups:

Error Type 1: Mistranslation of demonstrative, resulting in an expression with the target missing (which was relevant in pointing gestures). For instance that was automatically translated into la (the) instead of esa (that).

Error Type 2: the form of the Spanish verb automatically translated was not included in the knowledge base. As a result, the interpreter did not return any action.

Error Type 3: the resulting Spanish expression, even if correct, did not have the expected form and the interpreter was not able to provide a result.

Pointing Gesture Testing

The performance of the Pointing Gesture was already validated in the context of the EU funded EUROC project (EuRoC). This project aims at 'sharpening the focus of European manufacturing through a number of application experiments, while adopting an innovative approach which ensures comparative performance evaluation'. As part of the evaluation, each participant team proposed different FreeStyle experiments and the corresponding metrics. Then, external evaluators verified the achievements. The Pointing Gesture was proposed as one of the experiments and the validation took place in September 2016 at Fraunhofer IPA facilities.

Experiment description

In flexible manufacturing cells, tombstones are used to mount parts that have to be machined. In the context of future human-robot collaboration, the operator will need to indicate the position within the tombstone on which the robot has to operate, performing tasks such as loading/unloading parts, inspecting or measuring the part on that position, etc.

An RGB-D camera was placed above the working area to monitor the person's arms as well as the tombstone. The RGB-D camera was calibrated with respect to the robot base, and the point cloud obtained was referred to this new coordinate frame. Taking the robot's base as reference, two regions of interest (ROIs) were defined in the point cloud: the first one to identify the pointing gesture and the second one, containing the tombstone, to identify the cell

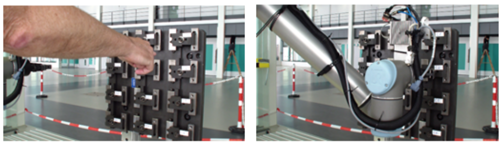


Figure 8. Pointing Gesture Identification validation setup

pointed at on it. The target cell was then obtained by the intersection of the pointing gesture projection and the point cloud corresponding to the tombstone.

There were 12 different positions in the tombstone arranged in 4 rows (labelled from 1 to 4) and 3 columns (labelled from A to C). The external evaluator chose on the spot which position had to be pointed at (e.g. A1, B3). The worker then pointed at it, and the system identified the target position and generated the robot movement to place the tool center point (TCP) in front of the corresponding position. This was repeated twice for each of the 12 possible positions, in a random sequence.

The sequence is presented in Figure 8: on the left, the worker is pointing at one of the 12 possible target positions in the tombstone; on the right picture, the robot is positioned on the cell pointed at.

Results achieved

Two metrics were used to measure the performance of the system: time needed to identify the pointing gesture and target accuracy.

According to (Nielsen (1993)), in human-machine interaction, 0.1 seconds is about the limit for having the user feel that the system is reacting instantaneously, meaning that no special feedback is necessary except to display the result. In addition, 1 second is about the limit for the user's flow of thought to stay uninterrupted, even though the user will notice the delay. Normally, no special feedback is necessary during delays of more than 0.1 but less than 1.0 second.

The system succeeded in identifying all the target points without any error, employing 0.49 seconds as an average value for target identification with a standard deviation of 0.11 seconds. From this, we can already conclude that the performance observed for pointing gesture detection in our implementation was satisfactory for natural interaction. It is important to underline that only 3 out of 24 of the gesture identifications required of more than 0.5 seconds. Although we did not record the different styles of performing gestures utilized by the workers, It is possible that this parameter had some influence on the recognition time variations that we observed.

Conclusions and future work

We have presented a semantic-driven multimodal interpreter for human-robot safe collaborative interaction, with a focus on industrial environments. The interpreter relies on text and gesture recognition for request processing. It deals with the analysis of the complementary/contradictory aspects of both input channels, taking advantage of semantic technologies for a more accurate interpretation due to the reasoning capabilities that it provides.

This paper has presented the validation of each individual mechanism and the fusion of *complementary* information provided by both communication channels i.e., voice and pointing gestures. Nine out of the eighteen well-interpreted voice commands can only be disambiguated by using the pointing gesture as a complementary source. For instance, the command '*remove that burr*' can only be completely disambiguated if the fusion module uses the target position provided by the pointing gesture module. In the coming months, and in the context of this project, the use of gestures other than just pointing will be integrated and used for interactions in which the two sources of information are *redundant*, such as when saying '*stop*' while raising the hand.

This validation of the whole system performance and its benefits will be measured in real working conditions, in the pilot application at ALFA.

The use of semantic technologies to describe robot characteristics and capabilities as well as the context of the scenario, makes this approach generic and scalable: by including the scenario context information and the robot capabilities in the KB, the solution will be ready to reuse without any code modification or re-compilation. Even if in very different collaborative scenarios a deeper KB extension would be necessary, potentially no code modification would be necessary to get it working. This approach is, therefore, generic and it can be applied in different industrial scenarios.

Acknowledgment

FourByThree has received funding from the European Union's Horizon 2020 research and innovation program, under grant agreement No. 637095.

EuRoC has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 608849

References

FourByThree. URL http://fourbythree.eu/.

A. Bannat, J. Gast ans T. Rehrl1, W. Rösel, G. Rigoll, and F. Wallhof. A multimodal human-robot-interaction scenario: Working together with an industrial robot. In *In Proceedings of HCI PartII: Novel Interaction Methods and Techniques*, pages 303–311, 2009.

T. Fong, R.N. Illah, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3): 143–166, 2003.

M.A. Goodrich and A.C. Schultz. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3):203–275, 2007.

R. Stiefelhagen, C. Fugen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel. Natural human-robot interaction using speech, head pose and gestures. In *Intelligent Robots and Systems, 2004. (IROS 2004)*, pages 2422 – 2427 vol.3, 2004.

B. Burger, I. Ferrane, and F. Lerasle. *Towards multimodal interface for interactive robots: challenges and robotic systems description*. INTECH Open Access Publisher, 2010.

S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers., F. Dellaert, D. Fox, D. Hähnel, G. Lakemeyer, C. Rosenberg, N. Roy, J. Schulte, D. Schulz, and W. Steiner. Experiences with two deployed interactive tour-guide robots. In *Proceedings of the International Conference on Field and Service Robotics*, 1999.

K. Gunhee, C. Woojin, K. Munsang, and L. Chongwon. The autonomous tour-guide robot jinny. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3450–3455, 2004.

J. Suarez and R. Murphy. Hand gesture recognition with depth images: A review. In *Proceedings of RO-MAN, IEEE, 2012*, pages 411–417, 2012.

Xsens. URL https://www.xsens.com/products/xsens-mvn/.

Myo. URL https://www.myo.com/.

Bin Fang, Fuchun Sun, Huaping Liu, and Di Guo. A novel data glove using inertial and magnetic sensors for motion capture and robotic arm-hand teleoperation. *Industrial Robot: An International Journal*, 44(2):155–165, 2017.

Y. Wang, C. Yang, X. Wu, S. Xu, and H. Li. Hand gesture recognition with depth images: A review. In *4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pages 274–279, 2012.

M.H. Yan, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 24, pages 1061–1074, 2002.

T. Winograd. Procedures as a representation for data in a computer program for understanding natural language. Technical report, DTIC Document, 1971.

M. MacMahon, B. Stankiewicz, and B. Kuipers. Walk the talk: Connecting language, knowledge, and action in route instructions. *Def*, 2(6):4, 2006.

T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *Proceedings of the International Conference on Human-Robot Interaction*, pages 259–266, 2010.

M. Tenorth, L. Kunze, D. Jain, and M. Beetz. Knowrob-map - knowledge-linked semantic object maps. In *Proceedings of 2010 IEEE-RAS International Conference on Humanoid Robots*, 2010.

T. Wang and Q. Chen. Object semantic map representation for indoor mobile robots. In *Proceedings of International Conference on System Science and Engineering*, pages 309–313, 2011.

S. Dobrišek, R. Gajšek, F. Mihelič, N. Pavešić, and V. Štruc. Towards efficient multi-modal emotion recognition. *International Journal of Advanced Robotic Systems*, 10(1), 2013.

H. Liu, Y. Yu, F. Sun, and J. Gu. Visual-tactile fusion for object recognition. *IEEE Transactions on Automation Science and Engineering*, 14(2):996 – 1008, 2017a.

E. Zheng, B. Chen, X. Wang, Y. Huang, and Q. Wang. On the design of a wearable multi-sensor system for recognizing motion modes and sit-to-stand transition. *International Journal of Advanced Robotic Systems*, 11(2):30, 2014.

H. Liu, F. Sun, B. Fang, and X. Zhang. Robotic room-level localization using multiple sets of sonar measurements. *IEEE Transactions on Instrumentation and Measurement*, 66(1):2–13, 2017b.

S. Rossi, E. Leone, M. Fiore, A. Finzi, and F. Cutugno. An extensible architecture for robust multimodal human-robot communication. In *International Conference on Intellegent Robots and Systems (IROS)*, pages 2208–2213, 2013.

J. Liu and M. Kavakli. A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games. In *Proceedings of the6th international conference on Multimodal interfaces*, pages 1564–1569, 2010.

P.K. Atrey, M.A. Hossain, A. El Saddik, and M.S. Kankanhalli. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Multimedia Systems*, 16(6):345–379, 2010.

H. Holzapfel, K. Nickel, and R. Stiefelhagen. Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3d pointing gestures. In *2010 IEEE International Conference on Multimedia and Expo (ICME)*, pages 175–182, 2010.

B. Burger, I. Ferran, F. Lerasle, and G. Infantes. Two-handed gesture recognition and fusion with speech to command a robot. *Autonomous Robots*, 32(2):129147, 2012.

J. Wagner, E. Andre, F. Lingenfelser, and J. Kim. Exploring fusion methods for multimodal emotion recognition with missing data. In *IEEE Transactions on Affective Computing, Volume: 2, Issue: 4*, pages 206–218, 2011.

L. Kessous, G. Castellano, and G. Caridakis. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3(1):33–48, 2010.

Y. Wu, E.Y. Chang, K.C. Chuan Chang, and J.R. Smith. Optimal multimodal fusion for multimedia data analysis. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 572–579, 2004.

J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and Y. Andrew. Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning*, page 10961103, 2011.

D. Di Marco, M. Tenorth, K. Hussermann, O. Zweigle, and P. Levi. Roboearth action recipe execution. In *In Frontiers of Intelligent Autonomous Systems*, pages 117–126, 2013.

A. Gonzalez-Agirre, E. Laparra, and G. Rigau. Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. In *In Proceedings of the Sixth International Global WordNet Conference (GWC12)*, 2012.

Stardog. Stardog is an enterprise data unification platform built on smart graph technology: query, search, inference, and data virtualization. Stardog. URL http://stardog.com/.

GoogleApi. Google Speech Api. URL https://console.developers.google.com/apis/api/speech/.

L. Padró and E. Stanilovsky. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, 2012.

S. Harris, A. Seaborne, and E. Prudhommeaux. Sparql 1.1 query language. *W3C Recommendation*, 21, 2013.

EuRoC. European robotic challenge. Euroc. URL http://www.euroc-project.eu/.

J. Nielsen. *Usability Engineering*. Morgan Kaufman, 1993.

Iñaki Maurtua, Aitor Ibarguren, Johan Kildal, Loreto Susperregi, Basilio Sierra. Human Robot collaboration in Industrial applications: safety, interaction and trust. International Journal of Advanced Robotic Systems. 2017 [93].

Journal Title
XX(X):1–9
©The Author(s) 2015
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Human Robot Collaboration in Industrial Applications: Safety, Interaction and Trust

Iñaki Maurtua[1], Johan Kildal[1], Loreto Susperregi[1], Aitor Ibarguren[2] and Basilio Sierra[3]

Abstract

Human-robot collaboration is a key factor for the development of factories of the future, a space in which humans and robots can work and carry out tasks together. Safety is one of the most critical aspects in this collaborative human-robot paradigm. This paper describes the experiments done and results achieved by the authors in the context of the FourByThree project, aiming to measure the trust of workers on fenceless human robot collaboration in industrial robotic applications as well as to gauge the acceptance of different interaction mechanisms between robots and human beings.

Keywords

Safe Human-Robot Collaboration, Collaborative robots, Multimodal Interaction, Natural Communication, Manufacturing

Introduction

Human-robot collaboration can contribute to the development of factories of the future, a space in which humans and robots can work and carry out tasks together. It allows human operators to focus on operations with high added value or demanding high levels of dexterity, thus freeing them from repetitive or potentially-risky tasks. However, some tasks can be too complex to be performed by robots or too expensive to be automated, as they may require engineering special tools and systems. Therefore, a collaborative environment in which humans and robots can work side by side and share tasks in an open and fenceless environment is a relevant goal to reach.

Safety and interaction are key success factors for this vision of collaboration between humans and robots. On the one hand, the safety of human beings around the robot must be guaranteed during the execution of tasks. As physical safeguards may be impractical for real cooperation, the use of either power (or force) limiting, or speed and separation monitoring, are possible according to ISO (2011). In the second approach, the robot must be constantly aware of what is happening around it and it has to monitor the workers' actions in order to change its behaviour (speed and/or trajectory) according to the separation distance. On the other hand, an effective bidirectional human-robot communication contributes towards doing the collaboration more effective and safe.

This paper presents the results of two experiments carried out to measure how workers trust the safety measures developed and implemented in the EU funded *FourByThree* project. It also seeks to validate the effectiveness and acceptance of alternative interaction mechanisms between human workers and robots in industrial human-robot collaborative environments.

The paper is organized as follows: It starts with an overview of the project and the proposed safety strategy and interaction mechanisms are presented. Then the objectives of the experiment, related work and the experiment design are described. Finally, the results achieved and conclusions are summarized.

PROJECT CONTEXT

Project overview

Since December 2014, the FourByThree Project ('Highly customizable robotic solutions for effective and safe human robot collaboration in manufacturing applications') FourByThree (2017) is developing a new generation of modular industrial robotic solutions that are suitable for efficient task execution in collaboration with humans in a safe way and are easy to use and program by the factory worker. The FOUR main characteristics of FourByThree are:

1) Modularity

FourByThree outcomes are packed as a 'kit' of hardware and software tools for the development of custom robotic solutions. The concept includes fundamental mechanical elements (four different size series-elastic actuators, brackets, flange), the control unit (incorporating advanced techniques for safe HRI) and additional auxiliary hardware/software modules integrated in a ROS based FourByThree control architecture.

2) Safety

Safety strategies and low cost mechanisms allowing intrinsically safe behaviour of the robot in the presence of humans are developed. The safety approach is centered on the design of the actuators with the capability to monitor

[1]IK4-Tekniker, Spain
[2]Tecnalia, Spain
[3] RSAIT Group, University of the Basque Country, Spain

Corresponding author:
Iñaki Maurtua, IK4-Tekniker Parke Teknologikoa, Iñaki Goenaga, 5 20600, Eibar Spain
Email: imaurtua@tekniker.es

the force and torque in each one, which provides the opportunity to implement variable stiffness strategies and reactive behaviour in case of contact/collision. The system also includes space monitoring using a projection and a vision system, which provide the information needed to modify the velocity of the robot according to its relative distance with respect to the worker.

3) Ease of use

FourByThree offers a set of multimodal interaction mechanisms that facilitate robot programming and control, e.g., voice based interaction, gestures, projection system and manual guidance.

4) Efficiency

Robots are intended to help workers in doing a task, to this aim they have to be reliable, maintainable and intrinsically safe. Performance metrics are established for each application addressed in the project, i.e., assembly, deburring, welding, riveting and machine tending, implemented in four challenging industrial Pilot Studies (aeronautics, sheet metal forming, investment casting and professional training).

Safety strategy

The safety strategy in FourByThree is based on five pillars:

- The serial elastic actuators allow measuring the force and torque values using and provide redundant torque and position estimation. The torque is computed by using two different sources: using the motor currents and using the spring deflection and its identified model. If both torque estimation values do not agree, the motor is disabled. The position sensors on motor and gear side are compared to each other to detect sensor failures.
- The robot design, by eliminating sharp edges, reducing trapping risks, etc.
- The external monitoring system, which consists of projection and vision systems, allowing to monitor the space around the robot to detect any possible violation by the worker.
- Adjustable stiffness control, that adjust the stiffness level based on different factors, such as relative distance between robot and worker or tasks at hand.
- The control architecture, as shown in Figure 1.

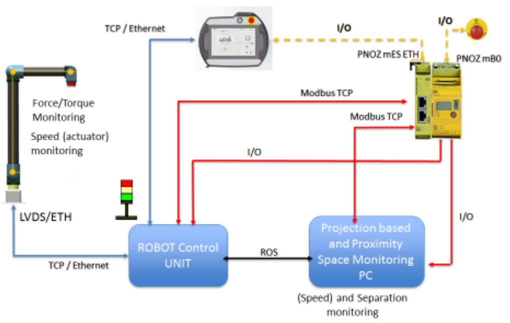


Figure 1. Safety strategy

The proper use of these features makes it possible to satisfy the operating conditions established in ISO10218 parts 1 and 2, and ISO/TS15066, once the mandatory Risk Assessment has been performed in each scenario.

In brief, the FourByThree safety strategy allows implementing Speed and Separation monitoring and Force Limiting collaboration modalities.

In fact, it is possible defining a protective area around the robot for co-existence and interference situations (i.e., when the human moves through the robot workspace but does not interact directly with the robot or when the human reaches into the robot working area or obstructs the robot workspace in a non-planned task). The projection Vogel et al. (2013) and vision systems are in charge of monitoring the robot workspace and triggering the safety signal when there is a violation in the area.

For co-operation activities (i.e., when the human has to interact with the robot in a productive way), the system's capability to monitor and limit the force and torque is used to guarantee the safety.

Interaction mechanisms

A requirement for natural human-robot collaboration is to endow the robot with the capability to capture, process and understand accurately and robustly human requests. Voice and gestures are key channels that humans use between them for natural communication. By analogy, they can be considered as relevant to achieve such a natural communication between humans and robots. In this multimodal scenario, the information coming from the different channels can be complementary or redundant:

- A worker says '*Take this*' while pointing at an object (complementary).
- The worker says '*Stop*' while performing a stop gesture (redundant).

The need for complementary channels is clear, but redundancy can also be beneficial Bannat et al. (2009), e.g., in industrial scenarios in which noise and variable lighting conditions may reduce the robustness of each channel when considered independently. The FourByThree project has designed and developed a semantic approach that supports multimodal (voice and gesture based) interaction between humans and robots in real industrial settings. For such semantic interpretation, four main modules have been developed: a *Knowledge-Manager* module that describes and manages the environment and the actions that are feasible for robots in a given environment, using semantic representation technologies; a *Voice Interpreter* module that given a voice request, extracts the key elements on the text and translates them into a robot-understandable representation, combining NLP and semantic technologies; a *Gesture Interpretation* module, mainly for resolving pointing issues and some simple commands, such as stopping or resuming movements; and a *Fusion Engine* for combining the output of both text and gesture modules and to construct a complete and reliable command for the robot.

These main modules are described in detail in the following subsections.

1) Knowledge Manager

The knowledge manager uses an ontology to model the environment and the robot capabilities, as well as the relationships between the elements in the model, which can be understood as implicit rules that the *reasoner* exploits to infer new information. Thus, the *reasoner* can be understood as a rule engine in which human knowledge can be represented as rules or relationships.

Ontologies are reusable and flexible at adapting to dynamic changes, thus avoiding to have to re-compile the application and its logic whenever a change is needed. Through ontologies, we model the industrial scenarios in which robots collaborate with humans. The model includes robot behaviours, actions they can accomplish and the objects they can manipulate/handle. It also considers features and descriptors of these objects.

2) Voice Interpreter

Given as input a human request in which a person indicates the desired action via voice, the purpose of this module is to understand exactly what the person wants the robot to do and, if the information is complete, to generate the corresponding command for the robot. For instance, if a worker says '*Remove the burrs from there*' the voice interpreter should interpret that the verb *remove* corresponds to the *deburring action* and check if it is a feasible action in the current collaborative robot application and generate the necessary information to do it.

3) Gesture Interpretation

The gesture interpreter module recognizes two different types of gestures: pointing gestures and command gestures used to request a specific and predefined action (e.g., *start/stop*). A RGB-D sensor providing depth information in a 2D image is used in both cases.

For the pointing gesture, a point cloud processing approach has been used. A RGB-D camera is used to acquire the information of the working environment and the worker (in particular the arms and hand). The camera is placed pointing towards the working area of the robot in the region above the human operator. The camera is calibrated with respect to the robot base and the point cloud referred to the robot frame. Two cuboid regions of interest (ROIs) are then defined in the point cloud, one for the operators pointing gesture (basically his/her forearm) detection and processing and a second ROI where the intersection area has to be identified.

In the ROI where the pointing gesture has to be detected, the forearm of the operator is modelled as a cylinder and its axis taken as the pointing line.

The pointing gesture is modeled as a straight line, while the intersection area can be assumed to be a planar surface (tables, working surfaces, etc.) in most of the cases. The intersection point (x, y and z) is obtained geometrically.

4) Fusion Engine

The fusion engine is in charge of merging the information provided by the voice interpreter, the gesture interpreter and the part identification module to identify the worker intention and send the corresponding command to the robot.

The engine considers different situations regarding the complementary and/or contradictory levels of both sources. It was decided that the text interpreter output will prevail over the gesture information. When no contradiction exists between the two sources, the gesture information is used either to confirm the text interpretation (redundant information), or to complete it (complementary information).

RELATED WORK

Human-robot collaboration has been a significant research topic since the beginning of robotics. The constant introduction of robots in industrial environments, the creation of new compliant robots and the sensors available nowadays in the market (cheaper and more accurate), make human-robot interaction an even more active and exciting research subject Heinzmann and Zelinsky (2003), Pervez and Ryu (2003).

The different approaches for safe human-robot interaction can be classified as either pre-collision or post-collision strategies.

Post-collision methods detect a collision as it occurs, and attempt to minimize the resulting damage. Commercial robots that are purposely designed for collaborative applications fall mainly in this category. The implemented methods may vary. The more common ones are: power and force limiting, use of Series Elastic Actuators that minimize the force of impact, and the use of protective skins that detect the collision or the proximity of the worker. These approaches have been used in collaborative robots (COBOTs) available in the market, such as in Nextage (KAWADA), LBR iiwa (KUKA), Roberta (ABB/GOMTEC), Yumi (ABB), Apas (BOSCH), UR3/5/10 (Universal Robots), CR-35iA (FANUC), Baxter (Rethink Robotics) or Franka.

In this field, it is very relevant the work described in Haddadin (2014) that is the result of other previous works such as Haddadin et al. (2007) and Haddadin et al. (2009). They focus mainly on the identification of limit values of force and power that a robot may exert upon a person without causing severe injuries derived from real human impact experiments. They were able to drive eight times faster and cause thirteen times higher dynamical contact forces than were suggested by the first version of the norm for the static case. As the impact experiments yielded such low injury risks, the question whether this standard was too conservative raised.

Pre-collision strategies attempting to prevent collisions by detecting them in advance are also very relevant, not only to avoid the collision itself, but due to the fact that there are more than 1.5 million industrial robots already in use worldwide, and there is a great interest in designing solutions that can turn those robots into human-safe platforms.

There are several approaches endowing robotic cells with sensors to determine that a human is present in the vicinity of the robot. Some add-on solutions have been developed for robots that have not been designed as inherently human-safe robots, such as ABB'where's SafeMove Behnisch (2008).

In Przemyslaw et al. (2014) a PhaseSpace motion capture system that has the drawback of requesting the worker to wear active led marks- was utilized to sense the position of the human worker within the workspace of an industrial robot with no built-in safety features and no compliant joints. In Lasota and Shah (2015) the same authors evaluate through human subject experimentation whether this motion-level adaptation leads to more efficient teamwork and a more satisfied human co-worker. The results indicated that people learn to take advantage of human-aware motion planning even when performing novel tasks with very limited training and with no indication that the robots motion planning is adaptive.

In Vogel et al. (2013), a projector emitting modulated light patterns into the shared human-robot workspace is used. The light reflected from the environment is detected by a camera, being able to detect any intrusion. The system is also used to provide visual information to the worker.

In PRybski et al. (2012), authors fuse data from multiple 3D imaging sensors of different modalities (two time-of-flight cameras and two stereo cameras) into a volumetric evidence grid and segment the volume into regions corresponding to background, robots, and people that have been previously modelled. The system allows slowing down the robot and even stopping the motion when people and robot approximate. Also relevant, Morato et al. (2014) presents a multiple Kinects based exteroceptive sensing framework to achieve safe human-robot collaboration during assembly tasks.

Different approaches Jaward et al. (2006), Benfold and Reid (2011) and Przemyslaw et al. (2014) propose the use of particle filtering and statistical data association for tracking multiple targets, adding robustness to the tracking process.

Factors affecting trust in human-robot interaction has been subject of analysis in some works, in particular when working in high-risk situations as military applications. In Hancock et al. (2011) they are evaluated and quantified the effects of human, robot and environmental factors on perceived trust in human-robot interaction (HRI). They concluded that factors related to the robot itself, specifically, its performance, had the greatest current association with trust, and environmental factors were moderately associated. There was little evidence for effects of human-related factors.

In PBainbridge et al. (2008) and Tsui et al. (2010) it is explained how the type, size, proximity, and behaviour of the robot affect trust. In Park et al. (2008) it is described that trust can be dynamically influenced by factors (or antecedents) within the robotic system itself, the surrounding operational environment, and the nature and characteristics of the respective human team members. In Sadrfaridpour et al. (2014a) and Sadrfaridpour et al. (2014b) it is proposed a model for dynamic trust of human to robot based on the robot performance and the human performance. They simulated the human performance and robot performance and the corresponding trust during a typical work day when they do a certain manufacturing collaborative task.

EXPERIMENT OBJECTIVES AND DESCRIPTION

Objective

The objective of the experimentation has been to obtain valuable information about two key aspects in a human-robot collaborative environment:

- **Safety**. How do workers perceive the safety aspects when working in the vicinity of an industrial robot without physical barriers? How do they perceive the safety level achieved with the measures that have been proposed and implemented in FourByThree.
- **Interaction**. What is the workers' feedback with respect to some of the interaction mechanisms implemented?

Most papers dealing with perceived safety and interaction satisfaction use a similar approach: participants are requested to execute a collaborative task with a robot and observation and questionnaires are used to measure different features. This is the case of the experiment described in Przemyslaw et al. (2014), in which participants worked with a robot to perform a collaborative task, placing 8 screws at designated locations; human satisfaction and perceived safety and comfort were evaluated through questionnaires. As recruiting neutral participants is difficult, they followed the common practice of using participants affiliated to the institution, in this case they were 20 MIT affiliates. In our experiment, participants had not any kind of relationship with the experimenters, on the contrary, they were attendants to the two fairs that accepted to participate in the experiment. A second differential aspect in these experiments was the fact that the proposed tasks demanded the participants to use both the interaction mechanisms and safety features to complete them.

Experiment design

FourByThree has had the opportunity to be present in two important Trade Fairs in 2016:

- TECHNISHOW. It is the largest and most important trade show in the field of industrial production technology, treatment and processing of metals, plastics, accessories and tools in the Benelux area
- BIEMH. This bi-annual fair is one of the most important industrial fairs in Spain and is devoted to the Machine Tool sector, including robotics and automation technology

In each one of these events, we conducted user experiments with attendees that agreed to take part voluntarily. We chose to conduct studies at these events because, in this way, we could gain access to bodies of participants professionally involved in industrial processes from different sectors, with a variety of levels of experience with automation technologies. Such a mix of profiles would have been difficult to recruit otherwise.

The common objective of these experiments was to gain insight into human attitudes with respect to fenceless robotics and some selected interaction mechanisms. By means of the tests, observation and questionnaires we obtained valuable feedback about the following aspects:

- Users' trust. Even if there are many factors that affect trust we focus on the following safety aspects:
  - Acceptance of vision-based human detection
  - Force control in case of collision
- Interaction mechanisms
  - Pointing Gesture
  - Manual guidance
  - Tapping
- Overall attitudes with respect to collaborative robots in the workplace

The technologies and tasks demonstrated in both experiments were similar, although with complementary focus on the interaction techniques used. In both cases, an introductory explanation was provided by the experimenter regarding the technologies employed and the sequence to be followed by the participant in the experiment.

1) TECHNISHOW: Study 1

This experimental study took place during the TECHNISHOW fair in Utrecht, May 2016, in the booth run by STODT, one of the partners in FourByThree. The experiment setup consisted of a Universal Robot, a RGB-D vision system to monitor the environment and capture data for the pointing gesture, a table and two trays with some parts on one of them. The participants in the experiment had to perform three tasks and answer to a questionnaire afterwards.

Pointing gesture

The first task demonstrated collaborating with the robot using a pointing gesture. There were 3 different parts on Tray-1, as soon in Figure 2. The participant pointed at one of them with the finger (extended arm). Following the gesture, the robot identified the part that was being pointed at, grasped it and placed it on the corresponding position on the second tray. The setup and pointing gesture are shown in Figure 3.

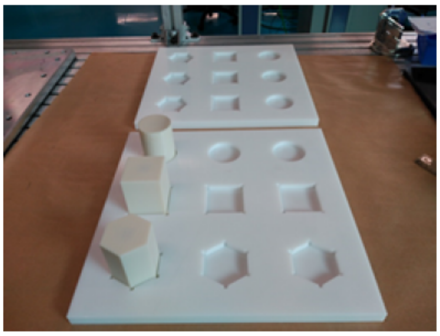


Figure 2. The three different objects in one of the two trays

Safety monitoring The second task demonstrated the safety feature of the robot interrupting its movement whenever the relative distance between a person and the robot was below a threshold. While the robot was moving a part, the participant was invited to reach out to the robot with their hands, without touching it. The monitoring system detected this intrusion of the robot's space and interrupted its movement, so as to avoid a possible collision.

Manual guidance The third task demonstrated programming the robot by moving it with the hands. The participant draged the robot arm's gripper to a position near one of the objects on one of the trays. The position was recorded and afterwards the robot executed the program going to the recorded point, taking the part and placing it on the second tray.

Collision In addition to the tasks above, participants were invited to be 'hit by the robot: the monitoring system was disabled and the robot collided against the participant's arm. The robot detected the collision (force exerted) and it stopped as a result.

2) BIEMH: Study 2

The setup prepared for the experimental study at the BIEMH fair (Bilbao, June 2016), a different setup was

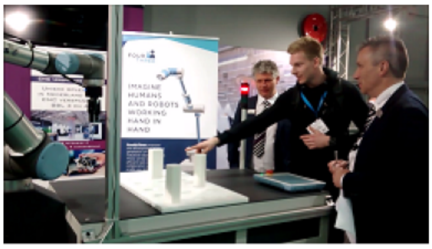


Figure 3. Experiment setup at Technishow and participant pointing at a target object

used consisting of a KUKA IIWA robot placed on a tabl two plastic storage bins and the RGB-D visual monitori system. This second experiment consisted of the followi tasks in the sequence described below.

Safety monitoring The sequence of the experiment was follows: while the robot moved one of the bins continuous from side to side (see Figure 4), the participant was invited bring their hands close to the robot. The monitoring syste detected that the hand was close to the robot and interrupte the movement. When the visitor moved his/her hand awa the robot resumed its movement.

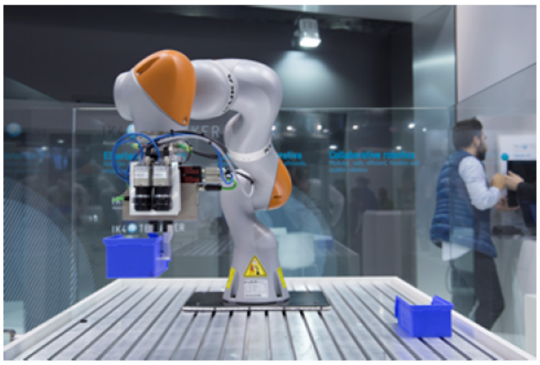


Figure 4. Robot moving a plastic drawer

Tapping and pointing gestures Once the robot was st after detecting the participant's hand nearby, she could app a light downward force with the hand on the robot ar (tapping gesture). The robot reacted to this by leaving tl bin on the table and withdrawing itself backwards to a restir position. There, the robot waited for the participant to sign which of the two bins present on the workbench (which cou be casually placed in a random arrangement) it had to gras The participant then selected one of the trays on the table t pointing at it with the finger (extended arm). The robot us the camera placed on its flange to analyze the actual positic of the bin that had been pointed at, computed a graspi strategy and took it from the surface of the workbench, liftir it up and resuming the continuous sideways movement whi carrying it. The next cycle of interaction was then ready begin with a new participant. Both interaction mechanisn are shown in Figure 5.

In order to understand the results presented in next sectio it is important to point out a difference introduced in th setup with respect to the previous one in Technishow: BIEMH (Study-2), we added to the base of the robot a bar of LEDs that was designed to inform the user about tl

Prepared using sagej.cls

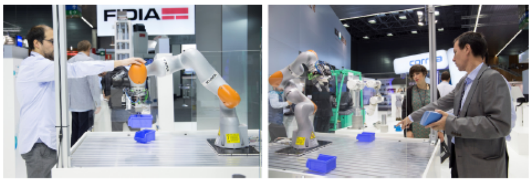


Figure 5. Visitor exerting a force on the robot to interact with it (left), and a participant pointing at a container bin that he wanted the robot to take (right)

robot's status in the interaction. In other words, to improve situational awareness for the human actor in the collaborative interaction. Specifically, this band lit up while the robot was awaiting for the person to perform a pointing gesture, and for as long as the gesture had not been recognized by the vision system.

For the rest of the cycle in the interaction, the luminous band remained switched off. This simple information was helpful for the person to understand when then robot was expecting an input gesture, as well as to know when a pointing gesture had been performed for long enough for the robot to have registered it. In both studies, participants had to fill in a questionnaire at the end of each session, including five point Likert scale and multiple choice closed-ended and open-ended questions divided into 4 sections: Demographics, Interaction, Safety and Usability.

In the demographic section participants were asked about pure demographic issues as age, gender, academic level, the company they work for (sector, size, main activities, current use of robots, plans for introducing robots) or experience in working with robots, as well as their opinion on the impact of robotics on cost, employment, working conditions, efficiency, etc., main barriers for the introduction of collaborative robots and the key requirements for collaborative robots. In the interaction and usability sections, workers had the opportunity to provide feedback on the interaction mechanisms implemented, i.e., pointing gesture (reaction time, naturalness, etc.), hand guiding (effort needed, speed of the movements, safety perception, etc.) and tapping (this was included only in Study-2). In the Safety section, participants evaluated which safety feature was the most relevant, safety perception and acceptance of fenceless collaborative environments.

Finally, participants were interviewed about the more relevant aspects that were observed in the experiments and regarding the most salient responses they had entered in the questionnaires.

Participants

Altogether, 115 participants took part in both studies: 38 participants in Study-1 (Technishow), 2 of which were women, and 77 participants in Study-2 (BIEMH), 13 of which were women (see Figure 6).

The prior experience of the participants in these studies is presented in Figure 7, showing relative proportions of type and extent of the experience. More specifically, 13% of the participants had more than ten years of experience as machine operators, production system designers, commissioning and maintenance of automation systems, 8% and 37% had worked between 6 and 10 years and between 1 and 5 years respectively with robots, and the remaining 48% had no previous direct experience with robots at work.

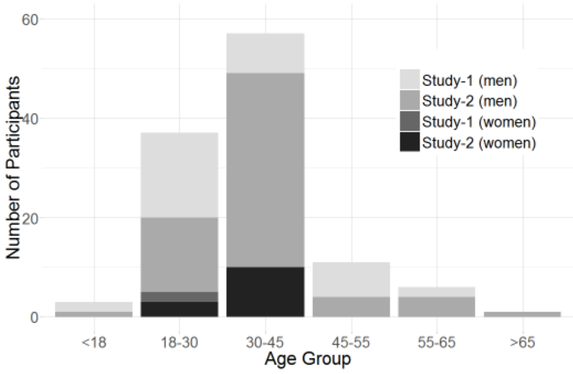


Figure 6. Demographics of participants

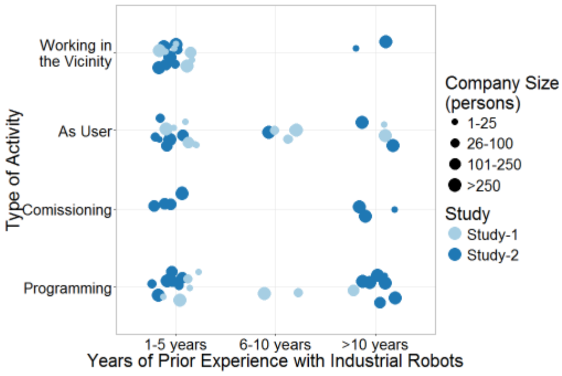


Figure 7. Participants' prior experience with robots

It is also worth noticing that significant number of the participants with previous experience working with robots had acquired their experience in the automotive sector (see Figure 8).

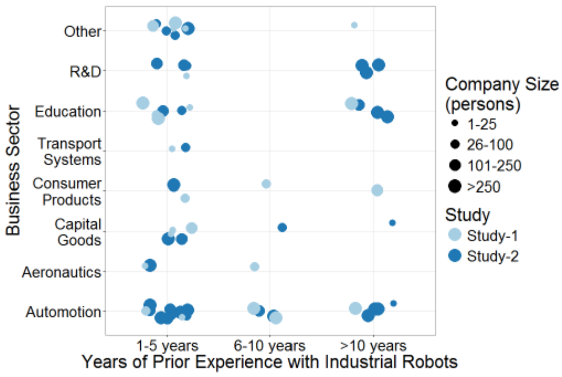


Figure 8. Participants' experience by business sector and size of company

EXPERIMENT RESULTS

Interaction mechanisms

Participants were asked about naturalness, reliability, usefulness, ease of use and response time of the Pointing Gesture interaction mechanism.

Figure 9 suggests a clear improvement between Study-1 and Study-2 in the users' subjective experience with respect to pointing gesture based interaction modality, and in particular the perceived response time. This improvement could be due to the introduction of the feedback mechanism introduced in Study-2 (the LED luminous band that designed for situational awareness). As explained before, in this second setup it was introduced a lighting system on the base of the robot that was switched on whenever the robot was waiting for a pointing gesture and was immediately switched off once the gesture was identified.

In both experimental setups, a period of time passed since the gesture was identified and before the actual movement of the robot started (during which the planning of the trajectory and the acceleration ramp took place). From the participants' subjective perspective, the response time in Study-1 included the complete time until the robot started to move. In contrast, in Study-2 only the actual pointing detection time was displayed to the participant through the lighting pattern, results in perceiving the system as showing faster response. This factor can also be the reason that explains the lower score for usefulness in Study-1. The answers recorded for the rest of the parameters (naturalness, ease of use and reliability) were very positive (Figure 9).

Figure 9. Feedback on pointing gesture

In Study-1, participants were asked to drag the robot arm's gripper to teach the grasping position. To do that, we used the gravity compensation feature of the Universal Robot. We asked the participants about three aspects of this hand guiding interaction mechanism (Figure 10): effort needed to drag the robot, usefulness and difficulty to perform. Although the answers were positive for the three factors, it should be stressed that there was a significant number of participants that found the effort needed to move the robot to be in the limit of acceptability. Based on this, we consider that there is still room for improving the control algorithms needed to implement this functionality.

In Study-2, we implemented the tapping functionality: users lightly tapping downwards on the robot arm to indicate that it could resume its movement and watch for a pointing gesture. The acceptance of the participants was overall positive (Figure 11) if we look at the answers to the same four questions (naturalness, reliability, usefulness and ease of use), even though, 8% of the participants did not feel this way of interaction to be natural. Nonetheless, as this is a feature that does not demand additional sensors (we used the force feedback provided by the collaborative robot), it is worth considering its use in the future as an additional interaction mechanism.

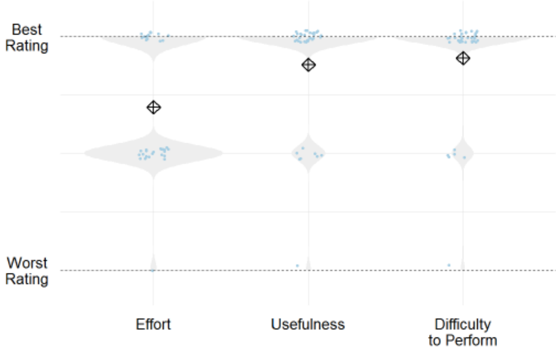


Figure 10. Feedback on hand guiding

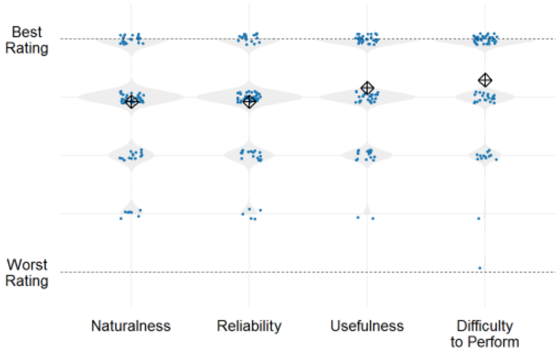


Figure 11. Feedback on tapping

Safety

Participants were asked about the safety perception when working and interacting with the robot without any physical barrier between them (see Figure 12). In both studies, the safety measure that created the 'safe' collaborative environment was the proximity monitoring system based on a RGB-D vision system, which measures the relative distance between the robot and a person and stopped the movement of the robot when such distance was below a threshold value.

It was encouraging to observe a clearly positive acceptance of the fact that collaborative robots will bring many benefits both to workers and to processes in which they integrate. However, it was striking to observe the widespread opinion that such robots will have a negative effect on jobs. It is clear that the robotics community needs to have a clear and convincing answer for this recursive general opinion, and the motivation to generate supporting evidence through the creation of success case studies.

While feedback was overall positive in both cases, it is worth analysing closer the near-unanimity (97%) achieved in Study-1. There are two possible reasons for that: first, the participants in that study were younger than in Study-2. In fact, 50% of them were aged below their thirties. In addition, in Study-1 it was demonstrated to the participants how the robot reacted in case of an unexpected collision (and some of them tried it first hand), i.e., how the robot stopped when it detected a collision (following the implementation of the Force and Torque limiting approach from ISO 10218).

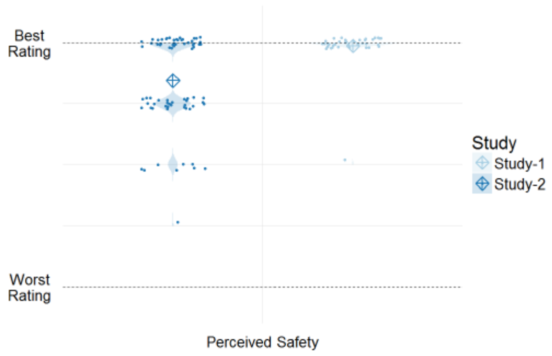


Figure 12. Feedback on perceived safety

Workers opinion

At the end of each session participants were asked about their opinion on the consequences of robot introduction in factories as well as the key elements to make this a success. The answers to the first question (see Figure 13) were as expected: we observed was the widespread opinion that robots will result in a reduction on the number of jobs. On the contrary, there was a broad consensus on the benefits that this introduction will bring to productivity, quality of production, competitiveness and working conditions of the workers.

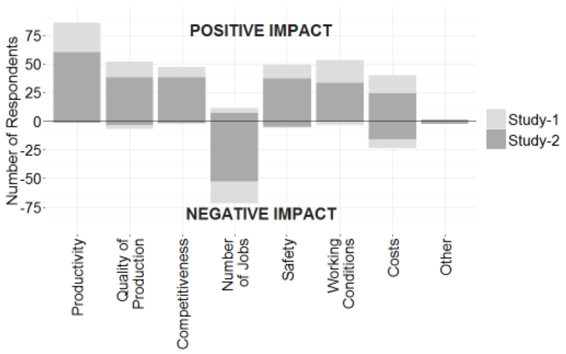


Figure 13. Negative and positive effects of introducing robots at work

Safety was considered the key requirement for succeeding in the introduction of the human-robot collaborative paradigm, followed by usability, flexibility and efficiency (see Figure 14).

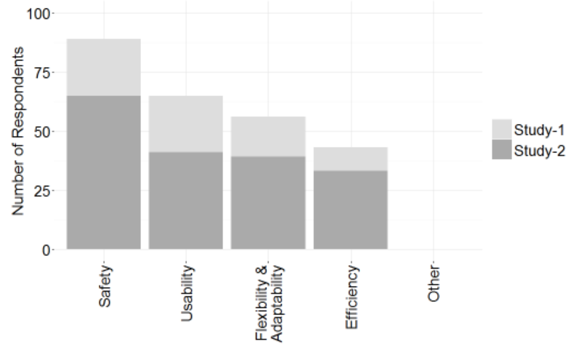


Figure 14. Key success factors for HRC

CONCLUSION AND FUTURE WORK

The perception of trust in the safety strategy tested in the experimental studies described above was overall positive. It seems that no special objections could be expected from users, according to our data. This finding is in line with the fact that 97% of the participants in Study-1 declared that in the future the collaboration between robots and workers will be possible and that they would accept working together with in this way. In addition, gesture-based interactions and hand-guiding interaction mechanisms were also rated positively by participants, as potential future users.

As a limitation of our work presented here, this analysis has to be considered for revision in more realistic scenarios, in which workers perform real tasks with the support of robots during a longer period of time. Such a study should be considered for follow up future work. In fact, authors will conduct a further analysis in five real industrial pilot studies in which workers and robots will have to accomplish deburring, assembling, welding, riveting and machine tending operations. This work will be part of the validation process of the FourByThree project.

Additionally, the projection-based safety mechanism that is also part of the FourByThree project, but that was not available by the time these experiments were done, will provide an additional mechanism that can improve the trust of workers.

We also expect that the two additional mechanisms that are included in FourByThree (using the projection system and the gestures to command basic actions of the robot, such as stop and resume, including voice based interaction) will contribute to a richer interaction experiences. Both aspects will be tested with a wide spectrum of real workers at STODT in the coming months.

Acknowledgment

The FourByThree project has received funding from the European Union's Horizon 2020 research and innovation program, under grant agreement No. 637095.

References

ISO. Iso 10218-1:2011: Robots and robotic devices safety requirements for industrial robots part 1: Robots. Technical report, International Organization for Standardization, 2011.

FourByThree, 2017. URL http://www.fourbythree.eu/.

C. Vogel, C. Walter, and N. Elkmann. A projection-based sensor system for safe physical human-robot collaboration. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

A. Bannat, J. Gast, T. Rehrl, W. Rsel, G. Rigoll, and F. Wallhof. A multimodal human-robot-interaction scenario: Working together with an industrial robot. In *Proceedings of HCI PartII: Novel Interaction Methods and Techniques*, page 303311, 2009.

J. Heinzmann and A. Zelinsky. Quantitative safety guarantees for physical human-robot interaction. *International Journal of Robotics Research*, 22(7):479–504, 2003.

A. Pervez and J. Ryu. Safe physical human robot interaction-past, present and future. *Journal of Mechanical Science and Technology*, 22(3):469–483, 2003.

Sami Haddadin. *Towards Safe Robots: Approaching Asimovs 1st Law*. Springer, 2014.

S. Haddadin, A. Albu-Schffer, and G. Hirzinger. Safety evaluation of physical human-robot interaction via crash-testing. In *Proceedings of Robotics: Science and Systems, Vol.3*, pages 217–224, 2007.

S. Haddadin, A. Albu-Schffer, M. Frommberger, J. Rossmann, and G. Hirzinger. The 'dlr crash report': Towards a standard crash-testing protocol for robot safety - part i: xresults. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pages 272–279, 2009.

Kevin Behnisch. White paper: Safe collaboration with abb robots, electronic position switch and safemove. Technical report, ABB, 2008.

A. Lasota Przemyslaw, F. Rossano Gregory, and A. Shah Julie. Toward safe close-proximity human-robot interaction with standard industrial robots. In *Proceedings of IEEE International Conference on Automation Science and Engineering (CASE)*, pages 18–22, 2014.

P. A. Lasota and J. A. Shah. Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration. *The Journal of the Human Factors and Ergonomics Society*, 57(1):21–23, 2015.

P. PRybski, P. Anderson-Sprecher, D. Huber, C. Niessl, and R. Simmons. Sensor fusion for human safety in industrial workcells. In *PInternational Conference on Intelligent Robots and Systems*, page 3612, 2012.

C. Morato, K.N. Kaipa, B. Zhao, and S.K. Gupta. Toward safe human robot collaboration by using multiple kinects based real-time human tracking. *Journal of Computing and Information Science in Engineering*, 14(1):011006, 2014.

M. Jaward, L. Mihaylova, N. Canagarajah, and D. Bull. Multiple object tracking using particle filters. In *Proceedings of the IEEE Aerospace Conference*, pages 8–16, 2006.

B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3457–3464, 2011.

Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *The Journal of the Human Factors and Ergonomics Society*, 53 (5):517–527, 2011.

W. A. PBainbridge, J. Hart, E. S. Kim, and B. Scassellati. The effect of presence on human-robot interaction. In *Proceedings of the 17th IEEE Symposium on Robot and Human Interactive Community*, page 701706, 2008.

K. M. Tsui, M. Desai, and H. A. Yanco. Considering the bystanders perspective for indirect human-robot interaction. In *Proceedings of the 5th ACM/IEEE International Conference on Human Robot Interaction*, page 129130, 2010.

E. Park, Q. Jenkins, and X. Jiang. Measuring trust of human operators in new generation rescue robots. In *Proceedings of the International Symposium on Fluid Power*, pages 489–492, 2008.

Behzad Sadrfaridpour, Jenny Burke, and Yue Wang. Human and robot collaborative assembly manufacturing: Trust dynamics and control. In *PRSS 2014 Workshop on Human-Robot Collaboration*, 2014a.

B. Sadrfaridpour, H. Saeidi, Y. Wang, and J. Burke. Modeling and control of trust in human and robot collaborative manufacturing. In *AAAI Spring Symposium Series*, 2014b.

Multiple target tracking based on particle filtering for safety in industrial robotic cells

Aitor Ibarguren, Iñaki Maurtua, Miguel Angel Pérez, Basilio Sierra: Multiple target tracking based on particle filtering for safety in industrial robotic cells. Robotics and Autonomous Systems 72: 105-113 (2015) [94]

Robotics and Autonomous Systems 72 (2015) 105–113

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Multiple target tracking based on particle filtering for safety in industrial robotic cells

A. Ibarguren [a,*], I. Maurtua [a], M.A. Pérez [a], B. Sierra [b]

[a] Autonomous and Intelligent Systems Unit, IK4-Tekniker, Spain

[b] Computer Sciences and Artificial Intelligent Department, University of the Basque Country, Spain

HIGHLIGHTS

- Security framework for collaborative industrial robotic cells.
- Laser based multiple target tracking system.
- Joint Probability Data Association Particle Filter implementation for security purposes.

ARTICLE INFO

Article history:
Received 18 December 2014
Received in revised form
21 April 2015
Accepted 4 May 2015
Available online 13 May 2015

Keywords:
Tracking
Robotics
Security
Particle filter

ABSTRACT

Human–robot collaboration is a key issue for the development of factories of the future, a space where humans and robots can work and carry out tasks together. Within this collaborative human–robot interaction paradigm, safety is one of most critical subjects. The presented paper describes a security framework for industrial robotic cells based on laser rangefinders able to detect and track obstacles around the robot, information that is used to modify robot's behaviour and guarantee safety. The system includes a Joint Probability Data Association Particle Filter to enhance the tracking process. The obtained results, implemented and tested in a real industrial robotic cell, show the suitability of the approach.

1. Introduction

Human–robot collaboration [1] is a key issue for the development of factories of the future, a space where humans and robots can work and carry out tasks together [2]. The introduction of robots in shop-floors allows sharing the different tasks of the production process between humans and machines, giving the chance to liberate human operators from repetitive and monotonous works to perform high added value operations. Even so, some tasks can be too complex to be performed by robots or too expensive to be automated as it requires the design and development of specific tools. Therefore, a collaborative environment where humans and robots can work side by side and share complex tasks in an open and fenceless environment is a significant goal to reach.

Within this collaborative human–robot interaction paradigm, safety is one of the most critical subjects [3,4] as it must be guaranteed that humans around the robot are not hurt during the execution of tasks. To this end, the robotic cell must be constantly aware of what is happening around it and monitor the humans' actions to change its behaviour when the operator is near the robot to ensure its safety.

This paper presents a security system for the detection and tracking of people in industrial robotic cells based on information received from laser rangefinders. The data sent from the laser measurement systems is initially analysed to detect the presence of people in the surroundings of the robot. This information is further used as input for a multiple target tracking algorithm based on a *Joint Probability Data Association Particle Filter* [5–7], which allows a continuous tracking and filtering of the detected obstacles. Finally, the system uses this filtered information to estimate the distance to the targets as well as their trajectory in order to modify robot's behaviour to ensure safety of human operators around it according to ISO 10218 standard. The laser based system is combined with a SafetyEYE [8], certified security sensor, enhancing its capabilities and creating a complete security framework. The proposed approach has been designed and developed within the scope

* Corresponding author.
E-mail addresses: aitor.ibarguren@tekniker.es (A. Ibarguren), b.sierra@ehu.es (B. Sierra).

http://dx.doi.org/10.1016/j.robot.2015.05.004

of X-ACT[1] European project, focused on cooperative industrial robots.

The main contribution of the presented paper is the implementation of a Joint Probability Data Association Particle Filter for security purposes, integrating it in the security framework of a real robotic cell, adding a continuous tracking of multiple mobile obstacles around the robot. This tracking process allows estimating the dynamics of the moving obstacles, offering a rich information source to the security framework. The experiments have also measured the estimation error as well as the computational cost of the algorithm, key point aspects for its integration in the security framework of a robotic cell.

The paper is organized as follows. In Section 2 the related work is presented. Section 3 poses the theoretical basis used in the presented paper. Section 4 is devoted to the proposed approach, including the hardware and the proposed architecture. In Sections 5–7 the different modules of the architecture are described in detail. Section 8 gives information about the implementation of the proposed approach. Section 9 shows the experimental results. Finally, Section 10 presents the conclusions as well as the future work to be done.

2. Related work

Human–robot collaboration has been a significant research topic since the initial steps of robotics. The constant introduction of robots in industrial environments, the creation of new compliant robots and the sensors available nowadays in the market, cheaper and more accurate, makes it an even more interesting research subject [3,4]. Different approaches can be found for safe human–robot interaction, from pre-collision to post-collision [9] approaches, using force sensors [9,10] or different vision systems [2,10,11], just to name few.

Focusing on industrial environments, Ribsky et al. [2] propose a real-time system based on multiple 3D imaging sensors, fusing their information to ensure the safety of people in close proximity to robots in industrial workcells. The system allows slowing down the robot and even stop the motion when people and robot approximate. On the other hand, Morata et al. [11] present a multiple Kinects based exteroceptive sensing framework to achieve safe human–robot collaboration during assembly tasks. Höcherl et al. [12] propose a system based on ultrasounds and two monocular cameras to detect human extremities during human–robot collaboration and adapt the robot behaviour to avoid physical contact. Lihui Wang et al. [13] developed a prototype system linked to robot controllers for adaptive robot control, with zero robot programming for end users. The system can alert an operator, stop a robot, or modify the robots trajectory away from an approaching operator.

Recently Pilz introduced the SafetyEYE [8] in the market, a system based on stereo vision and able to detect obstacles in different regions defined by the user. The system, designed in accordance with relevant norms and safety standards, allows the design of fenceless robotic cells. Augustsson et al. [14] presented a work using a safetyEYE to define flexible safety zones in shared working environments, reducing the robot speed and maintaining a safe distance to the human operator during the robot carries out nailing routines. Even so there are some drawbacks in the safetyEYE as this system only provides discrete information about the state of the robotic cell (safe, warning or alarm) and it cannot provide any information about the placement of the detected obstacles.

People detection and tracking is also a widely studied topic, used for a great variety of applications, from surveillance to human–robot collaboration. Within all the available literature, different approaches [15–17] pose the use of particle filtering and statistical data association for tracking multiple targets, adding robustness to the tracking process.

The presented paper proposes the use of multiple laser sensors to detect obstacles around the robot in industrial environments in order to modify robot's speed and ensure safety, adding a Joint Probability Data Association Particle Filter to enhance the tracking process.

3. Theoretical basis

This section describes the theoretical basis of the presented work, specifically the sequential Monte Carlo implementation of the JPDA algorithm. Appendices A and B include more information about the sequential Monte Carlo filtering [18,19] and Joint Probabilistic Data Association [5–7] (JPDA) respectively.

3.1. Sequential Monte Carlo JPDA

The presented paper implements the JPDA algorithm using a sampling method, specifically a sequential Monte Carlo filter as done in [15,16].

In this sequential approach, the state X^t is approximated by posterior distribution $P(X^t \mid Z^t)$. State of target i in time t is represented by a set S_i^t of M weighted particles. Therefore, the particle set is defined by the tuple $S_i^t = \{x_{i,k}^t, w_{i,k}^t\}_{k=1}^M$ of state $x_{i,k}^t$ and its associated weight $w_{i,k}^t$.

In each iteration of the algorithm, these weights are updated based on the observation Z^t using equation

$$w_{i,k}^t = \alpha \sum_{j=0}^{m_t} \beta_{j,i} P(z_j^t \mid x_{i,k}^t), \tag{1}$$

where α is a normalization constant and $\beta_{j,i}$ and $P(z_j^t \mid x_{i,k}^t)$ are probabilities presented in the previous subsection. Those weights are used in next steps to approximate state x_i^t.

The procedure of the implemented JPDA Particle Filter is given as:

1. Initialization
   - Set time $t = 0$
   - For each of the $i = 1 \dots N$ targets draw $k = 1 \dots M$ samples $x_{i,k}^0$ where weight $w_{i,k}^0 = 1/M$
2. For $i = 1 \dots N$ targets
   For $j = 0 \dots m_t$ observations
   - Calculate $\beta_{j,i} = \sum_{\theta \in \Theta_{j,i}} P(\theta \mid Z^t)$

   End for
   For $k = 1 \dots M$ particles
   - Calculate $w_{i,k}^t = \sum_{j=0}^{m_t} \beta_{j,i} P(z_j^t \mid x_{i,k}^t)$
   - Normalize weights $w_{i,k}^t = \frac{w_{i,k}^t}{\sum_{k=1}^M w_{i,k}^t}$

   End for
   - Estimate $x_i^t \approx \sum_{k=1}^M w_{i,k}^t x_{i,k}^t$
   - If *ESS* < *threshold* (*Effective Sample Size*), draw M samples with selection with replacement for target i
3. Set $t = t + 1$, go to Step 2.

[1] http://www.xact-project.eu/.

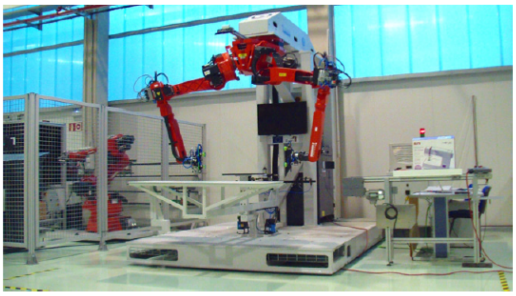


Fig. 1. Robotic cell.

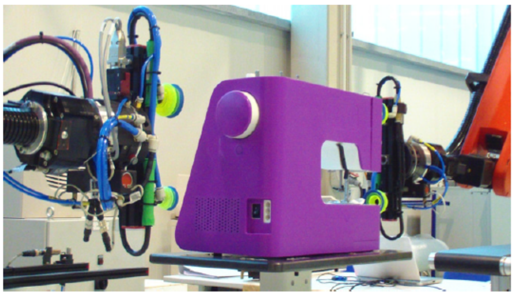


Fig. 2. Robot disassembling.

In this procedure EES [20] (Effective Sample Size) is calculated as

$$cv_t^2 = \frac{var(w_{i,k}^t)}{E^2(w_{i,k}^t)} = \frac{1}{M}\sum_{k=1}^{M}(M\, w_{i,k}^t - 1)^2, \qquad (2)$$

$$ESS_t = \frac{M}{1 + cv_t^2}, \qquad (3)$$

where M is the number of particles and $w_{i,k}^t$ is the weight of particle k of target i in time t.

Based on this discrete approximation of the posterior probability, the N objects are tracked along time.

4. Proposed approach

As stated in Section 1, the aim of the presented work is the design and development of an environment for collaborative industrial robots. To this end, a laser based system is proposed to track people around the robot and modify its behaviour and speed in order to ensure safety of human operators. This system, initially, gathers data from the sensors, processes this information in order to detect mobile obstacles, estimates the appropriate robot speed based on the detected obstacles and finally sends this information to the robot controller.

Besides the laser based system, which allows a smooth modification of the robot speed, a commercial security product is also included in the robotic cell to ensure safety, a SafetyEYE vision system from Pilz. This camera system allows creating safety and warning zones which are linked to the robot directly and can be used to stop or slow it down. This product gives digital signals as output and cannot provide information about the position of the obstacles, giving only the chance the define discrete security behaviours for the robot (*safe*, *warning* and *danger*). Moreover, it only provides a three colour beacon to sign its state.

The proposed approach complements the lacks of actual security products, creating a redundant security system with enhanced capabilities that is included in the overall safety strategy of the industrial robotic cell. On one hand, the proposed approach includes a multiple target tracking algorithm based on a Joint Probability Data Association Particle Filter, allowing a continuous tracking of the obstacles, feature that is not included in actual security systems like the SafetyEYE. On the other hand, the detection and tracking processes generate useful information like obstacles' position and velocity that this approach uses also for interaction purposes, offering this data through a visualization interface to the users, improving the collaborative experience from the users' point of view.

Next lines give information about the used hardware and the architecture of the proposed system.

Fig. 3. Sick LMS100 scanners.

4.1. Hardware

The central element of the industrial cell is a COMAU RML robot, see Fig. 1, a dual arm industrial robot used for assembly and disassembly tasks within X-ACT project. Besides the robot, the cell is completed with three workbenches and a tool exchanger with four different tools. These tools are used to disassemble a sewing machine, as shown in Fig. 2, a process that includes the unboxing of the sewing machine and the removal of different parts including small caps and bolts.

The detection of obstacles around the robot is based on two laser measurement systems, specifically two Sick LMS100 scanners, see Fig. 3. This indoor laser based sensor offers an operating range between 0.5 and 20 m, a field of view of 270°, angular resolution of 0.25° and a scanning frequency up to 50 Hz. Both sensors are placed on the base of the main workbench at a height of 45 cm (around the knees), facing opposites sides of the robotic cell, as shown in Fig. 4. This configuration allows covering 360° as well as avoiding occlusions near the workbench as readings are overlapped to a great extent. Moreover, laser scanners have been placed in different heights to avoid interferences between both sensors.

4.2. Architecture

The proposed approach divides the data processing in three different modules, converting laser data into robot speed commands in three steps, as shown in Fig. 5.

Next lines describe the features and behaviour of those modules:

- **Laser data analysis module**: This initial module gathers data coming from the two LMS100 laser scanners and analyses it in order to find people in the robotic cell. To this end, the received readings are compared with a previously recorded environment pattern to filter the static elements of the cell. For each received

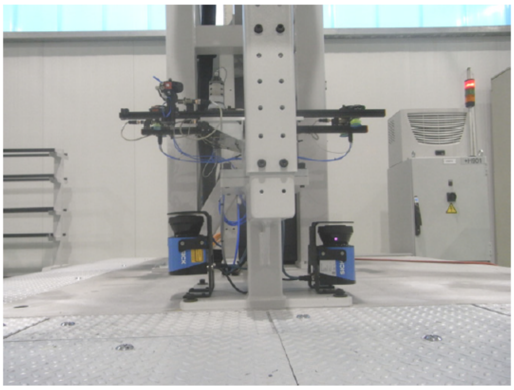


Fig. 4. Placement of laser scanners.

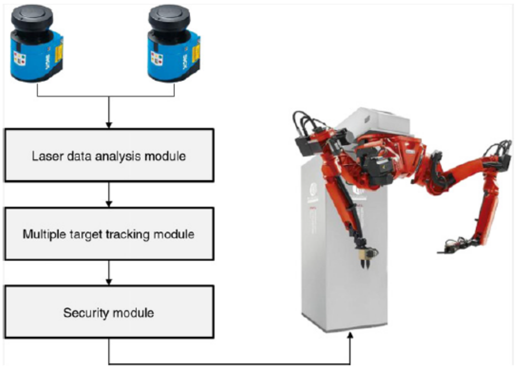


Fig. 5. Proposed architecture.

laser reading, a set of detected obstacles' coordinates are sent to the *multiple target tracking module* to be used as input for the tracking process.

- **Multiple target tracking module**: This second module receives as input the positions detected in the *laser data analysis module* and uses this information to perform a tracking process. Based on the scenario presented previously, where several people can enter the robot cell, a multiple target tracking algorithm is proposed, specifically a Joint Probabilistic Data Association Particle Filtering algorithm. The filtering allows performing the tracking of multiple objectives over time, overcoming the sensor and measurement noise. The output is the position and velocity of people around the robot that is sent to the next module.
- **Security module**: This last module receives the filtered position and velocity information of people around the robot and uses this input to modify the robot speed in order to ensure security and show a comfortable behaviour of the robot from the operator's point of view. Once the appropriate robot speed is estimated from the received position and velocity data, this information is sent to the robot controller to change its behaviour.

This architecture allows analysing laser scanner data, tracking people around the cell and estimating the appropriate speed for the robot to make humans feel comfortable around it. Besides, the addition of a Bayesian filtering method permits avoiding errors in measurements and sensor noise, an important aspect in an industrial environment.

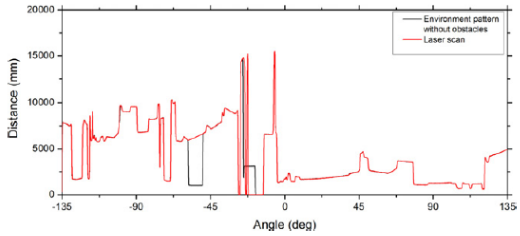


Fig. 6. Laser scan and environment pattern without mobile obstacles.

5. Laser data analysis module

The proposed approach bases people detection on data provided by laser measurement systems, as stated in Section 4. Based on the information provided by two scanners, this module detects and tracks people moving around the robot. Even so, there are static objects/obstacles around the robot which are not related with people and that are detected by the laser scanners, obstacles like work-tables, conveyor belts, walls or even the structure of the robot. In order to discriminate between the elements in the environment and people, an environment pattern has been introduced in the *laser data analysis module* which allows distinguishing static elements and mobile obstacles. This pattern of the environment is stored by the system and can be acquired at any time, to adapt this pattern to the changes in the layout.

Let us define r as a reading received from the laser scanner where $r = \{r_1, \dots, r_n\}$ is composed of n distances. On the other hand, let p denote a laser scanner reading containing the pattern of the environment (without any people or mobile obstacles around) where $p = \{p_1, \dots, p_n\}$. Fig. 6 shows the representation of a scan received from the sensor in red, as well as the environment pattern associated to the robot cell in black. Both readings show the same distances but in some small areas, zones where people or mobile obstacles are placed.

To extract the information about the mobile obstacles from the received laser reading, the difference between the scan and the static environment pattern is extracted. To this end, the information of the areas where the difference between them is less than a given threshold is filtered, as shown in the next equations

$$d(r, p) = \{d_i(r_i, p_i)\}_{i=1 \dots n}, \tag{4}$$

$$d_i(r_i, p_i) = \begin{cases} r_i & \text{if } |r_i - p_i| > \Delta \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where r_i and p_i are the ith distance measurement of the received scan r and the environment patter p respectively and Δ is the threshold value.

Fig. 7 shows the filtered signal obtained, where the pattern of the environment without any person around has been removed. The plot shows the mobile obstacles detected around the robotic cell, as well as some single outliers due to sensor noise.

Based on this filtered data, a segmentation process is applied in order to extract only the information related with the mobile obstacles. To this end, regions with a uniform non-zero distance measurement and a minimum length are identified, which allows removing sensor noise. Fig. 8 shows the laser data after the segmentation step, where two people can be observed, mobile obstacles that do not appear in the environment pattern used as reference.

To estimate the position of people around the robot cell, initially the mean angle α and distance l of each segment is calculated, a

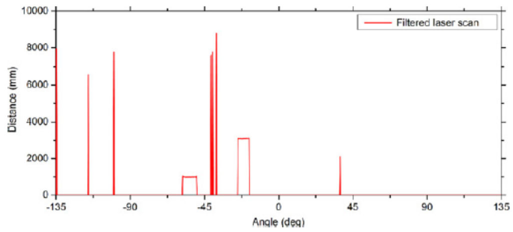


Fig. 7. Filtered laser scan.

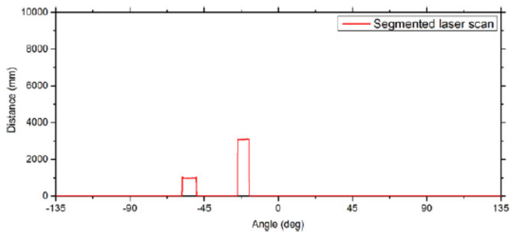


Fig. 8. Filtered and segmented laser scan.

pair of $\{\alpha, l\}$ of polar coordinates for each obstacle. These polar coordinates are then transformed to Cartesian space, obtaining a 2D Cartesian point $\{px, py\}$ for each polar coordinate. Therefore, the output of this *laser data analysis module* at time t is a set of m_t measurements:

$$Z^t = \{z_1^t, \ldots, z_{m_t}^t\}, \tag{6}$$

where each measurement z_j^t is a 2D point $\{px_j^t, py_j^t\}$ in Cartesian space

$$z_j^t = \{px_j^t, py_j^t\}. \tag{7}$$

To sum up, when a new laser scan arrives to the *laser data analysis module* it is filtered using an environment pattern and segmented to find the position of people around the robot. This position information is transformed in a second step to place the obstacles in the overall coordinate system. Finally, those Cartesian positions are sent to the *multiple object tracking module* for a further processing.

6. Multiple target tracking module

The proposed *multiple target tracking module* is based on sequential Monte Carlo filtering [18,19] and Joint Probabilistic Data Association [5–7] (JPDA). Sequential Monte Carlo filtering allows the tracking, overcoming noise effects inherent to measurements, while JPDA helps disambiguating the relation between the multiple targets and the acquired measurements. Next lines give information about the sequential Monte Carlo implementation of the JPDA filter.

6.1. System modelling

In the posed problem, a set of N targets will be moving around the robot at time t. Therefore, the state of the system at time t is defined by $X^t = \{x_1^t, \ldots, x_N^t\}$ where each target's state is described as

$$x_i^t = \{px_i^t, py_i^t\}, \tag{8}$$

where px_i^t and py_i^t are the 2D Cartesian coordinates of target i around the robot at time t respectively.

Additionally, the state transition is defined as

$$x_i^{t+1} = x_i^t + \Delta_t \dot{x}_i^t + V^t \tag{9}$$

$$\dot{x}_i^t = \{\dot{px}_i^t, \dot{py}_i^t\}, \tag{10}$$

where Δ_t is the time step, $\dot{x}_i^t$ is the dynamic part describing the variation of the state elements and V is an additive, zero mean Gaussian noise.

6.2. Likelihood evaluation

For likelihood evaluation, initially laser scans are analysed to detect possible obstacles around the robot, as exposed in Section 5. From this step, a set of m_t measurements are extracted, m_t possible obstacles. Those measurements form the observation Z^t, where each measurement z_j^t is defined by a 2D Cartesian coordinate of a possible target, as described by Eqs. (6) and (7).

In this stage of the implementation it is necessary to define the evaluation of probability $P(z_j^t \mid x_i^t)$ of observing measurement z_j^t given state x_i^t. To this end, initially the distance between the state x_i^t and the measurement z_j^t is calculated using the Euclidean distance of Cartesian coordinates:

$$dist\,(x_i^t, z_j^t) = \sqrt{(px_i^t - px_j^t)^2 + (py_i^t - py_j^t)^2}. \tag{11}$$

Finally the likelihood is calculated as the exponential of the distance as shown in the next equation

$$P(z_j^t|x_i^t) = e^{-\lambda\, dist\,(x_i^t, z_j^t)} \tag{12}$$

where λ is a weight factor.

Based on the presented likelihood evaluation, the sequential Monte Carlo JPDA algorithm estimates iteratively the process state.

7. Security module

This final layer calculates the appropriate robot speed based on the detected mobile obstacles' position. To provide a safety perception from the operators' point of view, the main idea is to decrease gradually robot speed as obstacles approximate the robot. As the described scenario considers multiple people walking around the robotic cell, it was decided to use the information about the closest obstacle to the robot as safety is the main criteria of the described approach. Taking into account ISO 10218 standard, when humans and robots work side by side robot speed must not exceed 250 mm/s, although a risk analysis must be carried out to study each robotic cell and application. In the presented work, this generic speed of 250 mm/s is considered the maximum speed when people approximate the robot. To ensure safety, the presented system is combined with the SafetyEYE in order to create a redundant safety architecture.

In the current layout, the maximum reach of the robot is below 1.75 m from the workbench although the typical working area of the robot is below 1 m. Taking it into account, three different distances have been considered; above 5 m as a safe area, between 3 and 5 m as a warning area and below 3 m as a risky area. This distribution allows a safe approach when a human enters the robotic cell.

Next lines describe the main features of the security speed calculation:

- The output of the JPDA Particle Filter is constantly analysed, looking for the closest obstacle from the centre of the robot. This distance is the one used to estimate the robot speed.

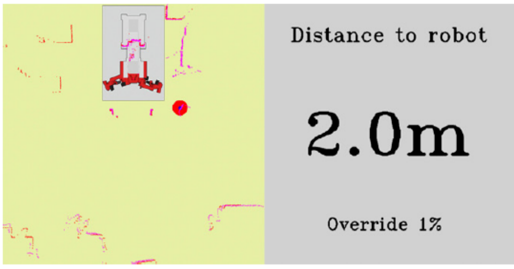


Fig. 9. Visualization interface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- The robot speed is reduced gradually as obstacles approximate the robot. While people maintain a distance above 5 m the robot speed is set to its maximum value. From 5 to 3 m the speed decreases linearly from its maximum to 250 mm/s, the suitable speed for collaborative spaces. Between 3 and 2 m the override is linearly reduced until a 1 mm/s, which practically maintains the robot almost stopped. Below 2 m the override is maintained in a 1 mm/s.
- The *Security module* constantly sends the new robot speed to the robot control, modifying it in less than 50 ms.

Through this last module, the output of the *Multiple Target Tracking Unit* is analysed in order to estimate a safe robot speed and modify it in real time via the robot controller.

8. Implementation

In a next step, the proposed approach has been implemented and integrated in the existing robotic cell at IK4-TEKNIKER's facilities. The algorithm has been developed in C++ and integrated in a security application which uses the results of the detection and tracking to calculate the proper speed of the robot and sends it to the robot control through Ethernet connection. All the applications have been executed in a laptop computer Intel Core i7 2.80 GHz and 4 GB of RAM, under Windows 7 operating system.

The developed application also includes a visualization tool which presents a map of the robotic cell with the robot and its platform on the upper part of the image, the readings of the laser scanners and the observations (in blue) and targets (in red), as shown in Fig. 9. Moreover, the visualization includes information about the nearest obstacle and the estimated robot override to fulfil the security requirements. In the case of the shown image, an obstacle can be seen at the right side of the map, near the right edge of the robot platform, as well as the distance to the robot of 2 m and the estimated override of the robot of 1%.

In terms of performance, the algorithm executes with a frequency of 25–30 Hz with 3–5 people walking around the cell, a usual situation in the shopfloor. The visualization updates each 5 cycles, 5–6 frames per second, as it became too time consuming with a higher frame rate.

The system has been thoroughly used during all the tests and developments carried out in X-ACT European project, integrating it in the security chain of the robotic cell.

9. Experimental results

To test the efficiency of the presented approach, different aspects of the algorithm have been measured in order to perform a complete evaluation of the system. Parameters like accuracy, behaviour with occlusions and execution time have been measured in different experiments. All the experiments have been performed while the robot disassembles a sewing machine, a process of around ten minutes where the sewing machine is unboxed and eight different parts of the sewing machine are removed, including small caps and bolts.

Table 1
Summary of results.

Summary of results		
Detection error	μ	32.02 mm
	σ	40.72 mm
Execution time	1 observ.	8 ms
	7 observ.	32 ms
	12 observ.	100 ms

In order to measure the accuracy of the JPDA Particle Filter, a set of different paths have been defined around the robot cell while the robot disassembles. Those paths were followed by different human operators in five different recording sessions, including from one to three moving targets in each session. These subjects followed the defined paths walking at different speeds, from slow walk to fast steps, while the laser scanners were receiving measurements acquired at a height of 45 cm (around the knees), as stated previously. Fig. 10 shows the different tracking results in red, green and blue as well as the nominal path followed by test subjects in black in three of the recording sessions.

Results show an oscillating tracking process along the nominal path. This detected oscillation is caused by the self occlusion that takes place while walking, as one leg hides the other one in most areas of the robotic cell. In the same way, the laser data segmentation and detection steps also increase this oscillation when the target is far from the centre of the robot, as the further the target is the less laser points the system receives to detect the observations. The analysis of the obtained results showed a mean position estimation error μ of 32.02 mm with a standard deviation σ of 40.72 mm, that corresponds to the distance between legs.

Besides the self occlusions, during the path following with multiple targets the JPDA Particle Filter managed the occlusions between them. In those cases, the algorithm could track the hidden target in almost all the cases, losing the target's position around 0.5 s in the worst cases, detecting these missing targets immediately after the occlusion.

The execution time of the different steps of the algorithm were also measured. Specifically two different execution times were recorded, the laser data analysis time and the tracking time through JPDA Particle Filter algorithm. During the experiments observations were added in the robot cell gradually, more possible targets in the environment to be detected. Fig. 11 shows the obtained execution times, the detection time in the analysis of laser readings in green and the execution time of the JPDA Particle Filter algorithm in red.

Results show, as expected, that the detection time keeps constant, below 5 ms, as observations are introduced. In the case of the execution time of the tracking algorithm, it remains nearly constant from 1 to 7 observations, between 8 and 32 ms. Then the execution time increases gradually, reaching an execution time of around 100 ms with 12 observations. Above this number of observations, the execution time increases quadratically, making it unfeasible to perform the tracking process. The number of joint association events, which grows near quadratically with the number of observations, causes this huge increase in the execution time, aside of the number of particles included in the algorithm. The experiment shows the poor scalability of the algorithm in crowded environments, making it necessary to apply a 'relaxed' creation of joint association events or a parallelization of the code to speed up calculus time to allow a real time execution of the algorithm under these conditions.

Finally, Table 1 summarizes the different results obtained during the experiments, including detection error distances and execution times.

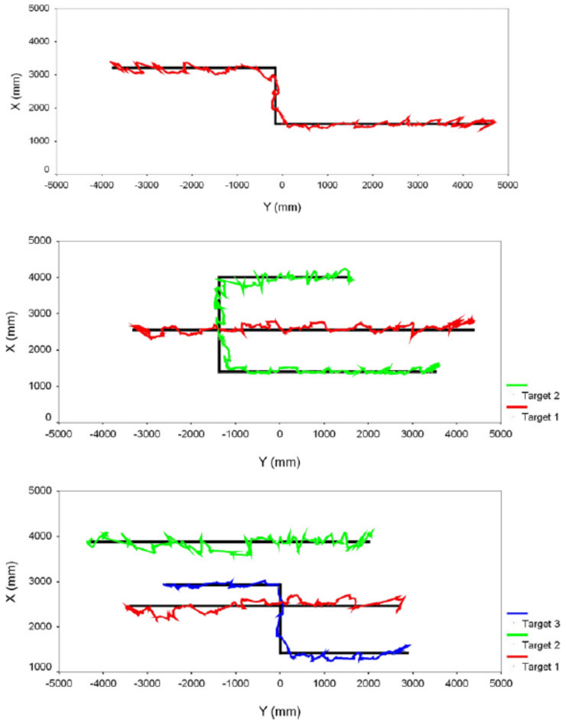


Fig. 10. Tracking process for one, two and three targets.

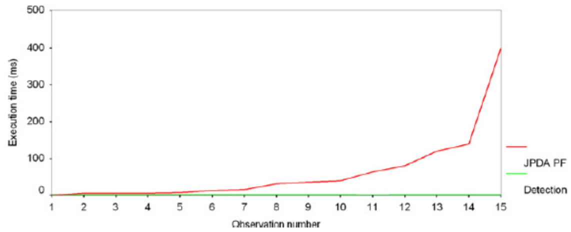


Fig. 11. Execution time of laser data analysis and JPDA Particle Filter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

10. Conclusions and future work

The presented paper describes a security framework for collaborative industrial robots. The presented system, based on laser rangefinders, is able to detect and track people in a robotic cell in real time and modify robot's behaviour in order to ensure safety of human operators. The data sent from the laser measurement systems is initially analysed to detect the presence of people in the robotic cell using an environment pattern to this end. This information is further used as input for a multiple target tracking algorithm based on a Joint Probability Data Association Particle Filter. Finally, the system uses this filtered information to estimate the distance to the targets as well as their trajectory in order to modify the robot's speed according to ISO 10218 standard.

To test the suitability of the approach, different experiments have been carried out. On one hand, the detection and tracking algorithms have been tested, measuring the robustness and precision of the approach. A mean detection error of around 30 mm have been measured during the different experiments, minimizing also the effects of occlusions when various targets are moving around the robot. In the same way, the scalability of the algorithm has also been tested. The results show the stability of the algorithm while dealing with up to 10 observations/targets. With a higher number of observations, the execution time starts to increase quadratically, making it infeasible its use in real time with the actual implementation. Although the management of 10 or less targets is enough for the presented scenario, this is an aspect to be improved in future developments. The parallelization of the JPDA particle filter code could be an interesting path to follow, as well as the modification of the JPDA algorithm to decrease the number of joint association events in crowded situations where more than 15–20 observations are detected.

Finally, the current implementation of the systems uses two laser rangefinders as environment data source. Future steps plan to include 3D vision sensors to allow discriminating between humans and different mobile obstacles in the environment. The use of Joint Probability Data Association Particle Filter allows an easy integration and fusion of both information sources, allowing the creation of a detection and tracking system which mixes different sensor types. Moreover, different tests are being designed within X-ACT project in order to evaluate the security strategy from the user's point of view, including aspects like safety perception, comfort or the best human–robot communication channels for industrial environments.

Acknowledgements

This work has been performed within the scope of the project X-ACT: Expert cooperative robots for highly skilled operations for the factory of the future. X-ACT is a Small or Medium-scale focused research project supported by the European Commission in the 7th Framework Programme (FP7-314355). For further information see http://www.xact-project.eu/.

Appendix A. Particle filter

Particle filters, also known as sequential Monte Carlo methods (SMC), are sequential estimation techniques that allow estimating unknown states x^t at time t from a collection of observations $z^{1:t} = \{z^1, \dots, z^t\}$ measured along time. The state-space model is usually described by state transition and measurement equations

$$x^t = f(x^{t-1}, v^{t-1}), \tag{A.1}$$

$$z^t = g(x^t, u^t), \tag{A.2}$$

where f and g are the state evolution and observation model functions respectively and v and u denote the process and observation noise respectively.

Based on these previous equations, particle filters allow approximating the posterior density (PDF) by means of a set of particles $\{x^t_{(i)}\}_{i=1,\dots,n}$ using equation

$$p(x^t|z^{1:t}) = \sum_{i=1}^{n} \omega^t_{(i)} \delta(x^t - x^t_{(i)}), \tag{A.3}$$

where each particle $x^t_{(i)}$ has an importance weight $\omega^t_{(i)}$ associated and δ is the Kronecker delta. These weights are computed following equation

$$\omega^t_{(i)} = \omega^{t-1}_{(i)} \frac{p(z^t|x^t_{(i)})p(x^t_{(i)}|x^{t-1}_{(i)})}{q(x^t_{(i)}|x^{0:t-1}_{(i)}, z^{0:t})}, \tag{A.4}$$

where $p(z^t|x^t_{(i)})$ is the likelihood function of the measurements z^t and finally $q(x^t_{(i)}|x^{0:t-1}_{(i)}, z^{0:t})$ is the proposal density function.

Based on the previously presented equations the particle set evolves along time, changing the weights of the particles and resampling them in terms of the observations.

Particle filtering provides a robust tracking framework when dealing with non-linear and non-Gaussian state and observation functions as it considers multiple state hypotheses simultaneously.

Appendix B. Joint probabilistic data association

Consider the problem of tracking N objects. $X^t = \{x^t_1, \dots, x^t_N\}$ represents the state of these N objects in time t, where X^t is a random variable over the entire state space. In the same way, $Z^t = \{z^t_1, \dots, z^t_{m_t}\}$ denotes a set of m_t measurements acquired at time t by the system. The main issue of the presented problem is how to assign the detected m_t measurements to the N targets to be tracked.

Let θ denote a joint association event composed by a set of pairs $(j, i) \in \{0, \dots, m_t\} \times \{1, \dots, N\}$, where each θ determines uniquely which measurement of time t is assigned to each target. In the algorithm, artificial measurement z_0 has been added to represent false alarms or clutter (e.g. there is no measurement for target n). Besides, Θ represents the whole set of θ events. Finally, let $\Theta_{j,i} \in \Theta$ denote all the valid joint association events where measurement j and target i are coupled.

Based on the defined elements, JPDA algorithm computes at each time step t the posterior probability that measurement j has been caused by target i as

$$\beta_{j,i} = \sum_{\theta \in \Theta_{j,i}} P(\theta \mid Z^t), \tag{B.1}$$

where $P(\theta \mid Z^t)$ represents the probability of assignment θ given the set of measurements Z^t. To compute it, the assumption that the problem is Markovian is performed and Bayes' rule is applied, as shown in the next equations

$$\begin{aligned} P(\theta \mid Z^t) &= P(\theta \mid Z^t, Z^{t-1}) \\ &\overset{\text{Markov}}{=} P(\theta \mid Z^t, X^t) \\ &\overset{\text{Bayes}}{=} \alpha\, P(Z^t \mid \theta, X^t)\, P(\theta \mid X^t), \end{aligned} \tag{B.2}$$

where α is a normalization constant, $P(Z^t \mid \theta, X^t)$ is the probability of observing measurements Z^t given assignment θ and state X^t and finally $P(\theta \mid X^t)$ is the probability of assignment θ given state X^t.

The estimation of term $P(Z^t \mid \theta, X^t)$ requires some considerations around the false alarms in the measurements. Let P_D and P_F denote the detection probability and the false alarm probability respectively. In the same way, the number of false alarms in an association event θ is given by $(m_t - |\theta|)$, where m_t is the number of measurements at time t and $|\theta|$ determines the number of non false alarm targets. The probability of $P(Z^t \mid \theta, X^t)$, assuming that each measurement is detected independently, is calculated as

$$P(Z^t \mid \theta, X^t) = P_F^{(m_t - |\theta|)} \prod_{(j,i) \in \theta} P(z^t_j \mid x^t_i), \tag{B.3}$$

while the probability of an assignment event θ conditioned to a state X^t is approximated by

$$P(\theta \mid X^t) = P_D^{|\theta|}\, (1 - P_D)^{(N - |\theta|)}\, P_F^{(m_t - |\theta|)}. \tag{B.4}$$

Finally, to estimate the probability $P(x^t_i \mid Z^t)$ of a target i at time t given the measurements Z^t, assignment probabilities $\beta_{j,i}$ are used as it is not known which measurement has been caused by each target. Therefore the probability $P(x^t_i \mid Z^t)$ is calculated as

$$P(x^t_i \mid Z^t) = \alpha \sum_{j=0}^{m_t} \beta_{j,i}\, P(z^t_j \mid x^t_i)\, P(x^t_i), \tag{B.5}$$

where α is a normalization factor, $P(z^t_j \mid x^t_i)$ is the probability of observation j given the state of target i at time t and $P(x^t_i)$ is the probability of state of target i at time t.

Based on the JPDA algorithm, the proposed approach needs to define the model for estimating $P(x^t_i \mid x^{t-1}_i)$ and $P(z^t_j \mid x^t_i)$ according to the posed tracking problem.

References

[1] H.A. Yanco, J.L. Drury, A Taxonomy for Human–robot Interaction, Tech. rep., in: Proceedings of the AAAI Fall Symposium on Human–Robot Interaction, 2002.

[2] P. Rybski, P. Anderson-Sprecher, D. Huber, C. Niessl, R. Simmons, Sensor fusion for human safety in industrial workcells, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2012.
[3] J. Heinzmann, A. Zelinsky, Quantitative safety guarantees for physical human–robot interaction, I, J. Robot. Res. 22 (7–8) (2003) 479–504.
[4] A. Pervez, J. Ryu, Safe physical human robot interaction-past, present and future, J. Mech. Sci. Technol. 22 (3) (2008) 469–483.
[5] Y. Bar-Shalom, T.E. Fortmann, Tracking and Data Association, in: Mathematics in Science and Engineering, vol. 179, Academic Press, 1988.
[6] I. Cox, A review of statistical data association techniques for motion correspondence, Int. J. Comput. Vis. 10 (1) (1993) 53–66.
[7] R. Karlsson, F. Gustafsson, Monte Carlo data association for multiple target tracking, in: IEE Target Tracking: Algorithms and Applications, 2001.
[8] Pilz safetyeye. URL http://www.pilz.com.
[9] S. Lu, J. Chung, S. Velinsky, Human–robot collision detection and identification based on wrist and base force/torque sensors, in: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005, ICRA 2005, 2005, pp. 3796–3801.
[10] A. De Santis, V. Lippiello, B. Siciliano, L. Villani, Human–robot interaction control using force and vision, in: C. Bonivento, L. Marconi, C. Rossi, A. Isidori (Eds.), Advances in Control Theory and Applications, in: Lecture Notes in Control and Information Sciences, vol. 353, Springer, Berlin, Heidelberg, 2007, pp. 51–70.
[11] C. Morato, K.N. Kaipa, B. Zhao, S.K. Gupta, Toward safe human robot collaboration by using multiple kinects based real-time human tracking, J. Comput. Inf. Sci. Eng. 14 (1) (2014) 011006.
[12] J. Höcherl, T. Schlegl, An image based algorithm to safely locate human extremities for human–robot collaboration, in: ICIRA (2), in: Lecture Notes in Computer Science, vol. 7507, Springer, 2012, pp. 164–175.
[13] L. Wang, B. Schmidt, A.Y. Nee, Vision-guided active collision avoidance for human–robot collaborations, Manuf. Lett. 1 (1) (2013) 5–8.
[14] S. Augustsson, L.G. Christiernin, G. Bolmsjö, Human and robot interaction based on safety zones in a shared work environment, in: Proceedings of the 2014 ACM/IEEE International Conference on Human–Robot Interaction, ACM, 2014, pp. 118–119.
[15] D. Schulz, W. Burgard, D. Fox, A. Cremers, Tracking multiple moving targets with a mobile robot using particle filters and statistical data association, in: IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA, vol. 2, 2001, pp. 1665–1670.
[16] M. Jaward, L. Mihaylova, N. Canagarajah, D. Bull, Multiple object tracking using particle filters, in: 2006 IEEE Aerospace Conference, 2006, p. 8.
[17] B. Benfold, I. Reid, Stable multi-target tracking in real-time surveillance video, in: CVPR, 2011, pp. 3457–3464.
[18] A. Doucet, N. De Freitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, 2001.
[19] J.H. Kotecha, P. Djuric, Gaussian particle filtering, IEEE Trans. Signal Process. 51 (10) (2003) 2592–2601.
[20] J. Liu, R. Chen, T. Logvinenko, A theoretical framework for sequential importance sampling and resampling, in: A. Doucet, N. De Freitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, 2001.

Aitor Ibarguren received his M.Sc. in Computer Sciences in 2004 and Ph.D. in Computer Sciences in 2008 at the University of the Basque Country. He is a researcher in the Smart and Autonomous System Unit in IK4-Tekniker. His research interest and expertise include robotics, artificial intelligence and artificial vision.

Iñaki Maurtua is Electrical Engineer with more than 25 years of experience. He is the Head of the Smart and Autonomous System Unit in IK4-Tekniker and has participated and coordinated several European and national projects dealing with robotics and machine vision, wearable computing, system monitoring and flexible manufacturing systems.

Miguel Angel Pérez is Electrical Engineer of the Smart and Autonomous System Unit in IK4-Tekniker. He has participated in many automation projects, most of them involving the use of robots. His research interest and expertise include robotics, automation and artificial vision.

Basilio Sierra is Full Professor in the Computer Sciences and Artificial Intelligence Department at the University of the Basque Country. He received his B.Sc. in Computer Sciences in 1990, M.Sc. in Computer Science and Architecture in 1992 and Ph.D. in Computer Sciences in 2000 at the University of the Basque Country. He is the director of the Robotics and Autonomous Systems Group in Donostia-San Sebastian. He is presently a researcher in the fields of Robotics, Computer Vision and Machine Learning, and he is working on the development of different paradigms to improve classification behaviours; he has written more than 25 journal papers in those fields, as well as more than 100 conference contributions and more than 30 book chapters.

7.2. Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques

Torsten Felsch, Gunnar Strauss, Carmen Perez, José M. Rego, Iñaki Maurtua, Loreto Susperregi, Jorge R. Rodríguez: Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques. Robotics 4(2): 103-119 (2015) [102]

Robotics **2015**, *4*, 103-119; doi:10.3390/robotics4020103

OPEN ACCESS

robotics

ISSN 2218-6581

www.mdpi.com/journal/robotics

Article

Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques

Torsten Felsch [1,*], Gunnar Strauss [1], Carmen Perez [2,*], José M. Rego [2], Iñaki Maurtua [3], Loreto Susperregi [3] and Jorge R. Rodríguez [4]

[1] Fraunhofer IFF, Sandtorstrasse 22, 39106 Magdeburg, Germany;
E-Mail: gunnar.strauss@iff.fraunhofer.de

[2] TECNATOM, Avda. de Montes de Oca 1, San Sebastián de los Reyes (MADRID) 28703, Spain;
E-Mail: jmrego@tecnatom.es

[3] IK4-Tekniker, Iñaki Goenaga 5, Eibar 20600, Spain; E-Mails: inaki.maurtua@tekniker.es (I.M.); loreto.susperregi@tekniker.es (L.S.)

[4] TORRESOL Energy, Avenida Zugazarte 61, Las Arenas 48930, Spain;
E-Mail: jorge.rodriguez@torresolenergy.com

* Authors to whom correspondence should be addressed; E-Mails: torsten.felsch@iff.fraunhofer.de (T.F.); cperez@tecnatom.es (C.P.).

Academic Editor: Huosheng Hu

Received: 22 December 2014 / Accepted: 18 March 2015 / Published: 27 March 2015

Abstract: Concentrated solar power (CSP) plants are expansive facilities that require substantial inspection and maintenance. A fully automated inspection robot increases the efficiency of maintenance work, reduces operating and maintenance costs, and improves safety and work conditions for service technicians. This paper describes a climbing robot that is capable of performing inspection and maintenance on vertical surfaces of solar power plants, e.g., the tubes of the receiver in a central tower CSP plant. Specifically, the service robot's climbing mechanism is explained and the results of the nondestructive inspection methods are reviewed. The robot moves on the panels of the receiver in the tower and aligns the sensors correctly for inspection. The vertical movement of the climbing kinematics is synchronized with the movement of the tower's crane. Various devices that detect surface defects and thickness losses inside the tube were integrated into the robot. Since the tubes are exposed to very high radiation, they need to be inspected regularly.

Keywords: service robot; climbing robot; solar power plant; maintenance; nondestructive testing; inspection

1. Introduction

Efficient and effective inspection of large capital-intensive systems such as chemical, steel and power plants establishes the prerequisites for high plant safety. Routine inspections that ensure scheduled availability and system efficiency generate know-how relevant to maintenance. Routine inspection thus has a significant impact on the maintenance costs throughout an industrial plant's operating period. Concentrated solar power (CSP) plants that generate electricity are typical examples of capital-intensive plants. Such power plants are expansive and contain an extremely high number of components and parts.

The inspection of large plants requires considerable time and labor. The high complexity and enormous dimensions of such plants translate into specific requirements and challenges, namely:

- dangerous or difficult to access work environments,
- tight inspection schedules,
- a large number of inspection points, and
- diverse inspection technologies.

Service robots that autonomously inspect CSP plants were developed in the MAINBOT project funded in the EU's Seventh Framework Programme for Research. The robots must execute various tasks, the most important of which are:

- safe and autonomous movement and navigation in a structured environment of horizontal and vertical inspection areas,
- mobile manipulation of different tools and testing equipment for maintenance and inspection, and
- sensor data fusion for comprehensive evaluation of the sensor data.

The service robots will be used in two different types of solar power plants: a CSP plant and a central tower CSP plant. Torresol Energy operates several such plants in Southern Spain. These expansive installations present a major challenge for inspection. Two different robot systems that inspect solar power plants were developed in the MAINBOT project:

- a mobile robot that inspects parabolic mirrors and
- a climbing robot that inspects the receiver of a central tower CSP plant.

This paper focuses on the climbing robot for the central tower of the GEMASOLAR plant (see Figure 1). The main function of the climbing robot is to transport and position the inspection system on the desired tube panel of the GEMASOLAR tower.

The GEMASOLAR solar power plant—a central tower CSP plant—has 2650 heliostats (a set of reflectors that follow the sun automatically to concentrate solar radiation in the receiver atop the tower), and a 140-meter-high receiver tower. A heat exchanger constructed of tubes that convert solar radiation into thermal energy is located atop the tower at a height of 130 m. The heat exchanger has a polygonal structure consisting of sixteen panels. Each panel consists of over sixty special metal tubes. Molten salt is pumped through the tubes to convert solar energy into thermal energy and to store the thermal energy in big tanks.

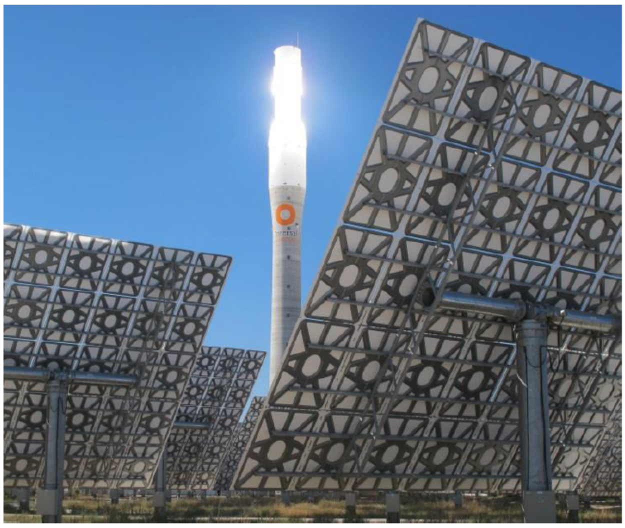


Figure 1. GEMASOLAR Power Plant, Torresol Energy property, © Torresol Energy.

Since the surface temperature of the tubes can reach hundreds of degrees Celsius and cause high stress on the material of the tubes and the components of the plant, the heat exchanger must be inspected regularly for possible surface coating wear and defects in the deeper structure of the tubes.

2. Related Work

The automation of inspection in industrial plants has been a major challenge for many years. Many approaches to robot design exist for a broad range of applications and environments. The automatic access of vertical surfaces of structures opens additional maintenance, repair and servicing capabilities. Vertical mobility is used in numerous commercial applications from window cleaning to pipeline, bridge and tank inspection. The wide range of potential applications has generated a variety of different methods for robot locomotion and adhesion. In the literature, three main types of climbing applications have been studied and developed: wall climbing, pole climbing and rope climbing robots.

Specific technologies that have been employed include wheeled robots with a frictional adhesion that move on cables to inspect or repair bridges and power lines [1]. Cable climbing robots are suitable for hanger ropes of long span suspension bridges. Goldman [2] presents a robot that moves on poles at construction sites and on scaffolds. The serpentine robot prototype climbs a pole by converting the oscillating motion of its joints into rolling motion of its entire body.

Several studies present climbing robots for flat vertical surfaces with magnetic properties. Fernández *et al.* [3] present a prototype wall-climbing robot for tank inspection. Similarly, Eich *et al.* [4] propose a lightweight crawler with magnetic wheels including hybrid legwheels and a passive magnetic tail, which can climb tall metallic walls and navigate small obstacles. Weld inspection of vessels is another application for robots with magnetic adhesion [5].

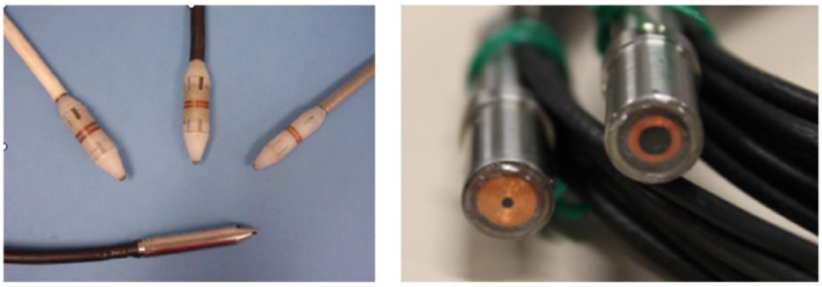
Several climbing systems using electroadhesion technology have been developed to enable wall climbing [5–8]. Inspired by climbing animals, these robots use van der Waals forces of attachment. This dry adhesion is useful for remote monitoring or inspection of concrete pillars or other structures such as bridges and tunnels.

The most common type of climbing robots uses vacuum suction cups to adhere to flat and homogenous surfaces, e.g., for automated cleaning of high-rise building facades [9,10]. Guimaraes *et al.* [11] propose a small, remote-controlled, lightweight climbing machine for walls, ceilings or rounded surfaces. Vacuum forces induced in a central vacuum chamber surrounded by a flexible foam seal, holds the robot to hard surfaces. Leibbrandt *et al.* [12] present a specific version of vacuum adhesion with a climbing robot for routine inspections of reinforced concrete structures. It generates suction by creating an air vortex in a central tube. Only its wheels need to be in contact with the surface being climbed.

This paper discusses a new variation of a climbing robot that employs suction adhesion and is specifically able to move and operate on the surface of the solar power plant's heat exchanger.

Eddy current (ET) inspection is often used to detect corrosion, erosion, cracking and other changes in tubing [13]. Heat exchangers and steam generators, which are used in power plants, have thousands of tubes that have to be kept from leaking. This is especially important in nuclear power plants where reused, contaminated water must be prevented from mixing with freshwater that is returned to the environment. Eddy current testing and related remote field-testing are high-speed inspection methods for such applications. Both electromagnetic methods are applied to thin tubes (up to 3 mm thick), specifically ferromagnetic materials (stainless steel, Inconel, *etc.*) in the former and nonmagnetic materials (carbon steel) in the latter.

The test is performed with a bobbin coil that produces an electromagnetic field in the tube. This enhances the sensitivity of inspection of the inside diameter of a tube where defects are most likely to occur. By using multiple frequencies, 100% of a tube's wall can be scanned to detect flaws at various depths. When the probe is moved across uneven surfaces, the electromagnetic field is distorted as a function of the size and location of the asperity relative to the probe. This distortion in the magnetic field alters the coil impedance relative to the asperity. This coupling effect of the eddy current probe with the material makes it essential that tubes be properly cleaned and their dimensions are known prior to inspection. The key to reliable detection of the extent and depth of such defects is correct selection of the probe material and precise calibration. Calibration must therefore be done with exactly the same material as that inspected later. Since they are accessible, tubes could be inspected from the inside. This is not possible in the receiver, however. That is why, the use of nondestructive testing (NDT) methods from the outside has been proposed in MAINBOT. This requires a different type of ET sensor. The sensor concept used in [14] (in-service oxide layer measurement in fuel rods) was modified for the inspection of receiver tubes. High frequency and low frequency ET probes were designed to measure coating and tube thickness, respectively (see Figure 2). Many other applications employ ET inspection from the outside surface (railway inspection, turbine and compressor blade inspection [15], *etc.*).



(a) **(b)**

Figure 2. (**a**) Conventional ET bobbin coil; (**b**) ET sensors proposed in MAINBOT.

3. Climbing Robot Design

3.1. Robot Specifications

The climbing robot is designed to inspect vertical structures and can be deployed in different areas of the plant. The prototype climbing robot built is intended to inspect the tubes of a central tower CSP plant's receiver.

Figure 3 pictures the architecture of the climbing robot as well as the existing prototype. The climbing robot is moved vertically on the tower by the crane atop the receiver tower. The robot system includes a climbing mechanism that can be synchronized with the crane to bring the robot in the desired position. The robot is attached to the tower by arrays of vacuum suction cups. Four arrays are attached to the robot frame (the outer contact elements) and one array is integrated in the climbing mechanism (the center contact element). Since every climbing action entails aligning every one of the oval suction cups with the tubes of the panels, an optical sensor (tube scanner) scans the profile of the panel to determine the position of the tubes.

To implement the climbing robot design effectively, its movement and performance were simulated on a virtual tower model. The undercut of the panels on the receiver tower demands special attention when the inspection system is brought into the start position.

The features of the MAINBOT climbing robot are:

- robot weight: ~280 kg
- robot dimensions: 2.3 m × 1.6 m × 0.8 m
- high payload (only limited by the crane)
- obstacle navigation (climbing mechanism)
- adaptability to different surfaces, material and structures (interchangeable and adjustable contact elements)
- accurate tool (sensor) positioning thanks to servo controlled motion system
- fully and semi-automated operation modes with Web UI interface
- synchronized movement of robot and crane

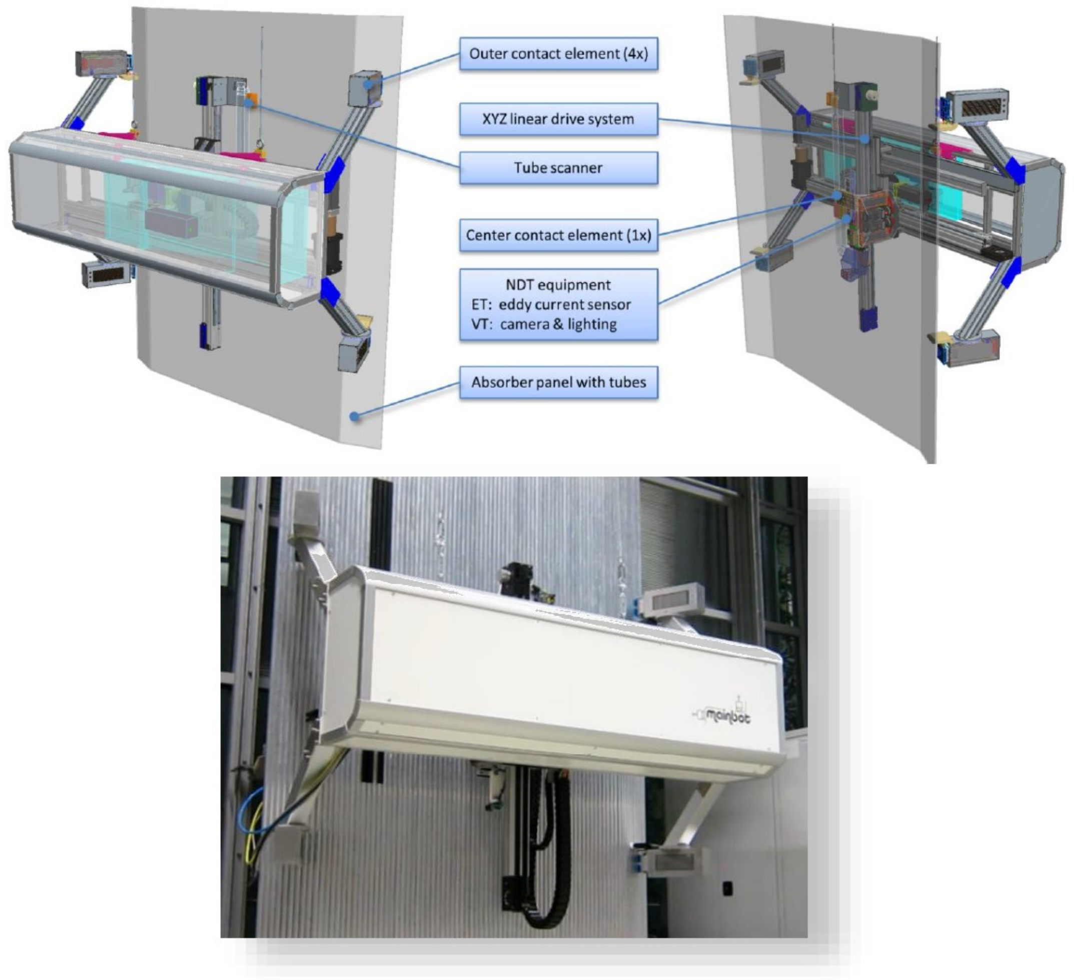


Figure 3. Service robot for absorber panel inspection (schematic view; picture of prototype).

3.2. Robot Climbing Mechanism

The robot includes a climbing mechanism with variable step size. The load of the robot is borne by the crane atop the tower (or the object being inspected). The robot is secured horizontally by its contact elements to prevent the system from swinging. The robot is in contact with both adjacent panels in order to leave one entire panel free for inspection.

The telescopic mechanism makes step width greater than the robot's height. The climbing kinematics is also used to move sensors. The diagram below illustrates the robot's climbing procedure (see Figure 4).

Arrays of vacuum suction cups at the contact elements establish contact to the tubes of the receiver.

The outer elements are in contact with the adjacent panels during inspection. The vacuum grippers are aligned with the tubes automatically in a two-stage contact process. Soft bumpers at the elements ensure collision-free movement of the robot to protect the tube surface (see Figure 5a). The contact with the adjacent panels leaves the entire area between the robot and the panel free, thus, allowing inspection by the NDT sensors (see Figure 5b).

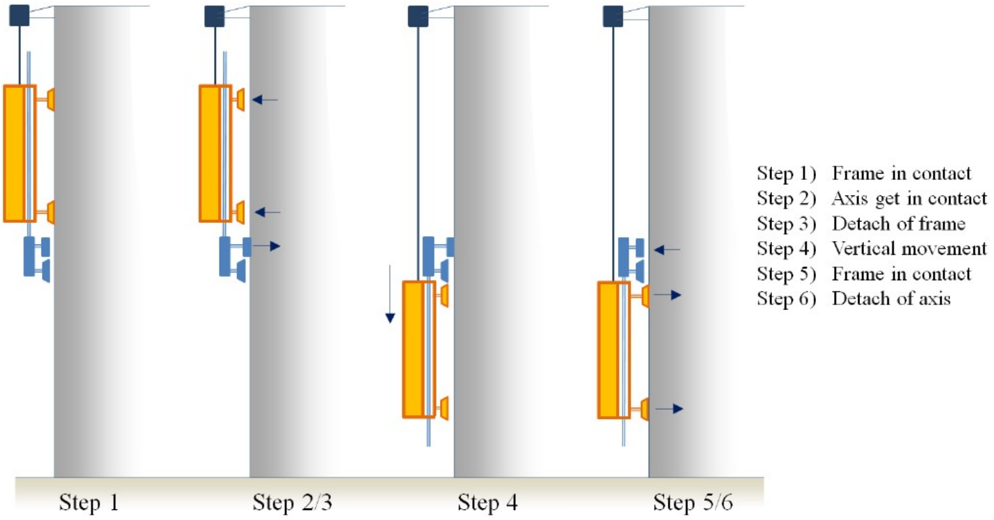


Figure 4. Stepping procedure of the climbing robot.

The central contact element can move in XYZ-direction relative to the robot frame. The element is in contact during vertical robot movement. A wire sensor measures any potential vertical deviation of the crane position and the vertical robot position, which can arise when their speeds differ. The position signal is used to synchronize the crane and the robot.

(a) **(b)**

Figure 5. (**a**) Contact element of the climbing robot; (**b**) Working area of the sensor system.

The performance of the climbing robot prototype was validated using an indoor mock-up of the receiver panel, which is composed of an array of aluminum tubes. Various tests were performed to validate the contact process, including the alignment method, the adhesion forces and the energy consumed to establish the vacuum.

Table 1 provides an overview of the maximum adhesion forces of each element in contact with the tube surface as well as the forces without vacuum contact and with contact to a flat surface. The forces without contact result from the minimal inclination angle of the winch rope.

When it is moving vertically to the next inspection area, the robot's central element is in contact with the panel. The robot runs in synchronous mode together with the winch. Figure 6 compares uncontrolled and controlled vertical robot movement. The deviation between the robot's vertical axis and winch position increases during uncontrolled vertical movement. Reasons for this are different target speeds and the swinging of the robot hanging from the winch. Uncontrolled movement is only possible for small vertical steps and only within the maximum range of deviation allowed.

Table 1. Horizontal adhesion forces of the contact elements.

Holding Forces [N]	@ Contact Element				
	Top right	**Top left**	**Down right**	**Down left**	**Center**
without vacuum contact	55	58	62	64	154
vacuum contact with the tube surface	209	223	209	214	365
vacuum contact with a flat, homogeneous surface	289	305	270	295	405

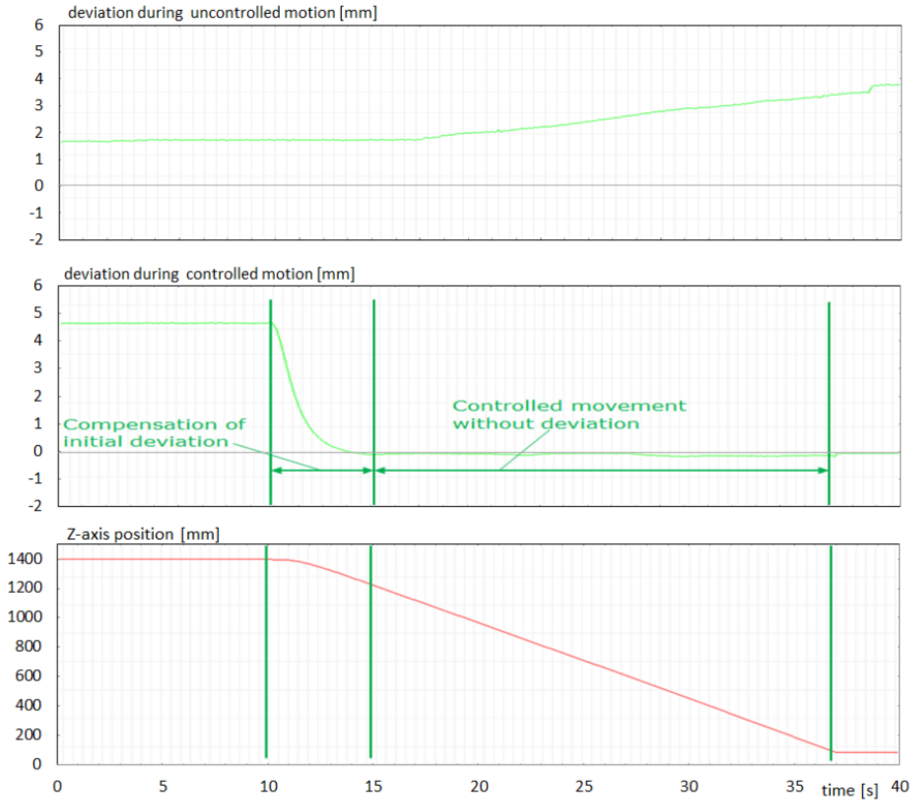


Figure 6. Comparison of uncontrolled and controlled vertical robot movement.

The control algorithm implemented compensates for the different speeds of the winch and the robot's Z-axis. The speed of the robot axis (slave speed) is controlled as a function of the speed of the winch (master speed). Furthermore, any initial deviation is compensated within a short time. Several tests were performed at different speeds, in different directions and under different initial conditions to verify the control algorithm.

A typical motion sequence during the inspection of individual tubes is presented in Figure 7. The total time required to inspect one panel area is about 27 min. Since increasing the climbing inspection speed is expected to reduce this time by as much as 50%, most of the individual robot actions were tested on the mockup at higher speeds. Table 2 shows the time durations of individual robot actions and complete receiver inspection, comparing the measured values with the times after optimization.

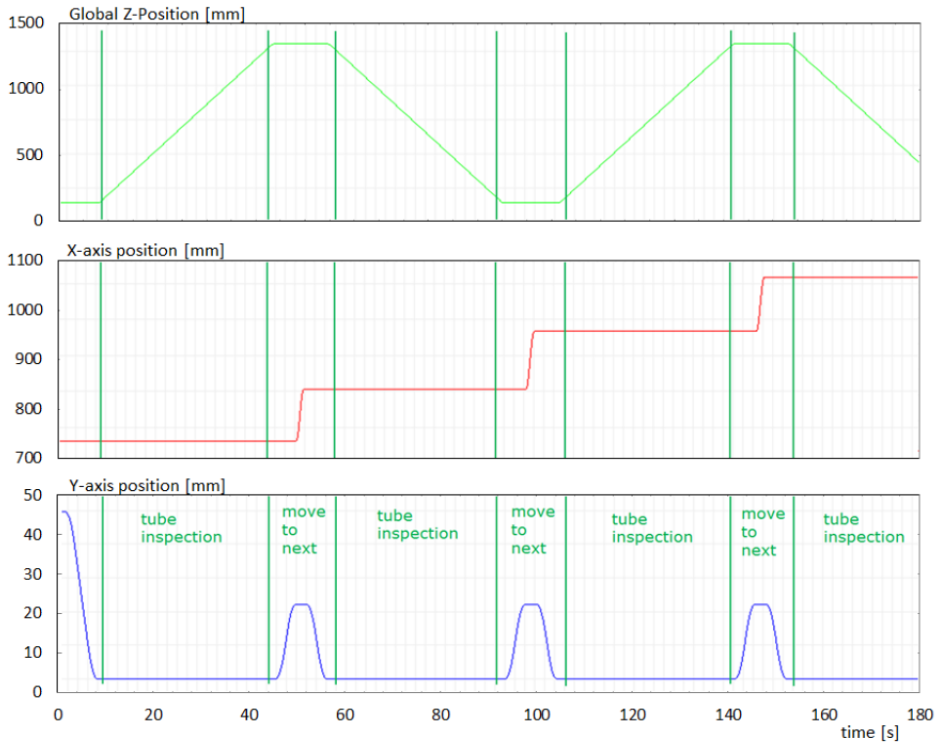


Figure 7. Robot movement during the inspection of the exchanger tubes.

4. Nondestructive Testing (NDT) of the Receiver Tubes

Two techniques are proposed for nondestructive inspection of the panels: visual inspection and eddy current testing. The first is used to assess coating degradation. Eddy current testing has the dual objective of measuring coating thickness and detecting internal corrosion (manifested as a reduction in tube thickness).

Table 2. Total inspection times of the climbing robot.

Robot Operation	Time Duration [s]	
	measured	after optimization
Move to panel	-	300 (estimated)
Remove from panel	-	300 (estimated)
Move to next area (performing one step)	130	70
sensor positioning	15	5
data acquisition (for tube section of 1.5 m):		
paint layer thickness and internal assessment with ET	50	25
external assessment with video	30	8
move to next tubes	10	5
Inspection of complete GEMASOLAR receiver		
with 2 sensors in parallel	95 h	50 h
with 8 sensors in parallel	30 h	16 h

Tube inspection is quite common in nuclear power plants. Different eddy current sensors (bobbin coils, rotatory coils, *etc.*) are used to inspect in-service steam generator tubes periodically for degradation. These sensors are inserted in the tube and run along its whole length to detect stress corrosion cracking and other mechanical degradation modes that could cause tube failure. The inspection of fuel rods to measure oxide layers is another nuclear application in which eddy currents deliver the best performance. Specialized sensors are used to inspect them from the external surface. Comprehensive inspection of such components requires a combination of advanced probe technology coupled with versatile instruments and robotic systems controlled by fast computers and remote communication systems. The in-service inspection company and manufacturer of eddy current sensor and data acquisition systems, Tecnatom has contributed its extensive experience and know-how in this field to the MAINBOT project.

Visual cameras and eddy current sensors were integrated into the service robot to assess the degradation of the receiver tubes: The coating thickness (μm) is measured by the eddy current testing (ET) to detect diminished heat transfer performance. Furthermore, tube thickness (<3 mm) is also measured by ET to detect internal degradation of the tubes (corrosion, deposits, *etc.*). A camera and lighting system (visual inspection or VT) detects external loss of coatings (≥3 × 3 mm). Figure 8 shows NDT instrumentation integrated next to the center contact element.

The robot's climbing kinematics position the sensors along the tubes. The robot remains in its current position while on a tube area is being scanned and only proceeds to the next area afterward. An on-board PLC controls all functions of the robot.

A higher-level control system is used to select the panel inspected and the inspection task. The General Control Manager uses Robot Manager software to command every robot movement and uses NDT Manager software to initiate NDT inspection. The individual inspection tasks, including on-line defect detection, are defined through the NDT Manager user interface and, then, executed automatically (see Figure 9).

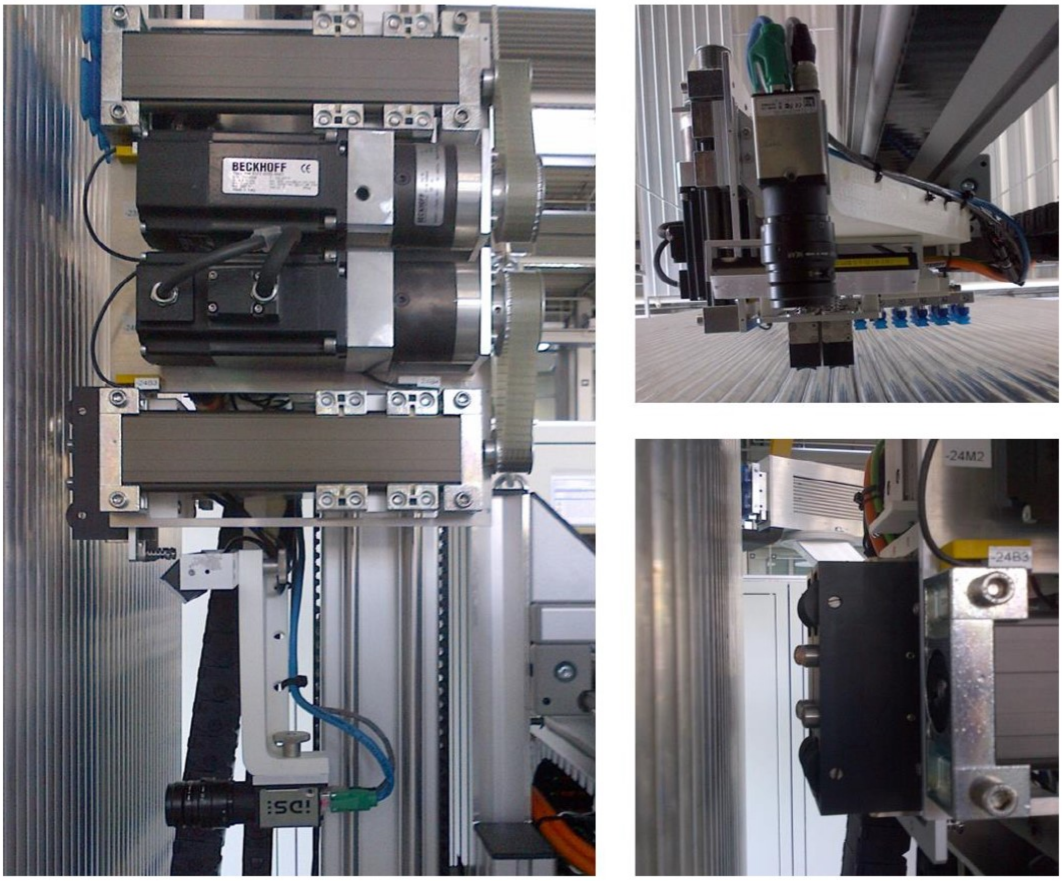


Figure 8. NDT integration and central contact element.

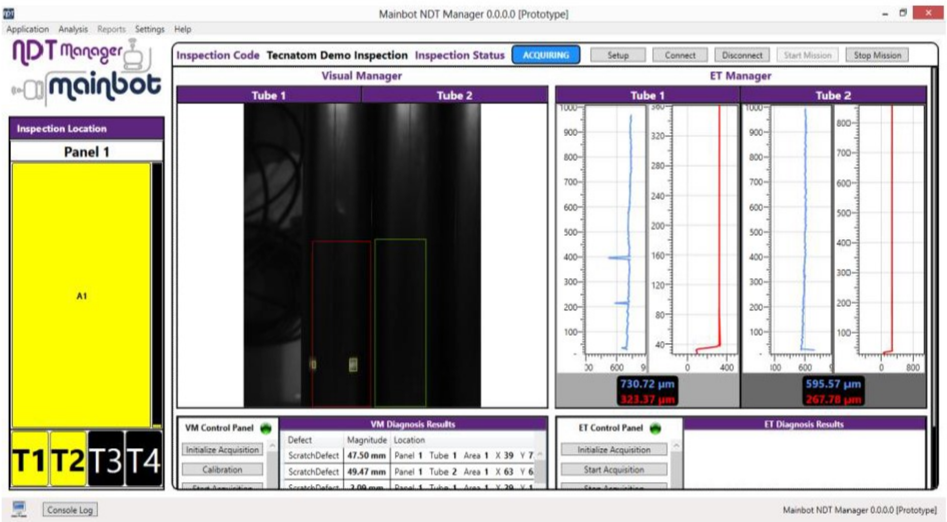


Figure 9. NDT Manager.

4.1. Eddy Current Testing of the Absorber Tubes

The ETbox2i containing all necessary ET hardware and software was developed by Tecatom. The main advantages of this system are its compatibility with different ET sensors, its automatic calibration of ET, and its online data processing on demand. All of the ETbox2i hardware has been optimized for size and performance. Two calibration tubes were made to test the coating sensors:

- The first tube has sections with different coating thicknesses (see Figure 10, top).
- The coating in the second tube has been removed to reproduce defects of predefined sizes (3×3 mm^2, 5×5 mm^2 and 10×10 mm^2) in different locations (see Figure 10, bottom).

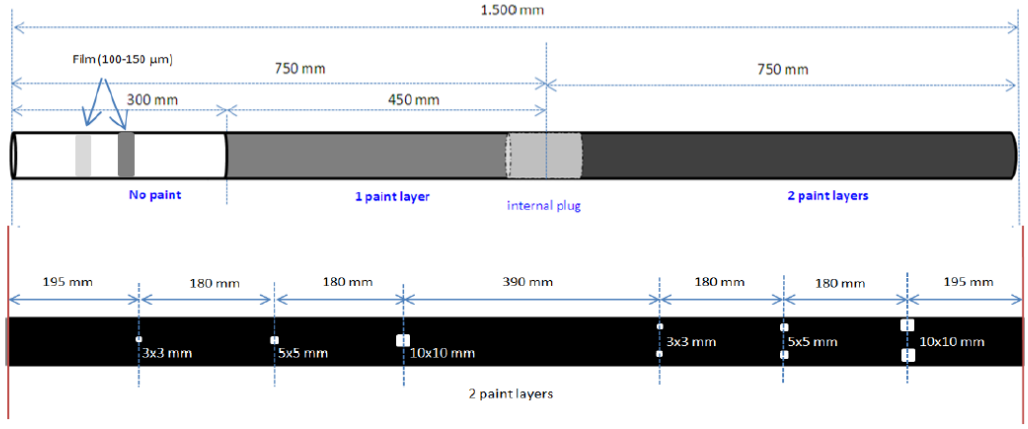
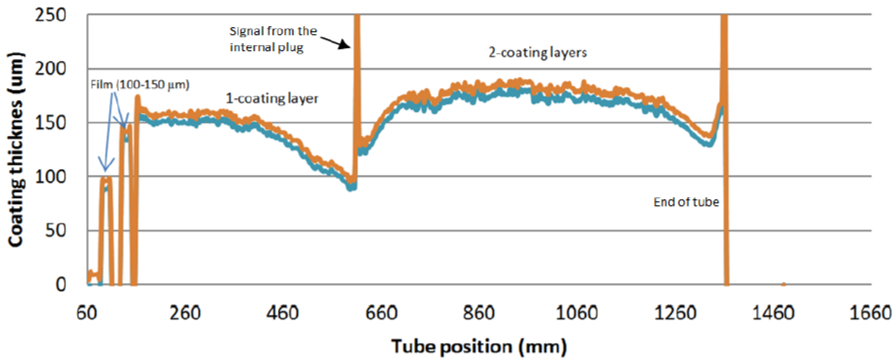


Figure 10. Calibration tubes for coating assessment.

The coating measurements reveal varying thickness along the tubes (see eddy current inspection of calibration tube no. 1 in Figure 11, top). Coating wear is only measurable when the ET sensor is perpendicular to the tube within a tolerance of ±30°. Figure 11 (bottom) presents the results of the inspection of calibration tube no. 2, the defects introduced in the tube (see Figure 10, bottom) being detected. Numerous tests validated the high precision of positioning and measurement.

A calibration tube with artificial internal defects was made to facilitate their detection (two sets of EDM notches representing 20%, 40%, and 60% reductions of tube thickness, see Figure 12).

The recorded ET data from the calibration tube are presented in Figure 13. Low excitation frequencies are required to increase penetration. Moreover, raw ET data are processed online to calculate tube thickness, to stabilize ET signals and to eliminate the influence of coating. To do so, data are recorded at two frequencies, "f", the optimum low frequency for the inspection of tube thickness, and "2f", a higher frequency for the inspection of surface defects. The signals of both frequencies are combined by applying several algorithms that eliminate undesirable effects. Specifically, the frequency "f" detects the influence of coating, whereas the frequency "2f" detects both the interference of coating and surface defects. The coating signal has to be removed from the higher frequency "2f" in order to be able to characterize surface defects. To do so, the coating signal of the lower frequency "f" is rotated and rescaled to be equal to the coating signal of the higher frequency "2f". The correct values of the rotation angle and the rescaling factor are computed using an optimization algorithm and a pure coating signal. The signal without coating influence is obtained by extracting to the higher frequency of "2f", the rescaled and rotated version of the lower frequency "f". The different properties of the magnetic field and the phase difference of the coating and tube surface guarantee that the processed signal only contains information on surface defects. Once they had been tested and validated, these algorithms were programmed in NDT Manager for automatic defect extraction.



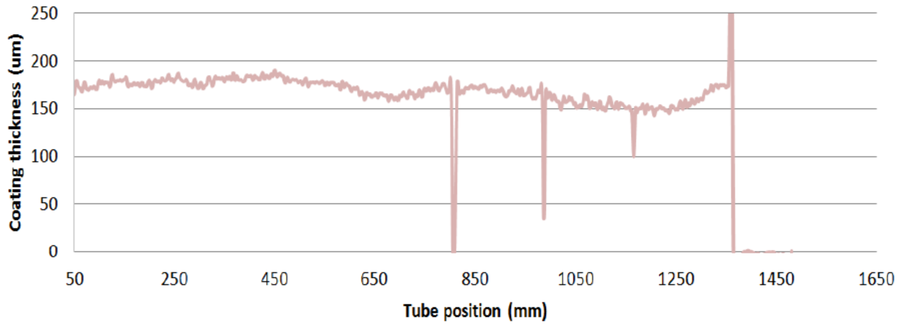


Figure 11. Coating measurements in calibration tubes.

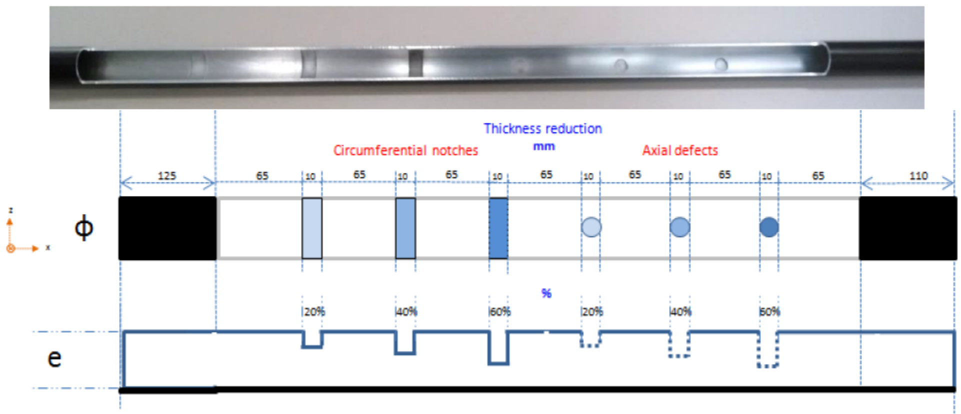


Figure 12. Calibration block for thickness measurement.

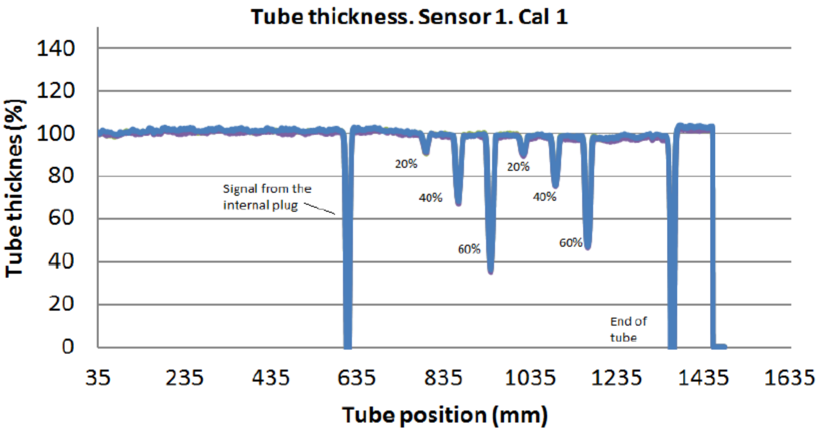


Figure 13. ET record with tube thickness measurements.

4.2. Visual Inspection of Absorber Tubes (VT)

The camera and lighting device integrated in the climbing robot can inspect up to four tubes simultaneously. Since the ET module only inspects two tubes, the calibration process for the VT allows the user to:

- select the area of interest for image processing, thus automatically assigning detected defects specific tubes and keeping low quality areas of the image from being processed,
- adjust sensitivity for defect detection in terms of grayscale values and defect size (*i.e.*, keeping defects smaller than 2 × 2 mm from being reported), and
- correct defect sizing, eliminating image distortions caused by tube curvature (Figure 14).

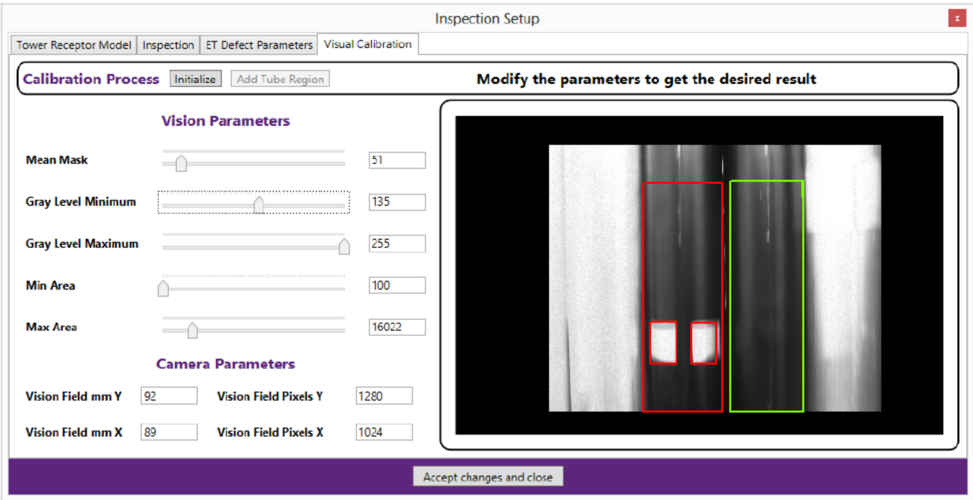
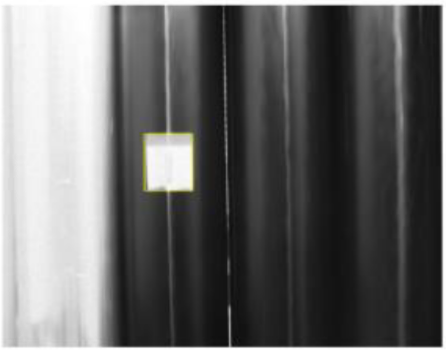


Figure 14. VT calibration process.

The camera effectively records the images and extracts the defects. The NDT Manager automatically generates reports with a table of defects and pertinent images (see Figure 15).

Id	Type	Panel	Tube	Area	Magnitude (mm^2)	X (mm)	Y (mm)	Width (mm)	Height (mm)
1	Scratch	1	1	1	90.1153	37.55	197.62	8.69	10.37
2	Scratch	1	1	1	94.6781	51.41	197.62	9.13	10.37
3	Scratch	1	1	1	28.4408	36.24	381.51	4.87	5.84
4	Scratch	1	1	1	24.3302	51.5	381.73	4.78	5.09
5	Scratch	1	1	1	7.7256	35.46	565.44	2.96	2.61
6	Scratch	1	1	1	7.3437	51.06	565.73	2.69	2.73
7	Scratch	1	1	1	132.0066	42.11	964.11	12.43	10.62
8	Scratch	1	1	1	37.3096	41.72	1147.64	6.26	5.96
9	Scratch	1	1	1	9.424	41.07	1331.21	3.04	3.1

Id	Type	Panel	Tube	Area	Magnitude (mm^2)	X (mm)	Y (mm)	Width (mm)	Height (mm)
7	Scratch	1	1	1	132.0066	42.11	988.11	12.43	10.62



Id	Type	Panel	Tube	Area	Magnitude (mm^2)	X (mm)	Y (mm)	Width (mm)	Height (mm)
4	Scratch	1	1	1	24.3302	51.5	381.73	4.78	5.09

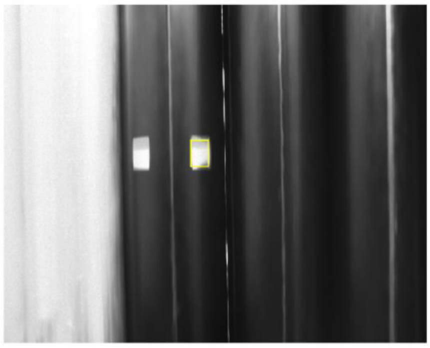


Figure 15. Automatic VT inspection report.

5. Conclusions

This paper presents a prototype climbing robot that inspects vertical surfaces, specifically a heat exchanger in a central tower CSP plant. The robot uses customized contact elements with vacuum suction cups to establish contact with the tube structure. The unique method of alignment and contact makes it possible to secure the robot on the desired surface with specific tube characteristics. The integration of two major functions (climbing and sensor guidance) in the telescopic XYZ kinematics reduces robot complexity and increases the inspection area. The integrated measurement instrumentation assesses the degradation of the tubes. A new design is proposed for the ET sensors. New electronics (ETbox2i) for the data acquisition system are integrated in the robot. Dedicated software was developed for automated system calibration, for data acquisition synchronized with robot movement, and for automated defect extraction and characterization combining visual inspection and eddy current testing.

Initial results of tests of climbing action and NDT inspection with the prototype on a mockup are presented. Future work will improve the robot structure to satisfy every certification requirement, including the performance of robot tests under real conditions.

Service robots allow comprehensive evaluation of large industrial plants. The proposed robot system can inspect the surfaces and the internal structures of heat exchanger tubes. Other applications in addition to the use case of the climbing robot on the receiver in a central tower CSP plant are conceivable, e.g., tank inspection. Furthermore, the robot can also be used to inspect other objects, e.g., dams, bridges, facades, or similar structures. The fully automated inspection robot increases the efficiency of maintenance work, reduces operating and maintenance costs, and improves safety and working conditions for service technicians.

Acknowledgments

The research project "MAINBOT. Mobile Robots for Inspection and Maintenance in Extensive Industrial Plants" is being funded by the European Union in its Seventh Framework Programme FoF.NMP.2011–2013 under grant agreement no. 285245 [16].

Author Contributions

All authors contributed equally to this work. T. Felsch and G. Strauss developed and tested the climbing kinematics. C. Pérez and J.M. Rego developed and tested the NDT technology. T. Felsch and C. Perez wrote the main paper, and G. Strauss wrote the supplementary information. All authors discussed the results and implications and commented on the manuscript at all stages.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Cho, K.H.; Jin, Y.H.; Kim, H.M.; Moon, H.; Koo, J.C.; Choi, H.R. Caterpillar-based cable climbing robot for inspection of suspension bridge hanger rope. In Proceedings of the 2013 IEEE International Conference on Automation Science and Engineering (CASE), Madison, WI, USA, 17–20 August 2013; pp. 1059–1062.
2. Goldman, G.J. Design Space and Motion Development for a Pole Climbing Serpentine Robot Featuring Actuated Universal Joints. M.S. Thesis, Department of Mechanical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.
3. Fernández, R.; González, E.; Feliú, V.; Rodríguez, A.G. A wall climbing robot for tank inspection. An autonomous prototype. In Proceedings of the IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; pp. 1424–1429.
4. Eich, M.; Vogele, T. Design and control of a lightweight magnetic climbing robot for vessel inspection. In Proceedings of the 2011 19th Mediterranean Conference on Control & Automation (MED), Corfu, Greece, 20–23 June 2011; pp. 1200–1205.
5. Leon Rodriguez, H.; Sattar, T.; Park, J.O. Wireless climbing robots for industrial inspection. In Proceedings of the 2013 44th International Symposium on Robotics (ISR), Seoul, Korea, 24–26 November 2013; pp. 1–4.
6. Wang, H.; Yamamoto, A.; Higuchi, T. Electrostatic-motor-driven electroadhesive robot. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 914–919.

7. Menon, C.; Murphy, M.; Sitti, M. Gecko Inspired Surface Climbing Robots. In Proceedings of the ROBIO 2004. IEEE International Conference on Robotics and Biomimetics, Shenyang, China, 22–26 August 2004; pp. 431–436.
8. Prahlad, H.; Pelrine, R.; Stanford, S.; Marlow, J.; Kornbluh, R. Electroadhesive robots—Wall climbing robots enabled by a novel, robust, and electrically controllable adhesion technology. In Proceedings of the IEEE International Conference on Robotics and Automation 2008 (ICRA 2008), Pasadena, CA, USA, 19–23 May 2008; pp. 3028–3033.
9. Sack, M.; Elkmann, N.; Felsch, T.; Bohme, T. Intelligent control of modular kinematics—The robot platform STRIUS. In Proceedings of the 2002 IEEE International Symposium on Intelligent Control, Vancouver, Canada, 27–30 October 2002; pp. 549–553.
10. Elkmann, N.; Felsch, T.; Sack, M.; Saenz, J.; Hortig, J. Innovative service robot systems for facade cleaning of difficult-to-access areas. In Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, Lausanne, Switzerland, 30 September–4 October 2002; Volume 1, pp. 756–762.
11. Guimaraes, M.; Lindberg, J. Remote controlled vehicle for inspection of vertical concrete structures. In Proceedings of the 2014 3rd International Conference on Applied Robotics for the Power Industry (CARPI), Foz do Iguassu, Brazil, 14–16 October 2014; pp. 1–6.
12. Leibbrandt, A.; Caprari, G.; Angst, U.; Siegwart, R.Y.; Flatt, R.J.; Elsener, B. Climbing robot for corrosion monitoring of reinforced concrete structures. In Proceedings of the 2012 2nd International Conference on Applied Robotics for the Power Industry (CARPI), Zurich, Switzerland, 11–13 September 2012; pp. 10–15.
13. Vázquez, J.; Guerra, F.J.; Vano, J. Automatic Process for Eddy Current Inspection of Steam Generator Tubes. In Proceedings of the 2nd International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurized Components, New Orleans, LA, USA, 24–26 May 2000.
14. Fernández, J.R.; Guerra, F.J. Fuel Rod Inspection System, SICOM-ROD. In Proceedings of the 6th International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurized Components, Budapest, Hungary, 8–10 October 2007.
15. Jansen, H. Eddy Current Testing: Profiled eddy current probes for complex shape inspection. In Proceedings of the 18th World Conference on Nondestructive Testing, Durban, South Africa, 16–20 April 2012.
16. MAINBOT—Mobile Robots for Inspection and Maintenance Activities in Extensive Industrial Plants. Available online: http://www.mainbot.eu (accessed on 3 August 2014).

7.3. Particle Filtering for Industrial 6DOF Visual Servoing

Aitor Ibarguren, José María Martínez-Otzeta, Iñaki Maurtua: Particle Filtering for Industrial 6DOF Visual Servoing. Journal of Intelligent and Robotic Systems 74(3-4): 689-696 (2014) [104]

J Intell Robot Syst (2014) 74:689–696
DOI 10.1007/s10846-013-9854-2

Particle Filtering for Industrial 6DOF Visual Servoing

Aitor Ibarguren · José María Martínez-Otzeta · Iñaki Maurtua

Received: 27 March 2013 / Accepted: 6 June 2013 / Published online: 16 June 2013

Abstract Visual servoing allows the introduction of robotic manipulation in dynamic and uncontrolled environments. This paper presents a position-based visual servoing algorithm using particle filtering. The objective is the grasping of objects using the 6 degrees of freedom of the robot manipulator in non-automated industrial environments using monocular vision. A particle filter has been added to the position-based visual servoing algorithm to deal with the different noise sources of those industrial environments. Experiments performed in the real industrial scenario of ROBOFOOT (http://www.robofoot.eu/) project showed accurate grasping and high level of stability in the visual servoing process.

Keywords Robotics · Visual servoing · Particle filtering

1 Introduction

Traditional industrial robotic applications, like part placement or spot welding, require precise information about the position of the objects to perform their task. Visual servoing [1, 2] can enhance those industrial applications allowing corrections on the robot trajectories. This technique would help to introduce robots in applications where the workpiece is moving or is placed in the working area in an unknown pose.

Even so, industrial environments raise their own challenges in the inclusion of visual servoing techniques, especially when the production line is not completely automated. Dirt, imprecision in the workpiece placement or changing lighting conditions are some of the problems that must be tackled in this kind of environments, introducing uncertainties in the trajectory correction process.

Particle filtering [3, 4], a sequential Monte Carlo algorithm, is a suitable choice to deal with uncertainties in the observed data. This technique offers a reliable way to estimate unknown states based on an observation sequence, as they are able to deal with multiple hypotheses in a simple and effective way. Due to those features, particle filters are a suitable technique that can be efficiently introduced to deal with the uncertainties identified in industrial environments.

This paper presents a position-based visual servoing algorithm using particle filtering. Based on the real industrial scenario of ROBOFOOT project, the paper proposes an algorithm to grasp a workpiece (a shoe last specifically) from a not constrained workshop, correcting the 6 degrees of freedom of the robot during the visual servoing process.

A. Ibarguren (✉) · J. M. Martínez-Otzeta · I. Maurtua
IK4-Tekniker, Iñaki Goenaga 5, 20600 Eibar,
Gipuzkoa, Spain
e-mail: aitor.ibarguren@tekniker.es

The paper is organized as follows. In Section 2 the related work is presented. Section 3 exposes briefly particle filters. Task specification and configuration is shown in Section 4. Section 5 is devoted to the proposed approach, while in Section 6 the experimental results are shown. Finally, Section 7 presents the conclusions as well as the future work to be done.

2 Related Work

Several approaches tackle the use of visual servoing in industrial environments, posing different industrial scenarios and approaches.

Sung-Hyun et al. [5] propose an image-based visual servoing based on stereo vision. The use of stereo vision allows guiding the robot manipulator to the desired location without giving such prior knowledge about the relative distance to the desired location or the model of the object.

Nomura et al. [6] describe a visual servoing system able to track and grasp industrial parts moving on a conveyor using a 6DOF robot arm. A hybrid Kalman Filter is also incorporated to track a moving object stably against visual data noise. Experiments are also presented, performing both 3DOF and 6DOF visual servoing.

Lippiello et al. [7, 8] presented visual servoing applications on Industrial Robotic cells. On their setup, composed of two industrial robot manipulators equipped with pneumatic grippers, vision systems and a belt conveyor, a position-based visual servoing is proposed. The system also uses Extended Kalman Filters (EKF) [9] to manage the occlusions during the multi-arm manipulation.

Finally, [10, 11] propose particle filtering for 3D-edge tracking as well as model-based tracking, allowing to manage multiple hypotheses during the process. It will allow to achive a more robust detection, overcoming noise problems of the image acquisition.

3 Particle Filter

Particle filters, also known as sequential Monte Carlo methods (SMC), are sequential estimation techniques that allow estimating unknown states x_t from a collection of observations $z_{1:t} = \{z_1, \ldots, z_t\}$. The state-space model is usually described by state transition and measurement equations

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{1}$$

$$z_t = g_t(x_t, u_t) \tag{2}$$

where f and g are the state evolution and observation model functions respectively and v_t and u_t denote the process and observation noise respectively.

Based on the previous equations, particle filters allow approximating the posterior density (PDF) by means of a set of particles $\{x_t^{(i)}\}_{i=1,\ldots,n}$ using equation

$$p(x_t|z_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{3}$$

where each particle $x_t^{(i)}$ has an importance weight $\omega_t^{(i)}$ associated and δ is the Kronecker delta. These weights are computed following equation

$$\omega_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(z_t|x_t^{(i)})\, p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})} \tag{4}$$

where $p(z_t|x_t^{(i)})$ is the likehood function of the measurements z_t and finally $q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})$ is the proposal density function.

Based on the previously presented equations the particle set evolves along time, changing the weights of the particles and resampling them in terms of the observations.

Particle filtering provides a robust tracking framework when dealing with non-linear and non-gaussian state and observation functions as it considers multiple state hypotheses simultaneously.

4 Task Specification and Configuration

Based on the needs of ROBOFOOT project, whose aim is the introduction of robots in the footwear industry, an object grasping task has been designed using real specifications of footwear workshops. One of the principal aspects of the project is the focus on minimizing the impact of the introduction of the robots in the existing production means, nowadays basically handcrafted in high added value shoe production. Taking it into account, the grasping scenario has been specified as:

– Lasts, material with the shape of a foot used to build shoes, are the object to be grasped. An iron piece (*grasping device*) has been added to lasts to allow a precise and stiff grasping, see Fig. 1a, as well as to protect the leather during the grasping process. Those *grasping devices* will be the objects to be identified during the visual servoing process.
– Lasts are carried in specific trolleys mounted on a manovia. The trolleys are designed to allow the placement of lasts of different shapes and sizes. Lasts are placed in the trolley by human operators. Due to those previous facts it is not possible to know the pose of the last in the trolley, as seen in Fig. 1a. Therefore, it forces the visual servoing system to correct its 6DOF to grasp the workpiece.
– A 6DOF robot arm with a gripper and a camera and lighting system mounted on the end-effector with an eye-in-hand configuration, as seen in Fig. 1b.
– Based on the design of the gripper and the *grasping device*, the grasping process requires a precision of around a millimeter and 1–2 degrees on each axis to grasp the last smoothly. In the same way the maneuver should take no more than 5–6 seconds.

Based on this scenario, the initial set-up of the system has raised some problems related with the pose estimation of the *grasping device*:

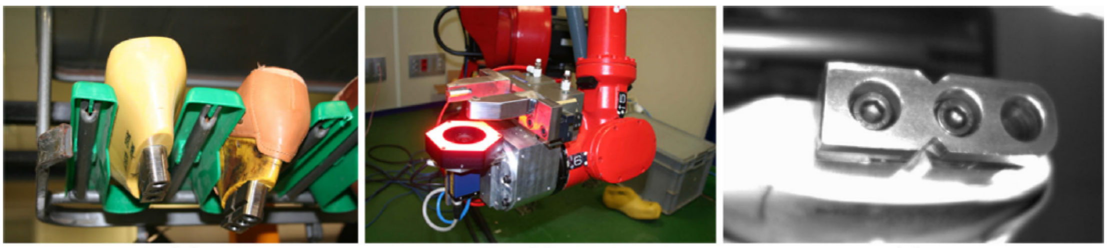
– Illumination is a key aspect in a vision system. In this industrial scenario is complicated to place a suitable external illumination, that is why it was decided to put a specific lighting system on the gripper. Even so, the metallic nature of the *grasping device* makes it difficult to get a good image due to the brightness, reflection and the impossibility of lighting all the image properly, as shown in Fig. 1c.
– Some of the tasks to be performed by both the human operators and robots involve the use of ink, wax or generate dust (roughing process). This dirt can be adhered to the *grasping device*, complicating the visual servoing process.

Those previous points will make it difficult to acquire clear images of the *grasping device*, adding uncertainties to the 6DOF pose estimation that will be the base of the visual servoing process.

5 Proposed Approach

Taking into account the needed precision, which implies a high resolution camera, and the demanding image processing due to the unstable conditions of the image, a dynamic look-and-move approach has been adopted.

A particle filter has also been added to the system to manage the uncertainties of the vision system. Problems on the illumination, the metallic nature of the *grasping device* and the possible



(a) (b) (c)

Fig. 1 Lasts with the *grasping device* on the trolley, gripper and image of the *grasping device* taken in real conditions dirt of the environment could be managed using a particle filter where each particle represents an hypothesis, whose likelihood is estimated and updated according to the noisy observations.

Next lines will describe the general structure of the system, as well as the vision module, likelihood estimation, the particle filter and the grasping algorithm.

5.1 System Modeling and Architecture

In the described scenario, the space can be represented by $P \in \Re^6$. In the same way, this scenario will be composed of two different frames, the robot frame r and the camera frame c. Given those two frames, the camera pose is represented as a 4×4 transformation matrix, denoted by ${}^cT_r$, which transforms poses from robot frame r to camera frame c as

$$P^c = {}^cT_r P^r, \quad {}^cT_r = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{5}$$

where $R \in SO(3)$ is a rotation matrix and $t \in \Re^3$ is a translation vector.

The error of the positioning task involved in the grasping process is represented by vector $E \in \Re^6$ which represents the difference between the pose of the object P_o^r in the robot frame and the pose of the end-effector P_e^r in the robot frame (Eq. 6). The grasping process can be seen as a minimization of this error that will be fulfilled when $|E| = 0$.

$$E = P_e^r - P_o^r = \begin{bmatrix} x_e^r - x_o^r \\ y_e^r - y_o^r \\ z_e^r - z_o^r \\ \alpha_e^r - \alpha_o^r \\ \beta_e^r - \beta_o^r \\ \gamma_e^r - \gamma_o^r \end{bmatrix} \tag{6}$$

where α, β and γ represent the Euler angles.

For pose estimation, position-based visual servoing systems extract features from the acquired images, estimate the pose of the object P_o^r and perform the corrections. Even so, the described scenario introduces uncertainties in the feature extraction step (illumination, metallic workpiece...), introducing errors in the pose estimation. To deal with this problem, the use of a particle filter is proposed. From each image, a set of n feature vectors $F_i = \{f_1, f_2, \ldots, f_m\}_{i=1\ldots n}$ will be extracted for the pose estimation, each of them related with a specific image analysis procedure. Each of those n vectors will be an observation hypothesis of the values of the m features used for the pose estimation, as it will not be possible to have a unique feature vector extracted from each image due to the uncertainties in the image.

Those n observation hypotheses, F_i, will be the observations of the particle filter, which will output the final pose estimation of the object in the robot frame P_o^r. This final pose will be used to calculate the error E between the object and the end-effector, used to calculate the next robot movement. Figure 2 shows the structure of the proposed visual servoing system.

Next lines will describe the feature extraction, likelihood estimation, particle filtering and grasping algorithm of the grasping process.

5.2 Feature Extraction

As stated before, one of the challenges of the presented scenario is the feature extraction for pose estimation. The metallic nature of the *grasping device* and the illumination problems make it difficult to detect the different features (edges, corners, holes...) precisely. Taking also into

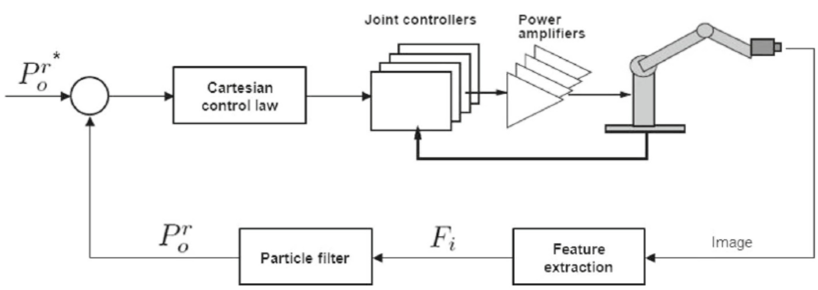


Fig. 2 Dynamic position based look-and-move structure with particle filter

account the perspective of THE camera through the grasping process, the image features used for pose estimation, shown in Fig. 3, are:

- The center of the three holes (1, 2, 3) of the *grasping device*. Only the pixel coordinates of the center of the holes are taken into account, excluding other information about the holes, due to the difficulties of extracting their contour precisely.
- The inclination of the left edge (4) of the *grasping device*.

To detect those image features different thresholds values and Canny edge detector have been used. As illumination is not constant over the image and each hole has a different grey value, it was decided to apply different threshold values to detect the holes as a dynamic thresholding showed poor results. Even so, in some images it is not possible to determine the exact position of the three holes' centres as there are various possible circular shapes in each position (ex. the inner screw, outer circle and dirt around it). In those cases it is not possible to define a universal rule to determine which the real contour of the holes is. To overcome this problem, this approach proposes to use all those possible centers of the three holes (left, central and right) obtained with the different threshold values, creating a set of observation hypotheses that will be used for pose estimation.

Once the centers of the holes and the left edge are detected, a feature vector will be calculated for each observation hypothesis as:

$$F_i = \{c_1, c_2, c_3, \phi_{\text{edge}}, \lambda\} \tag{7}$$

where c_1, c_2, c_3 are the coordinates in pixels of the left, central and right hole respectively, ϕ_{edge} is the angle of the left side of the *grasping device* and λ is a coefficient that measures the noise (quality) of the observation hypothesis based on the similitude of the circular shapes and their alignment and it is calculated as

$$\lambda = \frac{C_v\left(p_1, p_2, p_3\right) + \frac{|\phi_{12}-\phi_{23}|+1}{|\phi_{13}|+1}}{\text{Min}\left(r_1^{xy}, r_2^{xy}, r_3^{xy}\right)} \tag{8}$$

where p_i is the perimeter of the ith hole, C_v is the *coefficient of variation* of the perimeters, ϕ_{ij} is the angle between the holes i and j and r_i^{xy} is the xy axis ratio of the bounding box of the ith hole. It will allow to obtain a low value when three aligned holes of similar size are found in the image while the coefficient will grow when irregular circles with different sizes appear on the image. In other words, this coefficient captures information not available in the other features and summarizes its uniformity.

Those are the features that will be used to estimate the pose of the workpiece.

5.3 Particle Filter

Once the possible observation hypotheses are calculated, it is necessary to merge and fuse this information to perform the grasping process. To this end a particle filter is proposed, as it fits in this kind of non-gaussian problem.

Focusing on the posed problem, the state in time t will be defined as a pose of the object in the robot frame

$$X_t = \left[x_t, y_t, z_t, \alpha_t, \beta_t, \gamma_t\right]^T. \tag{9}$$

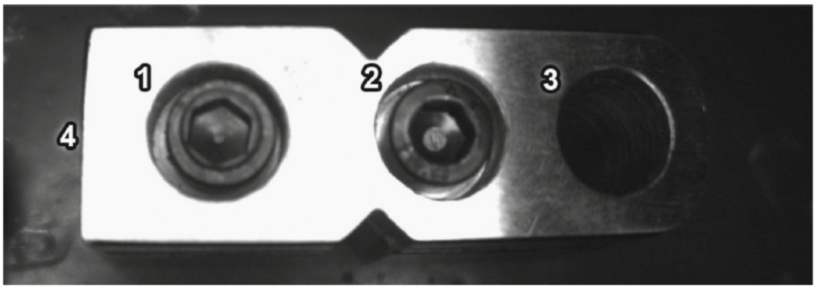


Fig. 3 Visual features for pose estimation

As it is not possible to model the pose estimation error a priori, the state transition is defined as

$$X_t = X_{t-1} + V_{t-1} \tag{10}$$

where X_{t-1} is the previous state vector and V_{t-1} is the process noise.

The observation, on the other hand, is defined by a set of n feature vectors of the object.

$$Z_t = F_{i=1..n}. \tag{11}$$

So based on this information source each particle is evaluated according to the observation hypotheses.

5.3.1 Likelihood Evaluation

Each particle represents a pose of the object in the robot frame which will be evaluated using the observed feature vectors $F_{i=1..n}$. To this end, initially the center of each hole and the left corner of each particle is projected in the image, taking into account the pose of the camera in the robot frame, as proposed in [10, 11]. Therefore, for each particle we will project the position of each hole and the left corner in the image, obtaining the three different pixel coordinates, one for each hole, as well as a theoretical angle of the left part of the *grasping device*, a set of features $F_p = \{c_{1_p}, c_{2_p}, c_{3_p}, \phi_{\mathrm{edge}_p}\}$ derived from each particle p.

To calculate the likelihood of each particle $P(Z_t|X_t)$, initially the distance between the features of each particle F_p and the observed features F_i is measured. To calculate the distance of the positions of the holes, the Euclidean distance in pixel coordinates is proposed, Eq. 12, while the distance between the angles of the left size of the *grasping device* is calculated using the quadratic error as shown in Eq. 13.

$$dist\,Pos(F_p, F_i) = \sqrt{(c_{1_p} - c_{1_i})^2 + (c_{2_p} - c_{2_i})^2 + (c_{3_p} - c_{3_i})^2} \tag{12}$$

$$dist\,Ang(F_p, F_i) = |\phi_{\mathrm{angle}_p} - \phi_{\mathrm{angle}_i}|^2 \tag{13}$$

In a next step, both distances are mixed as shown in Eq. 14, using also the λ_i noise coefficient of each observed set of features. α and β values have been introduced to weight the importance of the position and the angle.

$$dist(F_p, F_i) = \lambda_i + \alpha\, dist\,Pos(F_p, F_i) + \beta\, dist\,Ang(F_p, F_i) \tag{14}$$

Finally, the likelihood of each particle is calculated using the product of the exponential of the distances between the particle and each of the observed feature sets

$$P(Z_t|X_t) = \prod_{i=1}^{n} e^{-dist(F_p, F_i)} \tag{15}$$

where $dist(F_p, F_i)$ is the distance between the features of particle p and observation hypothesis i.

5.3.2 Particle Filtering Procedure

Finally the procedure of the particle filter is given as:

1. Find the *grasping device* in the initial image and initialise N particles $X_0^{(i)}$ with the different hypotheses randomly, where $w_0^{(i)} = 1/N$,
2. if *ESS* < *threshold* (*Effective Sample Size*), draw N samples with *selection with replacement*,
3. propagation of the particles $x_t^{(i)} = x_{t-1}^{(i)} + v_{t-1}$,
4. update importance weights $w_t^{(i)} = w_{t-1}^{(i)} P(Z_t|X_t)$,
5. normalize weights $w_t^{(i')} = w_t^{(i)} / \sum_{j=1}^{N} w_t^{(j)}$,
6. estimation of x_t, using best particle or robust mean estimators, and
7. set $t = t + 1$, goto step 2.

In this procedure EES [12] (Effective Sample Size) is calculated as

$$cv_t^2 = \frac{var(w_t^{(i)})}{E^2(w_t^{(i)})} = \frac{1}{N}\sum_{i=1}^{N}(Nw_t^{(i)} - 1)^2 \tag{16}$$

$$ESS_t = \frac{N}{1 + cv_t^2} \tag{17}$$

where N is the number of particles and $w_t^{(i)}$ is the weight of particle i in time t.

Based on this discrete approximation of the posterior probability, the object is tracked along the grasping procedure.

5.4 Grasping Algorithm

The feature extraction step has shown that the best images are acquired when the camera is perpendicular to the *grasping device* and the end-effector to a 30 mm distance, $E = [0, 0, 30, 0, 0, 0]^T$, as it solves in part the illumination problems. Taking it into account the grasping algorithm will try to minimize the error until this value is reached, adding an small tolerance of ± 1 mm in position and $\pm 1.5°$ in orientation to avoid an infinite loop. Once this error is reached the robot will make a final approach in just one axis and perform the grasping.

In the same way, in each robot position two loops of the particle filter have been executed before determining the most likely pose of the *grasping device*. It was decided to perform only two loops in every position in order speed-up the algorithm. Besides, each movement places the end-effector of the robot nearer to the desired final position, allowing to acquire better images which makes it possible to improve the quality of the observations, as stated previously.

6 Experimental Results

To test the performance of the proposed approach an experiment has been designed in order to measure its suitability. Specifically the test measures the success rate and time of the grasping process, as well as other meaningful information, using different parameters for the particle filter. Those are the specifications of the experiment:

- Six different particle filter configurations have been set-up, mixing different state estimation methods and number of particles. Specifically the state estimation methods are:
  - Best particle (the one with maximum weight)
  - Robust mean with the 3 particles with maximum weight (denoted as R.M. 3)
  - Robust mean with the 5 particles with maximum weight (denoted as R.M. 5)
- For each configuration, 250 repetitions have been performed using different shoes and *grasping devices* to ensure a variety of positions and orientations. In the same way most of the *grasping devices* have been dirtied up to include noise in the acquired image and simulate real conditions. Due to the structure of the manovia, the *grasping device* has been placed in a space of $200 \times 100 \times 200$ mm (depending on the shoe and its placement) and with a rotation of $\pm 15°$ in each axis.
- The grasping process will fail if does not achieve to pick up the shoe. There are two reasons for this fail; the *grasping device* has not been found in the initial image (ex. not well illuminated due to its orientation) or a wrong pose has been estimated which leads to a movement that leaves the *grasping devices* out of the scope of the camera.
- In the likelihood evaluation step, the same value for coefficient α and β have been used ($\alpha = \beta = 1$), giving the same weight to the position and orientation of the particle.

Table 1 shows the results of the experiment. The first column describes the number of particles and the estimation method, the second one the success rate, third and fourth columns show the mean (μ) and the standard deviation (σ) of the grasping time in seconds, fifth and sixth columns the mean (μ) and the standard deviation (σ) of the number of movements required to grasp the shoe and finally the last column shows the time required to process each cycle of the particle filter in milliseconds.

Results show a better performance of the system using the *robust mean* estimation method with 5 particles, both in success rate and in grasping time, as shown by bold entries. In the case of the success rate, in all the configurations a part of the fails were related with the search of the *grasping*

Table 1 Results of the experiment

	%	Time (s)		Mov.		ms/image
		μ	σ	μ	σ	
50—Best	97.4	3.34	1.29	10.39	3.95	27.53
100—Best	96.6	3.44	1.53	10.41	4.84	33.50
50—R.M. 3	98.2	3.46	1.35	10.36	4.49	33.79
100—R.M. 3	99	3.48	1.21	10.27	4.18	36.71
50—R.M. 5	**100**	3.28	1.44	10.25	4.55	**23.03**
100—R.M. 5	**100**	**3.03**	**1.09**	**9.66**	**3.49**	23.55

device in the initial image (more or less the same quantity for each configuration, around a 0–1 %). The rest of them are related with imprecise pose estimations that lead to camera positions that leave the shoe out of the camera scope. Results indicate that a robust mean estimation allows to avoid the noise of the pose estimation by means of the use of a particle distribution, giving the chance to recover the system from noise effects, achieving a 100 % of success with a lower grasping time.

In the same way, it seems that the addition of more particles does not help to improve the success rates in all the cases although it does not increase significantly the processing time of each visual servoing iteration.

7 Conclusions and Future Work

This paper presents a dynamic position-based look-and-move architecture to perform visual servoing with 6DOF in industrial environments. This kind of environments usually suffers from unstable conditions like changing lighting condition or dirt, introducing uncertainties in the visual servoing process. To overcome the above mentioned problem, this paper proposes the use of a particle filter to manage multiple hypotheses of the poses of the workpiece to grasp. The particle filter is integrated in the dynamic look-and-move structure, allowing to manage multiple hypotheses of the possible pose of the workpiece and fusing this information to avoid the noise of the process.

The results show a high success rate of the grasping system, reaching around a 99–100 % of success in the different experiments performed. The use of particle filtering allows the use and fuse of different observation hypotheses, overcoming the noise problems of the presented scenario. The system also performs the grasping process in a suitable time, increasing few processing time with the addition of the particle filter.

As further work, there are two interesting paths to follow. On one hand, to test this approach in similar scenarios (different workpiece, environment, noise source...) to test its suitability. On the other hand, one or more sensors could be attached to the end-effector as new data sources, using the particle filter to fuse the information received from the different sources as done in different robotic applications.

Acknowledgements This work has been performed within the scope of the project "ROBOFOOT: Smart robotics for high added value footwear industry". ROBOFOOT is a Small or Medium-scale focused research project supported by the European Commission in the 7th Framework Programme (260159). For further information see http://www.robofoot.eu.

References

1. Weiss, L., Sanderson, A., Neuman, C.: Dynamic sensor-based control of robots with visual feedback. IEEE J. Robot. Automat. **3**, 404–417 (1987)
2. Hutchinson, S., Hager, G., Corke, P.: A tutorial on visual servo control. IEEE Trans. Robot. Automat. **12**, 651–670 (1996)
3. Doucet, A., De Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, Berlin (2001)
4. Kotecha, J.H., Djuric, P.M.: Gaussian particle filtering. In: Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing, pp. 429–432 (2001)
5. Sung-Hyun Han, H., Seo, W.H., Yoon, K.S., Man-Hyung, L.: Real-time control of an industrial robot using image-based visual servoing. In: Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '99, vol. 3, pp. 1762–1767 (1999)
6. Nomura, H., Naito, T.: Integrated visual servoing system to grasp industrial parts moving on conveyer by controlling 6DOF arm. In: 2000 IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, pp. 1768–1775 (2000)
7. Lippiello, V., Siciliano, B., Villani, L.: An experimental setup for visual servoing applications on an industrial robotic cell. In: Proceedings, 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1431–1436 (2005)
8. Lippiello, V., Siciliano, B., Villani, L.: Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration. IEEE Trans. Robot. **23**(1), 73–86 (2007)
9. Julier, S.J., Uhlmann, J.K.: Unscented filtering and nonlinear estimation. Proc. IEEE **92**(3), 401–422 (2004)
10. Klein, G., Murray, D.: Full-3D edge tracking with a particle filter. In: Proc. British Machine Vision Conference (BMVC'06), vol. 3, pp. 1119–1128 (2006)
11. Teuliere, C., Marchand, E., Eck, L.: Using multiple hypothesis in model-based tracking. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 4559–4565 (2010)
12. Liu, J., Chen, R., Logvinenko, T.: A theoretical framework for sequential importance sampling and resampling. Technical report, Stanford University, Department of Statistics (2000)

7.4. Thermal Tracking in Mobile Robots for Leak Inspection Activities

Aitor Ibarguren, Jorge Molina, Loreto Susperregi, Iñaki Maurtua: Thermal Tracking in Mobile Robots for Leak Inspection Activities. Sensors 13(10): 13560-13574 (2013) [98]

Sensors **2013**, *13*, 13560-13574; doi:10.3390/s131013560

OPEN ACCESS

sensors

ISSN 1424-8220

www.mdpi.com/journal/sensors

Article

Thermal Tracking in Mobile Robots for Leak Inspection Activities

Aitor Ibarguren *, Jorge Molina, Loreto Susperregi and Iñaki Maurtua

IK4-Tekniker, Calle Iñaki Goenaga 5, Eibar 20600, Spain;
E-Mails: jorge.molina@tekniker.es (J.M.); loreto.susperregi@tekniker.es (L.S.);
inaki.maurtua@tekniker.es (I.M.)

* Author to whom correspondence should be addressed; E-Mail: aitor.ibarguren@tekniker.es;
Tel.: +34-943-206-744; Fax: +34-943-202-757.

Received: 31 July 2013; in revised form: 23 September 2013 / Accepted: 27 September 2013 / Published: 9 October 2013

Abstract: Maintenance tasks are crucial for all kind of industries, especially in extensive industrial plants, like solar thermal power plants. The incorporation of robots is a key issue for automating inspection activities, as it will allow a constant and regular control over the whole plant. This paper presents an autonomous robotic system to perform pipeline inspection for early detection and prevention of leakages in thermal power plants, based on the work developed within the MAINBOT (http://www.mainbot.eu) European project. Based on the information provided by a thermographic camera, the system is able to detect leakages in the collectors and pipelines. Beside the leakage detection algorithms, the system includes a particle filter-based tracking algorithm to keep the target in the field of view of the camera and to avoid the irregularities of the terrain while the robot patrols the plant. The information provided by the particle filter is further used to command a robot arm, which handles the camera and ensures that the target is always within the image. The obtained results show the suitability of the proposed approach, adding a tracking algorithm to improve the performance of the leakage detection system.

Keywords: thermal image; leak detection; particle filter; tracking

---

1. Introduction

Efficient and effective maintenance is crucial for all kinds of industries. In the case of capital intensive investment industries, such as petrochemicals, the steel industry or power generation plants, it is even more relevant and has an important impact on the operation costs during the long lifecycle of their production means.

Automating inspection activities in industrial plants, especially in extensive plants, poses strong requirements from different points of view: a huge number of elements to inspect (pipes, valves, switches, pumps, vessels, motors, vibrating machinery, chillers, ovens, *etc.*), handling multiple sensors or special non-destructive testing equipment to be used (visual, ultrasonic, vibration, radiography, thermography, eddy current, noise analysis, gas sensors, *etc.*) and extensive production facilities that spread out for thousands of square meters, either in the vertical or horizontal, and risky working conditions for maintenance personnel, due to the presence of hazardous materials.

This paper presents part of the work performed in the MAINBOT European project. This project aims at developing service robot applications to autonomously execute inspection tasks in extensive industrial plants. The objective is to develop a surveillance robotic system able to detect the leakage of fluids using a vision system in the thermal and visible ranges. To validate the proposed solution, a solar plant of cylindrical-parabolic collectors is used, testing the approach in a very demanding environment from a mobile manipulation point of view.

The paper is organized as follows. Section 2 gives information about related works. Section 3 introduces the inspection task to be performed by the system. Section 4 presents the proposed approach, the architecture for leak inspection in solar thermal plants and autonomous navigation. Sections 5 and 6 are devoted to the leakage detection algorithm and the tracking system, respectively. In Section 7, the experimental results of the system are shown. Finally, Section 8 poses the obtained results, as well as the future work to be done.

2. Related Work

Robots have been used for maintenance tasks in a wide range of applications and environments. From preventive maintenance of high-voltage transmission power lines [1] to inspection of cables [2] and nuclear reactor pressure vessels [3] in the nuclear industry, autonomous systems are used in many industries in an attempt to improve maintenance tasks. Focusing on pipeline inspection, Suzuki *et al.* [4] propose an autonomous robot for industrial pipeline inspection by means of ultrasonic diagnosis equipment. In the same way, Camerini *et al.* [5] present an underwater inspection robot for offshore pipeline inspection, using the pipeline itself for guidance purposes. Even so, a few works add some tracking tools as particle filters to improve the inspection task.

Several works have also been performed using thermal images for tracking and detection. Jiping *et al.* [6] propose a target detection and tracking system based on different morphological operations. Senthil Kumar *et al.* [7] pose the fusion of thermal images with 2D images for tracking Unmanned Aerial Vehicles (UAVs), adding optical flow techniques, such as LucasKanade and HornSchunck methods. Finally, Padole and Alexandre [8] propose the use of particle filters for human tracking with thermal images, using motion information to feed the particle filter. The presented work also proposes the use of particle filtering, although it uses the whole image information and dynamics to feed the particle filter, not only the motion information.

3. Task Specification

Valle 1 and 2 are two solar thermal power plants, whose promoter and owner is Torresol Energy Investments, S.A., and which are located in San José del Valle (Spain); see Figure 1. Valle 1 and 2 are two adjacent solar thermal power plants that generate electricity by means of cylindrical-parabolic collectors.

Figure 1. Valle 1 and 2 solar thermal power plants.

The solar field is composed of nearly 7,500 parabolic cylinder collectors. These collectors transport a Heat Transfer Fluid (HTF), which absorbs the solar energy. HTF circulates at a high temperature (around 390 °C) inside the absorber tubes, which is used after to heat the molten salts to generate steam in the Steam Generation System (SGS).

Swivel joints are critical points where leakages may happen (this is the point where the collector tube connects with the infrastructure of pipes that are deployed all over the plant). HTF leakages are not desirable, as oil losses may be unsafe, due to the high temperatures reached in the solar power plant. Early detection and prevention of leakages is a key issue for the maintenance of those kinds of facilities. Even so, the huge area of solar power plants makes it difficult to perform proper maintenance, due to the large amount of kilometers of collectors and pipelines, as well as the hazardous environment with the really high temperatures reached in its elements.

Nowadays, the inspection is performed by human operators using a thermographic camera while they travel along different parts of the solar plant by car through poorly asphalted and dirt roads. The inspection is carried out while the vehicle is moving, looking at the screen to analyze the thermal image and detecting leakages. Even so, the operator must inspect the image for long time periods (around 2 h) while correcting the the pose of the camera to keep the pipeline in the field of view when irregularities in the terrain appear, which may lead to leakages being left undetected.

4. Proposed Approach

Based on the task specifications posed in the previous section, this paper proposes an autonomous robotic system to perform the pipeline inspection for early detection and prevention of leakages. The autonomous robot patrols along the power plant while inspecting the collectors using thermographic images to identify the leakages. The inspection system continuously analyzes the images in order to monitor the status of the elements, the tube in this case, and highlight anomalies.

Initially, a path is defined for the robotic platform based on a hybrid map approach mixing topological graphs with local occupancy grids. Using the information provided by a GPS/IMU sensor, the robot executes the planned path. During this execution, the system takes advantage of a local metric planner to avoid obstacles and follows the initial path as close as possible.

Even so, irregularities in the terrain (poorly asphalted road and dirt road) make it difficult to perform the inspection using a fixed camera mounted on the robot, as the pipeline can be out of range when slopes and bumps are found on the road. Taking this into account, the addition of a robotic arm is proposed to allow the manipulation of the thermal camera and to track the pipeline through the inspection task. The aim is to establish a coordination between the thermographic inspection and the robot arm movements in order to keep the objective in the field of view of the camera.

The next sections will give information about the used hardware and system architecture, including the different software units defined within the system and the navigation system.

4.1. Architecture

The architecture presented in this paper is based on the specifications and work performed in the MAINBOT project. The main robotic platform used is a RobucarTT developed by Robosoft (http://www.robosoft.com/), designed for outdoor environments and with the Ackermann steering geometry. The robot includes a GPS/IMU sensor for localization purposes, which provides the robot position with an accuracy of 0.20 m. The platform also has a robotic arm attached to it, which is employed in this case to handle a thermographic camera, specifically, an FLIRThermoVision® A20. This thermographic camera is used for leakage detection, as well as for tracking purposes. Figure 2 shows the physical elements of the architecture.

Figure 2. Robotic platform and thermographic camera.

There are two main tasks to be performed by the robotic platform while the predefined path is being executed by the robotic platform: on the one hand, to inspect the pipelines to search for leakages; on the other hand, to track the pipeline in order to send movements to the robotic arm and to maintain the pipe in the field of view of the camera. To this end, as shown in Figure 3, two different modules have been defined in the software architecture:

- **Leak inspection unit:** the unit in charge of receiving images from the thermographic camera and analyzing them in order to detect leakages in the collectors. Based on different morphological operations in the image, the unit is able to detect the leakages in the collectors in an accurate way.
- **Thermal tracking unit:** the unit in charge of tracking the pipeline and commanding the robotic arm to keep the pipe in the field of view of the camera. A particle filter-based tracking system is proposed, because to its capacity to accurately model the underlying dynamics and its rapid adaptation to changing signal features.

Figure 3. Software architecture: leak inspection unit and thermal tracking unit.

Based on this architecture, the system is able to (a) inspect the pipelines and detect leakages and (b) maintain stable the detection process by means of the tracking process, overcoming the problems derived from the irregularities of the terrain, while the mobile robot executes the planned path for inspection.

4.2. Autonomous Navigation

The autonomous navigation approach of the MAINBOT project is based on the use of a hybrid map consisting of a topological graph overlaid with local occupancy grids. Since the workspace is a large area, the overall plan is formed on a topological graph, as planning in a large metric map quickly becomes unwieldy. However, local metric information is used for achieving precise localization (needed for some operations) and obstacle avoidance. Hence, the path planning is performed in two steps:

- The overall plan is created in the topological graph, using Dijkstra's algorithm. This is the basis for the low level planning.
- The robot navigates locally using local metric maps and a search-based planning algorithm.
  - The global metric planner integrated in MAINBOT generates a path from the current position to a desired goal by combining a series of short, kinematically feasible "motion primitives". Planning is done in x, y and theta dimensions, resulting in smooth paths that take robot orientation into account. This is especially important for a RobucarTT robot, as it has nonholonomic constraints (*i.e.*, due to the Ackermann configuration).
  - The local metric planner can be seen as a controller that drives a mobile base in the plane.
  - An execution component called "*move base*" links the global and local planners to achieve the metric navigation.

This navigation approach allows for creating paths for pipeline inspection in two steps, using the "*move base*" component to execute the defined path as accurately as possible. The autonomous navigation module has been developed using ROS (http://www.ros.org) (Robot Operating System) libraries and packages.

5. Leak Inspection Unit

The aim of this unit is the detection of leakages in the collectors based on information provided by a thermographic camera. To this end, initially, (a) the image is analyzed, searching for a pipeline section. Once the pipeline has been found; (b) the section is inspected to detect abrupt changes in the temperature, which indicate that there are leakages in the collectors.

In this process, as the first step, the object parameters of the thermal camera must be fixed. To extract the temperature information from an image, the output data of the thermal camera must be interpreted based on the correct fixing of parameters, such as emissivity, object distance or reflected temperature.

The emissivity is a surface property that states the ability to emit energy; it is expressed as the ratio of the radiation emitted by a surface to the radiation emitted by a blackbody. Emissivity is a unitless quantity and spans from zero to one.

Following the energy conservation law, all energy exchange is compensated mutually: the flux incident, Φ_i, is equal to the flux reflected, Φ_r', absorbed, Φ_a, and transmitted, Φ_t:

$$\Phi_i = \Phi_r' + \Phi_a + \Phi_t \tag{1}$$

In general cases, terms on the right side of Equation (1) are specifically weighted following particular radiative properties related to reflection (ρ), absorption (α) and transmission (τ).

These properties are linked together considering the flux exchanges on a semitransparent object in its environment for which:

$$\rho + \alpha + \tau = 1 \tag{2}$$

The general form of Kirchhoffs law provides a link between the absorption and emission processes and, thus, between emissivity and absorbance, since:

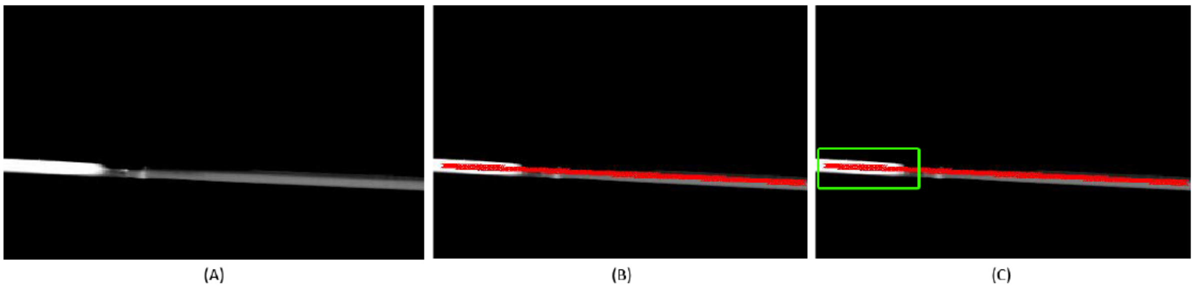
$$\varepsilon(\lambda, \theta', \phi') = \alpha(\lambda, \theta, \phi) \tag{3}$$

The object distance is defined as the distance from the camera to the surface. Finally, reflected temperature indicates the temperature reflected by the surface of an object.

In the case of the glass that covers the collector, it has a high transmissivity, around 94% (0.94), so the emissivity is fixed to 0.04. The estimated reflected temperature is fixed empirically to 10 °C and the object distance to 10 m, based on the features of the environment and the solar plant.

Once the settings are established, the camera processes the thermal information and provides an image where pixels give information about the temperature. These values of temperature are normalized to gray values performing a thermal adjustment between the minimum and maximum values defined (in our case, between 100 °C and 300 °C), as shown in Figure 4A.

Figure 4. (**A**) Thermographic image; (**B**) Pipeline detection; and (**C**) Leak detection.



This image is then thresholded, highlighting the pixels with temperatures above 100 °C, as they are related with the pipelines. Skeletonization/Medial Axis Transform [9] is applied to this image, detecting the longest straight section in the image, as illustrated in Figure 4B. This straight section is then analyzed to detect temperature changes along the pipeline.

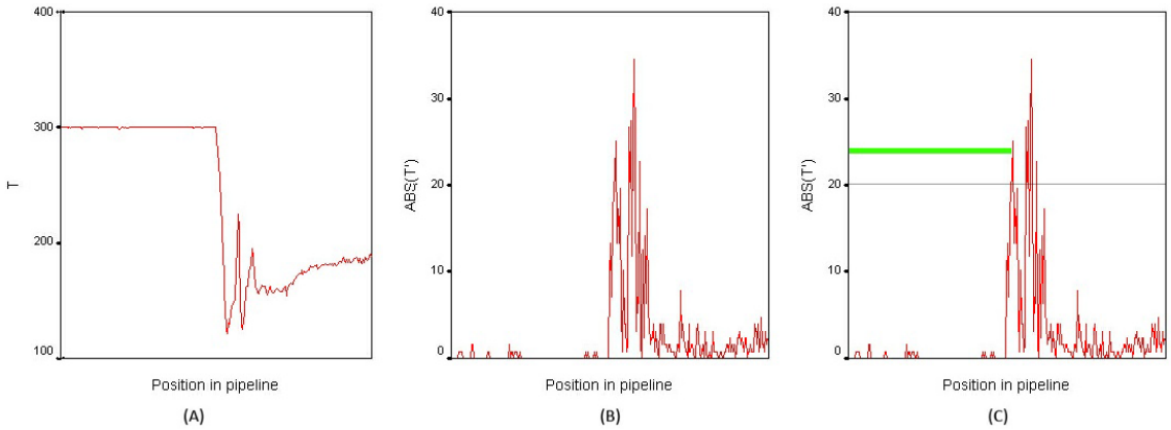
As the first step to detect the temperature changes along the collector, the temperatures of the previously obtained section are stored, as shown in Figure 5A, where the temperature along the collector is plotted. Based on these data, the absolute value of the first derivative is calculated as:

$$ABS(T') = |T_{i+1} - T_i|, \quad {}_{i=[1..N-1]} \tag{4}$$

where T_i is the temperature of pixel i of the pipeline section.

The information about the absolute value of the derivative is used to divide the section into different parts, which have similar temperature, defining a ΔT threshold (minimum change) to perform this division. An important consideration is the fact that the joints of different sections of the collector are metallic and have a high reflectivity, so the obtained temperature is unstable at this point and must be filtered. In order to filter them, an approximate resolution of the image is estimated and short parts with abrupt change of temperature are removed, defining a minimum width of the pipeline part. This allows us to filter short pipeline parts with abrupt temperature changes (joints), detecting real leakages, which fill a wider space on the image. Once the pipeline is divided, the mean temperature of each subsection is computed. The pipeline parts above the maximum temperature are labeled as leakages, as highlighted in green in Figures 4C and 5C.

Figure 5. (**A**) Temperature along the pipeline; (**B**) Absolute value of the first derivative; and (**C**) Detection on the plot.



Based on this algorithm, the *Leak Detection Unit* is able to raise alerts when leakages are detected while the robot is patrolling along the collectors.

6. Thermal Tracking Unit

Irregularities in the terrain can make it difficult to perform a correct leak inspection, as pipelines can be out of the field of view of the camera when slopes and bumps are found along the road. To overcome this problem, a tracking system is proposed based on particle filtering. This tracking system follows the target pipeline through thermographic image sequences, and it is able to generate movement commands when the target is reaching the edge of the image. The next lines will give information about all the elements of the particle filter for thermal tracking.

6.1. Particle Filter

Particle filters [10,11], also known as Sequential Monte Carlo methods (SMC), are sequential estimation techniques that allow estimating unknown states, x_t, from a collection of observations $z_{1:t} = \{z_1, ..., z_t\}$. The state-space model is usually described by state transition and measurement equations:

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{5}$$

$$z_t = g_t(x_t, u_t) \tag{6}$$

where f and g are the state evolution and observation model functions, respectively, and v_t and u_t denote the process and observation noise, respectively.

Based on the previous equations, particle filters allow for approximating the posterior density (PDF) by means of a set of particles, $\{x_t^{(i)}\}_{i=1,\ldots,n}$, using equation:

$$p(x_t|z_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{7}$$

where each particle, $x_t^{(i)}$, has an importance weight, $\omega_t^{(i)}$, associated with it and δ is the Kronecker delta. These weights are computed following equation:

$$\omega_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(z_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})} \tag{8}$$

where $p(z_t|x_t^{(i)})$ is the likelihood function of the measurements, z_t, and, finally, $q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})$ is the proposal density function.

Based on the previously presented equations, the particle set evolves along time, changing the weights of the particles and resampling them in terms of the observations.

Particle filtering provides a robust tracking framework when dealing with non-linear and non-Gaussian state and observation functions, as it considers multiple state hypotheses simultaneously.

6.2. Particle Filtering for Thermal Tracking

In the first step, a tracking process has been defined in an attempt to maintain the object to be analyzed in the field of view of the thermal camera. A particle filter-based tracking is proposed, allowing us to correct the position of the robotic arm to maintain the pipe in the center of the image for a further analysis. In an attempt to develop a thermal tracking system for multiple detection tasks, the particle filter has been generalized to allow the tracking of objects with different shapes and configurable for each tracking process (easy reconfiguration for similar scenarios), although the system modeling explained in the next lines is tuned for the presented environment and posed problem.

6.2.1. System Modeling

In the presented scenario, the tracking process is able to follow a pipeline in successive frames. The state of the process is defined as:

$$X_t = [x_t, y_t, l_t, \alpha_t]^T \tag{9}$$

where x_t and y_t are the x and y pixel coordinates of the center of the pipeline in the image, l_t is the length of the pipeline and α_t is the orientation of the pipeline in time t.

Additionally, the state transition is defined as:

$$X_{t+1} = X_t + \Delta_t \dot{X}_t + V_t \tag{10}$$

$$\dot{X}_t = [\dot{x}_t, \dot{y}_t, \dot{l}_t, \dot{\alpha}_t]^T \tag{11}$$

where Δ_t is the time step, $\dot{X}_t$ is the dynamic part describing the variation of the state elements and V_t is an additive, zero mean Gaussian noise.

6.2.2. Likelihood Evaluation

For the likelihood evaluation, initially, the thermal image is analyzed to highlight the parts with a predefined temperature range, in this case, the temperature of the pipe to be tracked. To this end, thresholding and Skeletonization/Medial Axis Transform algorithms have been used, as in the *leak inspection unit*. From this step, a set of N connected regions are extracted; N possible pipe sections. Those regions form the observation, Z_t, where each region, z_i^t, is defined by their center in pixel coordinates, length and angle:

$$Z_t = z^t_{i=1..N} \tag{12}$$

$$z_i^t = [x_i^t, y_i^t, l_i^t, \alpha_i^t]^T \tag{13}$$

For the likelihood evaluation, initially, the distance between the particle and each of the observed regions is calculated using the pixel coordinates of the center, length and angle as:

$$dist_i = \lambda \sqrt{(x_t - x_i^t)^2 + (y_t - y_i^t)^2} + \beta\, |l_t - l_i^t|^2 + \gamma\, |\alpha_t - \alpha_i^t|^2 \tag{14}$$

where λ, β and γ are coefficients to weigh the importance of the pixel coordinates, length and angle, respectively. The distance between the particle and the observation is then calculated as the minimum distance between the particle and the N regions found:

$$dist(X_t, Z_t) = min(dist_i), \quad {}_{i=[1..N]} \tag{15}$$

Finally, the likelihood is calculated as the exponential of the distance, as shown in the next equation:

$$P(Z_t|X_t) = e^{-dist(X_t,Z_t)} \tag{16}$$

Based on the presented likelihood evaluation, the particle filter estimates iteratively the process state as presented in the next paragraphs.

6.2.3. Particle Filtering Procedure

To initialize the process, when the first pipe is detected, a set of N random particles is drawn around its position and with its scale and orientation. Afterwards, the procedure of the particle filter is given as:

- Find the object in the initial thermal image and initialize N particles, $X_0^{(i)}$, with random samples around it, where $w_0^{(i)} = 1/N$;
- If $ESS < threshold$ (*effective sample size*), draw N samples with *selection with replacement*;
- Predict $x_t^{(i)} = x_{t-1}^{(i)} + v_{t-1}$;
- Update importance weights $w_t^{(i)} = w_{t-1}^{(i)} P(Z_t|X_t)$; 5: Normalize weights $w_t^{(i')} = w_t^{(i)} / \sum_{j=1}^{N} w_t^{(j)}$;
- Set $t = t + 1$, go to Step 2.

In this procedure, ESS [12] (effective sample size) is calculated as:

$$cv_t^2 = \frac{var(w_t^{(i)})}{E^2(w_t^{(i)})} = \frac{1}{N} \sum_{i=1}^{N} (N w_t^{(i)} - 1)^2 \tag{17}$$

$$ESS_t = \frac{N}{1 + cv_t^2} \quad (18)$$

where N is the number of particles and $w_t^{(i)}$ is the weight of particle i in time t.

Based on this discrete approximation of the posterior probability, the object is tracked along the inspection process.

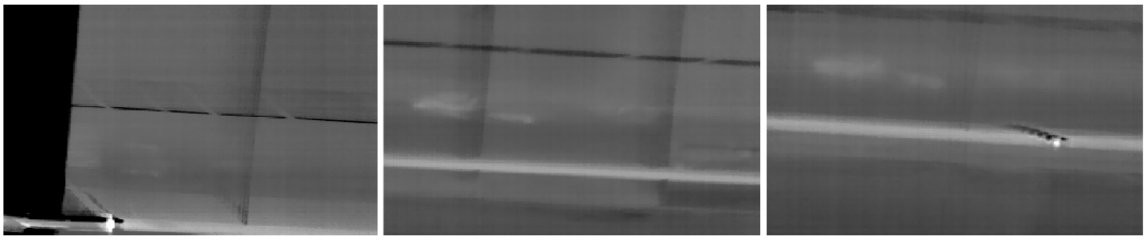
7. Experimental Results

To test the suitability of the proposed approach, a set of experiments have been designed trying to asses both the leakage detection algorithm and the tracking system. To this end, a database of image sequences was created in Valle facilities (see Figure 6) using actual production means and replicating the behavior of the maintenance robot:

- Recording of collectors and pipelines using a thermal camera 10 km at night, as performed now by human operators, divided in sequences of 150 m (half loop of collectors).
- Camera placed on a vehicle circulating at a speed of 20 km/h through a terrain with irregularities.
- Real leakages appearing in the images.

Based on this real data, two different experiments have been performed. The next lines give information about each experiment and the obtained results.

Figure 6. Sequence of thermographic images on Valle facilities.



7.1. Results of the Leak Detection

An efficiency of the leak detection has been tested using 60 different thermal sequences. In those 60 sequences, human operators found a total of seven leakages during the recording session, which were labeled in the database. Those 60 sequences were analyzed by the previously presented algorithm, searching for leakages. A threshold, ΔT, of 70 °C was established for the detection algorithm empirically based on the gathered data. In order to have numeric data, a mean value for each stretch of the tube with temperature change (filtering joints) was also saved.

Figure 7 shows an example of the output of the algorithm while performing the validation test, where each column is related with a sequence. The first row indicates the image sequence ID, while the next rows show the mean temperature of the pipe section along the collectors where abrupt changes of temperature can be observed. Pipe parts with a high temperature (leakages) are marked in red.

Figure 7. Example of changes in temperature in recorded image sequences.

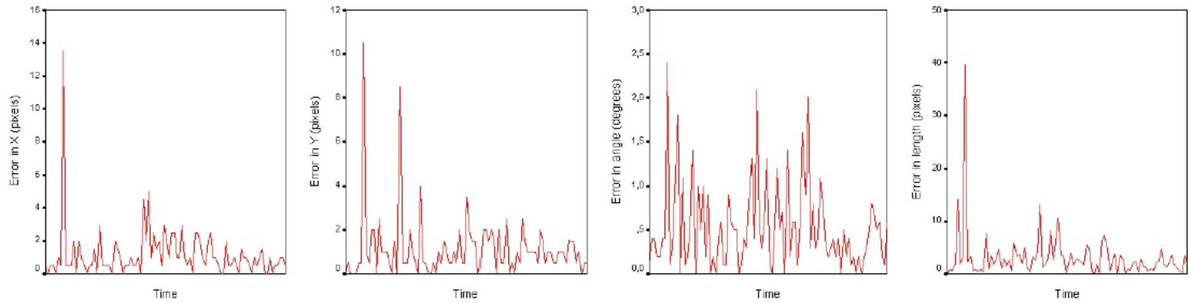
P1A2	P1A4	P1A5	P1A7	P1A8	P1A10	P1A11	P1A12	P1A16	P1A24	P1A25	P1A28	P1A35
136,07	136,07	143,13	135,29	142,35	160,39	156,47	147,84	134,5	142,35	129,8	128,23	149,41
132,15	132,15	132,15	158,82	132,15	142,35	137,64	143,92	142,35	138,43	158,03	119,6	133,72
127,45	127,45	132,15	299,21	144,7	174,5	189,41	153,33	172,15	246,66	207,45	271,76	138,43
299,21	299,21	192,54	298,43	193,33	260,78	202,74	292,94	237,25	217,64	189,41	138,43	205,88
293,72	293,72	299,21	300	292,94	249,8	272,54	252,15	238,03	132,94	122,74	138,43	209,01
290,58	290,58	296,07	298,43	246,66	152,54	152,54	260	170,58	129,01	122,74		139,21
292,94	292,94	141,56	296,07	217,64	148,62	154,11	141,56	146,27				125,09
150,19	150,19	135,29	299,21	252,15			138,43	152,54				128,23
140	140	136,07	299,21	136,07			139,21					
139,21	139,21		148,62	125,88								
	...		145,49									
	...											
	148,62											
	143,13											
	179,21											
	258,43											
	263,92											
	263,13											
	243,52											
	143,13											
	144,7											

The algorithm found all the leakages labeled by human operators, obtaining 100% sensitivity. Besides, six more leakages were found along the sequences, leakages that match exactly with the previous leakage patterns. Based on the similarity and after an analysis of the new leakages, they could be considered as real leaks that were missed by the human operators during visual inspection.

7.2. Results of the Tracking Process

To test the efficiency of the particle filter for tracking, 10 different image sequences have been used, each of them with 200 images approximately (a total of 2,000 images). For each sequence, the particle filter has been initialized using the first pipeline appearance, and the pipeline has been tracked through the rest of the images. In this experiment, the manipulation of the robotic arm has been left out of the scope of this manuscript, as the sequences have been recorded by a human operator from a car and it is not possible to simulate the arm movements. For each frame, the error between the output of the particle filter and the labeled images has been computed, as shown in Figure 8. Images show peaks in the error plots, derived from the image noise and irregularities in the terrain.

Figure 8. Tracking error during a recorded sequence in position in X and Y orientation and length.



Specifically, the experiment measures the tracking error in position, angle and length of the pipeline, using different parameters for the particle filter. Those are the specifications of the experiment:

- Six different particle filter configurations have been set up, mixing different state estimation methods and numbers of particles. Specifically, the state estimation methods are:
  - Best particle (the one with maximum weight);
  - Robust mean with the 10 particles with the maximum weight;
  - Weighted mean using the whole particle set;
- For each configuration, the previously cited image sequences have been used, tracking the pipeline through around 2,000 images; the mean error and standard deviation of the position, angle and pipe length has been measured for all the configurations;
- In the likelihood evaluation step, the same values for coefficients λ, β and γ have been used, applying values that give a similar weight to position, angle and length.

Table 1 shows the obtained results. The first column describes the number of particles and the estimation method, the second and third columns, the mean (μ) and the standard deviation (σ) of the error in the position and the fourth and fifth columns, the mean (μ) and the standard deviation (σ) of the error in the angle, and finally, the last columns show the mean (μ) and the standard deviation (σ) of the error in the length.

Table 1. Results of the Tracking Process.

	Pos. Error (pix)		**Angle Error (°)**		**Length Error (pix)**	
	μ	σ	μ	σ	μ	σ
500 - Best particle	1.80	1.27	0.96	0.86	3.16	2.58
1,000 - Best particle	**1.42**	**0.95**	0.72	0.63	2.94	**2.51**
500 - Robust mean 10	1.82	2.11	0.51	0.41	2.86	5.35
1,000 - Robust mean 10	1.57	1.83	**0.41**	**0.34**	**2.65**	5.66
500 - Weighted mean	6.34	5.39	0.43	0.36	3.67	7.68
1,000 - Weighted mean	6.11	4.67	0.35	0.38	3.45	6.64

Results show a better performance of the particle filter with a set of 1,000 particles, improving the mean error and standard deviation in almost all the configurations. The addition of a bigger set of particles could decrease the error rates, although it would be necessary to optimize the computational time as much as possible to ensure a suitable frame rate. In the case of the state estimation methods, the best particle and robust mean methods show very similar results, overcoming in any case the errors of the weighted mean.

8. Conclusions and Future Work

This paper presents an autonomous leakage detection system for maintenance tasks in extensive industrial plants, including a leakage detection system. Based on the information provided by a thermographic camera, the system is able to detect leakages in pipelines and collectors of thermal power plants. This system is enhanced by a particle filter-based tracking system, used to maintain the target in the field of view of the camera and stabilize the detection process when irregularities are found along the road.

The results show a high success rate of the leakage detection unit, reaching 100% sensitivity (on data labeled by operators) and detecting even more leakages on the recorded sequences. The use of thermographic information allows for detecting the fluids leaked from the collectors, taking advantage of morphological operations to highlight the leaks in the thermal images.

Besides, a tracking system has been added to manage the thermographic camera and to avoid the loss of the target. The particle filtering for tracking has shown a position error of less than 1.5 pixels and less than a 0.5° error in the angle. It also allows for modeling the system in a simple and effective way, adapting rapidly to the changing image features and data noise.

As further work, there are two main paths to follow: on the one hand, to add the robot arm movements and to mix the information extracted by the leak inspection unit and thermal tracking unit with the navigation module to create a complete system that takes into account the navigation information in the inspection task and *vice versa*. In the same way, it would also be interesting to test the tracking system in similar inspection tasks where different elements of the solar plant are analyzed (valves, vessels, *etc.*) and observe its suitability.

Acknowledgments

This work has been performed within the scope of the project "MAINBOT: Mobile robots for inspection and maintenance activities in extensive industrial plants". MAINBOT is a small or medium-scale focused research project supported by the European Commission in the 7th Framework Program (285245). For further information, see http://www.mainbot.eu.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Debenest, P.; Guarnieri, M.; Takita, K.; Fukushima, E.F.; Hirose, S.; Tamura, K.; Kimura, A.; Kubokawa, H.; Iwama, N.; Shiga, F.; *et al.* Expliner—Toward a Practical Robot for Inspection of High-Voltage Lines. In Proceedings of the Field and Service Robotics (FSR), Cambridge, MA, USA, 14–16 July 2009; pp. 45–55.
2. Lee, J.K.; Cho, B.H.; Jang, K.N.; Jung, S.C.; Oh, K.Y.; Park, J.Y.; Kim, J.S. Development of Autonomous Cable Inspection Robot for Nuclear Power Plant. In Proceedings of World Academy of Science, Engineering and Technology, Rome, Italy, 28–30 April 2010; pp. 314–318.
3. Luk, B.L.; Collie, A.A.; Cooke, D.S; Chen, S. Walking and Climbing Service Robots for Safety Inspection of Nuclear Reactor Pressure Vessels. In Proceedings of the Asia Pacific Conference on Risk Management and Safety, Hong Kong, 1–2 December 2005; pp. 432–438.

4. Suzuki, M.; Yukawa, T.; Satoh, Y.; Okano, H. Mechanisms of Autonomous Pipe-Surface Inspectic Robot with Magnetic Elements. In Proceedings of the IEEE International Conference on System Man and Cybernetics, Taipei, 8–11 October 2006; pp. 3286–3291.
5. Camerini, C.; Freitas, M.; Langer, R.A.; von Der Weid, J.P.; Marnet, R.; Kubrusly, A.C. A Rob for Offshore Pipeline Inspection. In Proceedings of the 9th IEEE/IAS International Conference (Industry Applications (INDUSCON), Sao Paulo, Brazil, 8–10 November 2010; pp. 8–10.
6. Xu, J.; Ikram-ul-haq; Chen, J.; Dou, L.; Liu, Z. Moving Target Detection and Tracking FLIR Image Sequences Based on Thermal Target Modeling. In Proceedings of the Internation Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsh China, 13–14 March 2010; pp. 715–720.
7. Senthil Kumar, K.; Kavitha, G.; Subramanian, R.; Ramesh, G. Visual and Thermal Image Fusic for UAV Based Target Tracking. In *MATLAB—A Ubiquitous Tool for the Practical Enginee* Ionescu, C.M., Ed.; InTech: Rijeka, Croatia, 2001; pp. 307–326.
8. Padole, C.N.; Alexandre, L.A. Motion Based Particle Filter for Human Tracking with Therm Imaging. In Proceedings of the 3rd International Conference on Emerging Trends in Engineerir and Technology (ICETET 2010), Goa, India, 19–21 November 2010; pp. 158–162.
9. Aichholzer, O.; Aurenhammer, F. Straight Skeletons for General Polygonal Figures in the Plane. Proceedings of the 2nd Annual International Conference on Computing and Combinatorics, Hoı Kong, 17–19 June 1996; pp. 117–126.
10. Doucet, A.; de Freitas, N.; Gordon, N. *Sequential Monte Carlo Methods in Practice*; Springe New York, NY, USA, 2001.
11. Kotecha, J.H.; Djuric, P.M. Gaussian Particle Filtering. *IEEE Trans. Signal Process.* **2003**, *5* 2592–2601.
12. Liu, J.S.; Chen, R.; Logvinenko, T. A theoretical Framework for Sequential Importance Samplir and Resampling. In *Sequential Monte Carlo Methods in Practice*; Springer: New York, NY, US/ 2001; pp. 225–246.

7.5. Wearable Technology in Automotive Industry: from Training to Real Production, Human-Computer Interaction

Iñaki Maurtua. Wearable Technology in Automotive Industry: from Training to Real Production, Human-Computer Interaction, Chapter 4, Inaki Maurtua (Ed.), InTech, (2009) DOI: 10.5772/7742. Available from: https://www.intechopen.com/books/human-computer-interaction/wearable-technology-in-automotive-industry-from-training-to-real-production [105]

Wearable technology in automotive industry: from training to real production

Iñaki Maurtua
TEKNIKER Technology Centre
Spain

1. Introduction

This paper describes two different research activities carried out in the context of the wearIT@work project to analyse the benefits and drawbacks of using wearable technology to support blue-collar workers in real situations. To this aim we describe the UCD process followed, the tests performed and the results obtained in several experiments performed at the SKODA production facilities in Czech Republic.
User Centred Design approach has been followed in the two scenarios identified: training and quality control activities.
Compared to stationary computer systems, mobile- and wearable computing technology have seriously caught up in performance, functionality, scalability. This makes training solutions based on mobile- and wearable computing an attractive consideration for industrial organisations. In this sense, the objective was to supplement the training procedures at Skoda with a context-sensitive wearable computing solution. The idea was that the trainees gained mobile access to the information to carry out their assembly tasks. In fact, the wearable system was used to recognize the context of performed work, and as a result to provide the trainee with the required information to adequately perform individual assembly tasks. The wearable solution was able to track and analyse the trainee's actions, while providing the end-user with means for error handling. As a result, semi-autonomous training of trainees in automotive production was possible.
In the second scenario we moved to real production environment, specifically to the CP8: check point 8, the place where cars are visually and manually inspected before being delivered to customers. There, two worker teams work in parallel in inspection tasks: examining lights and bumpers misalignment, identifying dents and scratches, checking the spaces between doors and windows, and any other kind of faults.
Each time an error is found, the worker reports in one of the three check-list forms they handle. The sheets have a matrix structure identifying the different parts in the car and the possible types of faults.
The objectives of using wearable technology in this scenario are multiple, among others: making the worker activity easier and more efficient, allowing a paperless and hands-free inspection, to guarantee that all verifications have been performed (avoid oversights-mistakes), allow permanent documentation access and to facilitate workers interaction.

2. Summary of activities carried out

The '*wearit@work: empowering the Mobile Worker by wearable computing*' project aimed to study wearable technologies and their application in the workplace to improve the conditions in which workers carry out their activities.
To this end we created the multidisciplinary consortium with 35 partners from 15 countries that have worked together for the four and half years that the project was officially programmed for.
The technological developments have been based on User Centred Design (UCD) and have been developed in 4 different scenarios:

- Maintenance. Led by Giunti Labs and EADS. The target market was made up of workers carrying out maintenance tasks in the aeronautical sector.
- Emergencies. Led by the Paris fire service and Fraunhofer FIT. The study focused on fire-fighters in emergency situations.
- Health. Led by GESPAG and SAP. This study focused on health workers on their daily ward rounds in hospitals.
- Production. Led by TEKNIKER and SKODA. This is the scenario that Tekniker has been most involved in and which will be described in more detail below.

The following common viewpoint was reached after internal discussions about the wearable computing concept compared with mobile computing:

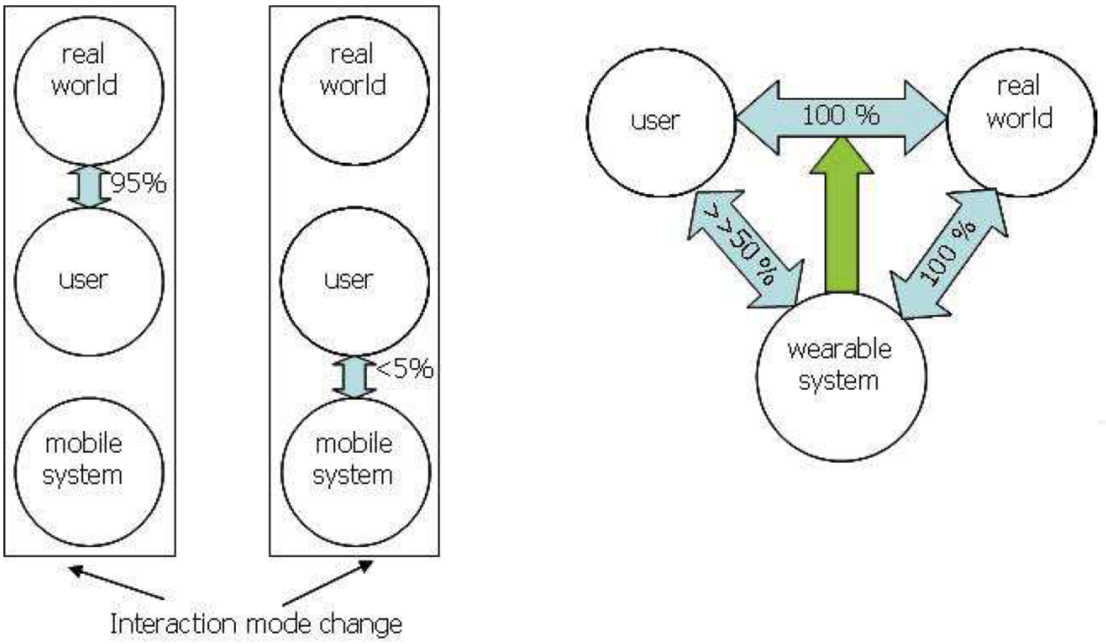


Fig. 1. Mobile vs Wearable

"Wearable computing is working with instead of on the computer, meaning: Using the computer is not the main activity but it is used while the activity is being carried out"

In the production scenario, TEKNIKER focused on designing prototypes and experimenting with end users in the two SKODA plants where the developments and pilot studies were validated.

- Pilot Case Study: Learning in the Vrchlabí plant
- Pilot Case Study: Quality Control in the Mladá Boleslav plant

3. Pilot Case Study: Learning

It took place at the SKODA plant in Vrchlabí. The facilities have what they call an 'E-factory', this being the certification process that must be passed by the workers before they start working on the production line. During this training period they are taught the basics of the work related to a particular task by means of a PowerPoint presentation and small questionnaires. After passing the test they then have a prototype car chassis on which they have to carry out the assembly task in question within a time limit while being supervised by an instructor. When they achieve the established production parameters they can join the assembly line.

Fig. 2. The learning island, the foreman and the trainee

This pilot case study was designed to create an environment that facilitated the worker training programme so that:

- The training period could be shortened
- It could be carried out by the trainees alone, without needing to have an instructor

When designing the prototype, different factors were taken into consideration to ensure that the wearable system did not interfere with the assembly work and that it could be accessed as naturally as possible.

The development process involved working directly with the plant workers and carrying out a usability experiment in the plant.

The prototype consisted of the following components:

- Hardware: Xybernaut V computer, Microoptical SV6 Head Mounted Display and a standard headset.
- Software. A programme that presented the 14 steps needed to carry out the chosen task in the proper sequence and allowed different forms of interaction. Each of the tasks was documented with the use of text, photo and video resources that could be selected by users according to their preferences.

Fig. 3. UI of the application

The operators had to use the application to 'learn' and practice assembling the right-hand front headlight.
As certain functionalities were not yet implemented, the Wizard of Oz (Kelley J.F., 1983) technique was used to simulate them. For example:

- We did not have an Automatic Speech Recognizer (ASR) for voice-based interaction, so we used a Czech speaker who understood the user and 'simulated' the interaction using a remote control application.
- To simulate task recognition, the 'Wizard' pressed the 'Next' button every time the operator completed a task, so that the information about the following task was displayed on screen.
- We wanted to test the validity of a virtual keyboard, so we created keyboard strips that were attached to the operators. They pressed the simulated keys and the 'Wizard' interacted with the application.

Fig. 4. Virtual fabric keyboard layout

The system recorded the most important process information: time used and mistakes.
The operators had to complete an evaluation questionnaire after carrying out the process.

The analysis of the experiment results allowed us to conclude that the system was well accepted in general terms, highlighting the nil use of the videos and the difficulty of accessing the information using the HMD.
As a result of the experiment, we decided to carry out more intensive experiments at TEKNIKER, in which a large number of individuals could take part without the limitations found in the Czech Republic caused by the lack of a common language for the participants and the TEKNIKER researchers.

3.1 Experiment: Validation of Wearable Technology for Learning Tasks

In order to carry out the experiments, we built a metal structure on which those taking part in the tests had to carry out the allocated assembly tasks. The following was taken into account when designing the structure:

- It should allow different manual tasks to be carried out
- All tasks should be of similar difficulty

Fig. 5. Platform for experiments at Tekniker

The task to be carried out by the operators consisted of:

- Assembling a set of metal shapes (three shapes) on a metal plate with 100 holes with different size threads. The assembly had to be carried out using Allen head screws and an Allen key.
- Make 3 pairs of connections using three wires on a panel with 16 BNC connectors.

The experiment was designed to:

- To validate the initial results of the experiment at Vrchlabí
- To measure user acceptance
- To analyse the improvement in learning in terms of time

40 workers divided into two groups took part in the experiment:

- With the first group, the aim was to measure and compare their efficiency in assembling when they accessed to information on paper compared to using wearable technology. To this end, the participants in the experiment had to carry out a single assembly as quickly as possible.
- The second group had to learn (memorise) the whole assembly process as quickly as possible, this being the parameter to be measured. In this case, the operators had to carry out the assembly as many times as was needed with the support of information until they managed to learn it and carry it out with no help at all. The information could be accessed on paper or by using wearable technology. In order to take into account the short-term memory effect, the assembly had to be carried out once again the following day.

In both cases, the operators had to perform the experiment twice: once using information on paper and once using wearable technology. In the second case, each operator was allocated a different mode of interaction:

- Access always involved using an Head Mounted Display (HMD)
- Interaction was in one of the following modes: textile keyboard on the sleeve of the working clothes, by voice or implicit interaction, this involving simulating that the system had detected completion of a task to present the information about the following task automatically.

The following parameters were measured during the experiment:

- Time
- Mistakes made
- Users acceptance, measured through open and closed questionnaires
- Measurement of mental load. We used the NASA TLX test (Hart & Lowell,1988), a subjective tool for evaluating the mental load of the operators when they carry out activities using machines
- Influence of learning styles. We used the VARK questionnaire (Fleming, 2001) to provide a learning preferences profile in accordance with four categories: Visual (learn through seeing), audio (learn through listening), Reading/writing (learn through reading) and kinaesthetic (learn through doing)

The workers used the following infrastructure for the wearable technology test:

- Microoptical VI head mounted display
- OQO computer

As at Skoda, we used the Wizard of Oz technique to simulate the three types of interaction (textile keyboard, voice and automatic recognition of activities).
The following fundamental conclusions were drawn from the experiments:

- Users improved their efficiency when they used recognition of activities as their source of automatic interaction. The operations were performed faster and with fewer mistakes: using implicit interaction (automatic recognition of activities) they used an average of 67 seconds less than when accessing the information on paper, the second best option.

- Users did not learn faster using wearable technology. They only achieved similar learning times to learning using paper when they used implicit interaction.
- Curiously, during the test carried out the following day, learning using paper gave the best results and implicit interaction the worst.
- Interaction using the voice was the option preferred by the users.
- Information in images was better accepted than that presented as text.
- In general terms, the workers rated the wearable technology based system as very useful for carrying out complex tasks, allowing them to work with their hands free and avoiding movements to access information

3.2 Experiment: Usability of Visual Information Access Devices

The second experiment was designed to compare the benefits of using Head Mounted Displays (HMDs) with using a large monitor near the workstation to access the information. 20 workers took part in the experiments and they had to perform four complete assembly procedures using the above-described infrastructure.
The devices compared were:

- A large monitor near the workstation
- HMD Carl Zeiss look-around binocular
- HMD, Carl Zeiss see-through HMD
- Microoptical VI monocular look-around.

The time used by each operator for assembly was measured and a usability questionnaire was used to measure the satisfaction and subjective opinions of the participants.
The experiment allowed us to conclude that:

- The best response in terms of times was obtained when using the big monitor
- The worst response was obtained using the binocular model HMD. However, we should point out that this access model obtained the best ratings in the satisfaction survey.

Fig. 6. Zeiss HMD used in the experiment

3.3 Wearable System Prototype

On the basis of the results obtained from the experiments and trials at the Skoda plant and at Tekniker, the first 'real' prototype was defined and built in collaboration with ETHZ, Univ. Passau and HP.

The prototype required the inclusion of different kinds of sensors:

- 5 FSR (force sensitive resistor) and 4 reed switches on the chassis of the car used for the learning process. These sensors detected the termination of certain assembly operations: screwing certain parts, fitting the headlight and checking its correct alignment with the bodywork.

The prototype also consisted of the following elements:

- An RFID reader on the back of the user's glove. This allowed the user to verify that the right tools are used for the fixing operations.
- The same glove was fitted with a triaxial accelerometer and a gyroscope to detect the movement caused when the preset torque was reached in certain fixing operations.
- All this information was sent via bluetooth to a central P.C. where the activity detection data was processed. Here we used the Context Recognition Toolbox, a Framework containing a set of frequently used data processing algorithms.

The wearable computing prototype used was that shown in the Figure and consisted of:

- For the wearable computer we used an OQO Model 01+ attached to the operator's belt.
- Microoptical VI monocular HMD
- A Bluetooth Sony Erickson HBH-300 headset
- A sensorised glove with the RFID reader, gyroscope, accelerometer and a set of FSR sensors to detect muscle activity

Fig. 7. Hardware prototype

The software architecture used is shown in the figure below:

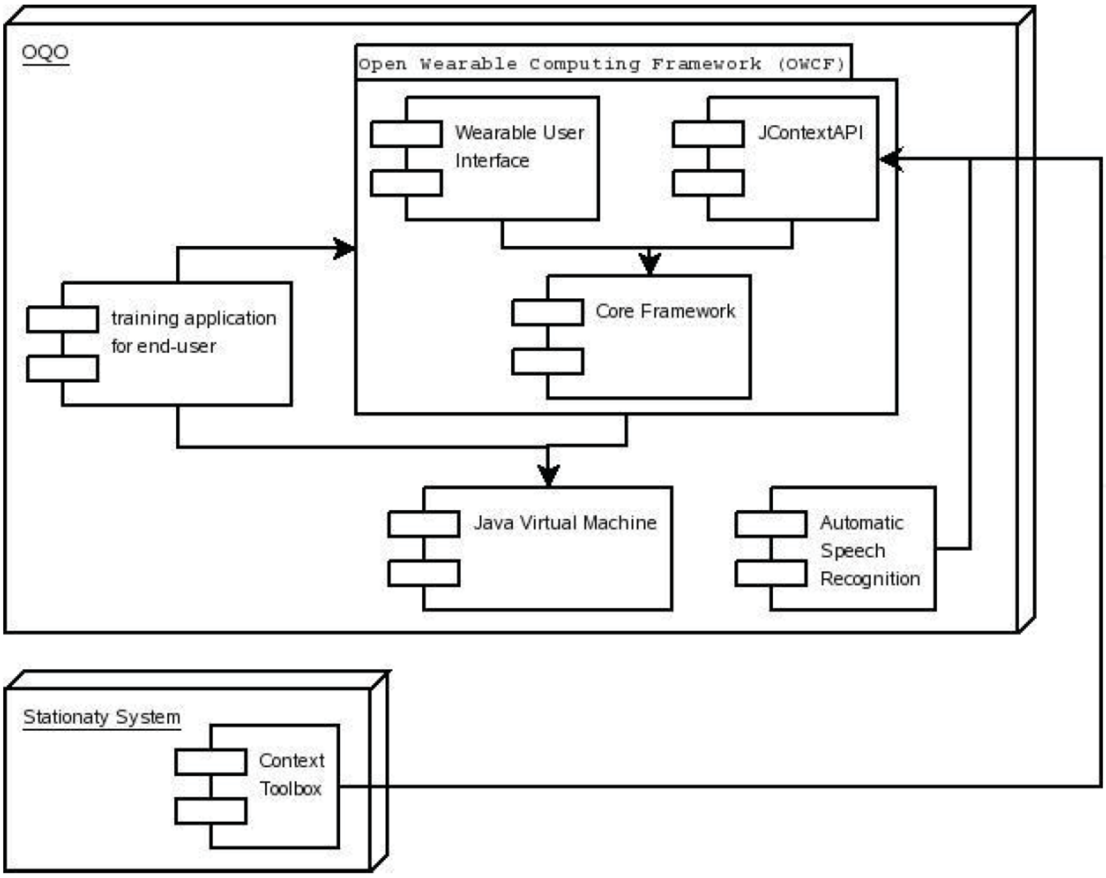


Fig. 8. Diagram showing the software architecture used for the prototype

The application was written in Java and used the Open Wearable Computing Framework (OWCF) developed in the project, more exactly the following components:

- Wearable User Interface (WUI)
- Context component (JContextAPI)

The application was modelled as a state machine in which each of the sub-tasks making up the assembly process were represented. The transitions were triggered by user actions, both explicit (e.g. by voice) and implicit (automatically detected by the system).
The application could monitor the actions of the user and inform him/her if there was a mistake detected.

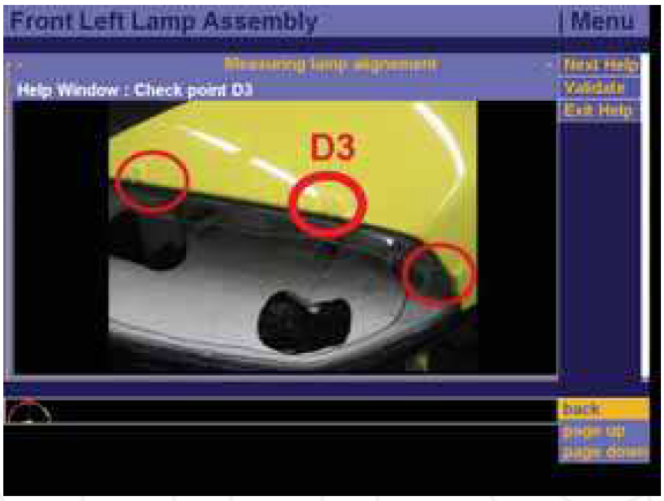


Fig. 9. Example of the UI, showing the verification positions

4. Pilot Case Study: Quality Control

After the initial phase, where we focused on the operator training process, we transferred our attention to the production line. After in-depth analysis of the possibilities of applying wearable technology, we opted to work on the last of the operations carried out on the cars just before they are sent to the dealers. Manual and visual inspection of the finish of the cars is carried out in this phase.

The process takes place on an inspection line, where a team of operators (two parallel lines of ten operators each) verifies different parts of the cars and, where necessary, repair small defects and misalignments. Basically, they detect impacts, marks on the paintwork, correct door opening and closing, headlight misalignment, etc. Hands and eyes are the tools used as inspection tools. During the inspection process they have to enter in the vehicle, stand, sit, crouch and walk around the cars.

During the process they carry a set of sheets with a matrix on them, on which they have to note down the errors detected. The procedure should not be regarded as a check-list. On the contrary, the operators use their own criteria to carry out the inspection and only mark the form when they find something wrong. This procedure can clearly lead to possible mistakes.

Once again, we followed the user-based design process to design the system aimed at helping operators to carry out their wok in a more usable fashion. To this end, we carried out two on-site studies that gave us an in-depth understanding of the process and gather the opinions of the potential users. This process took place in late 2006 and early 2007 and included video recordings and interviews with operators and line managers.

4.1 Social factors

Over the two days of our visit to the Mladá Boleslav plant, we used the following techniques to gain an understanding of the process from the social point of view:

- **Observation** – of the workers performing their normal tasks in the workplace, the interactions between them and the supervisors, time pressure, speed of the activity, rotations, atmosphere at work, etc.

- **Interviews** – semi-structured, based on a questionnaire and aimed at identifying the expectations of the workers as regards wearable computing, their attitude to new technologies and factors related to their normal activities in the plant. 8 people took part in the interviews (4 men and 4 women)

The study led to the following conclusions:

- The atmosphere at work was calm and relaxed. The environment was clean, tidy and well lit. There are common areas for relaxation and meetings. The staff is friendly and polite.
- The workers were continually interacting and helping each other. They rotated throughout every eight-hour shift to avoid the monotony of repetitive work. One person in each team acted as a "joker" to allow other members of the team to leave the line in case of need.
- At inspection point 8, the subject of the final pilot case study, the working atmosphere was even more relaxed.
- **Communication** – All the workers interviewed thought that wearable computing could be an interesting support tool. They didn't see it as a threat that would limit their capacity for interaction, but they didn't think it was either viable or desirable to use it permanently.
- **Privacy** – Although this was one of our major concerns, the interviewees didn't rate it negatively. In fact, they said that the presence of cameras in the plant was already more 'threatening' than the solution we were proposing.
- **Responsibility** – There were contradictory ratings regarding the possibility that wearable computing might affect the responsibilities assumed at the workstation.
- **Experience with the technologies** – except for one person, all the others were habitual users of computers. Obviously, none of them had experience with wearable computing.

In general terms, the workers were willing to use wearable computing because it could be of benefit to them in the following ways:

- To avoid errors and oversights
- To perform the work more quickly and efficiently
- To improve communication mechanisms

4.2 Activity recognition

During this stage of the study, we acquired data to create a prototype that would allow us to identify the activities performed by the workers.
The following sensors were used to capture the movements of the trunk and upper extremities:

- 7 **Xsens MTx** inertial sensors to detect body posture
- FSR (force sensing resistor) on the arms: 8 FSR of 4.7 x 4.7 cm. Fitted to each arm to detect gestures and activities
- A set of Ubisense sensors allowed us to locate the position of the operator of the car being inspected. In practice, they gave calibration problems

A worker was equipped with this equipment and the data was captured during an inspection process to allow later calibration of the algorithms for automatic recognition of activities.
Given the low quality and insufficient data obtained, the process was repeated in the laboratory with 8 individuals and 10 hours of data capture. See (Stiefmeier, 2008) for more details.

4.3 Prototype

The prototype for inspection post number 8 was designed to:

- Facilitate and make the activities of the operators efficient
- Allow paper-free inspection
- Guarantee verification of all points by avoiding oversights
- Provide permanent and easy access to the documentation
- Enhance interaction between the workers

On the basis of these goals, a prototype was created with the following functionalities:

- The worker and the car arriving at the inspection post are identified automatically
- Compliance with all programmed inspections is recorded
- Every time a worker is located in a different area around the car, he/she is presented with a list of possible tasks
- Voice is used as the interaction mechanism: using natural language or identifying the columns and rows that represent the verification document (check-list type)
- The OQO cursor is used for interaction whenever necessary
- The operator can consult the list of faults, related documents and establish a VoIP connection with other operators or the shift manager

Fig. 10. Inspection using the wearable prototype

The system consisted of the following Hardware:

- Zeiss HMD binoculars
- OQO
- Headset
- The above-mentioned system for recognising gestures and activities
- A specially-designed jacket to carry the hardware (Bo et al., 2006)

4.4 Test & Results

The prototype created was used with 8 Skoda workers to evaluate different factors:

- The validity of paper-free inspection
- Remote support
- Access to documentation
- Voice interaction
- Recognition of activities
- Usability of the HMD and the jacket

A document explaining the experiment was drafted and the workers were asked to collaborate voluntarily. Finally, 8 of them took part in the experiment.
The experiment took place without interrupting normal work processes. The process was as follows:

- They were told how the system worked (in Czech) and they were helped to adjust the hardware
- They carried out their work with the prototype as a support element
- They answered two questionnaires (open and closed questions with a 7 point Liekert scale)

The most significant results of the study were as follows (in brackets the mean value of their answers on a scale from 1 to 7, where 1 means completely disagree and 7 completely agree):

- The system is easy to use (3.7)
    - The users didn't have long to familiarise themselves with the system. A longer adaptation session would improve this perception
- I'll be able to carry out my work faster (4.8)
    - An interesting result that could be improved as they become more used to the system
- The system is comfortable (3.0)
    - Heat and the size of the jacket (one size only) could be the reason for this result
- The system is easy to learn (5.0)
    - It's clear that it's easy to learn, but they need to become more familiar with its use
- The information is effective for completing the work (4.8)
    - Positive response
- The font size is right (4.6)
    - A specific trial should be carried out to optimise this size

- I like the interface (5.5)
    - Positive response
- The system has the expected functionalities (5.3)
    - It responds to initial expectations. Continued use could lead to new expectations
- In general, I am satisfied with the system (4.4)
    - Positive response
- I can contact the supervisor easily (5.4)
    - Relevant
- I always knew where I was in the application (5.7)
    - Means that the UI was well designed to avoid operators getting lost between windows
- The commands are obvious (5.8)
    - They knew what the result of each command was going to be
- The system responds quickly to the commands (4.8)
    - This response is more positive than it may seem, as the Wizard of Oz was used for certain functionalities, something that creates an additional response delay
- I would recommend the system to my fellow workers (4.4)
    - Positive and could be improved by resolving certain aspects.
- The glasses are heavy (3.0)
    - The glasses used (Binocular HMD from Carl Zeiss) offer the best quality vision, but they are heavy and their use in longer sessions should be analysed
- The jacket makes my work more difficult (4.3)
    - Interesting, despite the problems inherent with using a single size for all
- The system is heavy (3.9)
    - Includes the weight of the complete unit: jacket + OQO + Cable + HMD battery
- I would use the current system if it were optional (2.7)
    - Very negative. Taking other answers into account, we believe that re-engineering the system would improve this rating.
- The on-screen image is large enough (6.8)
    - Corroborates the results of the experiment at Tekniker.
- I will make less mistakes in my work (4.0)
    - Positive but with room for improvement. The limited time of the experiment did not allow them to evaluate the benefits of the system when real oversights occurred.
- The voice interaction is simple (4.5)
    - Positive but with room for improvement through training
- The jacket is comfortable (3.3)
    - The size and stiffness of the OQO-HMD cable may lie behind this response.
- Paper-free inspection is easier than using paper (4.3)
    - Positive but with room for improvement through continued use of the system.
- The clothing is very hot (5.9)
    - This is the most critical factor. Most of the heat comes from the OQO and not from the clothing itself.

- It is easier to access the information on paper (5.0)
  - Negative response. The fact that we used a PDF document with no specific formatting may lie behind this.
- I prefer to access the information using the glasses than by using a large screen (3.8)
  - No clear conclusion, as they didn't evaluate access to such a screen from different positions in and around the car.
- I felt tense at times (5.3)
  - Not relevant. The presence in the surroundings of 5 members of the research team during the experiment may explain this response.
- The glasses made my work more difficult (4.7)
  - They sometimes forgot the fact that this model has an option allowing them to partially remove the glasses.
- I like the on-line supervisor option (5.9)
  - Positive.
- I felt controlled (3.8)
  - By the system or by the 5 members of the research team?
- I wouldn't like to use the system all day long (4.2)
  - Resolving certain aspects could improve this score.
- I felt that the system distracted me (4.9)
  - It was the first time and everything was new: from the hardware to the functionality
- The menus and the information were well organised (6.8)
  - This means that the UI was well designed.
- I felt stupid wearing the glasses (5.7)
  - Although negative, if everyone used the same system they wouldn't feel that way.
- It's easy to get used to the glasses (3.6)
  - They didn't have enough time to become familiar with them.
- The jacket made it difficult to get into the car (5.4)
  - The size and stiffness of the OQO-HMD cable may lie behind this response.
- I was afraid of breaking the system (3.3)
  - Positive, it was the first time and the system was not specifically designed not to be damaged or not to cause damage.
- I was afraid of damaging the car with the system (3.4)
  - Positive, it was the first time and the system was not specifically designed not to be damaged or not to cause damage.

To sum up, the evaluation was positive, although we detected certain aspects with room for improvement.

5. Final conclusions and acknowledgement

The wearIT project has highlighted the possibilities of wearable computing when it comes to improving the working conditions of certain groups of workers. It's obvious that there is no single valid option for all situations, but each case needs to be analysed with the help of the user groups to design a tailored solution.

In more general terms, we need to make more effort to design the hardware elements needed to exploit these benefits: lighter HMD (wireless where possible), more robust computing elements (the original QBIC solution continues to be an attractive option, due to its shape, although it has connection and computing power limitations), longer life batteries, more usable headphones, systems with better heat dissipation, etc.

We would like to thank the European Union for the funding that made it possible to carry out this research as part of the Sixth Framework Programme.
We would also like to thank all the Partners who collaborated on the project, especially Skoda, ETHZ, Univ. Passau, Univ. Cork, HP-Italy, Edna Pasher, Carl Zeiss, TZI-Univ. Bremen and BIBA

6. References

Fleming, ND (2001). *Teaching and Learning Styles: VARK Strategies,* Honolulu Community College, ISBN 0-473-07956-9, Place of Publication

Hart, S. ; Staveland LE. (1999). Development of NASA-TLX (Task load Index) : Results of Empirical and Theoretical Research, In: *Human Mental Workload,* Hancock & N. Meshkati, 239-250, North Holland Press, ISBN, Amsterdan

Stiefmeier, T. ; Roggen, D. ; Ogris, G. ; Lukowicz, P. & Tröster, G. (2008). Wearable Activity Tracking in Car Manufacturing, *IEEE Pervasive computing*, Vol. 7, No. 2 (April-June 2008), ISSN 1536-1268

Bo, G.; Lorenzon, A., Chevassus, N. & Blondie, V. (2006). Wearable Computing and Mobile Workers: The Aeronautic Maintenance Showcase in the WearIT@Work Project, *Proceedings of IFAWC 2006*,pp 33-44, ISBN 978-3-8007-2954-8 , Bremen, March 2006, Mobile Research Center, TZI Universität Bremen

Kelley, J.F. (1984). An iterative design methodology for user-friendly natural language office information applications, *ACM Transactions on Information Systems*, pp 26-41, ISSN 1046-8188, Boston, 1984, ACM, New York

7.6. Experimenting Wearable Solutions for Workers' Training in Manufacturing

Maurtua I., Unceta M., Pérez M.A. (2007) Experimenting Wearable Solutions for Workers' Training in Manufacturing. In: Jacko J.A. (eds) Human-Computer Interaction. HCI Applications and Services. HCI 2007. Lecture Notes in Computer Science, vol 4553 (663-671). Springer, Berlin, Heidelberg [106]

Experimenting Wearable Solutions for Workers' Training in Manufacturing

Iñaki Maurtua, Miren Unceta, and Miguel Angel Pérez

Production Engineering Department Fundación Tekniker Eibar, Gipuzkoa, 20600, Spain
imaurtua@tekniker.es

Abstract. This paper presents several experiments carried out by the AmILAB research group of Tekniker in the framework of the WearIT@work project (EC IP 004216) on the use of wearable technology base solutions for the training process of workers in a manufacturing environment. The description includes both the initial work at the SKODA production facilities in Czech Republic and the experiments with local workers at Tekniker. As an introduction, the authors of this paper shortly describe current training processes at Skoda, and derive the potential benefits and risks of applying wearable computing technology.

Keywords: wearable computing, automotive production, training, task recognition.

1 Introduction

In a manufacturing environment many situations exist where work is performed in a mobile manner (e.g. maintenance, work at the assembly line, etc.). Since the late 90's, applying Wearable Computing in industrial work environments has become an attractive approach to efficiently support mobile work processes [1].

The European research project WearIT@work investigates, among other scenarios, the impact of wearable computing in automotive production. Since the beginning of 2004, a thorough analysis was carried out, involving interviews, field studies and comprehensive process analysis at the Skoda production facilities in *Mlada Boleslav* and *Vrachlabi* (Czech Republic). The aim is to implement a wearable computing solution which is capable of supporting the training procedures of Skoda blue collar assembly line workers. The wearable prototype which derived from these field-studies offers semi-autonomous training by mobile and context-sensitive support of trainee personnel.

The two Skoda production facilities involved in the variant production showcase are signed by the fact that the personnel must receive theoretical and practical training using a real vehicle chassis before they are authorized to work at the assembly line.

Compared to stationary computer systems, mobile and wearable computing technology have seriously caught up in performance, functionality and scalability. This makes training solutions based on mobile and wearable computing an attractive consideration for industrial organisations. In this sense, one of the objectives of

J. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2007, LNCS 4553, pp. 663–671, 2007.

WearIT@Work was to supplement the training procedures at Skoda with a context-sensitive wearable computing solution. The idea was that the trainees gain mobile access to the information (e.g. instructions and required tools) to carry out their assembly tasks. In fact, the wearable system was used to recognize the context of performed work, and as a result provide the trainee with the required information to adequately perform individual assembly tasks. Concurrently, the wearable computing system tracks the trainees' activities and analyses them. While the workers perform their training, the supervisor is connected to all active wearable systems via his PC, and can monitor all activities.

The nature of the assembly activity itself made it necessary to design a system that does not restrict workers' freedom of movement, while allowing them to handle all necessary components and tools. It was especially crucial to take into account that workers had to adopt many different postures during the assembly process: crouching, standing, seated, inside and outside of the car. It can be assumed that one of the main advantages of a wearable training solution is that the constant direct presence of the supervisor is no longer required. Thus, the supervisor has the opportunity to observe a number of trainees at the same time via his PC. Since the supervisor has a continuous overview of real-time information such as performed activity, number of mistakes, and number of repetitions, he or she may interact with the trainees in difficult situations. Eventually, an immediate benefit of using wearable solutions in automotive production is that the time of training procedures may be reduced. On the other hand, this only applies when the threshold for getting acquainted with the wearable system is low enough. In order to ensure this, the wearable prototype is designed in line with the real requirements of the user (user-centred design). Additionally, the chosen setup does not introduce much additional effort on the user, because accessing required information is done without or only with minimized explicit interaction with the system. This way, there is no need for the user to be distracted e.g. by a stationary interaction device, such as keyboard or mouse that would impair the way the trainees do their job.

2 Usability Experiments at Vrachlabi

According to the User Centered Design approach the project is following, early field studies with real end-users were designed. This first prototype formed the basis to evaluate the different modalities of interaction with the assembly line workers under real conditions. The main objective was to obtain the users' feedback, regarding their preferences and attitudes towards several hardware, software, user interface and remote support-related-features. The chosen mechanisms for HCI were voice recognition, textile keyboard, non-explicit or task recognition-based interaction, and Head Mounted Displays. Finally, the results obtained had to be useful for future analysis on the impact of wearable technologies in the training process itself.

The front headlight assembly process was selected as a test case. Thus, the Front Light assembly represents a complex enough task which justifies the use of wearable technologies during training. After a detailed analysis, the process was broken down into 14 elementary steps that workers had to follow, using several tools, fixtures, accessories and measuring gauges. The focus of the test case was centered on aspects related to the interaction between the worker and the environment and on task detection. A simple software prototype was created that allowed the navigation through the assembly procedure.

Due to the fact that some of the functionalities were not available for the time the study was carried out, the technique known as Wizard of OZ was applied to simulate those features. During the tests, the workers wore as a wearable device a Xybernaut V computer unit (where the application is installed), a Microoptical SV6 Head Mounted Display, and a conventional headset. Besides, there was also a VNC application running on the Xybernaut. This allowed a remote control and interaction with the wearable application from an external computer by the Wizard. All 14 assembly procedure steps were documented using text, video and photos. The users had the freedom of selecting the information format they found most comfortable, while any kind of combination was allowed, such as text + pictures, text only or video only. All supporting material was structured using XML. Finally, the whole training process was recorded with some users learning in the traditional way, namely without wearable technology. After that, a new group of workers performed the whole training process using the prototype. The test runs were initiated with a short presentation and a pre-questionnaire dealing with ethnographic aspects, work experience and technical skills.

During the voice recognition based interaction, the workers navigated through the application using voice commands. Since no Czech recognition system was available at the time, the wizard interacted remotely with the application according to the workers' commands. Afterwards, the workers used a textile keyboard for interaction as illustrated in Fig. 1. As there was also no 'real' keyboard available, a set of keyboards, adhered to different parts of the workers' body, were constructed out of paper.

Fig. 1. The textile keyboard mockup

The wizard observed the worker and interacted with the application remotely using the VNC client. Finally, non-explicit interaction was tested. Although there was a first prototype for context detection available, it was not suitable for usability testing. Each time the worker finished one action, the wizard pushed the 'next' button remotely in order to present the information corresponding to the next step. In case of error, the wizard simulated the error detection and a message was shown on the screen. The duration time and number of errors during the assembly was recorded. Once finished the test, the workers were asked to complete a usability questionnaire.

The findings of the experiments were manifold. Generally, the usability test with the real end-users was rather complicated, as they had to be picked out from the

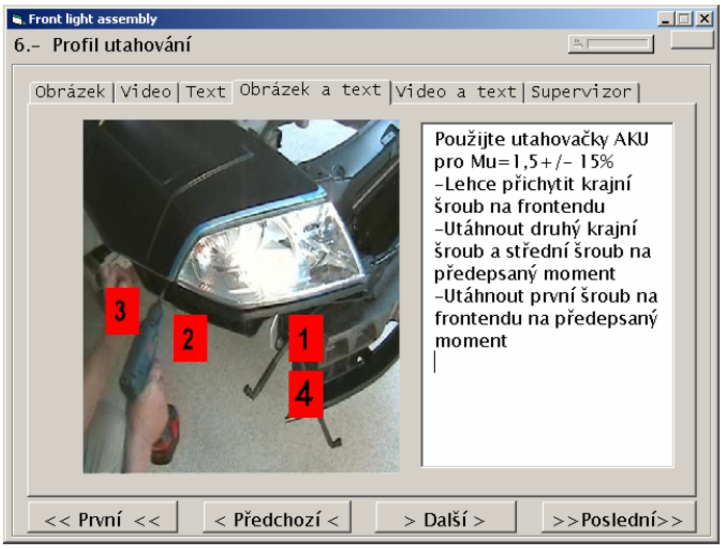


Fig. 2. The application used during the first demonstrator

running assembly line. Therefore, the time schedule for usability testing was very tight, making it impossible to extend the tests when required. On the other hand, it was not possible to recruit the amount of users needed in order to obtain statistically significant results. Dealing with real end-users, it was not feasible to try to apply a user-centred approach through a human translator. In fact, all end-users were unable to understand and speak English, which required the services of a professional Czech/English translator. These constraints prevented the usage of some effective techniques like ‘thinking aloud’, thus, interaction with the workers turned out to be un-natural and difficult. Additionally, it had to be dealt with the fact that the design-/developing team consisted of remotely located partners, what made it very complicated to guarantee a real iterative design process. Apart from these aspects, it was difficult to measure system performance. As a result, it was not easy to quantitatively evaluate the training process. Due to the “learning effects”, it was not legitimate to use the same worker to measure the same training process before and after the introduction of wearable technology. Furthermore, it was difficult to compare two different users because of their different skills and learning capabilities. Nevertheless, the first results of the experiment confirmed that the wearable system was well accepted. Regarding all theoretical concerns about lack of privacy and loss in autonomy, it was a surprise that one of the favourite features of the workers was the ability of the wearable system to monitor the task completion. In fact, when one of the workers made a mistake in the assembling process, the system detected it, and triggered an error message. Later on, during the post-questionnaire, it was one of the most valuable features the worker identified. It has to be underlined as well, that video support was not requested by the workers; on the contrary, they preferred pictures with aggregated information in comparison to simple text.

The main outcome of the Vrachlabi Usability Experiments was the decision to perform new usability experiments in Spain. The purpose was to involve a larger group of end-users located near the research team in order to overcome the constraints experienced at Skoda.

3 Usability Experiments at Tekniker

To overcome the constraints experienced at Skoda (difficulty for the recruitment of users located in the production line, language constraints and geographic distance which doesn't makes economically feasible a long-time research), it was decided to make an analysis based on local workers who were more accessible for the research team.

In order that the experiments were successful, an infrastructure was set-up to carry out assembly tasks. The prerequisites for such an infrastructure were rather simple:

- It should allow creating as many different tasks as were needed,
- All tasks had to be of a similar complexity degree,
- The assembly task should involve the use of manually or tool-assisted piece manipulation

The result was the platform highlighted in Fig. 3.

Fig. 3. The platform used in the experiments

The users had to assemble three different shaped metallic pieces into both sides of a panel where 100 holes of different metrics were drilled. The pieces were able to adopt different special orientations, and could be positioned at different relative distances among them. To check this distance a gauge had to be used.

On the other hand, there was a 16 BNC connector panel, where the user had to connect 3 pairs of wires.

Thanks to the platform it was possible to define multiple assembly tasks for the same user, and to compare the performances of the user when using different interaction modalities.

Two main experiments on usability have been carried out using this platform.

3.1 General Assessment of Wearables for Training

The aim of the first experiment was to extend the initial findings of the experiment made with the workers of Skoda at Vrchlabi: to measure the acceptance of the system, the performance in terms of memorability (how fast workers get trained), and in terms of task completion (time consumed and errors made). All in all 40 workers were recruited and divided into two groups of 20. With the first group of workers the aim was to measure and compare their performance in doing an assembly activity while they accessed explanatory paper-based information, and when this information was accessible through wearable technology. The workers had to perform the complete assembly task as fast as possible and only once.

By means of the second group of workers it was intended to evaluate how wearable technology can contribute to the training process. A prerequisite was that the workers had to learn how to complete the proposed activity. This involved that they had to perform the full process, until they were able to perform the activity without any kind of support. As the "short-term memory factor" was to be measured as well, the workers had to perform the same task one day later, without any kind of support.

In both cases the workers had to perform the experiment twice: once using paper-based support and second using one of the three interaction modalities which were proposed randomly: textile keyboard attached to the sleeve, speech commands and non-explicit or context based interaction.

Besides the strict performance measurement, i.e. required time and number of errors, other factors, such as user acceptance, mental workload and learning style influence were to be evaluated.

Before the actual experiment started, we assumed that the learning style of users might be in relationship with the performance they obtained in the training process. The VARK questionnaire [2] that provided users with a profile of their learning preferences was used. The preferences were divided into four categories: Visual, Aural, Read/Write and Kinaesthetic.

To measure the mental workload that users 'suffer' in doing the assembly task while using wearable technology, the NASA TLX test [3] was chosen. NASA TLX is a subjective tool to assess mental workload on operators working with human-machine systems.

In all cases the workers had to complete a usability questionnaire consisting of both close-end and open-ended questions.

When testing the wearable technologies, workers wore a Microoptical VI head mounted display attached to an OQO where the supporting application was running. As at Skoda, the Wizard of Oz was used to simulate all interaction modalities.

In summary the main findings were:

- Users improved their performance when using the wearable system with implicit interaction: The assembly tasks were performed faster and with less error. In fact, it took 67 seconds less in average than when paper-based information was used which was actually the second best alternative.
- Users did not learn faster using a wearable system. In fact, people were able to learn faster through paper-based support. Although the difference was neglectable when compared to those using context based interaction.

- In the test performed the day after, paper-based learning performed the best, while context-based interaction performed the worst.
- Voice recognition-based interaction was the preferred interaction modality by workers.
- Workers preferred graphical information to text.
- Workers found the system very useful when doing a complex task, allowing hands free access to information, avoiding dispensable movements in order to check information

3.2 Analysis of Different Mechanisms to Access Visual Information

The second experiment was aimed to compare the benefits of using Head Mounted Displays (HMDs) to access information, versus the presentation of the same information on a large screen near the working place.

Fig. 4. One user during the experiment

The 20 workers that took part in the new experiment (see Fig. 4) had to perform 4 new assembly tasks supported by the platform: At the first time, the information explaining the task was presented on a large monitor near the platform; the other three times, the task was explained by presenting the information on a HMD. The intention was to compare the performance and acceptance of three different HMDs, each of them based on different techniques, namely, Carl Zeiss' binocular look-around HMD, Carl Zeiss' see-through HMD and Microoptical VI monocular look-around.

The best performance was obtained when the information was presented on the large display, and the worst when accessing through the binocular HMD. However, when asked about the user's preferences, most of users chose the binocular HMD as the best choice.

4 Conclusions and Further Work

The results of the study have confirmed that effective training of personnel in automotive production is one of the most crucial factors which are responsible for securing production flexibility. The developed wearable computing prototype enables a context-sensitive provision of necessary information to the training personnel. The wearable solution was able to track and analyse the trainee's actions, while providing the end-user with actions for error handling. As a result, semi-autonomous training of trainees in automotive production was realised.

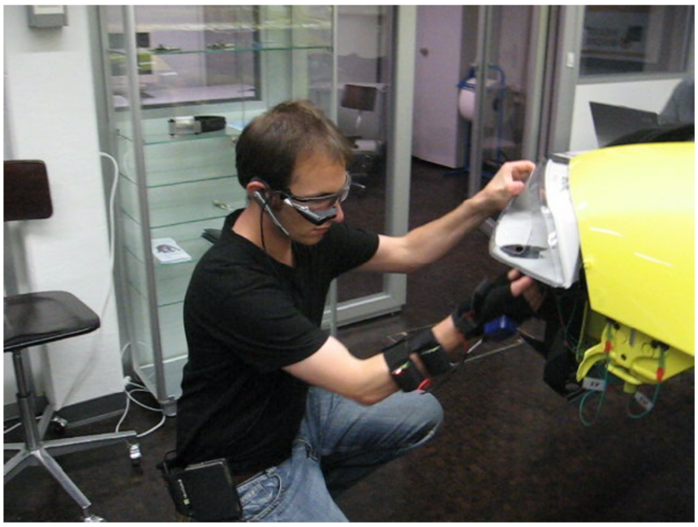


Fig. 5. Assembling the front light supported by the prototype

In the usage of wearable computing solutions for supporting training procedures (see Fig. 5), high benefits can be expected. However, at current stage there is not yet enough experimental data to draw clear conclusions on further benefits and issues of the proposed solution. Nevertheless the consortium will continue to refine the solution according to end-users feedback, and to conduct further tests and field studies within Skoda in order to gather enough knowledge to evaluate the prototype more comprehensively. In this respect there are some interesting features that have to be tested before the final prototype can be deployed in the factory, in 2007: a new wristband where to integrate sensors, collaboration mechanisms between trainer and trainee and different methods for event notification.

Besides the refinement of the final training prototype, WearIt@work envisages a second wearable prototype which will empower blue collar workers in selected stations of the assembly –line, not just for training but for real work activities.

Regarding the application of the User Centered Design approach, it came out that within an international integrated project, guaranteeing this kind of approach is not at all an easy task. In fact, the cultural and geographical distance makes it quite difficult to apply an orthodox approach. In WearIT@work, the production pilot team is following their own methodology: namely an initial requirement elicitation process with real end-users, usability tests with local users close to the research team, and final validation with final end-users. It must however be mentioned that it is nevertheless very difficult to get hold of the real end-users in a global company where production is highly dependent upon human resources.

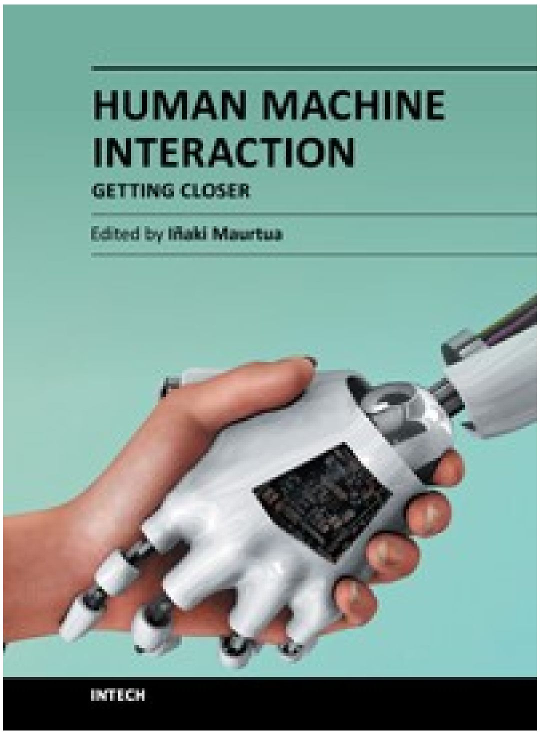
Acknowledgments. This work was partly funded under grant 004216 by the European integrated project "WearIT@work - Empowering the mobile worker by wearable computing". We thank our colleagues Pierre T. Kirisci (Bremen Institute for Industrial Technology and Applied Work Science at the University of Bremen), Thomas Stiefmeier (ETH Zürich, Institut für Elektronik), Marco Luca Sbodio (Hewlett-Packard, Italy Innovation Center) and Hendrik Witt (University of Bremen, TZI Wearable Computing Lab), for their collaboration in the implementation of the first prototypes and in writing this paper.

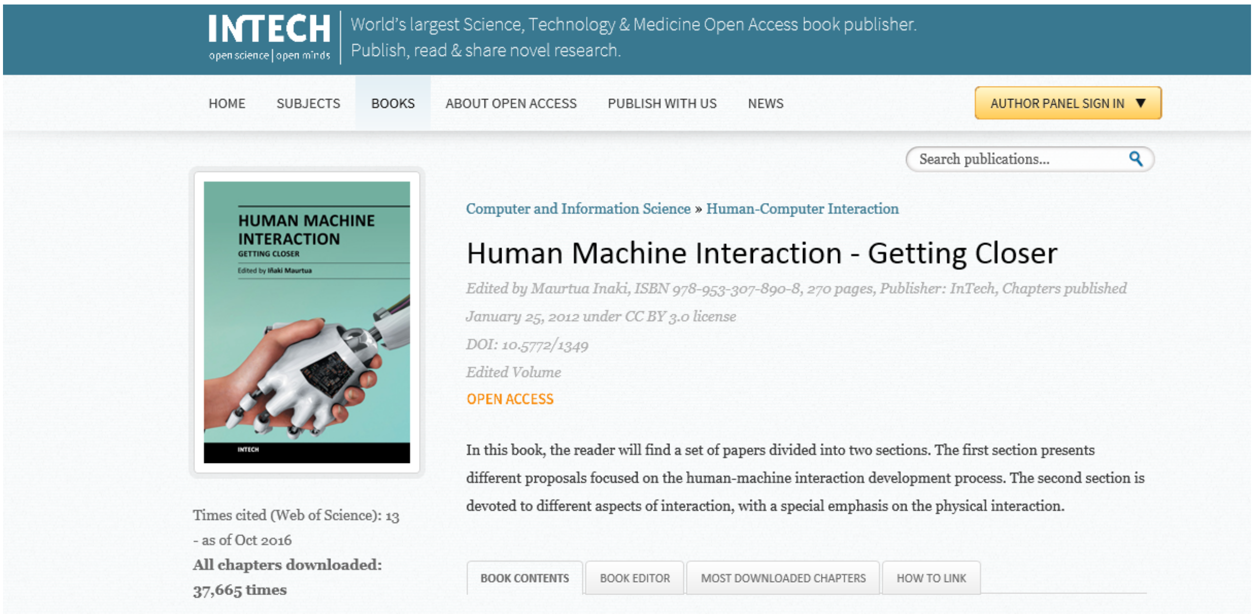
References

1. Siewiorek, D.P., Finger, S., Terk, M., Subrahmanian, E., Kasabach, C., Prinz, F., Smailagic, A., Stivoric, J., Weiss, L.: Rapid Design and Manufacture of Wearable Computers, COMMUNICATIONS OF THE ACM 39(2), 63–70 (1996)
2. Fleming, N.D.: Not Another Inventory, Rather a Catalyst for Reflection.To improve the Academy (1992)
3. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.)Human Mental Workload, North Holland B.V., Amsterdam (1988)

7.7. Human Machine Interaction - Getting Closer

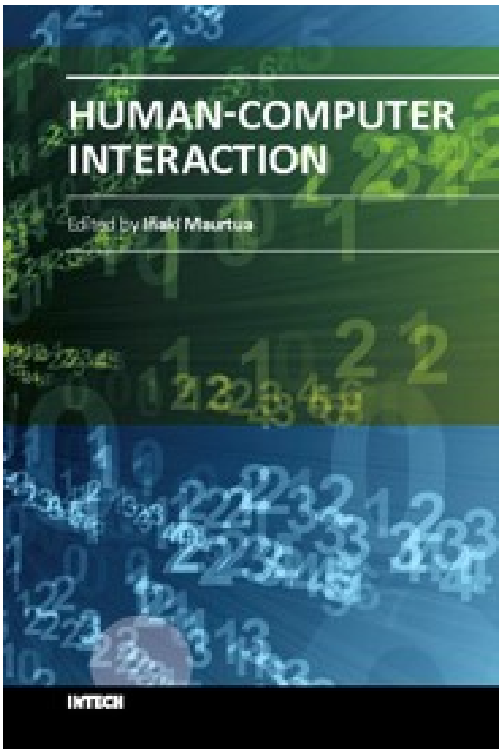
Human Machine Interaction - Getting Closer. Edited by Maurtua Inaki, ISBN 978-953-307-890-8, 270 pages, Publisher: InTech, Chapters published January 25, 2012 [107]


HUMAN MACHINE
INTERACTION
GETTING CLOSER
Edited by Iñaki Maurtua
INTECH


INTECH
open science | open minds
World's largest Science, Technology & Medicine Open Access book publisher.
Publish, read & share novel research.
HOME
SUBJECTS
BOOKS
ABOUT OPEN ACCESS
PUBLISH WITH US
NEWS
AUTHOR PANEL SIGN IN
Search publications...
Computer and Information Science » Human-Computer Interaction
Human Machine Interaction - Getting Closer
Edited by Maurtua Inaki, ISBN 978-953-307-890-8, 270 pages, Publisher: InTech, Chapters published January 25, 2012 under CC BY 3.0 license
DOI: 10.5772/1349
Edited Volume
OPEN ACCESS
In this book, the reader will find a set of papers divided into two sections. The first section presents different proposals focused on the human-machine interaction development process. The second section is devoted to different aspects of interaction, with a special emphasis on the physical interaction.
Times cited (Web of Science): 13
- as of Oct 2016
All chapters downloaded:
37,665 times
BOOK CONTENTS
BOOK EDITOR
MOST DOWNLOADED CHAPTERS
HOW TO LINK

7.8. Human-Computer Interaction

Human-Computer Interaction, Edited by Inaki Maurtua, ISBN 978-953-307-022-3, 570 pages, Publisher: InTech, Chapters published December 01, 2009 [108]


HUMAN-COMPUTER
INTERACTION
Edited by Iñaki Maurtua
INTECH


INTECH
open science | open minds
World's largest Science, Technology & Medicine Open Access book publisher.
Publish, read & share novel research.
HOME
SUBJECTS
BOOKS
ABOUT OPEN ACCESS
PUBLISH WITH US
NEWS
AUTHOR PANEL SIGN IN
Search publications...
Computer and Information Science » Human-Computer Interaction
Human-Computer Interaction
Edited by Inaki Maurtua, ISBN 978-953-307-022-3, 570 pages, Publisher: InTech, Chapters published December 01, 2009 under CC BY-NC-SA 3.0 license
Edited Volume
OPEN ACCESS
In this book the reader will find a collection of 31 papers presenting different facets of Human Computer Interaction, the result of research projects and experiments as well as new approaches to design user interfaces. The book is organized according to the following main topics in a sequential order: new interaction paradigms, multimodality, usability studies on several interaction mechanisms, human factors, universal design and development methodologies and tools.
Times cited (Web of Science): 5
- as of Oct 2016
All chapters downloaded:
108,394 times
BOOK CONTENTS
MOST DOWNLOADED CHAPTERS
HOW TO LINK

7.9. Studying Human-Robot Collaboration in an Artistic Creative Processes

Johan Kildal, Koray Tahiroglu, Juan Carlos Vasquez, Iñaki Maurtua. Studying Human-Robot Collaboration in an Artistic Creative Processes. 2017 Conference on Human-Robot Interaction (HRI2017), ReHRI'17 – International Workshop on reproducible HRI experiments: scientific endeavors, benchmarking and standardization. 2017, Viena [109]

Studying Human-Robot Collaboration in an Artistic Creative Process

Extended Abstract

J. Kildal
IK4-TEKNIKER
Iñaki Goenaga 5, Eibar 20600
Spain
johan.kildal@tekniker.es

K. Tahiroğlu
Department of Media, Aalto University
School of ARTS
Espoo, Finland
koray.tahiroglu@aalto.fi

J.C. Vásquez
Department of Media, Aalto University
School of ARTS
Espoo, Finland
juan.vasquezgomez@aalto.fi

I. Maurtua
IK4-TEKNIKER
Iñaki Goenaga 5, Eibar 20600
Spain
inaki.maurtua@tekniker.es

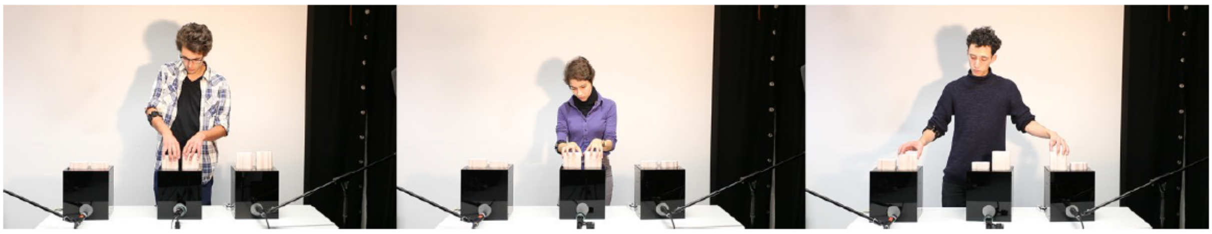


Figure 1: Musicians and network of robotic intelligent sonic agents creating music collaboratively

ABSTRACT

The design[1] and evaluation of a human-robot collaborative system requires following user centered methodologies and processes for its success. A cornerstone of this approach is to collect requirements from three main sources: technological factors, human factors and application-domain factors. In this paper, we discuss the relevance that the application domain has in the relative proportion of quantitative and qualitative metrics, for an adequate full evaluation. We discuss the particular application domain of collaborative artistic creation, with particular detail on a use case of human-robot collaborative music creation through improvisation. We highlight the higher proportion of qualitative metrics that is needed to evaluate the collaboration with a robotic system in such highly-subjective creative scenarios.

[1]It is a datatype.

WOODSTOCK'97, July 2016, El Paso, Texas USA

DOI: 10.1145/123_4

CCS CONCEPTS

• **Computing methodologies → Embedded and Cyber-Physical Systems; Human-Centered Computing → Human-Computer Interaction (HCI) → HCI design and evaluation methods; Applied Computing Robotics → Arts and Humanities → Performing Arts**

KEYWORDS

Human-robot collaboration, user studies, interaction, arts, creative process, autonomous intelligent agents, new interfaces for musical expression

ACM Reference format:

J. Kildal, K. Tahiroğlu, J.C. Vásquez, I. Maurtua. 2017. Human-Robot Collaboration in an Artistic Creative Process. In *Proceedings of ReHRI'17 – International Workshop on reproducible HRI experiments: scientific endeavors, benchmarking and standardization. 12th ACM Annual Conference on Human-Robot Interaction. Vienna, Austria, March 2017 (HRI'17)*, x pages.
DOI: 10.1145/123 4

1 INTRODUCTION

It is clear that collaborative scenarios between a robot and a person have to be designed, researched and developed following user-centered methodologies, with the aim to best realize envisioned scenarios and meet requirements [1, 2]. The primary subject of design in such interactive scenarios is the technological component (a robotic system or an autonomous intelligent agent) as well as the mechanisms through which interactions between robot and human take place. As the product of human design, they are also subject for iterative re-design and re-implementation, in the process towards meeting the goals envisioned for the collaboration. The restrictions and requirements that technology imposes on the design process relate to the state of the art in the technology that is available, its affordability and versatility, as well as the feasibility and complexity of implementing specific interaction techniques.

In sharp contrast, the non-technological partner in the interaction (i.e., the human actor) cannot be redesigned or otherwise altered easily in any substantial way. At most, humans engaged in collaborating with robotic systems can be required to acquire new skills or accept new paradigms, but they cannot be subjected to a design process. Instead, human factors that are relevant for the interaction with a robot need to be studied and understood, and these factors will be necessary additional input for the requirements that the interactive collaboration with the robotic system has to fulfill.

There is still a third main source of requirements for the design of collaborative human-robot system: the application domain in which the collaboration is framed, which encompasses additional requirements from stakeholders and from the environment. The application domain can modulate significantly the set of requirements that were introduced by the analyses of both technological and human factors. Very importantly, application domain factors also determine which metrics will be most relevant from the perspective of the experimental research of the interactive system.

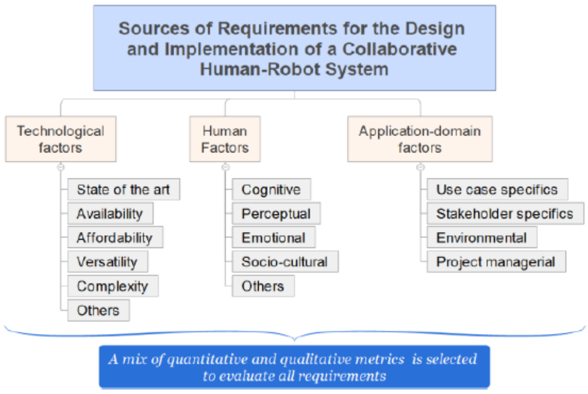


Figure 2: Sources of requirements for the design and implementation of a typical human-robot collaborative system.

The next sections in this paper will focus on discussing this last aspect: the selection of experimental research metrics based on the application domain. We will illustrate our points with a case study from research that we have conducted in a very specific application domain: collaborative musical creation between an expert musician and a robotic autonomous and intelligent instrument.

2 APPLICATION-DEPENDENT METRICS

When designing and developing a human-robot collaborative solution through user-centered methods, a set of various relevant metrics is generally needed in order to be able to evaluate all the important aspects about the system (well beyond the regular metrics used for the evaluation of robot developments alone [3]), with respect to the requirements that have been defined for it. The application domain has a deciding influence on the selection of such set of metrics.

As a general rule, it is a good practice to collect a mix of quantitative and qualitative data through experimental user studies at every stage of the development. In that way, it is easier to form a complete picture of the degree to which a solution adjusts to requirements from different origin (technology factors, human factors and application area). For specific aspects of the system, in our experience it is also a good strategy that the various selected metrics show a degree of redundancy in the data that is collected with them. For instance, if *time-to-complete-task* is considered to be important in a particular scenario, this can be measured both objectively (a quantitative measure) and subjectively (a qualitative measure). Quantitative data would usually be the absolute time actually elapsed till completion of the task, which would help evaluate the compliance with the efficiency requirements established for a particular application domain. Complementing this, a qualitative measure of the time to complete task could also be considered. For instance, the subjective perception that the person in the interaction loop has of the time elapsed until completion of the task. This second measure can help evaluate the user experience (UX) that the collaboration is capable of offering to the human collaborator. Considering both measures jointly, researchers could provide a nuanced and multi-perspective answer to the question, *is time to complete task fast enough in this collaboration*?

2.1 A Productive Application Domain

As already stated, forming a meaningful mix of quantitative and qualitative metrics depends heavily on the application domain. In most collaborative scenarios, robots are intended to perform sub-tasks with a degree of autonomy, while also maintaining coordination with the human they are collaborating with for the successful joint completion of an overall task. In such cases (for instance, in the collaborative assembly of flat-pack furniture), both the robot and the person can share the same detailed description of the work to be carried out. As a result, quantitative metrics tend to have a predominance in studies that evaluate the robot's performance and the collaboration as a whole. Metrics such as correctness, productivity, efficiency, effectiveness, cost-efficiency, waist (of time and material), resulting quality and similar others are commonplace. Beyond the objective and pragmatic measures just listed, qualitative measures are also necessary to account for factors that derive from the direct collaboration between person and robot, and the UX that can result from it. Important metrics that account for an overall UX can include perceived safety, mental workload, quality of rapport with their technological collaborator, and overall satisfaction with the resulting work done or service provided, to name some.

2.2 An Artistic Creative Application Domain

To illustrate the dependency of evaluation metrics on the application domain, it is useful to compare the productivity-oriented application domain example just outlined with another one in which the intended result of the collaboration cannot be known in detail by the robot (or even by the human actor in most cases, for that matter) until its completion: a scenario of collaborative artistic creation. The subjective nature of the artist's vision and creative agenda is a main defining feature of a collaborative artistic creation process. In such a context, the robotic collaborator may be aware of a general frame of collaboration previously agreed with the artist (e.g., the creative material to be used and, to some degree, the structure of the resulting work), but a detailed description of the steps to be taken and of the exact final result intended cannot exist beforehand, by the very nature of the artistic creative process.

This main difference between productive and artistic creative collaborative scenarios will be reflected in the mix of metrics that should be selected for evaluation. It is still true that there should be a mix of both quantitative and qualitative data sources for a complete analysis. However, there is not an objective "right or wrong" criterion that can be described easily in a mathematical way, and correctness of the final result will depend largely on purely artistic criteria. Having said that, the robot will still have clearly-defined tasks that have to be evaluated through quantification (against an *objective correctness* reference). Therefore, while both quantitative and qualitative data are still necessary for evaluation, it is likely that the mix of metrics shifts towards qualitative in this kind of application domain.

3 CASE STUDY: HUMAN-ROBOT COLLABORATIVE MUSIC CREATION

To illustrate our discussion in the previous section, we offer a brief outline of our NOISA project, as a case study of a collaborative robotic system designed and developed for the application domain of collaborative improvised music performance [4–6]. NOISA (Network of Intelligent Sonic Agents) consists of a set of three networked robotic autonomous agents, each of which is a musical instrument (see Figure 3 for an external and internal view of one of the agents).

3.1 The Rules of the Collaboration

The objective sought by the collaboration between the human performer and its instrument is to amplify the human musician´s creativity and to maximize her capacity to complete a musically-satisfactory improvised composition.

As a musical instrument, each NOISA agent provides an interface in the form of two handles that can be slid up and down by hand, producing synthesized sound that the musician can model and control with a great level of precision (Figure 1 shows stills of three musicians performing on NOISA.)

From the perspective of the musician, the three instruments (each with its signature sound space) provide a rich, expressive and non-trivial instrument that allows the performer to develop mastery over time, through practice and experimentation.

As robotic autonomous agents, each agent can decide to actively move its own motorized handles, displacing them physically in space, and producing its own sonic output (self-performing capacity). However, and crucially, the network of autonomous agents does not take the initiative in the performance. Instead, it stands respectfully and discretely in the background, acting as a mere passive musical instrument in the hands of its performer. From that position, it observes the musician perform and it observes itself being played. With the data collected during the performance, it learns from the musician's musical discourse, identifying the main motives used by the artist and the artist´s discourse as it unfolds.

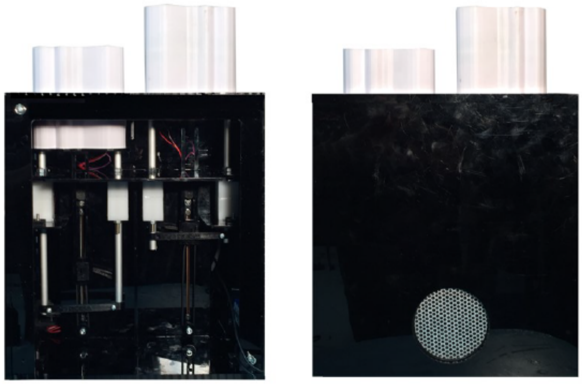


Figure 3: One of the robotic intelligent and autonomous agents in the NOISA networked instrument.

A further key functionality is enabled by sensor data obtained from multiple sensors on the instrument itself (position sensors, touch sensors on the handles), on the musician (EMG sensors on the forearm) and in the environment (external RGB and depth cameras analyzing the musician's movements and face expressions). The system is able to combine all the sensor data with the learning from the musician's performative discourse, and estimate the *level of engagement* that the musician has with the process of improvisation at each moment during the performance of a piece. With this information, when NOISA detects a drop in the level of engagement of the musician with the performance activity, the robotic instrument kicks in as a collaborator. It carefully executes physical (and consequently also musical) actions that are based on the motives and discourse learned from the artist through observation. These interventions are intended as cues that the collaborating system offers the musician to regain engagement with the performance and help maintain her musical discourse. When NOISA observes that the level of engagement has recovered, the system retreats again to continue observing and learning from the behavior of passive instrument.

3.2 Evaluation Metrics

Like with the design and development of any human-robot collaborative system, the evaluation of NOISA required a set of metrics that included both quantitative and qualitative ones. The themes for requirements derived from technological factors, as outlined in the introduction (see also Figure 2) were still relevant for the construction of the hardware system, the musical instrument and the programming of the intelligent autonomous agents. Similarly, regular human factors requirements applied, in particular for the design of the interaction. When compared with a productivity-centric collaborative system, the main differences in the metrics employed arose from the application domain. Some of the classic production and productivity-related metrics still applied, when considering the musical production as *the product*. However, some of the quantitative metrics have less relevance of become fussier to apply. For instance, productivity and efficiency of production were less relevant, as a performance session could be seen like a single unit production batch, where trial and error was not a plausible approach. Instead, quality of the outcome was still important, but the way of assessing it had a marked qualitative weight, as quality ratings would depend neatly on subjective artistic criteria. Evaluating system performance also showed interesting challenges, in particular when evaluating the engagement estimation engine. Such engine was developed to estimate an evolving curve of engagement as accurately and faithfully as possible (a quantification of engagement curves that could serve for decision making for the system). However, the assessment of such AI engine relied on the self-assessed levels of engagement provided by musicians over many performances. These master references were, in turn, highly subjective.

4 CONCLUSIONS

Like with any interactive technology, a cornerstone of the user centered processes for the design and development of collaborative human-robot systems is the comprehensive capture of relevant requirements. They are the ultimate checklist for the evaluation of a system, which should satisfy requirements to the largest possible extent. This is observed by selecting an adequate mix of metrics for the data collection in experimental studies.

The mix of metrics usually compiles both quantitative and qualitative data, necessary to respond to the varied nature of requirements that is always found in one such hybrid human and robotic environment.

We distinguish three broad areas as the main sources for the requirements to be followed for design and implementation: technological factors, human factors and application-domain factors. Of these three sources, in this paper we have stated that the application-domain factors have the largest influence on the relative proportion of quantitative and qualitative evaluation metrics in an optimal experimental design. We have illustrated this by outlining a typical productive human-robot collaboration scenario (the collaborative assembly of the parts of an object) and another one in which robot and person collaborate in developing artistic creative activity. Through a use case based on our NOISA research project, we have shown that, in an artistic creative domain, the proportion of relevant qualitative metrics needed to evaluate as system is necessarily higher than the proportion of quantitative metrics.

REFERENCES

[1] Karen Holtzblatt, Jessamyn B. Wendell, Shelley Wood. 2005. Rapid contextual design: a how-to guide to key techniques for user-centered design. *Elsevier*.
[2] Julie A. Adams. 2005. Human-robot interaction design: Understanding user needs and requirements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications. 447-451.
[3] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz and Michael Goodrich. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, Salt Lake City, Utah. 33-40.
[4] Koray Tahiroğlu,, Thomas Svedström, Valtteri Wikström. 2015. NOISA: A Novel Intelligent System Facilitating Smart Interaction. In *CHI EA '15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. Seoul. 279–282.
[5] Koray Tahiroğlu, Thomas Svedström, Valtteri Wikström. 2015. Musical Engagement that is Predicated on Intentional Activity of the Performer with NOISA Instruments. In *Proceedings of the international conference on New Interfaces for Musical Expression* (NIME '15). 132–135.
[6] Koray Tahiroğlu, Juan C. Vasquez, Johan Kildal. 2016. Non-intrusive Counter-actions: Maintaining Progressively Engaging Interactions for Music Performance. In *Proceedings of the international conference on New Interfaces for Musical Expression* (NIME '16). 444-449.

7.10. Enhancing safe human-robot collaboration through natural multimodal communication

Iñaki Maurtua, Izaskun Fernández, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibarguren: Enhancing safe human-robot collaboration through natural multimodal communication. ETFA 2016: 1-8 [110]

Enhancing Safe Human-Robot Collaboration through Natural Multimodal Communication

Iñaki Maurtua, Izaskun Fernandez, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibarguren
Email: {inaki.maurtua, izaskun.fernandez, johan.kildal, loreto.susperregi, alberto.tellaeche, aitor.ibarguren}@tekniker.es
IK4-TEKNIKER
Parke Teknologikoa
C/ Iñaki Goenaga, 5 - 20600 Eibar
Basque Country, Spain

***Abstract*—This paper presents a semantic multimodal interaction approach between humans and industrial robots to enhance the dependability of the collaboration in real industrial settings. Although this generic approach can be applied in different industrial scenarios, this paper explains in detail how it is implemented in a real case to enhance the accuracy of requests interpretation in order to achieve a more efficient, easy to scale and maintain collaboration between humans and robots.**

***Index Terms*—Safe Human-Robot Collaboration, Collaborative robots, Multimodal Interaction, Natural Communication, Semantic Web Technologies, Reasoning**

I. Introduction

In modern industrial robotics, the safe and flexible co-operation between robots and human operators can be a new way to achieve better productivity when performing complex activities. Introducing robots within real industrial settings makes the interaction between humans and robots gain further relevance. The problem of robots performing tasks in collaboration with humans poses three main challenges: robots must be able to perform tasks in complex, unstructured environments, and at the same time they must be able to interact naturally with the humans they are collaborating with, always guaranteeing the safety of the worker.

Let us imagine an industrial collaborative robot and an operator in a deburring collaborative process. While the robot is deburring a piece, the operator has started with another one. The robot finishes the piece and, as the end of the workday is approaching, the operator decides to ask the robot for finishing the piece he had started, while he finishes other tasks. There are several ways to make this request:

- Moving the robot manually to the initial position of the deburring operation.
- Pointing at the area of the piece where the deburring has to be done.
- Voice request identifying the task and the target deburring area.
- Combining voice and gesture. The first for indicating the action (i.e *Remove the burrs from this area*) and the second for defining the area

978-1-5090-1314-2/16/$31.00 © 2016 IEEE

These communication possibilities are developed in the context of the project *H2020 FourByThree* [1], which aims at creating a new generation of modular industrial robotic solutions that are suitable for efficient task execution in safe collaboration with humans and are easy to use and program. The project will allow system integrators and end-users to develop the custom robot that best answers to their needs. To achieve this, the project will provide a set of hardware and software components, ranging from low level control to interaction modules as it is shown in Figure 1. The results will be validated in 4 industrial settings: Investment Casting, Aeronautical sector, Machining and Metallic Part Manufacturing, in which industrially relevant applications will be implemented: assembly, deburring, welding, riveting and machine tending.

The present work describes the natural communication approach within the *FourByThree* European project. A requirement for natural human-robot collaboration is to endow the robot with the capability to capture, process and understand accurately and robustly requests from a person. Thus, a primary goal for this research is to analyze the natural ways in which a person can interact and communicate with a robot and go towards a natural, robust and reliable communication framework.

The most relevant channels for natural communication between humans and robots are voice and gestures. In this multimodal scenario, the information coming from the different channels can be complementary or redundant. Redundancy can be beneficial [2] in real industrial scenarios in which noise and variable lighting conditions represent a challenge.

In this paper, we present a semantic approach that supports multimodal interaction between humans and industrial robots in real industrial settings. This generic approach can be applied to different industrial scenarios by modifying the information about the environment in which the communication takes place, as described below.

II. Case Study

As explained before, FourByThree project includes four industrial scenarios of human-robot collaboration. For an initial validation of the semantic multimodal interpreter, it has been selected the application scenario of Investment Casting lead by the Spanish ALFA company that includes two collaborative

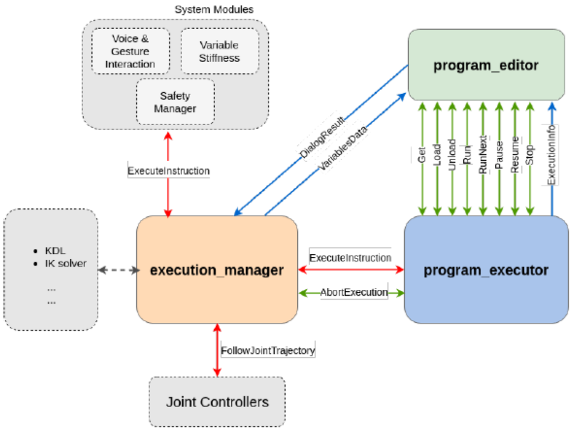


Fig. 1. FourByThree project architecture

tasks: die assembly (it includes screwing and unscrewing operations) and deburring of wax pieces. Due to the characteristics of the actors involved, the solution is focused on Spanish language.

In the case of the assembly task, the human and the robot work independently (un)screwing bolts on different parts of the die. In the deburring activity, the human and the robot perform sequential tasks on the same workpiece in a synchronized manner: meanwhile the person glues and removes difficult to access burrs, the robot deburrs the rest of them.

III. Related work

Over the last two decades, a considerable number of robotic systems including Human-Robot Interaction (HRI) capabilities have been developed [3], [4]. Although recent robot platforms integrate advanced human-robot interfaces (incorporating body language, gestures, facial expressions and speech) [5], [6] their capabilities to understand human speech semantically remain quite limited. Endowing a robot with semantic understanding capabilities is a very challenging task. Previous experiences with tour-guide robots [7], [8] show the importance of improving human-robot interaction in order to ease the acceptance of robots by visitors. In Jinny's HRI system [8], voice input is converted to text strings, which are decomposed into several keyword patterns and a specialized algorithm finds the most probable response for that input. For example, two questions like 'Where is the toilet?' and 'Where can I find the toilet' are interpreted in the same way, since the keyword pattern of 'where' and 'toilet' are extracted from both cases.

Human-robot natural interaction has also been tackled in industrial scenarios. For instance, Bannat et al. [2] introduced an interaction that consisted of different input channels such as gaze, soft-buttons and voice in an industrial scenario. Although voice constituted the main interaction channel in that use scenario, it was solved by command-word-based recognition.

SHRDLU is an early example of a system that was able to process instructions in natural language and perform manipulations in a virtual environment [9]. Later on, researchers extended SHRDLU's capabilities in real world environments. Those efforts branched out into tackling various sub-problems, including Natural Language Processing (NLP) and Robotics Systems. Notably, MAc Mahon et al. [10] and Kollar et al. [11] developed methods to follow route instructions given through natural language. Tenorth et al. [12] developed robotic systems capable of inferring and acting upon implicit commands using knowledge databases. A similar knowledge representation was proposed by Wang and Chen [13] using semantic representation standards such as the W3C Web Ontology Language (OWL) to describe an indoor environment.

A generic and extensible architecture is described in [14]. The case study presented there included gesture and voice recognition, and the evaluation showed that interaction accuracy increased when combining both inputs (91%) instead of using them individually (56% in the case of gestures and 83% for voice). Furthermore, the average time for processing both channels was similar to the time needed for speech processing.

Our work is based on this extensible architecture, combining gesture and speech channels and adding semantic aspects to their processing.

IV. Multimodal Interaction Semantic Approach

The approach proposed in this work aims at creating a safe human-robot collaborative environment in which interactions between both actors happen in a natural way, i.e. communication based on voice and gestures. Some examples of this natural interaction could be: human operator asking the robot for a certain task; robot asking for clarification when a request is not clear; or in some specific scenarios in which human intervention is necessary during an automatized robot task execution, the robot can ask the operator for accomplishing some task and once completed, the operator informs the robot to resume its task. This natural communication facilitates the coordination between both actors, enhancing the safe collaboration between robot and humans.

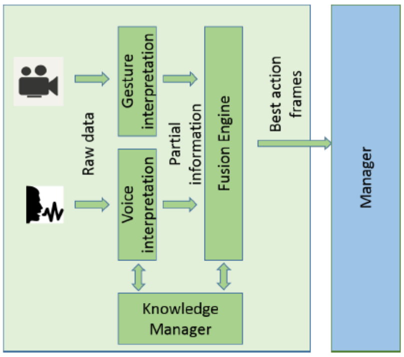


Fig. 2. Multimodal semantic approach architecture

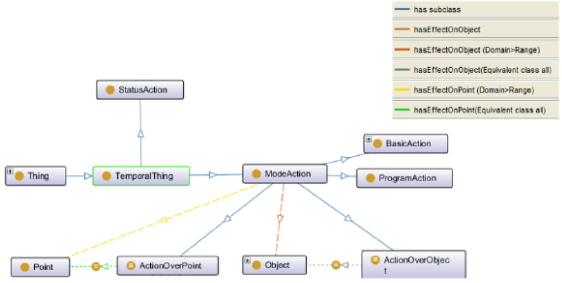


Fig. 3. An excerpt of the Knowledge Manager Ontology

To address such a natural communication, we propose a semantic multimodal interpreter that is able to handle voice and gesture-based natural requests from a person, and combine both inputs to generate an understandable and reliable command for industrial robots, facilitating a safe collaboration. For such a semantic interpretation, we have developed four main modules, as it is shown in Fig. 2: a *Knowledge-Manager* module that describes and manages the environment and the actions that are affordable for robots, using semantic representation technologies; a *Voice Interpreter* module that, given a voice request, extracts the key elements on the text and translates them into a robot-understandable representation, combining NLP and semantic technologies; a *Gesture Interpretation* module to resolve pointing gestures and some simple orders like stopping an activity (out of the scope of the work presented in this paper); and a *Fusion Engine* for combining both mechanisms and constructing a complete and reliable robot commanding mechanism.

These modules are described in detail in the following subsections.

A. *Knowledge Manager*

The knowledge manager comprises ontologies that model the environment, including the robot capabilities as concepts. In addition to the concepts, ontologies allow to model the relationships between concepts. These relationships are implicit rules that can be exploited by reasoners in order to infer new information from the ontology. As a result, reasoners can work as rule engines in which human knowledge can be represented as rules or relations.

Ontologies are reusable and flexible at adapting to dynamic changes, thus avoiding to have to re-compile the application and its logic whenever a change is needed. Being in the cloud makes ontologies even more reusable, since different robots can exploit them, as it was demonstrated in RoboEarth [15] project.

Through ontologies, we model the industrial scenarios in which industrial robots collaborate with humans, in terms of robot behaviours, task/programs they can accomplish and the objects they can manipulate/handle. We distinguish two main kinds of actions: actions that imply a change on the status of the robot operation, e.g., *start* or *stop*, and actions involving the robot capabilities, e.g., s *screw*, *carry* or *deburring*. This is shown in Fig 3.

Using OWL equivalentClass built-in property, *ActionOverPoint* and *ActionOverObject* classes have been also defined. Both are *ModeAction* sub-classes, with a particular restriction: for the first, the action must be related to a point (defined by *hasEffectOnPoint min 1 Point* restriction in the ontology); and for the latter, the action must be related to an object (defined by *hasEffectOnObject min 1 Object* restriction in the ontology). This way, applying a semantic reasoner able to interpret this OWL statements, it will infer that a *ModeAction* instance like *deburring* with *hasEffectOnObject burr* property also belongs to the *ActionOverObject* class. Same inference effect for *ModeAction* instances with *hasEffectOnPoint* property defined, that will be also considered instances of *ActionOverPoint*.

For each individual action or object, we include *tag* data property for listing the most common expression(s) used in natural language to refer to them, including reference to the language used. An automatic semantic extension of those tags exploiting Spanish WordNet [16] is done at initialization time. In this way, we obtain different candidate terms referring to a certain concept.

Besides task/programs and objects, the ontology also includes relations between the concepts, as it is shown in Fig 4. These relations are used by the interpreter for disambiguation at runtime. This ability is very useful for text interpretation because sometimes the same expression can be used to refer to different actions. For instance, people can use the expression *remove* to request the robot to *remove a burr*, but also to *remove a screw*, depending on whether the desired action is *deburring* or *unscrewing* respectively. If the relationships between the actions and the objects over which the actions are performed are known, the text interpretation is more accurate; it will be possible to discern in each case to which of both options the expression *remove* corresponds. Without this kind of knowledge representation, this disambiguation problem is more difficult to solve. These relations are formally modelled in the ontology as it is shown in Fig 4

For the current implementation, the two contexts described in the Case Study section have been considered. The possible

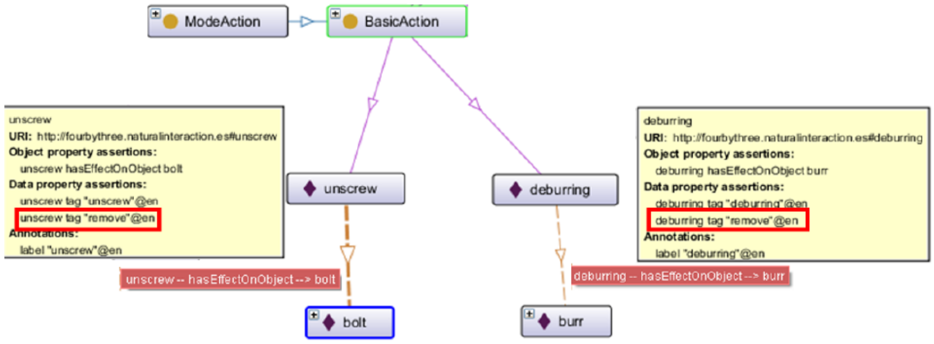


Fig. 4. An excerpt of the FourByThree Semantic Interpreter Knowledge Base

tasks the robot can fulfill in both scenarios have been identified and a knowledge base (KB) created, populating knowledge manager ontology with instances representing those tasks. The knowledge base also includes the elements that take part in both processes as instances of *Object* and *Point* classes, as well as the relationships they have with respect to the tasks. This knowledge base is published in StarDog 4.0.5 Community version [17] and extended with WordNet as explained before.

B. *Voice Interpreter*

Given as input a human request like *Remove the burrs from there* in which a person indicates the desired action via voice, the purpose of this module is to understand exactly what the person wants and if it is feasible to generate the necessary information for the robot. For instance, for the just mentioned example the voice interpreter should interpret that the verb *remove* corresponds to the deburring action and check if it is a feasible action for the current collaborative robot. For such an interpretation, the module follows three main steps: speech recognition step, dealing with voice to text transcription; a rule-based step for key elements extraction from the transcribed text; and a final matching between the key elements and the feasible tasks for the robot, based on the KB.

The speech recognition step is based on Google Speech API [18]. A recorded audio file including the request of the operator is sent to Google Speech API, which returns the corresponding text.

For the second step, which aims to extract the key elements in the transcribed text of the previous step, Natural Language Processing techniques are used. The main idea is to use syntactic information by means of rules for key elements extraction. In the current implementation, the Spanish version of FreeLing, an open source suite of language analysis tools [19], has been used for this structural analysis.

For rules definition, we have applied Freeling for morphosyntactic analysis and dependency parsing to a set of request examples obtained from different people. We have revised the complete information manually and there have been identified the most frequent morphosyntactic patterns that are relevant for extracting the key elements: elements denoting actions, objects/destinations (target onward) and explicit expressions denoting gestures, such as *there* and *that*. And finally, we have implemented those patterns as rules. In this manner, in execution time, given a Spanish FreeLing-tagged sentence it is able to extract the key elements on it. In the future this process can be used to extend the interpreter to other languages.

The five rules that have been defined for action extraction can be grouped in two classes:

- rules defining sequences of words tagged as verb groups, e.g. *Quita...(Remove...)* or *Sigue atornillando...(Continue screwing...)*
- rules to extract noun phrases parsed as direct object or prepositional objects like *Empieza con el atornillado (Start screwing)* or *Empieza el desbardado (Start deburring)*.

For the objects/destinations and elements denoting gesture extraction, we have defined a set of 10 rules. Those rules model different noun phrases and adverb phrases of different lengths. If a word sequence matches two different rules, the extraction will assign the longest one.

Once the key elements are extracted, it is necessary to identify which one of the tasks that the robot is able to perform suits the request best. We undertake this last step by making use of the KB information described above. First, we verify if the identified actions are among the feasible tasks described in the KB, accessing to the tag data property of the actions in the KB by the semantic query language, SPARQL [20]. Then, we apply a disambiguation step using the target information, as explained before (taking advantage of the OWL logic inference capabilities).

For instance, for the Spanish request *Quita las rebabas* (Remove the burrs) it will check, via Sparql query, which actions in the KB contain *tag* data property with *quitar (remove)* value, obtaining *unscrew* and *deburring* as potential actions. During the key elements extraction step *burrs* element is recognized as target element. We exploit this information going back again to the KB via Sparql and checking if any of the two candidate tasks is related with it, getting that *deburring* is directly related with it. So the module discards *unscrew* as potential action, leaving *deburring* as unique potential task.

There are other situations in which the disambiguation capability is exploited. For instance, let's suppose these two similar but different requests: *Acércate* (*come here*) and *Acerca la caja* (bring the box closer). In both sentences the element to look for among *tag* data properties of the tasks represented in the KB is *acercar*, which matches with *go* and *bring* tasks. As it is shown in Fig 5, both belong to the *ActionOverPoint* class, while the latter also belongs to *ActionOverObject*. Taking advantage of this information, and although the box is not being explicitly represented in the KB, as the previous step has identified it as target element, the interpreter is able to infer that *bring* task fits better with this request than *go*.

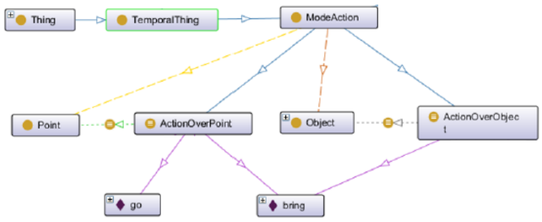


Fig. 5. An excerpt of the FourByThree Semantic Interpreter Knowledge Base

Moreover, when the verb to match with the tasks in the KB do not fit with anyone or in the case in which no verb is identified in the request, but some target or pointing information is extracted in the previous step, the interpreter is able to generate task suggestions, based on the relations and equivalences defined in the KB. For instance, if *tornillo* (*bolt*) is the only information extracted from the voice request, the interpreter is able to generate *screw* and *unscrew* as potential task candidates. This functionality will be exploited in the future allowing the robot to request the worker to carry out the disambiguation.

The voice interpreter final output consists of frames, one for each potential task candidate, including information denoting gestures, if any exists.

For the shake of illustration, Fig. 6 presents the example of asking a robot for deburring a piece starting from a certain point (*Remove the burrs from here*).

C. Gesture Interpretation

Two kinds of gestures are addressed within the *FourByThree* project: pointing gestures and gestures for simple commands such as stop/start. This paper only deals with pointing gestures that are recognized by means of point-cloud processing. The system must be able to the different pointing gestures that have been executed within a certain period of time, providing the *x,y,z* coordinates for each of them.

The setup used consists of a collaborative robot and a sensor capable of providing dense point clouds, such as the ASUS Xtion sensor, the Microsoft Kinect sensor or the industrial-grade Ensenso system by IDS. The sensor is placed above the

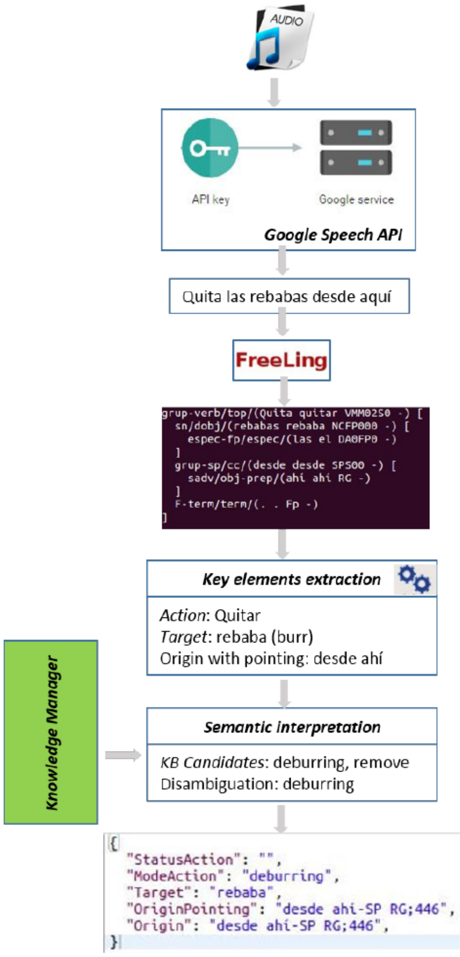


Fig. 6. Voice interpreter execution sequence

Fig. 7. Pointing gesture functional demonstrator

human operator and orientated towards the working area of the robot, so that the point cloud obtained resembles what the human operator in perceiving in the working environment (see Fig. 7).

The point cloud is initially divided into two regions of interest (ROI), the first one corresponding to the gesture detection area, and the second one defining the working area of the robot where the pointing gesture is done.

Two main problems need to be solved for the interaction between the person and the robot to succeed:

1) Robust estimation of the pointing gesture: The ROI for the pointing gesture detection is initially defined by specifying in the environment a cuboid with respect to the reference frame. In this case, the reference frame is the sensor frame, but it can also be defined using another working frame, provided a tf transformation exists between the frame used and the sensor frame. For robustness, the pointing gesture is defined using the forearm of the human operator. To identify the arm unequivocally, an Euclidean cluster extraction is performed.

2) Intersection of the pointing gesture with the working area of the robot: The main objective of a pointing gesture is to determine the point on the working area that is being pointed at. To identify this point, the points in the cloud corresponding to the pointing line are selected, from the furthest one all the way to the origin of the line that corresponds to the pointing arm. For each one of the points, a small cuboid is defined, and the ROI of the working area of the robot is filtered with it. If more than N points of the working area are present inside the small centered cuboid defined in the points of the projection line, an intersection has been found. The final intersection point that is published is the closest one to the origin of the projection line. As a threshold, a minimum Euclidean distance value is defined in order to avoid detecting intersections corresponding to the proper point cloud of the arm that generates the pointing gesture.

When detecting gestures in a time frame, a spatial filtering approach has been implemented to distinguish among real stable pointing gestures and natural arm movements. The system monitors the intersection points obtained by the algorithm, and once a valid intersection point is obtained, the spatial filtering monitoring is launched. To detect a stable gesture, N consecutive intersection points must be contained in a defined cube whose centroid is the first intersection point obtained. The number of consecutive intersection points and the edge of the filtering cube are defined as parameters. A pointing gesture is considered stable and valid if it fulfils the previous condition. If not, the points of the last filtering operation are discarded. Valid points are queued during the time frame, and dispatched at the end of the acquisition time according to the format described below.


```
{"points": [
    {"x": "x1","y": "y1","z": "z1"},
    ... ,
    {"x": "xN","y": "yN", "z": "zN"}
]}
```


D. Fusion Engine

The fusion engine aims to merge both the text and the gesture outputs in order to deliver the most accurate request to be sent to the Execution Manager, the element in the robot control architecture in charge of controlling the execution of commands. The engine considers different situations, according to the complementary and/or contradictory levels of both information sources.

As a first approach, it has been decided that the text interpreter output prevails over the gesture information. In this way, when a contradictory situation occurs, the final request is based on the text interpretation output. When no contradiction exists between both sources, the gesture information is used either to confirm the text interpretation (redundant information), or to complete it (complementary information). An example of this complementarity is, for instance, to use the voice to determine an action and a gesture to indicate the location of the object the action applies to. In contrast, if voice interpretation delivers a request that includes only one pointing gesture, but the gesture module identifies more than one gesture, only the first will be considered, discarding other points information.

As a result, the fusion engine will send to the Execution Manager the most coherent and reliable request that is understandable by the robot.

In the current implementation, only the pointing gesture is included and for that reason we have only tested the complementary functionality of the fusion engine. In the sample used in this paper the output is the deburring program having as parameter the coordinates of the point to start from (as showed in the box below).


```
{"program": "deburring",
   "from": {"x": "x1","y": "y1","z": "z1"} ,
   "to":{}
}
```


V. Semantic Multimodal Interpreter in Action

This section summarizes the results achieved after testing the Semantic Multimodal Interpreter with different types of requests comprising different voice and gesture inputs.

The initial setup consists of a button used to trigger the semantic interpreter, a microphone for voice input, a sensor capable of providing dense point clouds and a collaborative robot and one human operator performing the collaboration tasks introduced in the Case Study section.

Every time the operator wants to request the robot to do something, he/she has to click the button and keep it pressed until the end of the request. When the system detects the click on the button, it starts recording the voice and triggers the gesture module, which starts recognizing points until the operator releases the button. Then, the recorded audio is sent to the voice interpreter and, in parallel, the gesture module starts delivering the coordinates of the points that have been pointed at. When both modules return the potential candidates in terms of actions and points, the fusion engine is triggered to estimate the full command to be sent to the Execution Manager. The process sequence is shown in Fig. 8.

The initialization process and the interpretation of some sample requests have been carried out in a laboratory environment to validate relevant aspects such as the performance

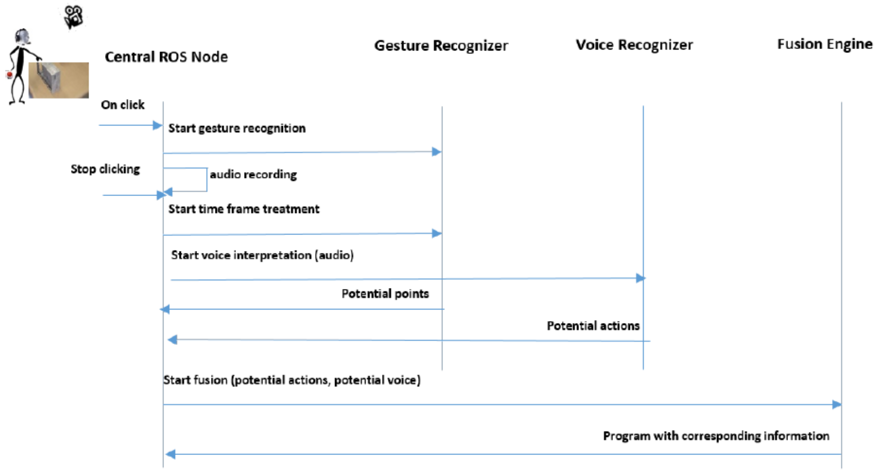


Fig. 8. Multimodal Interaction process sequence

in terms of processing time and to try to identify the most common errors during the interpretation.

The initialization (mainly knowledge base creation) takes around 15 seconds. Regarding the time required for gesture and voice interpretation, between 1 and 2 seconds are necessary for gesture recognition, while for voice interpretation times vary depending on the complexity, ranging from 3 to 3.5 seconds. Table I summarizes the processing time of some relevant examples, showing that one of the most critical steps within the voice interpreter module is the Google Speech API (GSA in TableI) that takes around 1.7-2.2 seconds to process each petition. Followed by FreeLing which response times remain stable in all cases, whilst the required time for the identification of the key elements and their interpretation (SemInt in the table) varies depending on the amount of elements to manage and if the disambiguation step is required, but is significantly lower than the time required by previous steps.

TABLE I
VOICE INTERPRETER DISAGGREGATED TIMES(SECONDS)

Voice Request	GSA	Freeling	SemInt	Total
Quita ese tornillo (Remove that screw)	1.66	1.13	0.14	2.93
Quita esa rebaba (Remove that burr)	1.695	1.1	0.185	2.98
Empieza a atornillar la pieza (Start to screw the piece)	1.87	1.12	0.25	3.24
Comienza el mecanizado de la pieza redonda de allí (Begin with the machining of the round piece that is there)	2.19	0.98	0.26	3.43
Desatornilla de aquí a allí (Unscrew from here to there)	1.83	1.06	0.12	3.01
Detén el mecanizado (Stop the machining)	1.75	1.167	0.25	3.167

VI. Conclusions and Future Work

We have presented a semantic-driven multimodal interpreter for human-robot safe collaborative interaction focused on industrial environments. The interpreter relies on text and gesture recognition for request processing, dealing with the analysis of the complementary/contradictory aspects of both input channels, taking advantage of semantic technologies for a more accurate interpretation due to the reasoning capabilities it provides. The use of semantic technologies to describe robot characteristics and capabilities and the context of scenario, makes this approach generic and scalable: by including the scenario context information and the robot capabilities in the KB, the solution will be ready to reuse without any code modification or re-compilation. Even if in very different collaborative scenarios a deeper KB extension would be necessary, potentially no code modification would be necessary to get it working.

This approach is generic and it can be applied in different industrial scenarios, although for evaluation purposes we are focusing on the human-robot collaborative assembling and deburring tasks. We intend to measure the whole system accuracy as well as the benefit of a multimodal system against a mono-modal one in industrial environments. In addition, we will assess the usability and the benefits of such a system in industrial scenarios, as part of the advancement towards natural communication in human-robot collaborative work.

Acknowledgment

The FourByThree project has received funding from the European Union's Horizon 2020 research and innovation program, under grant agreement No. 637095

References

[1] http://fourbythree.eu/.

[2] A. Bannat, J. G. ans T. Rehrl1, W. Rösel, G. Rigoll, and F. Wallhof, "A multimodal human-robot-interaction scenario: Working together with an industrial robot," in *In Proceedings of HCI PartII: Novel Interaction Methods and Techniques*, 2009, pp. 303–311.

[3] T. Fong, R. Illah, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3, pp. 143–166, 2003.

[4] M. Goodrich and A. Schultz, "Human-robot interaction: a survey," *Foundations and trends in human-computer interaction*, vol. 1, no. 3, pp. 203–275, 2007.

[5] R. Stiefelhagen, C. Fugen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel, "Natural human-robot interaction using speech, head pose and gestures," in *Intelligent Robots and Systems, 2004. (IROS 2004)*, 2004, pp. 2422 – 2427 vol.3.

[6] B. Burger, I. Ferrane, and F. Lerasle, *Towards multimodal interface for interactive robots: challenges and robotic systems description*. INTECH Open Access Publisher, 2010.

[7] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers., F. Dellaert, D. Fox, D. Hähnel, G. Lakemeyer, C. Rosenberg, N. Roy, J. Schulte, D. Schulz, and W. Steiner, "Experiences with two deployed interactive tour-guide robots," in *Proceedings of the International Conference on Field and Service Robotics*, 1999.

[8] K. Gunhee, C. Woojin, K. Munsang, and L. Chongwon, "The autonomous tour-guide robot jinny," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004, pp. 3450–3455.

[9] T. Winograd, "Procedures as a representation for data in a computer program for understanding natural language," DTIC Document, Tech. Rep., 1971.

[10] M. MacMahon, B. Stankiewicz, and B. Kuipers, "Walk the talk: Connecting language, knowledge, and action in route instructions," *Def*, vol. 2, no. 6, p. 4, 2006.

[11] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions." in *Proceedings of the International Conference on Human-Robot Interaction*, 2010, pp. 259–266.

[12] M. Tenorth, L. Kunze, D. Jain, and M. Beetz, "Knowrob-map - knowledge-linked semantic object maps," in *Proceedings of 2010 IEEE-RAS International Conference on Humanoid Robots*, 2010.

[13] T. Wang and Q. Chen, "Object semantic map representation for indoor mobile robots," in *Proceedings of International Conference on System Science and Engineering*, 2011, pp. 309–313.

[14] S. Rossi, E. Leone, M. Fiore, A. Finzi, and F. Cutugno, "An extensible architecture for robust miultmodal human-robot communication," in *International Conference on Intellegent Robots and Systems (IROS)*, 2013, pp. 2208–2213.

[15] D. Di Marco, M. Tenorth, K. Hussermann, O. Zweigle, and P. Levi, "Roboearth action recipe execution," in *In Frontiers of Intelligent Autonomous Systems*, 2013, pp. 117–126.

[16] A. Gonzalez-Agirre, E. Laparra, and G. Rigau, "Multilingual central repository version 3.0: upgrading a very large lexical knowledge base," in *In Proceedings of the Sixth International Global WordNet Conference (GWC12)*, 2012.

[17] http://stardog.com/.

[18] https://console.developers.google.com/apis/api/speech/.

[19] L. Padró and E. Stanilovsky, "Freeling 3.0: Towards wider multilinguality," in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, 2012.

[20] S. Harris, A. Seaborne, and E. Prudhommeaux, "Sparql 1.1 query language," *W3C Recommendation*, vol. 21, 2013.

7.11. FourByThree: Imagine humans and robots working hand in hand

Iñaki Maurtua, Nicola Pedrocchi, Andrea Orlandini, Jose de Gea Fernandez, Christian Vogel, Aaron Geenen, Kaspar Althoefer, Ali Shafti: FourByThree: Imagine humans and robots working hand in hand. ETFA 2016: 1-8 [111]

FourByThree: Imagine humans and robots working hand in hand

Iñaki Maurtua[1], Nicola Pedrocchi[2], Andrea Orlandini[3], Jose de Gea Fernández[4], Christian Vogel[5], Aaron Geenen[6], Kaspar Althoefer[7], Ali Shafti[8]

[1]IK4-TEKNIKER, Eibar, Spain
[2]CNR-ITIA, Milano, Italy
[3]CNR-ISTC, Roma, Italy
[4]DFKI-RIC, Bremen, Germany
[5]FhG IFF, Magdeburg, Germany
[6]ZeMA, Saarbrücken, Germany
[7]QMUL, London, United Kingdom
[8]KCL, London, United Kingdom

***Abstract*— Since December 2014, FourByThree Project ("Highly customizable robotic solutions for effective and safe human robot collaboration in manufacturing applications") is developing a new generation of modular industrial robotic solutions that are suitable for efficient task execution in collaboration with humans in a safe way and are easy to use and program by factory workers. This paper summarizes the key technologies that are used to achieve this goal.**

***Keywords*—modular, safety, manufacturing, collaboration, usability**

I. Introduction

Industrial robots have demonstrated their capacity to answer to the needs of many industrial applications, offering a high degree of dexterity, accuracy and efficiency. Their use is extended to all kinds of applications, but it is in the case of large production batches, repetitive operations, risky or unpleasant working conditions where their introduction has been most significant.

However, when the application requires collaboration between the robot and the worker, including workspace sharing, it is not feasible to use standard industrial robots due to safety restrictions. Recently, new robotic products have appeared on the market claiming to be safe when used in the vicinity of humans – examples include the Universal Robots UR3/UR5/UR10 [1] , the Light Weight Robot from KUKA [2], Yumi from ABB [3], the arms from Rethink Robotics[5] or the FRANKA robot [4] presented in the Hannover Messe 2016. These robots offer good solutions for some specific applications where close proximity between humans and robots is a must, making it possible to control the force exerted in case of collision. However, they either lack in flexibility (in terms of possible physical configurations) or are very expensive - some of them are three times more expensive than the counterpart standard ('non-safe') version.

Since December 2014, the FourByThree Project ("Highly customizable robotic solutions for effective and safe human robot collaboration in manufacturing applications") is developing a new generation of modular industrial robotic solutions that are suitable for efficient task execution in collaboration with humans in a safe way and are easy to use and program by the factory worker. The *FOUR* main characteristics (Modularity, Safety, Usability and Efficiency) of FourByThree are:

1) *Modularity*

FourByThree outcomes are packed as a 'kit' of hardware and software tools for the development of custom robotic solutions. The concept includes fundamental mechanical elements (four different size series-elastic actuators, brackets, flange), the control unit (incorporating advanced techniques for safe HRI) and additional auxiliary hardware/software modules integrated in a ROS based FourByThree control architecture.

2) *Safety*

Safety strategies and low cost mechanisms allowing intrinsically safe behaviour of the robot in the presence of humans are developed. The safety approach is centered around the design of the actuators with the capability to monitor the force and torque in each, providing the opportunity to implement variable stiffness strategies and reactive behaviour in case of contact/collision. The system also includes space monitoring using a projecting system and a vision system, which provide the information needed to modify the velocity of the robot according to its relative distance with respect to the worker.

3) *Ease of use*

FourByThree offers a set of multimodal interaction mechanisms that facilitate the programming and control of robots, e.g., voice based interaction, manual guidance. These multimodal interaction mechanisms are complemented by human-oriented automatisms ensuring intuitive and safe HRI.

4) *Efficiency*

Robots are intended to help workers in doing a task, to this aim they have to be reliable, maintainable and intrinsically safe. Performance metrics are established for each application addressed in the project, i.e., assembly, deburring, welding, riveting and machine tending, implemented in four challenging

978-1-5090-1314-2/16/$31.00 ©2016 IEEE

industrial Pilot Studies (aeronautics, sheet metal forming, investment casting and professional training).

In the following chapters, the key technologies developed in the project are described.

II. Modular Design

A. Actuators

The actuators are complete modules including the motors, gears, sensors, elastic element, and the embedded electronics (together with its software) required to drive and control a single elastic joint. They offer, as well, some of the functionalities needed in the safety strategy, i.e. speed, force and torque monitoring.

Mechanics. It was initially decided to build four different actuator sizes (with torques 28 Nm, 50 Nm, 120 Nm and 300Nm, respectively, at link side) to cover a wide range of possible arm configurations and scenarios. The initial list of requirements contemplated among others: maximum link-side torques of around 28Nm, 50Nm, 120Nm and 300Nm, mechanical deflection of around 5° at the maximum respective torque, compact, modular and lightweight design, link-side speeds of around 15rpm and the use of safety brakes.

Two actuators have been built for the initial prototypes: Type I (28Nm) and Type II (50Nm). The design of the Type III (120Nm) and Type IV (300Nm) is currently being finished. All actuators are based on previous modular actuators designed at DFKI (see example reference in [6]). They combine Robodrive brushless DC motors with Harmonic Drive gears. Additionally, in-house developed motor electronics consisting of four PCBs are embedded within the housing of each actuator. The main difference of these actuators with respect to older versions is that they include an elastic element in series with the motors.

Previous developments of the project CAPIO [7] - which were already using an elastic element - were taken as an starting point for building the actuators of Type I (28Nm). The elastic element is a combination of small disc springs placed at both sides of a lever rotating with the motor (see Fig. 1-left). In contrast to the CAPIO actuators, these new actuators include embedded electronics entirely based on FPGA (previous design consisted of a hybrid solution using a microcontroller and an FPGA), several mechanical optimizations, and a fourth electronics board acting as 'electronic brake'.

Fig. 1. Actuator Type I: 28Nm.

For the actuator Type II (50Nm), a new spring element based on coil springs has been developed (see Fig. 2-left). The spring coupling has a progressive characteristic: initially it exhibits a linear characteristic until approx. 5° of deflection and, after that, a more abrupt increase of stiffness is introduced. The purpose is to avoid having the spring completely compressed at the maximum torque, but to rather have it stiffer while reaching the maximum torque. This has been achieved with the introduction of a second harder spring placed inside the 'main' spring.

Fig. 2. Actuator Type II: 50Nm.

Embedded Electronics. The basic electronics stack is composed of three PCBs that incorporate all sensors that are required to monitor and control the actuators: motor current sensors are integrated in the low phases of the three-phase H-bridges, and absolute encoders with 19-bit resolution before and after the gear measure the motor position. Additionally, a third absolute encoder is placed after the elastic element to measure the link position. All mentioned sensors as well as current, speed, and position controllers are processed by a Spartan6 FPGA from Xilinx.

Moreover, the actuator electronics have been enhanced in this project with two additional electronic boards: a board for enabling/disabling the mechanical brakes of the Type II actuators (named as 'BrakeBoard') which also monitors the motor phase currents as an additional motor current measurement, and a board for short-circuiting the motor phases of the Type I actuators (the so-called 'electronic brake') to be used as an electrical brake.

Low-level Control. The FPGA-based robot joint controller developed and used previously at DFKI has been extended for the control of the spring deflection. Using a cascaded controller for position, velocity, and motor current, an additional PID control loop regulates the deflection of the spring element of the elastic actuators by either acting on the velocity controller input or by directly acting on the motor current controller input.

Furthermore, the model of the spring deflection and its relative output torque is required for being able to control the actuator torque. The torque-spring deflection is thus modeled by using joint probability densities that are represented by a dynamic Gaussian mixture model (DGMM) [8]. Initial experiments are being carried out to validate the results.

B. Robot design

The FourByThree robot system is designed for human-robot collaboration (HRC). Compared to traditional industrial robot systems, the modular robot system can be optimally adapted and used for different tasks and applications. Following this modularity objective, four different robots have been designed to answer the specific requirements of each of the four Pilot studies that are used to validate the concept. They include welding, riveting, handling, machine tending and assembly applications. The robot design provides a modular construction kit developed according to the norms and directives for collaborating robot systems.

Fig. 3. First robot prototype.

The robot construction kit consists of key basic elements: 'base', 'joints', 'link elements' and 'flange'. Thus, depending on the needed applications, different kinds of robot systems can be configured using the construction kit elements. The setup of the robot system is based on virtual modeling of kinematics, derived from the workspace analysis results of each application. Using simple calculations, it is possible to determine the custom configuration of the robot system based on the construction kit, for later assembly by system integrators or end users.

C. *Control Architecture*

A three-layer software architecture is proposed:

- Low level: which includes the drivers and the joint controllers. This level offers an interface with the motors (commands and information retrieval). Lowest control modules, e.g. impedance control and manual guidance, are also included in this level.
- Medium level: which is in charge of controlling the execution of user programs and any other action coming from the higher level.
- High level: which includes the system's high level modules, user applications and the Dynamic Task Planner.

ROS is used as core framework for the two higher levels.

D. *Robot identification and low level control*

Modularity and Compliancy are powerful instruments that will open a wide spectrum of applications for the novel generation of the FourByThree robot. However, Modularity and Compliancy are critical aspects in the design of the robot motion and control. Static/dynamic accuracy and repeatability are demanding when standard control strategies are deployed in combination with compliant robots. Step-changes are indeed necessary to address the challenges: (i) preserve the motion smoothness in all the working conditions, (ii) preserve the motion performances compared to standard rigid robots, (iii) auto-tuning of control parameters to overcome changes in working conditions by learning procedures. To face such challenges, the FourByThree solution will provide a set of innovative motion and control modules.

By considering (i) and (ii) an *elastic-input-output inversion centralized closed-loop controller* (ELIO) will guarantee the maximization of the controllable bandwidth through the integration and the inversion of the completed elasto-dynamic model of the robot [15][16]. The controller will take into account zero-dynamics behavior reducing the control effort. The ambition is to mitigate the elasticity of the robot up to 5Hz reaching therefore the typical performance of standard robots of similar payload. However, to overcome the well-known control effort problem with the input-output system inversion, the FourByThree robot will be endowed with two motion modules: a high-order motion planner (HELIOS) and an innovative Elasto-Dynamic Identification Tool (EDIT).

HELIOS is a motion planner based on a double concept: it works as a motion filter generating a smooth trajectory [13], and the core is based on an optimal constrained predictive control methodology [17], with an optimization window limited to preserve the necessary calculus performance. HELIOS results in a high-versatile motion planner that can be used off-line to plan the smooth trajectory, and on-line as input filter to smooth all the signals sent to the robot control (ELIO).

EDIT is of utmost importance in the FourByThree ecosystem: the tool allows high accuracy in the identification of dynamic properties of the system (friction, masses, inertia, joint stiffness etc.). The tool takes into account the nonlinear estimation of the inertial parameters, and the high-frequency parasitic modes of vibrations. The two effects are decoupled by using a projection method [18]. Furthermore, EDIT integrates in the estimation the minimum analytical representation of the system dynamics [14]. Such a feature is fundamental, especially for modular robots, in order to guarantee maximum accuracy in parameter estimation, avoiding observability issues of the dynamic system.

Finally, considering (iii), a control procedure with multiple learning levels will be proposed to compensate for friction at joint level (the most relevant problem in robotics applications [19]) and to compensate for robot-environment coupled interaction dynamics [20]. In fact, taking into account industrial interaction robotized tasks, such effects might result in instabilities, force overshoots and task failures. On top of compliance control, the proposed approach consists of two main control levels: a) iterative friction learning and b) iterative force tracking learning, both relying on the reinforcement learning procedure. While a) allows to locally improve the robot dynamics compensation (needed if working conditions change) through iterative and continuous estimation of the joints' friction parameters, b) allows to improve the interaction task execution, compensating for the elasticity of the interaction environment and avoiding force overshoots, adapting the force tracking control gain and the compliance control damping.

E. *Variable Stiffness*

The FourByThree robot's impedance control is handled through methods described in §II.D and based on the task at hand and its relevant program. However, it is also necessary to have a higher level mechanism to adjust the robot's impedance based on safety considerations. While basing the robot's impedance solely on the task at hand and the predefined program is fine to achieve the required results for the job, it does not take into consideration anomalies and mistakes in the human or robot's behavior which can lead to safety concerns.

To circumvent this, a higher level stiffness adjustment module is proposed. This module will consider the position and velocity of the robot and the human with respect to each other in order to keep the robot arm's stiffness at a safe level throughout. High velocity and low distance result in a high risk of collision, which means the robot arm's stiffness should be reduced in order to minimize damage.

A fuzzy algorithm approach is considered to implement this. Factors affecting stiffness are distance between the human and the robot, the direction and the velocity of motion. The decision algorithm which is applied through fuzzy logic will map these input values to an appropriate stiffness adjustment output. These input values are obtained through sensors provided as modules in the FourByThree system's architecture. These include RGB-D cameras which allow human and environment monitoring in real-time.

The fuzzy logic mechanism will feed the above sensor data to its fuzzification module. This will use triangular membership functions to map distance and velocity values to fuzzy definitions with their respective μ values representing possibility. The output of the fuzzification module is fed to the inference module. This relies on the product operation to infer the fuzzy stiffness μ values based on the fuzzified input μ values. Thus, the μ values for the fuzzy input are multiplied for distance and velocity to obtain the μ value for stiffness. The fuzzy stiffness value itself is obtained through a lookup table that maps different fuzzy distance and velocity values to individual fuzzy stiffness values ranging from 0 to 5. For defuzzification, the center average method is used. Thus, based on the fuzzy stiffness value achieved above, and the relevant μ values, the final stiffness value is obtained.

This system allows for a real-time, continuous adjustment of the robot arm's stiffness based on safety concerns. The safe range of the stiffness value is constantly published by this module. The low level impedance controller will consider a stiffness range with which the task at hand can continue. It will then change the stiffness within this range based on the suggestions offered by the high-level stiffness adjustment module. If the suggested change in stiffness falls outside the low-level controller's range for the task, then the circumstances have resulted in the task no longer being safe. In this case, the task will need to stop for the stiffness value to change appropriately.

F. *Dynamic Task Planning*

The FourByThree control architecture has been endowed wit a dynamic task planner designed and developed to implement continuous task synthesis features, ensure safety-critical properties at execution time, and endow the overall system with user modeling abilities for adapting tasks to the different humans at work collaborating with the robot. The integration of plan synthesis and continuous plan execution has been demonstrated both for timeline-based planning (e.g., [21]) and PDDL based (e.g., [22]). In scenarios of human-robot interaction important problems have been addressed: (a) "human aware" planning has been explored for example in [23], (b) the interaction of background knowledge for robotic planning in rich domain (addressed for example in [24]), (c) synthesis of safety critical plans to guarantee against harmful states relevant in co-presence with humans (addressed in [25] and [26]).

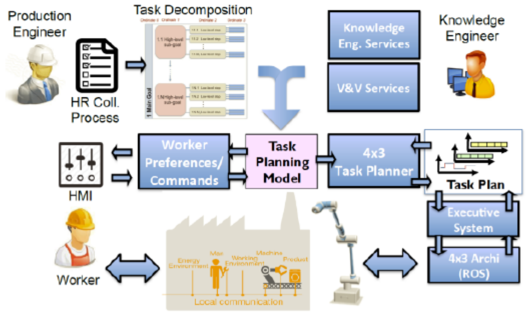


Fig. 4. Dynamic task planning framework.

Within the FourByThree project, a timeline-based planning approach is pursued relying on the APSI-TRF [27], developed for the European Space Agency and exploited in several missions. Then, the envisaged planning framework is to deploy a continuous task planning and adaptation system with humans in the loop [29]. The overall framework is depicted in Fig. 4. A Production Engineer is in charge of defining the Human-Robot Collaborative (HRC) production process characterizing each task according to specific HRC settings (i.e., interaction modalities). Then, a Knowledge Engineer is to encode such information in a task planning model following a hierarchical decomposition and leveraging the features provided by a Knowledge Engineering Environment for planning with timelines [28], that integrates *classical* knowledge engineering features with Verification and Validation (V&V), formal techniques to perform domain model validation, planner validation, plan verification, etc. The integration of Planning and technology with V&V techniques is key to synthesize a safety critical controller for the robot. The Task Planning Model can be, then, adapted also according to the preferences of the Human Worker that is supposed to interact with the robot during the production process. A FourByThree Task Planner then generates a temporally flexible task plan to be dispatched to the robot through an Executive System (integrated in the general ROS-based architecture). The dispatched tasks are then to be actually executed on the robot activating the proper control of motion actions and actuator activation signals. During the production process, the Executive System is also in charge of monitoring the plan execution and, in case of need (e.g., a specific command issued by the human worker), ask the task planner to dynamically face modifications of the production environment. It is worth underscoring that the task planning and other modules are intended to be tightly coupled; as motion planning modules are to provide temporal bounds for robot movements while safety modules such as, for instance, variable stiffness module, will leverage the outcome of the dynamic task planning system to better tailor robot settings while interacting with the human worker.

III. Safety

The safety strategy in FourByThree is based on five pillars:

- The actuators, which allow measuring the force and torque values using two different physical principles, resulting in a safe approach.
- The robot design, through emphasis on the elimination of sharp edges, reduction of trapping risks, etc.
- The external monitoring system, which consists of a projection system and a vision system, allowing to monitor the space around the robot to detect any possible violation by the worker.
- Adjustable stiffness control.
- The control architecture.

The proper use of these features makes it possible to satisfy the operating conditions established in ISO10218 parts 1 and 2, and ISO/TS15066, once the mandatory Risk Assessment has been performed.

A. *Architecture*

The Safety strategy in 4x3 is outlined in Fig. 5.

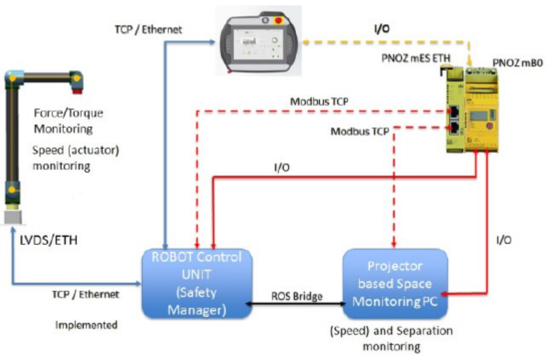


Fig. 5. Safety architecture.

In brief, the FourByThree safety strategy allows (1) defining a protective area around the robot for co-existence and interference situations (i.e., when the human moves through the robot workspace but does not interact directly with the robot or when the human reaches into the robot working area or obstructs the robot workspace in a non-planned task).

The projection system is in charge of monitoring the robot workspace and triggering the safety signal when there is a violation in the area; (2) for co-operation activities (i.e., when the human has to interact with the robot in a productive way) the system's capability to monitor and limit the force and torque is used to guarantee the safety.

The safety strategy and the different components are analyzed in collaboration with an external certification body and the certification roadmap will be established by the end of the project.

B. *Projection based space monitoring*

The projection-based monitoring system is responsible for ensuring the human's safety in applications that will not allow a contact between human and robot. This will be the case for instance if the robot moves with high speed, uses dangerous tools for grasping or handles risky workpieces. For monitoring such human-robot cooperation scenarios the Fraunhofer IFF developed an innovative sensor system that is based on projection and camera techniques [9][10]. The sensor system is capable of establishing safety spaces of arbitrary shapes by projecting light from the projector directly onto the environment. Violations of these safety spaces caused by disruptions to the emitted light are robustly detected by surrounding cameras. By incorporating the current joint positions and velocities of the robot, the safety spaces can be dynamically adapted to enclose the robot minimally at any point of time (see Fig. 6).

As there is no need for a complex computation of three-dimensional data of the environment, the implemented algorithms for image processing and collision detection lead to minimal reaction times of the system. Furthermore, the robustness and availability of the system is enhanced through synchronization of projectors and cameras. Here, the cameras are adapted to the frequency of the light emitted from the projector, reducing the influence of environmental light conditions on the collision detection process.

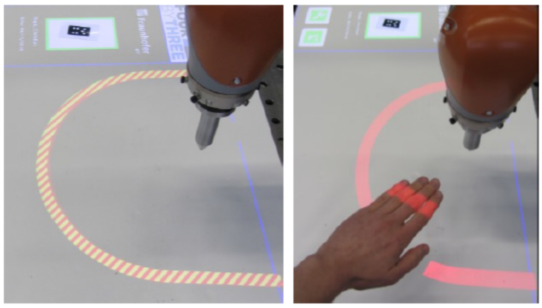


Fig. 6. Dynamically established safety space without safety violation (left) and with safety violation (right).

In "FourByThree" this technology is deployed for applications in real industrial environments. To meet the requirements of various industrial conditions, the projection-based monitoring system has been adapted according to availability, safety and modularity. Here, single modular projection units have been developed that provide higher flexibility and customizability. Each unit comprises one projector and two cameras that can be adjusted individually. By configuring several units to work together, it will even be possible to operate in difficult environmental conditions with low ceilings or large monitoring areas.

Besides technical improvements that include the enhancement of response time, detection capabilities and robustness, Fraunhofer IFF is working on the evaluation of the sensor system according to safety certifications.

IV. Interaction

Natural communication between humans and robots can happen through several channels, the main of which are voice and gestures. In this multimodal scenario, the information can be complementary between channels, but also redundant. However, redundancy can be beneficial [11] in real industrial scenarios where noise and low lighting conditions are usual environmental challenges that make it difficult for voice and visual signals to be captured with clarity.

FourByThree proposes a semantic approach that supports multimodal interaction between humans and industrial robots in real industrial settings.

A. *Voice and gesture based interaction*

The approach aims at creating a safe human-robot collaborative environment in which interactions between both actors happen in a natural way i.e. communication based on voice and gestures. We propose a semantic multimodal interpreter prototype that is able to process voice and gesture-based natural requests from a person, and combine both inputs to generate an understandable and reliable command for industrial robots, enhancing safe collaboration. For such semantic interpretation, four main modules are developed, as shown in Fig. 7: a Knowledge-Manager module that describes and manages the environment and the actions that are feasible for robots in a given environment, using semantic representation technologies; a Voice Interpreter module that given a voice request, extracts the key elements on the text and translates them into a robot-understandable representation, combining NLP and semantic technologies; a Gesture Interpretation module mainly for resolving pointing issues and some simple orders like stopping an activity; and a Fusion Engine for combining the output of both text and gesture modules and construct a complete and reliable order for the robot.

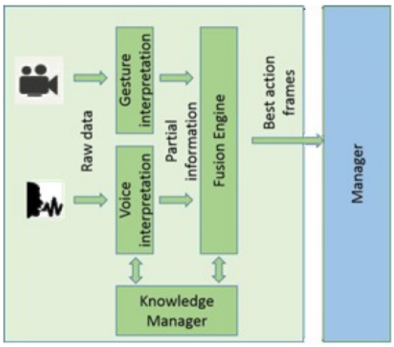


Fig. 7. Multimodal semantic approach architecture

These main modules are described in detail in the following subsections.

1) *Knowledge Manager*

The knowledge manager comprises ontologies that model environmental information of the robot itself, including its own capabilities. In addition, the knowledge manager allows modeling the relationships between the concepts. These relationships are implicit rules that can be exploited by reasoners in order to infer new information from the ontology. As a result, reasoners can work as rule engines in which human knowledge can be represented as rules or relations.

2) *Voice Interpreter*

Given as input a human verbal request, the purpose of this module is to understand exactly what the person wants and if it is feasible to generate the necessary information for the robot. The first step concerns speech recognition. The second step is based on superficial information, in the sense that it does not take into account the meaning of words in the context. The only purpose is to extract the key elements from the given order.

The last step attempts to identify the action that is asked for, considering the key elements in the given context. The module output consists of frames, one for each potential task candidate, including information denoting gestures, if any.

3) *Gesture Interpretation*

Two kinds of gestures are addressed within the FourByThree project: pointing gestures and gestures for simple commands such as stop/start. In the case of pointing gestures, they are recognized by means of point-cloud processing. In this context, the system must be able to not only recognize the pointing gesture, but also deliver within a certain period of time how many different pointing gestures have occurred and which ones those are, in terms of x, y and z coordinates.

The initial setup consists of the collaborative robot and a sensor capable of providing dense point clouds, such as the ASUS Xtion sensor, the Microsoft Kinect sensor, or the industrial-grade Ensenso system by IDS. The sensor is placed above the human operator and orientated towards the working area of the robot, so that the point cloud obtained resembles what the human operator in perceiving in the working environment.

4) *Fusion Engine*

The fusion engine aims to merge both the text and the gesture outputs in order to deliver the most accurate request to send to the executive manager. The engine considers different situations regarding the complementary and/or contradictory levels of both sources.

As a first approach, it is decided that the text interpreter output will prevail over the gesture information. When no contradiction exists between the two sources, the gesture information is used either to confirm the text interpretation (redundant information), or to complete it (complementary information).

B. *Projection based interaction*

Besides the safety aspect of the projection-based monitoring system the technology provides interaction and visualization capabilities. Here, the system can visualize relevant information to support the user at work but it also allows the user to offer input and information back to the robotic system. This means that the projection system is capable of providing buttons or simple menus that can be used to control the robot, task or process. The shape of these interaction areas and the reaction upon triggering can be configured individually.

At present, two interactive buttons that control the application's workflow have been implemented. A screenshot of these buttons is depicted in Fig. 8. The first one activates the manual task which enables the workpiece detection process and visualizes some additional task-related information. The second button activates the robot task. Thus, the safety space monitoring is enabled and the robot starts its motion and processes the workpiece autonomously.

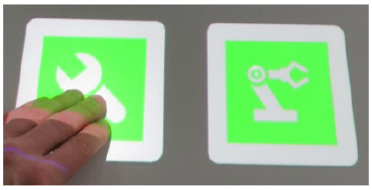


Fig. 8. Interactive buttons control task and robot.

In addition to the interaction possibility, the system performs an access control that offers different interaction buttons regarding the access rights of the user. For this, we implemented an identification area that detects the user's card and processes the user's rights accordingly (see Fig. 9).

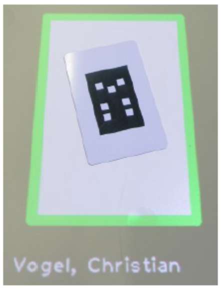


Fig. 9. User identification area with card and identified user.

V. Programming

Industrial robot manufacturers offer their own proprietary programming language that allow typical robot control of movements and I/O management. Most of them offer software packages for the application of domain-specific tasks (e.g., welding, gluing, handling, machining) that contain a set of additional instructions that can be used to program specific tasks. It is becoming very common to offer the possibility of using general purpose languages, such as C, C# or Java to customize and develop applications by end-users. This is the case with KUKA that provides the Sunrise API in its Sunrise Controller in JAVA [12] that, unfortunately, demands high programming skills.

In 4x3, the programing approach allows using both programing by demonstration and standard textual programing.

A. *Standard programming*

As there is not any widely accepted robot programming language, FourByThree proposes a simple to use language that allows programmers to access functionalities.

This standard textual programming includes:

- Open language definition. It has some commonalities with languages used by industrial robots. It includes movement commands, mathematical operators, I/O instructions, flow control primitives, logical operators, etc.
- Easy to use editor, including a syntactic analyzer. The lexical analysis is the process of converting a sequence of characters into a sequence of tokens, i.e. meaningful character strings. This process is generally combined with a syntactic analysis which takes a list of tokens and analyses them conforming to the rules of a formal grammar. A program or function that performs lexical analysis is called scanner, lexer or tokenizer, and the software component that takes the list of tokens and checks for the syntax correctness is called parser.
- Program executor. A component that interprets the content of the program and translates it into robot understandable instructions and sequences.

Scanner and parser functionalities will be implemented using existing tools, Flex and Bison.

B. *Programming by demonstration*

The FourByThree robot is programmable through its proprietary programming framework and compiler as described in §V.A. However, to create an easier interface for workers with no programming experience, a learning module is also considered. This module will enable the worker to program the robot through manual guidance and gesture/voice recognition. Additionally, the module allows the robot's behavior to be tailored to the worker by observing the worker's real-time kinematic behavior and focusing on ergonomics. Thus, a task is divided into coarse and fine movements of the robot, with the coarse movements being 'taught' through manual guidance and the finer movements which are dependent and specific to each worker 'learnt' by observing that particular worker's behavior.

Comfort and ergonomics are familiar terms with typically subjective definitions. Each person has their own thoughts on what is comfort and ergonomics to them, making these parameters hard to assess and compare objectively. Work related musculoskeletal disorders (WMSDs) are the result of issues in these same parameters in the workplace left unnoticed and unattended. There are methods and techniques proposed and currently in use for ergonomics assessments. These range from subjective questionnaires to observation-based measurement and scoring of joint angles involved in a posture and task, and are used regularly in clinical and industrial environments alike and are popular due to their ease of use and lack of a requirement for specific expertise. However, this simplicity has the downside of lack of objectivity and/or thoroughness. A more thorough and objective understanding of comfort and ergonomics can be achieved by relying on sensed data from a human rather than their subjective opinion. These can provide precious information about human behavior and allow assessment of different activities in terms of health and comfort. Furthermore, a real-time objective assessment of comfort will allow for better interaction between robots and humans.

The Rapid Upper Limb Assessment (RULA) method assigns scores to each part of the upper body based on the joint angles associated with it. These scores are then combined together using a look-up table in order to reach one final score, with higher numbers meaning a less ergonomic state. Using orientation sensors consisting of accelerometers and gyroscopes, or RGB-D camera systems, it is possible to obtain real-time values for the worker's joint angles. These joint angles can then be used to reach a real-time ergonomics score based on RULA. This score will then be the basis for the robot's reactive behavior. A high number indicating ergonomics risk will prompt the robot to move into a position that will affect the worker's posture positively, by forcing him/her to move to a more ergonomic state. This is implemented by identifying the different ergonomics states for the worker and using them to create a rewards function for the robot's learning module which will enable it to respond accordingly.

Acknowledgment

The FourByThree project has received funding from the European Union's Horizon 2020 research and innovation program, under grant agreement No. 637095.

References

[1] Universal Robots Website, http://www.universal-robots.com/ (retrieved on 27.04.2016)

[2] Kuka IIWA website, http://www.kuka-lbr-iiwa.com/ (retrieved on 27.04.2016)

[3] ABB robotics website, http://new.abb.com/products/robotics/yumi (retrieved on 27.04.2016)

[4] Franka website, https://www.franka.de, (retreived on 27.04.2016)

[5] Rethink Robotics Website, http://www.rethinkrobotics.com/ (retrieved on 27.04.2016)

[6] Hilljegerdes, P. Kampmann, S. Bosse and F. Kirchner, Development of an intelligent joint actuator prototype for climbing and walking robots, in International Conference on Climbing and Walking Robots (CLAWAR-09), 2009

[7] M. Mallwitz, N. Will, J. Teiwes and E. A. Kirchner, The CAPIO active upper body exoskeleton and its application for teleoperation, in Proceedings of the 13th Symposium on Advanced Space Technologies in Robotics and Automation. ESA/Estec Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA-2015), (ESA, 2015).

[8] M. Edgington, Y. Kassahun, F. Kirchner. Dynamic Motion Modelling for Legged Robots. Editors: Nikos Papanikolopoulo, Shigeki Sugano, Stefano Chiaverini, Max Meng. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, (IROS-09), 11.10.-15.10.2009, St. Louis, Missouri, pages 4688-4694, Oct/2009.

[9] Vogel, C.; Poggendorf, M.; Walter, C.; Elkmann, N.: "Towards safe Physical Human-Robot Collaboration: A Projection-based Safety System", Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, San Francisco, USA, 25-30. Sept. 2011

[10] Vogel, C.; Walter, C.; Elkmann, N.: "A Projection-based Sensor System for Safe Physical Human-Robot Collaboration", Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, Tokyo, Japan, 03.-07. Nov. 2013

[11] A. Bannat, T. R. J. Gast, W. Rosel,¨ G. Rigoll, and F. Wallhof, "A multimodal human-robot-interaction scenario: Working together with an industrial robot," pp. 303–311, 2009.

[12] http://www.kuka-robotics.com/en/products/software/kuka_sunrise.os/

[13] R. Zanasi, C.Guarino Lo Bianco, A. Tonielli, Nonlinear filters for the generation of smooth trajectories, Automatica, Volume 36, Issue 3, March 2000, Pages 439-448, ISSN 0005-1098, http://dx.doi.org/10.1016/S0005-1098(99)00164-8.

[14] Beschi, Manuel, et al. "A general analytical procedure for robot dynamic model reduction." Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015.

[15] A. Boekfah and S. Devasia, "Output-Boundary Regulation Using Event-Based Feedforward for Nonminimum-Phase Systems," in IEEE Transactions on Control Systems Technology, vol. 24, no. 1, pp. 265-275, Jan. 2016. doi: 10.1109/TCST.2015.2432153,

[16] A. de Luca and P. Lucibello, "A general algorithm for dynamic feedback linearization of robots with elastic joints," Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on, Leuven, 1998, pp. 504-510 vol.1. doi: 10.1109/ROBOT.1998.677024

[17] E. F. Camacho, "Constrained generalized predictive control," in IEEE Transactions on Automatic Control, vol. 38, no. 2, pp. 327-332, Feb 1993. doi: 10.1109/9.250485

[18] U. Forssell and L. Ljung, "A projection method for closed-loop identification," in IEEE Transactions on Automatic Control, vol. 45, no. 11, pp. 2101-2106, Nov 2000. doi: 10.1109/9.887634

[19] Armstrong-H´elouvry, B., Dupont, P., and De Wit, C. C. (1994). A survey of models, analysis tools and compensation methods for the control of machines with friction. Automatica, 30(7):1083–1138.

[20] Roveda, L., Iannacci, N., Vicentini, F., Pedrocchi, N., Braghin, F., and Molinari Tosatti, L. (2016). Optimal impedance force-tracking control design with impact formulation for interaction tasks. Robotics and Automation Letters

[21] F. Py, K. Rajan, and C. McGann. A Systematic Agent Framework for Situated Autonomous Systems. In *AAMAS-10. Proc. of the 9th Int. Conf. on Autonomous Agents and Multiagent Systems*, 2010.

[22] M. Cashmore, M. Fox, T. Larkworthy, D. Long, and D. Magazzeni. AUV mission control via temporal planning. In *IEEE International Conference on Robotics and Automation, (ICRA 2014)*. 2014.

[23] E. A. Sisbot and R. Alami. A human-aware manipulation planner. *IEEE Trans. Robotics*, 28(5):1045–1057, 2012.

[24] S. Lemaignan and R. Alami. Explicit knowledge and the deliberative layer: Lessons learned. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013*, pages 5700– 5707, 2013.

[25] T. Abdellatif, S. Bensalem, J. Combaz, L. de Silva, and F. Ingrand. Rigorous design of robot software: A formal component-based approach. *Robotics and Autonomous Systems*, 60(12):1563–1578, 2012.

[26] A. Orlandini, M. Suriano, A. Cesta, and A. Finzi. Controller synthesis for safety critical planning. In *IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013)*. IEEE, 2013.

[27] A. Cesta, G. Cortellessa, S. Fratini, and A. Oddi. Developing an End-to-End Planning Application from a Timeline Representation Framework. In *IAAI-09. Proc. of the 21st Innovative Application of Artificial Intelligence Conference, Pasadena, CA, USA*, 2009.

[28] A. Orlandini, G. Bernardi, A. Cesta, and A. Finzi. Planning meets verification and validation in a knowledge engineering environment. *Intelligenza Artificiale*, 8(1):87–100, 2014.

[29] A. Cesta, G. Bernardi, A. Orlandini, A. Umbrico. Towards a Planning-based Framework for Symbiotic Human-Robot Collaboration. In Proc. of the 21st IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2016), 2016.

7.12. Non-destructive inspection in industrial equipment using robotic mobile manipulation

Iñaki Maurtua, Loreto Susperregi, Ander Ansuategui, Ane Fernández, Aitor Ibarguren, Jorge Molina, Carlos Tubio, Cristobal Villasante, Torsten Felsch, Carmen Pérez, Jorge R. Rodriguez, and Meftah Ghrissi: Non-destructive inspection in industrial equipment using robotic mobile manipulation, AIP Conference Proceedings 1734, 130013 (2016) [101]

Non-destructive Inspection in Industrial Equipment Using Robotic Mobile Manipulation

Iñaki Maurtua[1, a)], Loreto Susperregi[1, b)], Ander Ansuategui[1,c)], Ane Fernández[1, d)], Aitor Ibarguren[1, e)], Jorge Molina[1, f)], Carlos Tubio[1, g)], Cristobal Villasante[1, h)], Torsten Felsch[2, i)], Carmen Pérez[3, j)], Jorge R. Rodriguez[4, k)], Meftah Ghrissi[5, l)]

[1] *IK4-Tekniker, Engineer, Main researcher, Iñaki Goenaga 5, Eibar 20600, Spain, +34943206744.*
[2] *Fraunhofer IFF, Sandtorstrasse 22, 39106 Magdeburg, Germany*
[3] *TECNATOM, Avda. de Montes de Oca 1, San Sebastián de los Reyes (MADRID) 28703, Spain*
[4] *TORRESOL Energy, Avenida Zugazarte 61, Las Arenas 48930, Spain*
[5] *ROBOSOFT, Voie Technopole Izarbel, Bidart 64210, France*

[a)]inaki.maurtua@tekniker.es, [b)]loreto.susperregi@tekniker.es, [c)]ander.ansuategui@tekniker.es, [d)]ane.fernandez@tekniker.es, [e)]aitor.ibarguren@tekniker.es, [f)]jorge.molina@tekniker.es, [g)]carlos.tubio@tekniker.es, [h)]cristobal.villasante@tekniker.es, [i)]torsten.felsch@iff.fraunhofer.de, [j)]cperez@tecnatom.es, [k)]jorge.rodriguez@torresolenergy.com, [l)]meftah.ghrissi@robosoft.com

Abstract. MAINBOT project has developed service robots based applications to autonomously execute inspection tasks in extensive industrial plants in equipment that is arranged horizontally (using ground robots) or vertically (climbing robots). The industrial objective has been to provide a means to help measuring several physical parameters in multiple points by autonomous robots, able to navigate and climb structures, handling non-destructive testing sensors. MAINBOT has validated the solutions in two solar thermal plants (cylindrical-parabolic collectors and central tower), that are very demanding from mobile manipulation point of view mainly due to the extension (e.g. a thermal solar plant of 50Mw, with 400 hectares, 400.000 mirrors, 180 km of absorber tubes, 140m height tower), the variability of conditions (outdoor, day-night), safety requirements, etc. Once the technology was validated in simulation, the system was deployed in real setups and different validation tests carried out. In this paper two of the achievements related with the ground mobile inspection system are presented: (1) Autonomous navigation localization and planning algorithms to manage navigation in huge extensions and (2) Non-Destructive Inspection operations: thermography based detection algorithms to provide automatic inspection abilities to the robots.

INTRODUCTION

MAINBOT project has developed service robots applications to autonomously execute inspection tasks in extensive industrial plants on equipment that is arranged horizontally (using ground robots) or vertically (climbing robots). MAINBOT has used already available ground robot and arm and develop a new climbing robot to deploy innovative solutions in order to fulfil project industrial objectives: to provide a means to help measuring several physical parameters in multiple points by autonomous robots able to navigate and climb structures, handling sensors or special non-destructive testing equipment.

Robots are being used in different maintenance applications: in power distribution line monitoring [1], nuclear power plants inspection [2], pipes inspection [3] or underwater pipes inspection [4]. In general, they are ad-hoc mechatronic solutions. There are also some general purpose wheeled platforms, such as [5] for offshore inspection, [6] for remote dangerous area inspection or [7] used in Fukushima. They are usually remotely controlled.

To define the requirements of this type of industries two validation scenarios were selected: a Parabolic Through collector technology (PT) solar plant (50Mw, seven hours Molten salts Thermal Storage) and a Central Receiver technology (CR) solar plant (19.9 Mw, 15 hours of molten salts TS capacity). Both plants pose strong challenges in

SolarPACES 2015
AIP Conf. Proc. 1734, 130013-1–130013-8; doi: 10.1063/1.4949223
Published by AIP Publishing. 978-0-7354-1386-3/$30.00

terms of the number of elements to be inspected, the size of the elements, the working conditions, etc. Some figures can present an idea of the magnitude of the problem in extensive plants:

- 400.000 mirrors, with a total of 1.200.000 m2 of surface in PT.
- About 90km of absorber tubes to be inspected (180 km) in PT.
- 2.650 heliostats (10 meters high and 11 meters width) composed of 35 facets in CR.
- A tower of 140 m with receiver tubes area of 11m height and 8m diameter at 120m above the ground

Based on the selection criteria (positive impact in plant, novelty, feasibility, risk) several operations to be performed autonomously by the robots were selected.

- **Ubiquitous sensing.** Measurement of reflectivity is done manually by operators using a special purpose sensor, the reflectometer. A global field reflectivity index is obtained statistically using specific measurements in selected mirrors of the solar field. In the project, the ground robot places a reflectometer on the points of the SCE specified by the plant Operator, touching the mirrors and recording data.
- **Leakage detection**. In PT plants, Heat Transfer Fluid (HTF) circulates at high temperature (around 390ºC) inside the absorber tubes. HTF leakages are no desirables because oil must be replaced and this operation needs to put the SCE's out of service for several hours. Robots using thermography inspection techniques can perform this detection.
- **Surface defects detection in vertical structures**. In CR plants a receiver located at the top of a tower heats molten salts. The receiver is a polyhedral structure composed of several panels of pipes. Receiver pipes have an external coating in order to improve radiation absorption. This coating has a thickness of microns. The climbing robot moves on top of those panels performing eddy current inspection, to assess the status of the coating by measuring its thickness. Moreover, a visual camera records external surface to detect loss of coating.
- **Surface defects detection in horizontal structures**. Ground robots in the plant look for broken mirrors since early detection can contribute to improve this efficiency. In addition, the ground robot patrolling at night and using thermography inspection can be used to identify any kind of loss of vacuum in receiver tubes.
- **Internal defects detection**. Detection of corrosion and internal defects in general (cracks, etc.) is required in many components in a power plant. The climbing robot can test the presence of this kind of possible defects in the collector tubes.

In order to validate the results achieved, several qualitatively and quantitatively metrics were defined and used during the experimental test.

Experiments related to the abovementioned operations have been performed by the robotic platform in a thermal solar plant using Parabolic Through technology in the south of Spain (VALLE 1 and 2, property of Torresol Energy Investment) and are described below in this paper, as well as the evaluation of navigation algorithms.

The results of the climbing robot are described in [8].

Ground Robotic Platform Overview

The ground platform is compound of two main elements: a wheeled robot based on the robucarTT platform and a 6 DoF robuArm robotic arm mounted on it. Both elements have been based on commercial-off-the-shelf products developed by Robosoft, one of the participants in the project. The modifications have included the modification of the battery system to increase its autonomy and the inclusion of a DGPS system to obtain a more accurate localization system.

The robucarTT platform uses a hidroneumatic damping system that absorbs high and low frequency vibrations, making it suitable for outdoor use. It has an Ackerman configuration with 4 driving electrical wheels that can be controlled independently. The robuArm mounted on the platform has 2,5kg payload and 1mm repeatability. Laser rangefinders, ultrasound sensors are used to ensure safety.

The software to control the mobile manipulator has been developed using the Robotic Operating System (ROS).

The robotic arm manipulates the two sensing devices used for non-destructive testing:

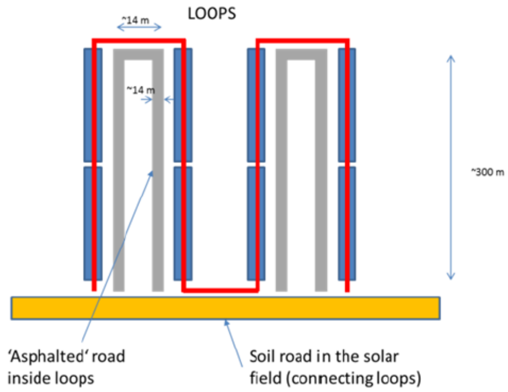
- A reflectometer
- A thermographic camera

EVALUATION OF NAVIGATION INSIDE A LOOP

In this experiment the robotic platform had to navigate autonomously inside a loop at the Valle facilities detecting any possible obstacle and maintaining a predefined distance with respect to the mirrors.

The navigation is mainly done at night, but the experiments were conducted during the day.

(a)



(b)

FIGURE 1. Interior of a loop (a) and scheme of loops (b)

The platform moved at a target speed of 0.65m/s when entering in the loop and a target speed of 0.5m/s while moving straight alongside the mirrors. This lower speed was required to obtain a more accurate final positioning.

During the navigation, all available data from the platform was recorded for later analysis. The platform repeated the maneuver five times inside a loop and four different areas have been considered: the loop entrance, alongside the mirrors, the hairpin turn and the loop exit.

The length of the trajectory performed by the robot and the time required to do it was registered. Using both values, the mean speed was calculated. The efficiency in time is measured as the difference between the theoretical time needed to complete the path given a target velocity and the real time employed.

The error between the final goal pose and the one measured by the DGPS system is used to measure the accuracy.

As it can be seen in the table below for the movement inside a loop the trajectories executed and real velocity are very close to those planned.

During the experiment the platform completed the goals all times it was sent to navigate inside the loop. The table below summarises the accuracy in the goal and the time needed to reach the goal.

TABLE 1. Errors during the navigation inside a loop

Navigation area	Planned path length (m)	Performed path length (m)	Duration (s)	Target_speed (m/s)	Mean speed (m/s)	Time diff (s)
Loop entrance	22,28	22,29	31,64	0,65	0,70	-2,2
Alongside the mirrors	289,58	289,81	585,08	0,5	0,49	-4,3
Hairpin turn	23,77	24,09	33,5	0,65	0,71	-1,9
Loop exit	14,58	14,59	23,42	0,65	0,62	-1,6

EVALUATION OF NAVIGATION OUTSIDE A LOOP

Three types of tests were performed:

- Transition from one loop exit to the adjacent loop entrance. During the transition from one loop to another the robot must move through an unpaved road
- Navigation through the solar plant It includes the transition from one terrace to another, i.e. climbing steep slopes through an irregular terrain, crossing bridges as in the picture below, etc.
- Taxing from the robot park area to the solar field (to the nearest loop entrance). It includes some paved path used by the staff to access to some points of the plant, although most of the navigation is done outside this asphalted road as it can be seen in Figure 2.

The platform had to navigate autonomously five times in each area, trying to reach a target position, meanwhile all data available was recorded for offline analysis of the error in X, Y position, the Yaw error and the time difference with the theoretical value.

(a)

(b)

FIGURE 2. Robot navigating form one loop to another (a) and navigating in the plant (b)

The robot completed the goal successfully all times.

The table below summarises the accuracy on average in the goal and the time needed to reach the goal in the case of taxing. The errors are mainly due to the position algorithm that tries to optimize the time to converge.

TABLE 2. Errors during the navigation taxing from car park to the loop

Goal error in X (m)	Goal error in Y (m)	Goal error in Yaw (rad)	Time diff (s)
0,1398	0,486	0,1788	-10,377

VACUUM LOSS DETECTION

HTF circulates inside the absorber tube, which is compound by an inner metallic pipe and an outer glass cover. Between these two elements there is vacuum. Broken glass and vacuum loss lead to a heat loss and as a consequence to a reduction of plant efficiency. MAINBOT proposes the ground robot patrolling at night and using thermography inspection to identify this kind of problems.

Two different experiments have been performed to test vacuum loss algorithms, based on the information provided by the thermography camera.

In the first one, they were analyzed thermal images corresponding to 13 km of tubes from Valle 1 and Valle2. The images were acquired at night (from 23:00 pm to 4:30 am) from a car traveling at 20 Km/h.

In the second one, the analyzed data was acquired from the camera mounted on the robot arm at the target speed. It was selected a loop in which vacuum loss presence was known: six vacuum losses, three of them in consecutive tubes. The arm was positioned so that the tube remained horizontally in the center of the thermal image. The robot travelled 3 times alongside the loop at 0.5 m/s (1.8 km/h) speed. The operation was done at daylight and at night.

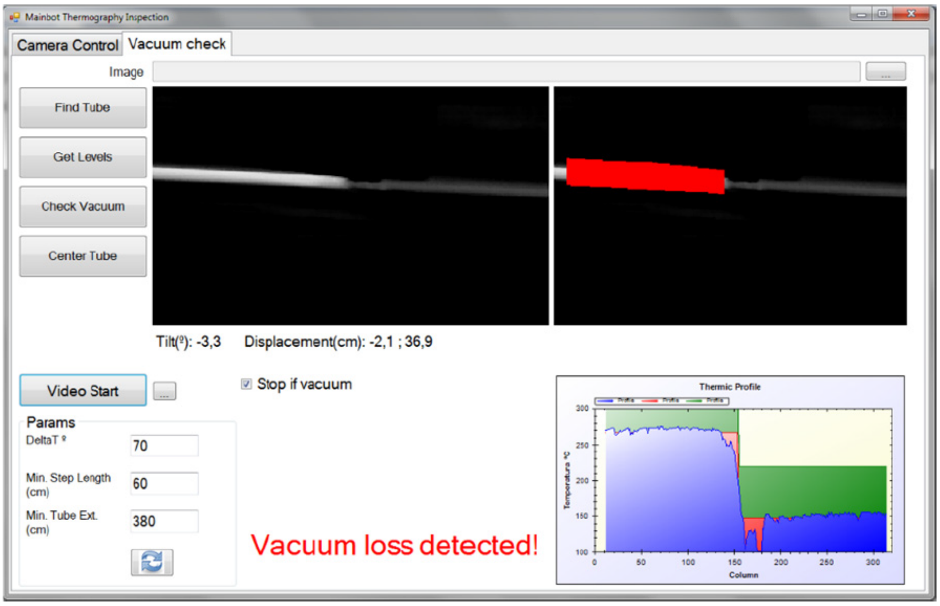


FIGURE 3. Application developed for vacuum loss detection

Figure 3 shows the application developed to automatically identify the vacuum loss, the thermal image, the area with the problem and the gradient of temperature.

The algorithm detected all of the vacuum losses present without giving false positives or false negatives, obtaining a detection ratio of 100%.

To the success of the operation, the thermography camera had to be maintained at a proper orientation and distance with respect the object to be inspected, in order to maintain the target object in the field of view of the camera. Several tests were successfully carried out as described in [9].

BROKEN MIRRORS DETECTION

The heuristic used to detect broken mirrors is based on the analysis of thermal images, using the background as temperature reference.

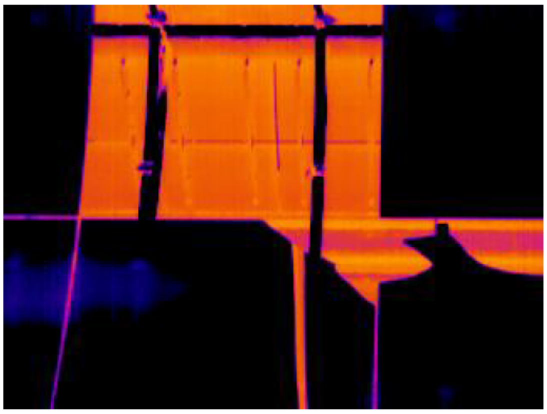
The developed algorithm is based on finding the facets present in a thermogram and analyzing them (other strategies, such as tracking the border of each mirror were discarded due to the huge amount of time required to inspect all the mirrors in the plant.). The analysis includes:

- To convert the thermal information present in the thermogram into gray scale image.
- A Hough lines algorithm is applied to find each of the facets present.
- With each facet a dynamic threshold is applied.
- Finally the resulting blobs are filtered based on area dimension, shape and circularity criteria.

Due to the reflective feature of mirrors, the thermal camera receives the reflection from the mirror in the direction of the camera and the radiation reflected by the neighbor elements (facets, tubes and other structural components). These reflections are perceived as an element with a higher temperature.

Taking this into account, the algorithms has to follow different approaches depending on the position of the facets: facets on top of the SCE have the sky as background meanwhile facets on the lower part of the SCE have not.

In the first case (background sky) both the detection of each facet and the search for breakages is performed with high detection rate because of the great contrast between the surface of the mirror and the rest of the scene. Next picture shows an example of the image obtained from the thermal camera and the result provided by the application. In this case, the breakages in the mirror are seen as part of the rest of the scene, and a geometrical discontinuity in the facets provides the information to detect the breakage.



(a)

(b)

FIGURE 4. Thermal image of a broken facet (a) and application for automatic detection (b)

In order to test the algorithm, two different thermal sequences of a loop were acquired (upper part and lower part of the SCEs). It is noteworthy that breakages are more frequent in the lower part of the SCEs.

The results are quite different depending on the mirror position. In upper facets breaks with an area greater than 90cm^2 from a distance of about 12 meters were found. In the case of lower facets in the SCE, the algorithm is able to detect breakages of the same area but with an horizontal length longer than 8cm and a circularity greater than 0,2, to avoid typical structural reflections present in the mirrors, assuming that breakages present a proportional aspect (based on experience).

LEAKAGES DETECTION

Heat Transfer Fluid (HTF) circulates at high temperature (around 390ºC) inside the absorber tubes. HTF leakages are not desirable because the repairing implies putting the SCE out of service during a long period of time. Robots using thermography inspection techniques are proposed to detect this problem.

Leakages can be detected with infrared cameras sensitive in the wavelength where the gas has absorption peaks. With the appropriate infrared camera, leakages can be perceived as steam. Taking this into account it has been implemented an algorithm based on optical flow which is able to detect the presence of steam in a scene.

The validation of this algorithm is only possible using a camera in the range of the HTF emission. This type of cameras are very expensive and were not available. As a consequence, it was decided to test the algorithm with a different gas. An experiment was done using a vaporizer. The camera focused the output of steam of the vaporizer. The objective was to detect the movement of steam particles.

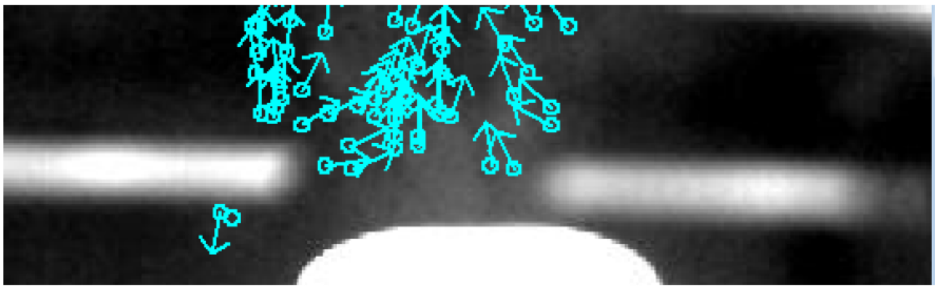


FIGURE 5. Optical flow used to detect steam

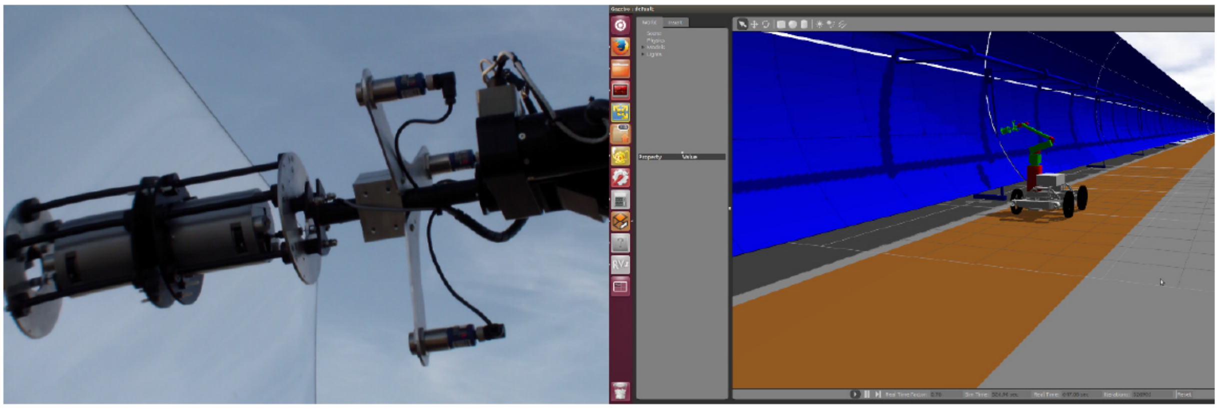
Once the algorithm was verified with the vaporizer and knowing that without mid-wave infrared sensitive equipment it was not possible to validate with HTF gas, it was found a non visible innocuous gas to emulate the ral situation more realistically. The chosen gas was acetic acid that is visible with the available long-wave thermal camera at ambient temperature.

In the test performed in laboratory conditions, the presence of steam was successfully detected.

REFLECTIVITY MONITORING

The reflectivity index of the plant is a parameter of paramount importance in order to establish an optimal cleaning policy. Measurement of reflectivity is currently done manually by operators using a reflectometer. The MAINBOT project has tested the ground robot placing the same reflectometer on the points of the SCE that TORRESOL uses as reference and recording the acquired values.

The objective of this experiment was to doublecheck the accurate and safe placement of the reflectometer on the surface of the mirror. To achieve this objective, the platform navigates up to the mirror to be inspected, the reflectometer toolholder is positioned parallel to the mirror and the approach manoeuvre is performed until the sensor touches the mirror. The signals provided by 3 ultrasound sensors in the toolholder are used for trajectory control.



(a) (b)

FIGURE 6. Robot placing the reflectometer on the mirror (a) and simulation environment (b)

During the experiments the sensor was successfully positioned on the mirror, without any kind of damage, neither on the mirrors nor on the reflectometer. The time required for reflectivity measurement was about 7min for each SCE in the experimental conditions.

As all loops have the same layout, it was possible to design an optimized navigation algorithm to reach the target measurement positions. The procedure is as follows:

- To setup the input point of the loop. The robot is placed at that target point and the position and orientation provided by the DGPS is registered.
- Calculation of the five inspection positions of the ground robot in a SCE.
- To establish the configuration of the manipulator to perform the measurement on the mirror.

Once the target measurement points are defined the reflectivity measurements operation is autonomously performed: the Robot moves to the initial position in loop, then the robot navigates to the first measurement point and places the tool over the mirror. The procedure is repeated along the rest of SCEs. Once the measurements are performed, the robot exits the loop to continue with the next one.

CONCLUSIONS

By means of the experiments, all selected operations have been validated and the feasibility of automated inspection operations has been demonstrated.

Thermography technology is a very useful tool in extensive plants inspection due to three main reasons.

- It is a non-contact technique
- It allows affording many different kind of problems: vacuum loss, surface defects and leakages.
- It provides rich data that can be processed very quickly.

In the case of **vacuum loss** high detection ratio (100%) has been achieved and the detection algorithm has been automated so that it has been integrated in the robot control.

Broken mirrors detection has been also solved with thermal analysis, however during final validation at Valle results were different depending on the position of the facets. Better performance was achieved in the upper part of the SCEs.

Leakages detection has been validated in laboratory conditions (it was not possible to validate it in a solar plant due to the mismatch between the absorption spectrum of the HTF and the camera available for the experiments) implementing optical flow algorithms that can detect vapor particles movement in the images.

Reflectivity measurement has been performed using the manual reflectometer (the most widely used in this kind of facilities). The key problem (positioning of the sensor on top of a fragile surface, i.e. the mirror) has been validated. A real implementation will demand a different approach to include the reading of the sensor values to close the positioning loop and a procedure for periodic re-calibration of the sensor.

From the **navigation** point of view, the combination of the implemented planners (SBPL, line follower, and hairpin) has provided efficient trajectories. On the other hand the robot was able to successfully deal with the rough terrain encountered at the end user facilities (slopes and irregular unpaved paths in some areas).

There is an additional obvious use for robots in this kind of plants, i.e. cleaning of mirrors. Although this was out of the scope of the project (focused on inspection), it would be of great interest to combine both applications, inspection and cleaning, in the same robot that would need to be of a complete different morphology.

ACKNOWLEDGMENTS

This work has been partially funded by the European Community"s Seventh Framework Programme (FoF.NMP.2011-3) under grant agreement no 285245.

REFERENCES

1. P. Debenest, M. Guarnieri, K. Takita, E. Fukushima, S. Hirose, K. Tamura y A. Kimura, «Toward a Practical Robot for Inspection of High-Voltage Lines,» de In Proceedings of the Field and Service Robotics (FSR), Cambridge, 2009.
2. J. Lee, B. Cho, K. Jang, S. Jung, K. Oh, J. Park y J. Kim, «Development of Autonomous Cable Inspection Robot for Nuclear Power Plant,» de In Proceedings of World Academy of Science, Engineering and Technology, Rome, 2010.
3. M. Suzuki, T. Yukawa, Y. Satoh y H. Okano, «Mechanisms of Autonomous Pipe-Surface Inspection Robot with Magnetic Elements,» de In Proceedings of the IEEE International Conference on Systems,Man and Cybernetics, Taipei, 2006.
4. C.Camerini, M. Freitas, R. Langer, J. von Der Weid, R. Marnet y A. Kubrusly, «A Robot for Offshore Pipeline Inspection,» de In Proceedings of the 9th IEEE/IAS International Conference on Industry Applications, Brazil, 2010.
5. B. a. P. K. a. S. H. Graf, «Mobile Robots for Offshore Inspection and Manipulation,» de IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009.
6. Sensabot, «Sensabot,» [En línea]. Available: http://www.rec.ri.cmu.edu/projects/shell/.
7. Quince, [online]. Available: http://www.jsme.or.jp/English/awards/awardn12-2.pdf.
8. Torsten Felsch, Gunnar Strauss, Carmen Perez, José M. Rego, Iñaki Maurtua, Loreto Susperregi and Jorge R. Rodríguez, Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques. Robotics 2015, Volume 4, Issue2, p103-119
9. Aitor Ibarguren, Jorge Molina, Loreto Susperregi and Iñaki Maurtua. Thermal Tracking in Mobile Robots for Leak Inspection Activities. Sensors 2013, 13(10), 13560-13574; doi:10.3390/s131013560, 9 October 201.

7.13. Use of machine vision in collaborative robotics: An industrial case

Alberto Tellaeche, Iñaki Maurtua, Aitor Ibarguren: Use of machine vision in collaborative robotics: An industrial case. ETFA 2016: 1-6 [100]

Use of machine vision in collaborative robotics: An industrial case.

Alberto Tellaeche, Iñaki Maurtua, Aitor Ibarguren
IK4 - TEKNIKER
Parke Teknologikoa
C/ Iñaki Goenaga, 5 - 20600 Eibar, Spain
Email: {alberto.tellaeche, inaki.maurtua, aitor.ibarguren}@tekniker.es

***Abstract*—Robotic applications are evolving to a paradigm of collaborative robotics, where human workers and compliant robots work together to solve complex tasks, until now done fully manually. These tasks also present another challenging issue for the use of robotics, variability in part and tool positions, in robot placement and even in the targets for the robot operation. To comply with all this uncertainties while solving the work efficiently, the robot needs to be equipped with sensors that allow it perceive the environment. For this, machine vision techniques in all its variants (2D, 3D, point clouds) becomes fundamental.**

This paper outlines a real industrial case of collaborative robotics, and the details of use of machine vision techniques to cope with variability and uncertainties. The industrial case presented has been developed as part of the EuRoC European project, under the 7th European Framework.

Keywords—collaborative robotics; machine vision; human robot interaction.

I. Introduction

In modern industrial robotics, the safe and flexible cooperation between robots and human operators can be a new way to achieve higher productivity when performing complex tasks. Introducing robots within real industrial settings makes the interaction between humans and robots gain further relevance. The problem of robots in these setups poses important general challenges: robots must be able to perform tasks in complex, unstructured environments, that present high variability and uncertainties in positions of parts, tools and workers, while assuring the safety of the humans in the working environment.

To assess the security in the working area, state of the art human robot interaction takes advantage of the tools provided by the artificial intelligence, to solve tasks such as human detection, motion planning, workspace reconstruction or compliant behavior using force control.

978-1-5090-1314-2/16/$31.00 ©2016 IEEE

For the second problem, i.e., the variability in an unstructured environment, machine vision becomes fundamental to cope with the uncertainties in positions of tools and parts needed to carry out the industrial task. Depending on the problems to solve, different machine vision techniques are applicable, such as point cloud processing to correct 3D poses and matching of pieces, depth information used to pick objects or 2D vision for color processing and segmentation.

The present article describes the work done related to machine vision and perception, within the *EuRoC* european project (7th European Framework), more specifically in the *Challenge 1:Reconfigurable Interactive Manufacturing Cell.* This challenge has presented a benchmarking task focused on the assembly of a car door, presenting the following problems associated to flexible manufacturing tasks:

- Need of adaptive perception and cognition skills to cope with illumination changes and position tolerances.
- Robust assembly capabilities to cope with part tolerances.
- Safe human robot interaction in small workspaces

This paper is organized as follows, section 2 presents the current state of the art in perception related to robotic applications, section 3 explains more in depth and technically the cell layout and the assembling problem proposed in the *EuRoC Challenge*. The developments for solving the perception tasks are explained in section 4 and, finally, section 5 presents the results achieved and the conclusions that can be extracted from them.

II. State of the Art

Machine vision and perception play a key role in many robotic applications, both for performing automated and fixed industrial operations such as measuring or quality control testing, and for more advanced tasks like pick and place operations, object detection, etc.

With the appearance of compliant robots, that make possible and safe the human robot interaction, and the flexibility of the new robotic cells, perception of the environment has become even of further importance.

Among all the different tasks and problems that require the use of cameras in the robotic cells under the Industry 4.0 paradigm, in the specific case of assembling and manufacturing tasks, object recognition and reconstruction, and tolerance estimation for manipulation are the main problems that arise in any layout.

Object recognition is not a trivial task and many approaches are currently being tested depending on the availability of sensors (source of data) and possibility of scene reconstruction (placement of the sensors, possibility of movement, etc).

In [1], dense SLAM techniques are used for object recognition and pose estimation, however, this approach is not feasible in industrial tasks because it requires many different views from the same scene and the reconstruction on the environment is very demanding computationally.

The works presented in [2][3] and [4] establish different techniques and approaches for 3D object recognition and fetching using robots. As common points, all of them use CAD model based object recognition and 3D cameras for scene retrieval.

The use of point clouds and optimization algorithms, such as maximum likelihood estimation for object detection is presented in [5]. Using the surface features of the scene detection, it is possible to provide a solution fast enough to use with manipulation tasks.

Reference [6] uses also point cloud processing and global features extracted from the cloud, while in [7] a local descriptor is used to recognize and match partially occluded objects using CAD models.

As it can be concluded from the different approaches, object recognition and matching presents difficulties in processing time and precision of results that must be solved in an optimum way depending on the application and the task to be performed. In addition, industrial tasks present requirements of robustness and speed that make this operation even more challenging.

III. Problem Description

The present section explains technically the flexible cell layout provided by the Fraunhofer IPA Institute for the *Challenge 1* benchmarking, within the *EuRoC* European project [8].

The task to solve is the automatic assembling of a car door, i.e. placing the inner control unit (see Fig. 1) on the door frame (see Fig. 2), including the placement and screwing. The accomplishment of this operation presents some partial problems, these are:

1. Location of the position of the inner plastic module of the door: This module is placed on a table in an unknown position. The part must be picked up by the robot automatically, no matter the initial position it may present.

Fig. 1. Inner module of the door and tolerance marks in its position

2. Position estimation of the metallic door frame: The main metallic door frame is hold by a vertical structure, but its position may also present variability. Precise pose estimation of the door is needed to correct the initial position of the force control based insertion procedure of the inner module.

Fig. 2. Detail of the door frame and the plastic module once inserted

3. Screwing task: Once the inner module is inserted in the door frame, it is necessary to screw it to finish the operation. The screws are placed in a box, also presenting rotation tolerances in its position (see Fig. 3). The screws must be precisely located to pick them up with the screw driver tool.

Fig. 3. Detail of the screw box

The robot available to carry out all these operations is a UR10 [9] from Universal Robots, equipped with a Robotiq FT150 force sensor [10], and a tool interchange, vacuum and screw driver. From the machine vision point of view the robot has a VRmagic D3 stereo camera [11], and a PMD nano time of flight (ToF) camera [12]. The mounted cameras can be seen in Fig. 4.

Finally, to interact with the tools and force sensor, a digital I/O Phidgets board is available.

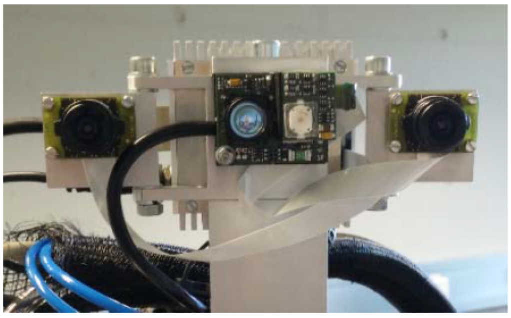


Fig. 4. Machine Vision systems mounted on the robot

The overall system architecture is based on the Robot Operating System, ROS [13], the current standard for research robotics applications. The long term supported Indigo version has been used in this research. A core PC provides all the data coming from the sensors as topics, and a full *tf* tree publishing the transformations of all the sensors and robot joints. Part of the *tf* tree showing the transformations of the cameras with respect to the robot flange is presented in Fig. 5:

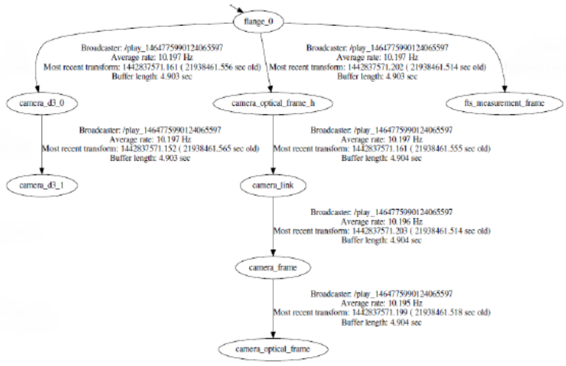


Fig. 5. *tf* tree corresponfing to the machine vision systems mounted on the UR10

The cell layout for the assembling operation can be seen in Fig. 6:

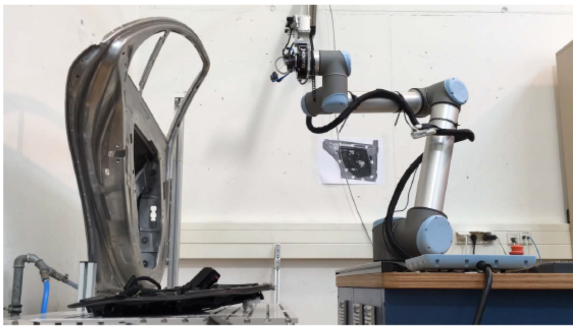


Fig. 6. Overall aspect of the proposed assembling cell.

Section 4 explains how the problems presented in this section have been solved using machine vision techniques, and the overall procedure to solve the complete assembling operation.

IV. Task solving procedures using machine vision

From the machine vision point of view, there are different problems of different nature to solve. Depending on the problem different techniques have been applied, such as, point cloud processing, 3D stereo reconstruction, camera calibration or model shape matching based on CAD.

Always under ROS environment, two main image processing libraries have been used: Point Cloud Library (PCL) [14], and MvTec Halcon 12.0 [15].

A. *Inner door black module position estimation.*

This task demands the precise position calculation of the module that has to be inserted, placed in an unknown location on the table. The main challenge this problem presents is that although the module has an irregular interior geometry, its black color makes not possible the stereo reconstruction of its surface.

The algorithm developed uses the point cloud given by the PMD nano ToF camera to identify precisely the white parts inside the module, the only sections inside the module that are possible to scan with the camera. The rest of the geometry does not provide any data because of its black color.

In the teaching phase, a nominal position has been defined, to refer the picking operation of the robot to that position. This nominal position is recorded as two points, (x_1,y_1,z_1) and (x_2,y_2,z_2), coinciding with the positions of the white parts in the default position of the module.

In addition, two nominal robot positions have been defined to start the search of the two white parts, one for each part. The following pseudo-code explains how the position of each of the white parts is detected:

1. To go to a nominal position (1st or 2nd white part)
2. To obtain point cloud from ToF camera.
3. TO process point cloud to extract parameters (Filtering, Clustering, obtain size, obtain gravity center)

4. To check point cloud parameters.
5. If parameters within limits:
   a. White part located, return gravity center
6. Else:
   a. Move a step in an ellipsoidal movement (search movement)
   b. Go to step 2

When actual positions of both white modules have been obtained, translation and rotation of the module can be estimated by the difference (Euclidean distance) of the current points to the nominal ones. As the inner black module is placed on a table, the robot inspects the white modules at a fixed distance z above the table, the correction of the module position is then planar.

The processing of the point cloud uses several modules of the PCL library, such as outlier and spatial filtering.

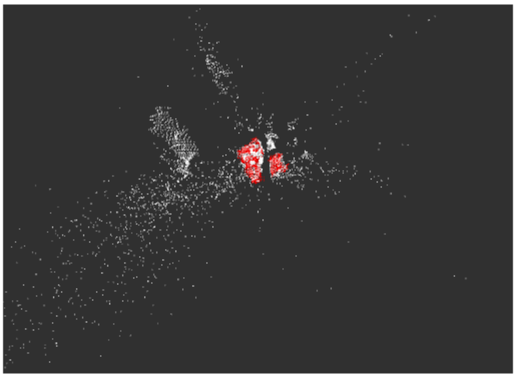


Fig. 7. Original point cloud from the ToF camera (in white), inner white module segmented and located element (in red)

B. Door reconstruction and position correction

Due to the size of the car door, the reconstruction of the full door must be accomplished by registering different partial views of the door obtained by stereo vision.

The partial views are obtained by using the stereo vision system provided by the VRmagic 3D camera, and using the ROS node *stereo_image_proc* to deal with parameters and reconstruct the stereo pair using the *Semi Global Block Matching* (SGBM) algorithm [16]. This algorithm offers much better precision and a more dense reconstruction than the less refined *Block Matching* (BM) algorithm.

Each of the partial views obtained are then transformed from the original *left_camera_frame* to the *ur_base* frame, and registered together to form the complete point cloud corresponding to the car door. The complete registered point cloud is then further processed for spatial filtering and downsampling.

This procedure is done twice. The first scanning is needed to store the point cloud of the door in its nominal position. This point cloud is then stored as a PCD format file model for later pose correction.

The second scanning is performed during the assembling process. The robot scans the door from the same previously recorded points following the described method. With the door scanned during run time, and the nominal model previously loaded a 3D robust matching process based on shape matching algorithms is carried out, obtaining the correction of the current pose of the door with respect to the nominal pose. The correction is expressed as an affine transformation matrix.

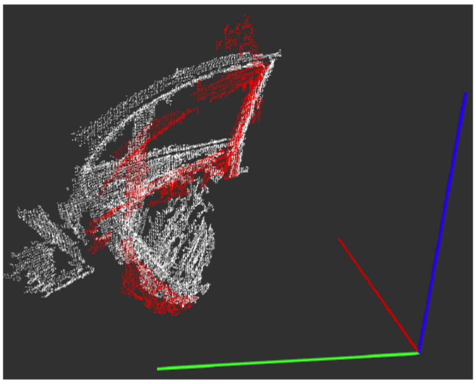


Fig. 8. Door in nominal postion (in white), moved door within tolerance (in red).

C. Screw detection

The last task where the machine vision has relevant importance is the screw detection when it is necessary to pick them up with the screw driver. In this case 2D vision and a CAD based shape model matching has been used to detect precisely the heads of the screws. Due to the fact that the detection must be very precise, the matching threshold for the algorithm has been set to a 90%, only returning good enough matches.

The camera used for this operation is the left camera of the stereo pair, and the coordinates obtained are directly connected to the tool, i.e., the screwdriver, through the *tf* tree in ROS.

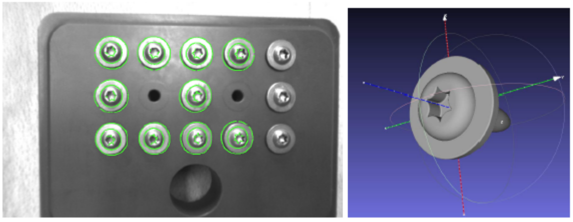


Fig. 9. Screws detected in the 2D rectified image, and STL model of the screws, used for detection.

D. Sequence of the overall assembling operation

The whole assembling operation of the door can be divided in two basic steps:

1. Teaching and configuration stage.
2. Assembling stage.

1) *Teaching and configuration stage*

In this stage, the moving robot base is fixed in a position and nominal robot positions are recorded for the later scanning processes, these are: two nominal points to detect and record the position of the inner module, ten points to reconstruct the point cloud of the door and one more point to match the screws. These points are recorded for posterior assembling operations, and a point cloud model of the door is stored in the nominal position. Other points are also recorded for the screwing operations once the module is inserted in the door.

With the recording of the robot points and the storage of the door model, the configuration stage is finished.

2) *Assembling stage*

In this stage, the inner module of the door, the door frame and the screws are placed within an area but in an unknown position to complete the assembling process, the robot follows the next sequence:

1. The robot scans the door frame without any tool to correct the actual door frame position with respect to the nominal position.
2. The robot mounts the vacuum tool to pick the inner part and scans the white parts on the inner module starting from the two nominal points previously recorded. Once the white parts are located, correction is applied to the picking position and the module is picked up.
3. The robot corrects the starting pose for module insertion, using the correction obtained in step 1. The module is inserted in the door applying a force-control algorithm.
4. Tool change leaving vacuum system and taking the screwdriver.
5. The robot detects the real positions of the screws using the 2D shape matching and picks the best match.
6. The robot corrects the screwing point using the correction value obtained in step 1, and inserts the screw. Steps 5 and 6 are done several times until all the screws are inserted.

V. Obtained Results and Conclusions

The assembling process has been tested and evaluated in the EuRoC Challenge 1 benchmarking phase. The process has been divided in different subtasks:

1. module grasping (task 1.1)
2. module insertion (task 1.2)
3. screwing (pick three screws and screw them in 3 positions out of 9, selected in advance)(task 2).

Each subtask has been executed three times, to assess the robustness of the method. The following table summarizes the results obtained in the final official evaluation:

TABLE I. Assembling task results.

Task number	Task results		
	Completed cycles (out of 3)	*Best Time (s)*	*Comments*
Task 1.1	2	452 s	Not controlled force error in the UR 10
Task 1.2	3		
Task 2	2	813 s	One screw insertion failed in second run of Task 2

Apart from the results presented in the previous table, several conclusions and statistics can be extracted for each of the tasks, taking into account the tests performed during the development.

A. *Inner module grasping.*

The accuracy of the picking operation of the module has proven to be robust. Translation corrections present less than 1cm of error and rotations less than 5 degrees of error in more than 90% of the operations.

The force-control algorithm supports bigger errors in translations and rotations, so the results provided by the correction algorithm for module grasping are perfectly valuable.

B. *Module insertion.*

Due to the high error tolerances that the force-based algorithm supports for module insertion, and the correct results provided by the grasping module correction, the module has been correctly inserted in the door frame in the 96% of the tests performed. Time for insertion varies from the 100 s in the best cases (very precise module grasping correction) to the more than 300 s when the initial position is not so accurate. These long times would be the principal aspect to strengthen in a real industrial setup.

C. *Screwing task.*

This task can be divided in screw picking and screwing.

The screw picking task is very precise, picking up the screw correctly with an accuracy of over 97% of the times.

The screwing task depends of the position correction of the door frame. If the angle correction presents an error of less than 5 degrees of error, the screwing operation in correctly finished more than 92% of the times. In the position angle presents a higher error, only a 60% of success is achieved.

D. *Final conclusions.*

The overall method has proven to be robust enough performing the task, but it is still quite slow for an industrial application. This is the main point that needs to be improved.

In the context of the *EuRoC* project, this approach, developed by the *PIROS* team (IK4-Tekniker and CNR-ITIA) ended up winning the benchmarking competition.

Acknowledgments

EuRoC: European Robotics Challenges" is a project founded by the European Commission in the same 7th Framework Programme (FP7-608849). Its website for further information is http://www.euroc-project.eu/.

IK4-Tekniker and CNR-ITIA would like to thank Fraunhofer IPA the help, the equipment and the setup for performing the tasks in the benchmarking phase of the project.

References

[1] K. Tateno, F. Tombari and N. Navab, "When 2.5D is not enough: Simultaneous Reconstruction, Segmentation and Recognition on dense SLAM" . 2016 IEEE International Conference on Robotics and Automation (ICRA), May 2016.

[2] R.C. Luo and C.W. Kuo, "Intelligent Seven-DoF Robot With Dynamic Obstacle Avoidance and 3-D Object Recognition for Industrial Cyber–Physical Systems in Manufacturing Automation". Proceedings of the IEEE, vol. 104, pp. 1102-1113, April 2016

[3] R.C. Luo and C.W. Kuo, "A scalable modular architecture of 3D object acquisition for manufacturing automation". 2015 IEEE 13th International Conference on Industrial Informatics (INDIN) , pp. 269-274, July 2015.

[4] R.C. Luo, C.W. Kuo and Y.T. Chung, "Model-based 3D object recognition and fetching by a 7-DoF robot with online obstacle avoidance for factory automation" 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 2647-2652, May 2015.

[5] H.G. Dantanarayama and J.M. Huntley, "Object recognition in 3D point clouds with maximum likelihood estimation", Proceedings. SPIE 9530, Automated Visual Inspection and Machine Vision, June 2015.

[6] K. Alhamzi, M. Elmogy and S. Barakat, "3D Object Recognition Based on Local and Global Features Using Point Cloud Library" International Journal of Advancements in Computing Technology (IJACT), vol.7, Number 3, May 2015.

[7] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R.B. Rusu and G. Bradski, "CAD-Model Recognition and 6DOF Pose Estimation Using 3D Cues", 2011 IEEE International Conference on Computer Vision Workshops, pp. 585-592, Nov. 2011.

[8] http://www.euroc-project.eu/

[9] UR10 robot. http://www.universal-robots.com/es/productos/robot-ur10/

[10] Robotiq Force Torque Sensors. http://robotiq.com/products/robotics-force-torque-sensor/

[11] VRmagic D3 smart camera. https://www.vrmagic.com/imaging/camera-platforms/d3-intelligent-platform/

[12] PMD nano Time of Flight camera. http://www.pmdtec.com/products_services/reference_design.php

[13] Robot Operating System. http://www.ros.org/

[14] Point Cloud Library. http://pointclouds.org/

[15] MvTec Halcon Library. http://www.halcon.com/halcon/version12/

[16] E. Dall'Asta, R. Roncella, "A comparison of Semiglobal and Local Dense Matching Algorithms For Surface Reconstruction". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-5, June 2014.

7.14. Interacting with collaborative robots in industrial environments: A semantic approach

Iñaki Maurtua, Izaskun Fernandez, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibarguren: Interacting with collaborative robots in industrial environments: A semantic approach. ICAPS 2016 Workshop on "Planning, Scheduling and Dependability in Safe Human-Robot Interactions", 2016, Londres. [112]

Interacting with collaborative robots in industrial environments: A semantic approach

Iñaki Maurtua, Izaskun Fernandez, Johan Kildal, Loreto Susperregi, Alberto Tellaeche, Aitor Ibargure
{inaki.maurtua, izaskun.fernandez, johan.kildal, loreto.susperregi, alberto.tellaeche, aitor.ibarguren}@tekniker.es
IK4-TEKNIKER
Iñaki Goenaga 5, 20600 Eibar
Basque Country, Spain

Abstract

This paper presents a semantic approach to support multimodal interactions between humans and industrial robots in real industrial scenarios. This is a generic approach and it can be applied in different industrial scenarios. We explain in detail how to apply it in a specific example scenario and how the semantic technologies help not only with accurate natural request interpretation but also their benefits in terms of system maintenance and scalability.

Introduction

In modern industrial robotics, the safe and flexible cooperation between robots and human operators can be a new way to achieve better productivity when performing complex activities. Introducing robots within real industrial settings makes the interaction between humans and robots gain further relevance. The problem of robots performing tasks in collaboration with humans poses two main challenges: robots must be able to perform tasks in complex, unstructured environments, and at the same time they must be able to interact naturally with the humans they are collaborating with.

The current work is carried out in the context of the *H2020 FourByThree*[1] project, which aims at developping a new generation of modular industrial robotic solutions that are suitable for efficient task execution **in collaboration with humans** in a safe way and are easy to use and program by the factory worker. The project will allow system integrators and end-users to develop their own custom robot that best answers to their needs. To achieve it, the project will provide a set of hardware and software components, ranging from low level control to interaction modules. The results will be validated in 4 industrial settings Investment Casting, Aeronautical sector, Machining and metallic part manufacturing, in which relevant applications will be implemented: assembly, deburring, riveting and machine tending in a collaborative context.

A requirement for natural Human-Robot Collaboration including interaction is to endow the robot with the capability to capture, process and understand accurately and robustly requests from a person. Thus, a primary goal for this research is to analyze the natural ways in which a person can interact and communicate with a robot.

[1] `http://fourbythree.eu/`

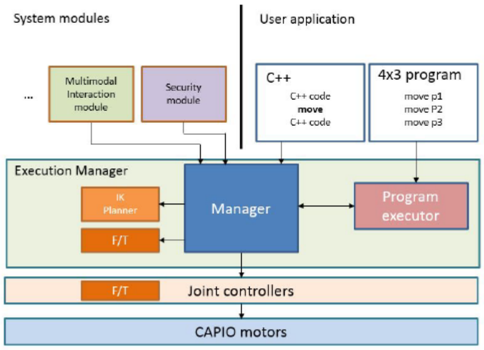


Figure 1: FourByThree project Architecture

Natural communication between humans and robots can happen through several channels, the main of which are voice and gestures. In this multimodal scenario, the information can be complementary between channels, but also redundant. However, redundancy can be beneficial (Bannat et al. 2009) in real industrial scenarios where noise and low lighting conditions are usual environmental challenges that make it difficult for voice and visual signals to be captured with clarity.

In this paper, we present a semantic approach that supports multimodal interaction between humans and industrial robots in real industrial settings that are being studied within the *FourByThree* European project. As mentioned earlier, the approach that we present is generic in the sense that it can be applied to different industrial scenarios by modifying the information about the environment in which communication takes place. After the approach description we introduce a case study corresponding to a real industrial scenario.

Related work

Over the last two decades, a considerable number of robotic systems have been developed showing Human-Robot Interaction (HRI) capabilities (Fong, Illah, and Dautenhahn 2003; Goodrich and Schultz 2007). Though recent robot platforms integrate advanced human-robot interfaces (incorporating body language, gestures, facial expressions, and speech) (R. Stiefelhagen and Waibel 2004; Burger, Ferrane, and Lerasle 2010) their capabilities to understand human speech semantically remains quite limited. To endow a robot with semantic understanding capabilities is a very challenging task. Previous experiences with tour-guide robots (Thrun et al. 1999; Gunhee et al. 2004) show the importance of improving human-robot interaction in order to ease the acceptance of robots by visitors. In Jinny's HRI system (Gunhee et al. 2004), voice input is converted to text strings, which are decomposed into several keyword patterns and a specialized algorithm finds the most probable response for that input. For example, two questions like 'Where is the toilet?' and 'Where can I find the toilet' are equally interpreted since the keyword pattern of 'where' and 'toilet' would be extracted from both cases.

Human-robot natural interactions have also been developed in industrial scenarios. For instance, in (Bannat et al. 2009) the interaction consisted of different input channels such as gaze, soft-buttons and voice. Although the latter constituted the main interaction channel in that use scenario, it was solved by command-word-based recognition.

SHRDLU is an early example of a system that was able to process instructions in natural-language and perform manipulations in a virtual environment (Winograd 1971). Researchers followed on that work towards extending SHRDLU's capabilities into real world environments. Those efforts branched out into tackling various sub-problems, including Natural Language Processing (NLP) and Robotics Systems. Notably, (MacMahon, Stankiewicz, and Kuipers 2006) and (Kollar et al. 2010) developed methods for following route instructions given through natural language. (Tenorth et al. 2010) developed robotic systems capable of inferring and acting upon implicit commands using knowledge databases. A similar knowledge representation was proposed by (Wang and Chen 2011) using semantic representation standards such as the W3C Web Ontology Language (OWL) for describing an indoor environment.

A generic and extensible architecture was described in (Rossi et al. 2013). The case study presented there included gesture and voice recognition, and the evaluation showed that interaction accuracy increased when combining both inputs (91%) instead of using them individually (56% in the case of gestures and 83% for voice). Furthermore, the average time for processing both channels was similar to the time needed for speech processing.

Our work is based on this extensible architecture, combining gesture and speech channels and adding semantic aspects to the processing.

Multimodal Interaction Semantic Approach

The approach proposed in this work aims at creating a human-robot collaborative environment in which interactions between both actors happen in a natural way (understanding by 'natural' the communication based on voice and gestures). We propose a semantic multimodal interpreter prototype that is able to process voice and gesture based natural requests from a person, and combining both inputs to generate an understandable and reliable command for industrial robots. For such a semantic interpretation we have developed four main modules, as shown in Figure2.

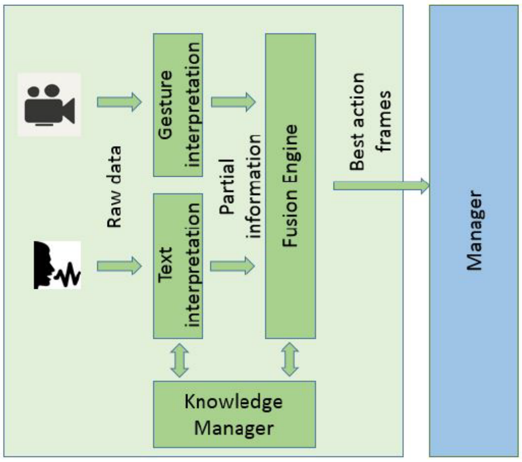


Figure 2: Multimodal Semantic Approach Architecture

A *Knowledge-Manager* module that describes and manages the environment and the actions that are feasible for robots in a certain environment, using semantic representation technologies. A *Text Order Interpretation* module that given a (request) text, it extracts the key elements on the text and translate them to a robot understandable representation, combining NLP and semantic technologies. A *Gesture Interpretation* module mainly for resolving pointing issues and some simple orders like stop an activity. And a *Fusion Engine* for combining the output of both text and gesture modules and construct a complete and reliable order for the robot.

These main modules are described in detail in the following subsections.

Knowledge Manager

The knowledge manager comprises ontologies that model environmental information of the robot itself including its own capabilities. In addition, the knowledge manager allows us to model the relationships between the concepts. These relationships are implicit rules that can be exploited by reasoners in order to infer new information from the ontology. As a result, reasoners can work as rule engines in which human knowledge can be represented as rules or relations.

Ontologies have many practical benefits. They are very reusable flexible at adapting to dynamic changes reuse, thus avoiding to have re-compile the application and its logic whenever a change is needed. Being in the cloud makes ontologies even more reusable, since different robots can exploit them, as was the case with e.g., RoboEarth (Di Marco et al. 2013).

Through ontologies, we model the industrial scenarios in which industrial robots collaborate with humans, in terms of robot behaviors, task/programs they can accomplish and the objects they can manipulate/handle from an interaction point of view. We distinguish two kinds of actions: actions that imply a status change on a robot operation, like *start* or *stop*, and actions related the robot capabilities such as *screw*, *carry*, *deburring* and so on.

Relations between all the concepts are also represented, which adds the ability for disambiguation during execution. This ability is very useful for text interpretation, since different actions can be asked from the robot using the same expression. For instance people can use the expression *remove* to request the robot to *remove this burr*, but also to *remove this screw*, depending on whether the desired action is *deburring* or *unscrew* respectively. If the relationships between the actions and the objects over which the action are performed are known, the text interpretation will be more accurate, since it will be possible to discern in each case to which of both options the expression *remove* corresponds. Without this kind of knowledge representation, this disambiguation problem is far more difficult to solve.

For task/programs we make an automatic semantic extension exploiting wordnet (Gonzalez-Agirre, Laparra, and Rigau 2012) each time the robot is initialized. In this way, we obtain different candidate terms referring to a certain task, which is useful for text interpretation mainly, as it is described bellow.

Text order interpretation

Given as input a human request in which a person indicates the desired action in natural language, the purpose of this module is to understand exactly what the person wants and if it is feasible to generate the necessary information for the robot. The module is divided into two main steps:

- The first step is based on superficial information, in the sense that it does not take into account the meaning of words in the context. Its only purpose is to extract the key elements from the given order.
- The second step attempts to identify the action that is asked for, considering the key elements in the given context.

For the first step, we apply natural language processing techniques using FreeLing, an open source suite of language analysis tools (Padró and Stanilovsky 2012). In particular, we apply a morphosyntactic and dependency parsing to a set of request examples from different people. In this way, we obtain the morphosyntactic information of every element and about the request itself. We revise the complete information manually and identify the most frequent morphosyntactic patterns. From them, we extract elements denoting actions, objects/destinations (target onward) and explicit expressions denoting gestures, such *there*, *that*, and so on. Following, we implement those patterns as rules, obtaining a set of rules that, given a FreeLing tagged sentence, is able to extract the key elements on it. Following, we implement those patterns as rules, obtaining a set of rules that are able to extract the key elements from the tagged sentence returned by FreeLing.

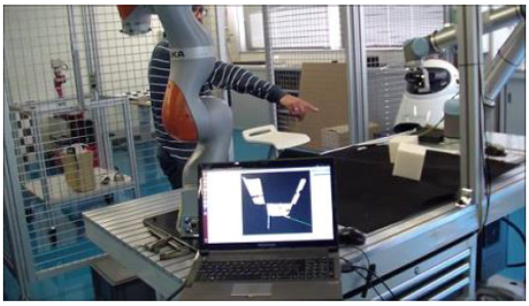


Figure 3: Pointing gesture mockup

The aim of the second step is to identify which one of the tasks the robot is able to perform suits the request best, considering the key elements in it. We undertake this step by making use of the knowledge-base information described above. First, we verify if the identified actions are among the feasible tasks described in the knowledge base, and then we apply a disambiguation step using the target information, as explained before. This process results in the delivery of the best fits for the given input, from among the potential tasks obtained from the previous step.

The module output consists of frames, one for each potential task candidate, including information denoting gestures if it exists.

Gesture Interpretation

Two kinds of gestures are addressed: pointing gestures and gestures for stop/start commands. The case presented in this paper deals with pointing gestures, that are recognized by means of point-cloud processing.

The initial setup consists of a collaborative robot and a sensor capable of providing dense point clouds, such as the ASUS Xtion sensor, the Microsoft Kinect sensor, or the industrially-available Ensenso system by IDS. The sensor is placed above the human operator and orientated towards the working area of the robot, so that the point cloud obtained resembles what the human operator in perceiving in the working environment (see Figure3).

The point cloud is then initially divided into two regions of interest (ROI), the first one corresponding to the gesture detection area, and the second one defining the working area of the robot where the pointing gesture will be applied.

With this setup, two main problems need to be solved for the interaction between the person and the robot to succeed:

1. Robust estimation of the direction of the pointing gesture.
2. Intersection of the pointing gesture with the working area of the robot.

Robust estimation of the pointing gesture The ROI for the pointing gesture detection is initially defined by specifying a cuboid in space with respect to the reference frame. In this case, the reference frame is the sensor frame, but it can also be defined using another working frame, provided a tf transformation exists between the frame used and the sensor frame. For robustness, the pointing gesture is defined using the forearm of the human operator. To identify the arm unequivocally, an euclidean cluster extraction is performed.

Intersection of the pointing gesture with the working area of the robot The main objective of a pointing gesture is to determine the point on the working area that is being pointed at. To identify this point, the points in the cloud corresponding to the pointing line are selected, from the furthest one all the way to the origin of the line that corresponds to the pointing arm. For each one of the points, a small cuboid is defined (by default, with an edge of 0.02 m), and the ROI of the working area of the robot is filtered with it. If more than N points of the working area are present inside the small centered cuboid defined in the points of the projection line, an intersection has been found. The final intersection point that is published is the closest one to the origin of the projection line. As a threshold, a minimum euclidean distance value is defined in order to avoid detecting intersections corresponding to the proper point cloud of the arm that generates the pointing gesture.

Fusion Engine

The fusion engine aims to merge both the text and the gesture outputs in order to deliver the most accurate request to send to the executive manager. The engine consider different situations regarding the complementary and/or contradictory levels of both sources.

As a first approach, we have decided the text interpreter output to prevail over the gesture information. In this way, when a contradictory situation occurs, the final request will be based on the text interpretation. When no contradiction exists between both sources, the gesture information is used either to confirm the text interpretation (redundant information), or to complete it (complementary information). For instance, using both voice and a gesture to stop a specific action provides redundant information through both channels. In contrast, using voice to determine an action and a gesture to indicate the location of the object that should suffer that action provides complementary information through both channels. In the second case, the knowledge base is used to check if the gesture information makes sense for a given task, discarding incoherent frame generation.

As a result, the fusion engine will send to the executive manager the potential, coherent and reliable requests that are understandable for the robot. The executive manager will then be in charge of task-planning issues considering those potential requests.

Case Study

In the context of the FourByThree project in which the work presented here is inscribed, there are several industrial scenarios that include human-robot collaboration via natural communication. For an initial validation of the semantic multimodal interpreter, we have selected a scenario that involves two such collaborative tasks that are carried out via interaction between a person and a robot. One task involves the collaborative assembly/disassembly on the same dies, handling different parts of the dies and (un)screwing bolts as required. The other task involves a collaborative deburring operation of wax patterns that requires managing different parts adequately in order to build a mould.

In the case of assembly task, the human and the robot work independently (un)screwing bolts on different parts of the die, and then they work together simultaneously (un)screwing different bolts on the same die cover. For the deburring activity, the human and the robot perform sequential tasks on the same workpiece in a sychronized manner, where the person glues and positions parts on the workbench while the robot deburrs them.

Considering these two contexts, we have identified the possible tasks the robot can fulfill and we have created a knowledge base starting from the knowledge manager ontology. We have also included in the knowledge base the elements that take part in both processes, together with the relations they have with respect to the tasks.

We have simulated the robot initialization to check for correct functionality. Currently we are carrying out a laboratory experimentation for evaluating the performance of the multimodal semantic interpreter.

Conclusions and future works

We have presented a semantic driven multimodal interpreter for human-robot collaborative interaction focused on industrial environments. The interpreter relies on text and gesture recognition for request processing, dealing with the analysis of the complementary/contradictory aspects of both input channels, taking advantage of semantic technologies for a more accurate interpretation due to the reasoning capabilities it provides.

This approach is generic and it can be applied in different industrial scenarios. However, in order to evaluate the approach, we are working on a specific scenario that includes the human-robot collaborative activities of assembling and deburring. We intend to measure the whole system accuracy as well as the benefit of a multimodal system against a mono-modal one in industrial environments. In addition, we will assess the usability and the benefits of such a system in industrial scenarios, as part of the advancement towards natural communication in human-robot collaborative work.

Acknowledgment

The FourByThree project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 637095

References

Bannat, A.; J. Gast, T. R.; Rösel, W.; Rigoll, G.; and Wallhof, F. 2009. A multimodal human-robot-interaction scenario: Working together with an industrial robot. 303–311.

Burger, B.; Ferrane, I.; and Lerasle, F. 2010. Towards multimodal interface for interactive robots: Challenges and robotic systems description.

Di Marco, D.; Tenorth, M.; Hussermann, K.; Zweigle, O.; and Levi, P. 2013. Roboearth action recipe execution. 117–126.

Fong, T.; Illah, R.; and Dautenhahn, K. 2003. A survey of socially interactive robots. 143–166.

Gonzalez-Agirre, A.; Laparra, E.; and Rigau, G. 2012. Multilingual central repository version 3.0: upgrading a very large lexical knowledge base.

Goodrich, M., and Schultz, A. 2007. Human-robot interaction: A survey. 203–275.

Gunhee, K.; Woojin, C.; Munsang, K.; and Chongwon, L. 2004. The autonomous tour-guide robot jinny. 3450–3455.

Kollar, T.; Tellex, S.; Roy, D.; and Roy, N. 2010. Toward understanding natural language directions. 259–266.

MacMahon, M.; Stankiewicz, B.; and Kuipers, B. 2006. Walk the talk: Connecting language, knowledge, and action in route instructions.

Padró, L., and Stanilovsky, E. 2012. Freeling 3.0: Towards wider multilinguality.

R. Stiefelhagen, C. Fugen, P. G. H. H. K. N., and Waibel, A. 2004. Natural human-robot interaction using speech, head pose and gestures. 2422 – 2427 vol.3.

Rossi, S.; Leone, E.; Fiore, M.; Finzi, A.; and Cutugno, F. 2013. An extensible architecture for robust miultmodal human-robot communication. 2208–2213.

Tenorth, M.; Kunze, L.; Jain, D.; and Beetz, M. 2010. Knowrob-map - knowledge-linked semantic object maps.

Thrun, S.; Bennewitz, M.; Burgard, W.; Cremers., A.; Dellaert, F.; Fox, D.; Hähnel, D.; Lakemeyer, G.; Rosenberg, C.; Roy, N.; Schulte, J.; Schulz, D.; and Steiner, W. 1999. Experiences with two deployed interactive tour-guide robots.

Wang, T., and Chen, Q. 2011. Object semantic map representation for indoor mobile robots. 309–313.

Winograd, T. 1971. Procedures as a representation for data in a computer program for understanding natural language.

7.15. Revisiting the end user's perspective in collaborative human-robot interaction

Revisiting the end user's perspective in collaborative human-robot interaction: Proceedings of the 19th International Conference on CLAWAR 2016. In book: Advances in Cooperative Robotics, pp.196-204 [113]

REVISITING THE END USER'S PERSPECTIVE IN COLLABORATIVE HUMAN-ROBOT INTERACTION

JOHAN KILDAL, IÑAKI MAURTUA

Intelligent and Autonomous Systems, IK4-TEKNIKER
Iñaki Goenaga 5, 20600 Eibar, Basque Country, Spain

The disciplines of Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) have evolved rather independently since their early days, mainly because early computers and early robots shared little common ground. However, as computers have jumped off the desktop to pervade the physical world, and as robots extend themselves from the physical world into the cloud, traditional boundaries between these two technological entities have blurred out due to the increasing complexity of their natural habitats. In this paper, we take a snapshot at these converging evolutions, and enquire about the benefits that the human user in HRI can derive from applying HCI methods for R&D, including viewing the collaborative robot itself as an interface.

1. Introduction

By definition, the disciplines of Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) share the common element of a *human actor*. Around that human actor (i.e., a person, end user of the technology), there is a *computer* in the first case and a *robot* in the second. The human actor *interacts* with each one of these technologies in order to achieve certain *goals*, which in turn dictate *requirements* that the technology has to satisfy.

The elements outlined in the paragraph above, constitute the foundational framework on top of which these two disciplines of study (HCI and HRI) operate, centered on the end user's needs and goals. Interaction designers that contribute to HCI, HRI or both, have the role to design, test and propose new interaction techniques that best serve the end user's needs and task requirements when interacting with computers or robots. They outline roadmaps for technology developers to produce new enablers needed in their envisioned interactions, and their work is fueled by advances in enabling technologies that make it to the labs

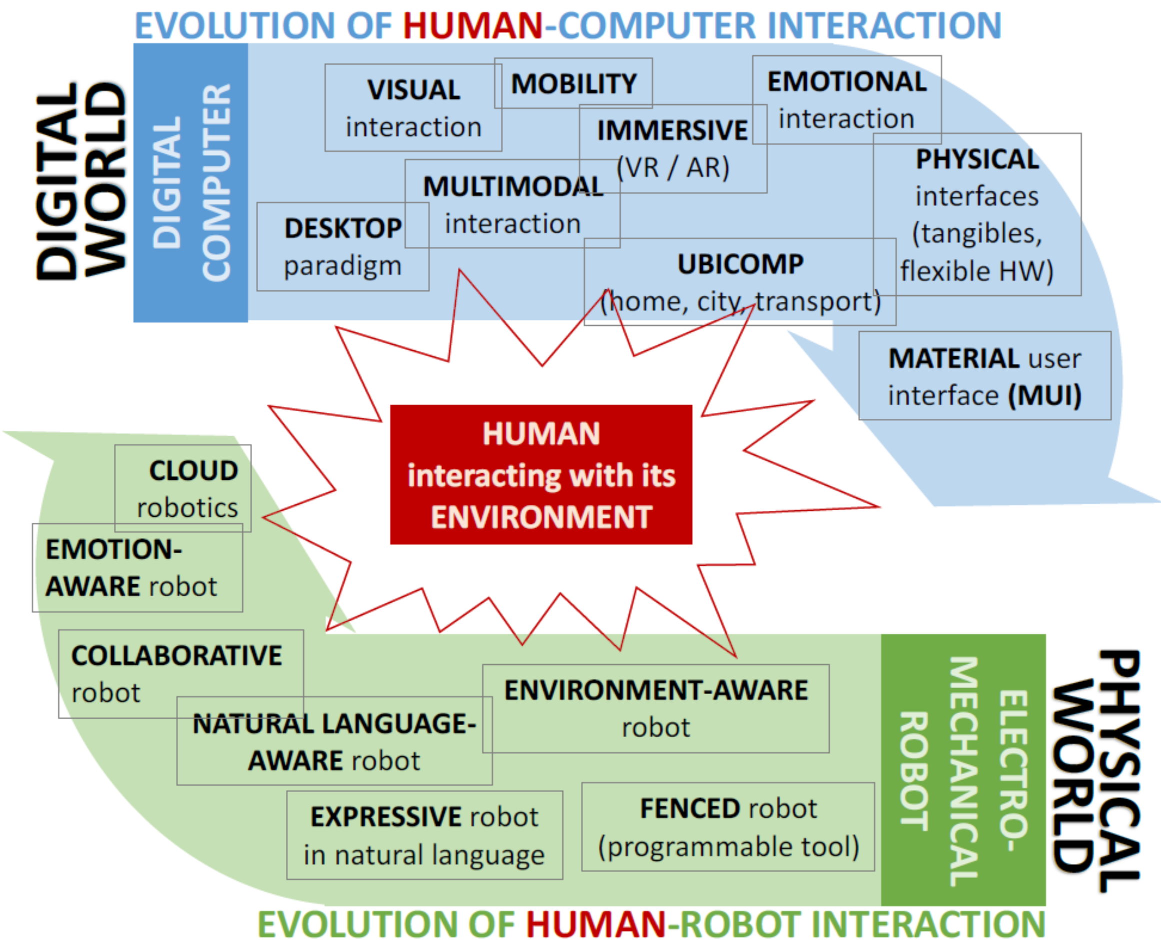


Figure 1. Converging evolution of HCI and HRI around the human end user

and to the market. When made available, such enablers are implemented across the board on technology-based systems and devices (including robotic systems), intended for any kind of user and use scenario. In contrast, the human factors taken into account are often dependent on whether the system is a computer or a robot, a personal consumer device or a productivity-oriented industrial system. This is due to the fact that HCI and HRI are the result of different traditions that originated from distant initial points (early computers and robots were very different animals), which have evolved through rather independent paths towards the current common scenario around the end user (Figure 1).

In this paper, we focus on the convergence point of the evolution paths that HCI and HRI have followed. We do so from the perspective of the interaction with collaborative robots (COBOTS), which should occupy that convergence point. As part of the H2020 FourByThree* project, we propose that considering the robot itself as the interface can help COBOTS claim that convergence point

* http://fourbythree.eu/

between disciplines, while supporting a truly collaborative relationship between a person and a robot that share the same task in the same workspace.

In the following sections, we outline the main milestones that have led to this convergence of disciplines. We then make a case for the concept of user experience (UX) in HRI, as a metric originated from HCI that should be central in the development of the robot-as-interface concept.

2. The converging evolutions of HCI and HRI

Computers and robots are immersed in an ongoing fast-paced train of transformation, and both HCI and HRI are constantly adapting to changes in technology. The ever-further miniaturization of digital electronics has democratized computational capacity. Benefiting from that, disciplines such as data science, artificial intelligence, and materials science, are providing engineers and designers with algorithms, hardware solutions, and exotic new sensors that are driving the evolution of everything computerized around us, including robots. As a result, a technological convergence has taken place between computers and robots, making it increasingly difficult to differentiate them from the perspective of human-technology interaction. Reviewing each of these paths of evolution can help with understanding these differences and identifying opportunities.

2.1. *The evolution of HCI*

The main stem of mainstream HCI emerged from the use of the very first digital computers. With the advent of the Graphical User Interfaces (GUIs) and Direct Manipulation [1], the Desktop Metaphor got consolidated as the dominant interaction paradigm. For a long time, psychologists were first in studying the human part of HCI, and they focused heavily on the study of visual perception and cognition, attention and memory. From that time, research methods from psychology became the norm in HCI research. In those visual-display-centric systems, InfoVis (the discipline that studies the visual representation of information) [2], found fertile ground for its development via rich interactive visualizations on computer displays. Later on, but already nearly three decades ago, multimodal interfaces incorporated new sensory channels to the development of new interactions that made a broader use of the person's perceptual and cognitive capacities. Around the same time, auditory displays [3] studied new forms of complex data representations that were intrinsically non-visual. Haptic interfaces experienced a rapid development in areas such as force feedback, with which remote or virtual objects could really be felt with the hands [4]. This marked, within HCI, the beginning of embodied interactions with the physical (albeit virtualized) world. The evolution of multimodality brought about the first Virtual Reality systems (VR) [5], and later on the more demanding Augmented Reality (AR) ones [6], in which the user consumed and interacted with a fusion of real and computer-generated data that were linked to the physical environment around the user. It is only in recent years that the enabling technologies have reached maturity for the mainstream adoption of VR and AR.

Mobility, and more recently wearability, of personal computers freed the computer user from his desk, allowing for interactions anywhere and on the go, through the use of new mobile interaction techniques. Mobile HCI gave a more prominent role to research methods that were better suited to evaluations in the field and in social contexts, rather than within the lab [7]. Thus, disciplines such as ethnography and sociology contributed their well-established research methods to the corpus of methods used in HCI. Almost paradoxically, the desktop metaphor still sticks around (e.g., in smartphone operating systems), and mobile interfaces have remained stubbornly vision-centric. Auditory interfaces are only testimonial and relegated to fancy alarms, and haptic experiences have been reduced to minimalistic expressions.

Two strands in HCI have recently shown significant development towards overcoming the WIMP (Windows, Icons, Menus, Pointer) [8] legacy. First, with research methods from linguistics and through new machine-learning-based enablers, interfaces based on natural language are reaching enough maturity through semantic technologies [9] to partially free the user's vision and support multitasking. Second, a trend for physically-rich interfaces in the form of physical user interfaces (such as tangibles [10], deformable computers [11, 12], and the physicalizations of information [13]), which bring new prominence to the physical properties of the interface. Thus, interfaces are objects that increasingly live in the three-dimensional physical space, together with the humans that use them, in the same way physical robots do.

2.2. *The evolution of HRI*

The statement in this paper that HCI and HRI have converged into a common ground of interaction with the human actor, has already been shown in the previous paragraph. In it, the description of the later stages in the evolution of HCI lists current interaction trends and enabling technologies that receive just as much attention from HRI in its latest developments. This is coherent with the fact that both computers and robots have come to occupy the same physical space and the same distributed connectivity (i.e., the same space in the cloud).

Considering the evolution of interaction through the prism of industrial HRI, in the beginning there was not a common physical space for interaction between a person and an active robot, primarily for safety reasons [14]. The interaction between them was limited to the programming phase, and even then, it was mediated by desktop human-computer interfaces. Later on, robots evolved senses that made them aware of their environment (machine vision, proximity and contact sensing). With improved and additional sensing, robots became reliable in identifying human presence in their space of action, and Human-Robot Interaction gained a new meaning in terms of collaboration in the same physical space (COBOTS) [15]. From that point on, sociability became an element of paramount importance for safety and collaborative productivity. Currently, some of the challenges for interaction design with collaborative robots lie on providing the human partner with situational awareness about the robot, its state, how it is interpreting the collaborative context, and anticipation of its immediate future actions [16]. In the collaborative relationship between the person and the robot, the skill of the robot to keep its human partner informed can be seen as an etiquette requirement without which the user has to maintain taxing levels of extra-vigilance for execution mistakes or even risk situations that might arise at any point. However, even when the robot can perform its physical tasks well, and is aware of (and respectful with) human presence sharing the same workspace, it struggles keeping that same human actor aware of its own understanding of the situation and of its imminent action plans.

3. In the point of convergence: the robot as an interface

HRI arrives at the point of convergence with many and sound legacy strengths in physical manipulation and contact interaction with humans in the 3D space, assisted by advanced sensing and interpretation of the environment, and with tested collision-safety standards. In turn, HCI brings to the point of convergence important complementary contributions that assist technology in embracing the end user. One contribution stands out in particular: the well-established concept of *User Experience* (UX). Applied to robotics, UX is the result of a holistic assessment of the objective and subjective imprint left on the end user by every aspect surrounding the relationship between the person and the robot. When this relationship is collaborative, aspects that are likely to influence the outcome in a UX assessment will include the efficiency and effectiveness of the collaboration, safety and perception of safety, overall workload, fluency of interaction, ergonomics and aesthetics, among a longer list.

As mentioned above, if the robot struggles to keep its human partner '*in the loop*', aware of its intentions for immediate action and about its understanding of the course of the collaboration, the end user develops a sense of uncertainty, which damages the end user's experience (her UX). Surprise and unexpected movements are at the root of the majority of accidents with robots [17, 18]. Even when robots are deemed safe for collaboration, research shows that less experienced users tend to be extra-vigilant and weary of the action decisions and action executions of robots, which has the counter-productive effect of stealing attention resources from the end user for vigilance and supervision, rather than freeing them for true parallel collaborative work [19]. In practice, there is a shared level of uncertainty between robot and person in the moments that precede the robot's actions. The person is uncertain about whether the robot will fulfil the actions expected from it, or even if it will perform any actions at all. On the robot's side, there is a level of uncertainty in the action decisions that it takes, which originates from the ambiguity of the user's instructions and from the uncertainty of the algorithms that have led to taking those decisions [20, 21]. Learning from HCI, the negative effect that uncertainty can have on UX can be reduced if the end user can refer to an interface that (*i*) keeps the user informed of what is going on in the system, and (*ii*) allows the user to cancel or modify ongoing processes or actions, thus granting him an enhanced sense of control [22].

As discussed earlier, for a truly collaborative interaction with a robot in the same workspace, we propose that the robot itself is the interface. For this purpose, the robot's embodiment is required to include (*i*) a display with real-time information about the robot's confidence and understanding of the status of the interaction, plus its plans for immediate action, and (*ii*) an input device to modify, modulate or cancel altogether the robot's next actions. This display and input device have to merge naturally with the interaction flow during collaboration. The goal is that they provide a level of naturalness comparable with instructing the robot via spoken language and deictic gestures. As an important additional requirement, the robot being the interface should not increase the cognitive workload, but it should free mental and perceptual resources in an end user that does not have to monitor the robot so closely due to uncertainty.

As also mentioned, the H2020 FourByThree project is taking steps towards contributing to this goal. As examples of input modalities on the robot, the project is investigating natural input techniques such as hand guiding, voice input using natural language, plus pointing gestures. Semantic technologies are being employed on the robot side in order to obtain a coherent fusion of the complementary input information that arrives simultaneously from these input channels. As an example of output on the robot interface, the FourByThree project is investigating the use of projection systems that outline the perimeter of action of the robot for safety purposes, and can be used for interaction and visualization. In fact, it is possible to visualize relevant information to support the worker's task and also provide virtual buttons or simple menus that the worker can use to control the robot. As future steps, we will follow a user-centered methodology inspired by interface design practices from HCI. The process will include the capture of requirements, the identification of perceptual, cognitive and physical resources available from the user, and iterative design, validation and evaluation cycles.

4. Conclusions

The converging evolutions of HCI and HRI have met in interactive scenarios that share important commonalities: the end user utilizes technology in the physical space and with physical actions, while technology monitors and strives to understand the end user, with the help of sensors and pervading cloud connectivity. Relatively recent collaborative robots can benefit from the use of HCI design methods that seek to model the end user's UX, by improving feedback communication and improving situational awareness in the user. We propose that viewing the collaborative robot itself as an interface can accommodate natural input and output techniques that can ease uncertainty, grant control through anticipation, and improve the outcome of collaborative tasks, including the user's reported UX.

Acknowledgment

The FourByThree project has received funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No. 637095).

References

1. Shneiderman, B., The future of interactive systems and the emergence of direct manipulation. Behav. Inf. Technol. **1**, 237–256 (1982)
2. Tufte, E.R., Graves-Morris, P.R., The visual display of quantitative information. Graphics press Cheshire, CT (1983).
3. Kramer, G., An Introduction to Auditory Display. In: Kramer, G. (ed.) Auditory Display: Sonification, Audification, and Auditory Interfaces. pp. 1–77. Addison-Wesley (1994).Asdas
4. Brewster, S., Murray-Smith, R., Haptic Human-Computer Interaction. Lecture Notes in Computer Science **2058**, Springer (2001).Asd
5. Choi, S., Jung, K., Do Noh, S., Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. Concurr. Eng. **23**, 40–63 (2015).Asd

7.16. Human robot interaction in industrial robotics. Examples from research centers to industry

Alberto Tellaeche, Iñaki Maurtua, Aitor Ibarguren: Human robot interaction in industrial robotics. Examples from research centers to industry. ETFA 2015: 1-6 [114]

Human Robot interaction in industrial robotics

Examples from research centers to industry.

Dr. Alberto Tellaeche Iglesias
Smart and Autonomous Systems Unit
IK4 - Tekniker
Eibar, Spain
alberto.tellaeche@tekniker.es

***Abstract*—Current state of the art in industrial robotics presents highly autonomous production plants, where an undefined number of robots work in cooperation to perform complicated tasks in an automatic way. However, these tasks are repetitive and generally do not allow a minimum change or variation in the overall process to obtain a successful result. In addition to this, these layouts do not allow any fast or easy change in the programmed operation, being not flexible at all. Contrary to this typical industrial situation, current research activities in industrial robotics are focused on human robot interaction and safety, to carry out added value operations that need to be performed both the human and the robot working together. This paper outlines real research cases of this latest type, which have as the final objective their set up in industrial facilities, presenting a review of the current state of the art in human robot interaction to explain, in following sections, successful industrial cases developed in the framework of European Commission Projects.**

***Keywords*—industrial robotics, human robot interaction**

Introduction

I.

Human-robot interaction (HRI) has been a constant concern from the very beginning of robot use, as an academic speculation before, and as an active field of research nowadays. According to [1], the HRI problem can be defined as *"to understand and shape the interactions between one or more humans and one or more robots"*.

Over all the problems and questions that HRI brings up, one of them mainly summarizes all the others: the idea of safe interaction. The closer the human and robot need to work, the more the risk of an accident. Until now, in advanced industrial set ups, this problem has been solved forbidding humans and robots to share a common workspace at any time. This is achieved by using protection fences, and an important number of sensors to detect potential risky situations.

Actual HRI research takes advantage of the tools provided by the artificial intelligence, to solve tasks such as human detection, motion planning, workspace reconstruction or compliant behavior using force control.

978-1-4673-7929-8/15/$31.00 ©2015 IEEE

One of the most common operations needed in HRI is building a 3D model of the environment through vision. Doing this allows the robot to have information about the surrounding area relative to a reference and control the presence of humans or another obstacles. The sensors typically used to collect information of the environment are RGB-D sensors, laser scanners, etc. 2D machine vision can also be used.

Motion planning and control is another of the main fields of fundamental research for HRI. Industrial robots, with typically six degrees of freedom, can already offer good results in this aspect.

In this paper, several cases of real HRI applied to industry are outlined. Among the typical operations that will be explained are: flexible object detection in a collaborative environment, human gesture detection to interact with the operations of the industrial robot, or even operator following of a big robot-crane.

All these cases are developed within the framework of the European Commission and pursue their final real implementation in industrial end-users.

II. State of the Art

In modern industrial robotics, a safe and flexible cooperation between robot and human operator can be a new way to achieve better productivity when performing complex activities. In [2] two main ways of cooperation between humans and robots were identified. First, robot controlling some degrees of freedom and the human controlling the rest, and second, the robot doing the process and finishing it roughly, and the human performing the final adjustment. In any case, both approaches present many technological questions and challenges to be solved.

In recent years, several researches have dealt with human-robot interaction, for example in [3][4][5] or [6], trying to solve different safety issues. These problems can be classified as a variation of one or more points in the following list:

1. 3D perception.
2. Human-robot interaction interfaces or information exchange.

3. Use of sensors for workspace monitoring.

At the present time, HRI is an active field of research with conferences [7] and journals [8] specifically dedicated to the field, but there is still a gap between research and real application to industry problems. To solve this aspect many European commission funded projects are nowadays trying to adequate the academic research results to industry.

III. Human Robot Interaction in Industry

In this section, results of three european projects are presented. These projects present different problems to solve, in which HRI plays a fundamental role. The problems related to HRI that are solved in each project are stated in the following list:

X-Act project [9]: Human and dual-arm robot collaboration, object detection in a 3D environment, disassembly of complex objects.

Robo-partner project [10]: Automatic guidance of a robot-crane using people detection algorithms.

EuRoC project [11]: Human robot collaboration in a rivetting task, gesture recognition and interaction.

A. X-Act project: Dual-arm robot and human collaboration for complex disassembling tasks.

The main objective of the X-Act project is *"to utilise advanced cooperative robotic systems within European manufacturing and assembly facilities."*.

In one of the cases of use, the main objective is to dismount a sewing machine, the product of one of the end users participating in the project. The dismounting of this product is necessary for repairing and maintenance operations and, nowadays, this laboriuous and complex process is performed completely manually by an operator. The unmounting procedures have been changed to use a dual-arm robot in collaboration with an operator, to speed-up the whole process and make it more simple.

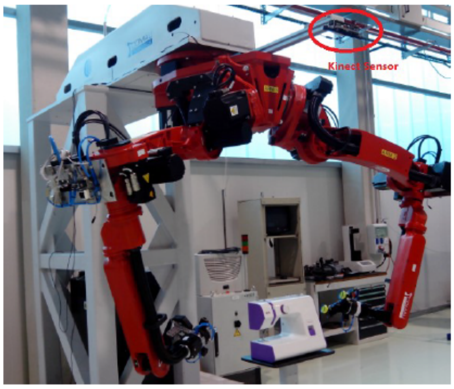


Fig. 1. Layout of the experiment, dual-arm robot, with the sewing machine in the workspace, with the Microsoft Kinect sensor in an azimutal position.

To introduce a dual arm robot in this operation, several stages in the process must be taken into account, that imply close human and robot interaction. These are:

1. Initial positioning of the sewing machine in the robot workspace: The box containing the sewing machine is initially placed by an operator in an imprecise way.
2. Operator collaboration during the process, to retire certain parts of the sewing machine, while the robot is holding it.
3. Continous monitoring of the working area, to avoid accidents with unexpected presences of operators during the unassembling process.

1) Initial positioning of the sewing machine in the robot workspace.

In the way the process is defined, the object, in this case, the sewing machine, is placed in an imprecise way in the working area, so it is fundamental its location and referencing with respect to the robot coordinates to carry out the dismounting operations.

To locate precisely the object and to obtain its 6DOF coordinates with respect to the robot, a CAD model of the sewing machine and a Kinect sensor to obtain the workspace pointcloud were used. Using 3D machine vision libraries such as MvTec Halcon [12] or PCL [13], the pose of the object with respect to robot coordinates was obtained with errors below 4mm in euclidean distance and 2.5° in rotation error [14].

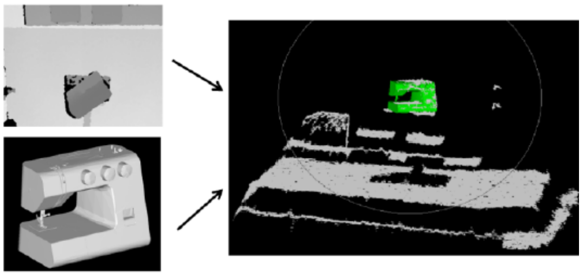


Fig. 2. CAD model of the sewing machine, 3D image and matching of the object in the workspace point cloud.

2) Operator collaboration with the unassembling process

In certain stages of the disassembling operation, the operator must collaborate directly with the robot, that is, entering the workspace to remove parts of the sewing machine, while the dual-arm robot is holding it. In this case, the presence of the operator must be continuously monitored to avoid un expected accidents.

When the operation of the dual-arm robot reaches a point where human intervention is needed, the safety measures monitoring the workspace allow human intervention.

Once the human worker has finished the operation, the workspace monitoring system is rearmed again.

In next section, the safety system for monitoring the working area will be explained.

3) Continuous monitoring of the working area.

For monitoring the working area of the robot, to avoid accidents and to decrease the speed of the robot in case someone approaches the surroundings, a redundant safety system has been implemented using laser range finders and the Safety Eye device [15], from Pilz.

The Safety Eye can be configured, to define 3D warning areas and stop areas. In the warning areas, the robot reduces speed to a 15% of its nominal speed to a secure operation speed, while in the stop areas, the robot stops operation completely to avoid accidents.

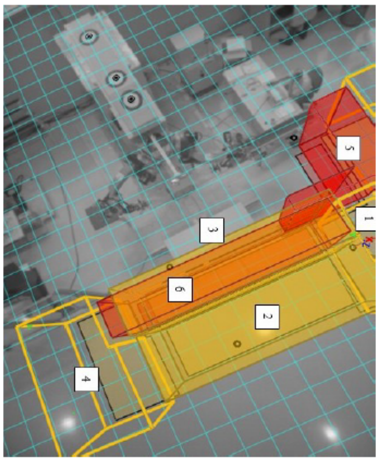


Fig. 3. Safety area definition with the safety eye

Complementing the Safety Eye, two laser range finders are monitoring the presence of people in the workspace. In this case, the reduce in the robot speed is linear, ranging from a 100% nominal speed when there is no human in the surroundings, to a complete stop when someone is close enough.

By combining both safety measures, occlusions and shadow areas in the workspace are prevented, offering a robust security system, always prevailing the most restrictive output for a given situation.

B. Robo-partner project: Automatic guidance of a robot-crane using people detection algorithms.

Robo-partner proposes *"Seamless Human-Robot Cooperation for Intelligent, Flexible and Safe Operations in the Assembly Factories of the Future "*

Among other activities, it is necessary to guide in a natural way a robot mounted in a crane inside a workshop, Hercules, designed to lift heavy weighted parts, offering a payload up to 3000 kg.

Fig. 4. Detail of the Hercules robot.

The parts to manipulate in this project are metal rings of 3 m in diameter, and with a weight of 2000 kg. Fig. 4 gives a detail of this operation in a simulation environment.

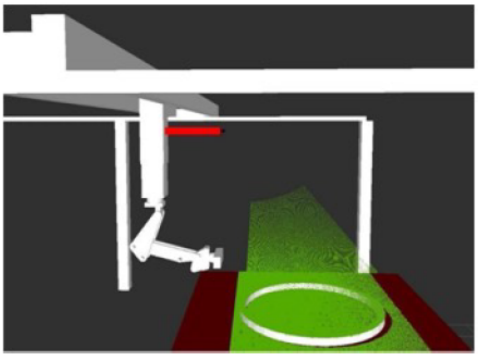


Fig. 5. Task to solve in the Robo-partner project, in a simulation environment.

In this case, HRI comes from the need of an easy method to guide the robot in big workshops. A person tracking algorithm has been developed, and Hercules robot performs a follow-me operation to reach the point where the operator is going to.

To solve this task, an static infaestructure of 2D color cameras has been set up in the workshop, covering the expected working area for the robot. These cameras are calibrated so that real position of the operator in the work shop can be precisely calculated. To position the operator in the workshop, an hybrid approach has been used for robust people detection in open spaces. This hybrid algorithm has been developed using OpenCV machine vision library [16], and is composed by the combination of three different algorithms:

1. Background segmentation and motion detection.
2. Color segmentation.
3. Use of Kalman Filters for trajectory estimation.

1) Background segmentation and motion detection

The first approach consists in detecting the movement by using background segmentation. This is carried out by substracting the background (mean of the last N images) from the foreground (last image acquired). This substraction gives all the inmediate changes occurred in the last frame acquired.

Further processing (image opening and dilation with an ellipse as a morphological element) allows to remove small detected changes corresponding to changes in lighting conditions, etc.

2) Colour segmentation

Color segmentation is performed using the HSV color model, so that segmentation is invariant to changes of light. A vest of a modelized yellow color is used:

$$H = [20, 40]$$
$$S = [100, 230] \qquad (1)$$
$$V = [200, 255]$$

This color is used as a reference to segment the scene. The resulting position of the person is the second center value obtained, used later to obtain the final center point and position of the operator.

3) Kalman Filters

As a third approach a first order Kalman filter [17] has been implemented taking as inputs the centers of the valid contours obtained with the background segmentation. This filter assumes that next position depends on the velocity in valid position changes, the velocity remains constant, and there is an absence of white noise in the measurements of position and velocity. Being p_k the position of the person in a given state k and v_k the velocity in the same state k:

$$p_k = p_{k-1} + v_{k-1}$$
$$v_k = p_k - p_{k-1} \qquad (2)$$

The use of the Kalman filter allow the occurrence of partial occlusions of the operator, offering a robust method for trajectory estimation.

4) Combination of different approaches

By combining the results of the three approaches it is possible to detect the position of the operator in the workshop very precisely, avoiding positioning errors due to other elements moving in the workshop, changes in environmental lightning, partial occlusions, or even the fact of a static position of the operator.

Fig. 6. Operator detected while walking. The centered circle corresponds to the Kalman estimated position.

C. EuRoC project: Human robot collaboration in a rivetting task, gesture recognition and interaction.

EuRoC (European Robotics Challenges) Project is focused on providing advanced robotics solutions to the european manufacturing industry. To achieve this, the project proposes three industry-relevant challenges:

1. Reconfigurable Interactive Manufacturing Cell.
2. Shop Floor Logistics and Manipulation.
3. Plant Servicing and Inspection.

Each challenge presents application experiments intended to solve problems present in real HRI operations. The first challenge proposes there operations related to the collaboration of an operator and the robot. These are:

1. Detect the accurate position of the part used in the rivetting operation.
2. Detect the pointing gesture of the operator
3. Detect the precise point in the part which the operator is pointing to.

The sensors to monitor the activity in the workspace consist of two Kinect cameras, providing pointclouds and color 2D images, one placed in an azimutal position above the table where the piece is going to be manipulated, and another one monitoring the workspace shared by robot and operator.

The first of the operations is very similar to the one explained in the X-Act case, and will not be explained again. The operations related with the gestures of the operator, are explained in the following sections.

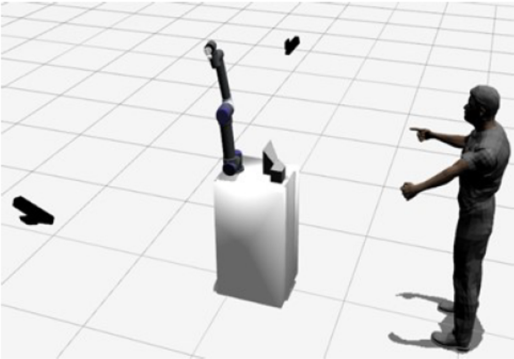


Fig. 7. Overall scene from EuRoC project, challenge 1

1) Gesture detection

Gesture detection is one of the fundamental tasks to be solved in HRI problems. This problem presents various previous challenges that need a robust solution, such as:

- Analysis of depth images for person segmentation.
- Skeleton detection and joint identification.
- Gesture recognition.

The case presented in EuRoC projects allows the monitoring of a working area near the part where the person performs the gesture. To detect the pointing gesture with precise results, three aspects have been taken into consideration:

- Monitoring of the person arm, in a region of interest.
- Analysis of movement and its stability (stable pointing position)
- Coordinates of the hand, to estimate where the gesture is pointing to.

Finally to solve these issues a mixed processing with the 2D color image and the depth image was programmed, using the data given by the Kinect sensor placed in a lateral position.

This algorithm presents a detection ratio of over 90% correct gestures detected.

Fig. 8. Operator´s pointing gesture detected.

2) Detection of the precise point in the part pointed by the human pointing gesture.

Using the Kinect placed in an azimutal position over the part as shown in fig.7, it is possible to detect the exact point over the piece where the operator is pointing to. The solution to this problem can be used to indicate to a robot an exact point of operation without contact.

The problem has been solved following these steps:

- Inspection of the point cloud given by the Kinect. Detection and segmentation of the operator finger.
- Adjust the finger model to a basic geometric shape, a cylinder.
- Obtaining of the cylinder equations. Principal axis equation, diameter.
- Project the principal axis of the cylinder (finger) to the part.
- Detect the intersection of the principal axis with the surface of the part and select that intersection point as the pointed point by the operator.

As in the previous case, this algorithm presents a high accuracy in point detection with errors below 3 mm in the euclidean distance between the obtained point and the theoretical result.

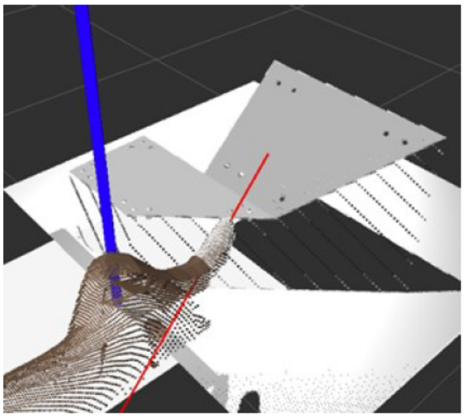


Fig. 9. Obtaining of the precise point over the surface of the part that has been pointed by the operator

IV. Conclussions

Human Robot Interaction in robotics is still a field under active development, and reaching industrial and robust solutions for many industrial tasks is still a big challenge, mainly in terms of reliability and precision, however, and thanks to the European funded projects with the principal objective of applying research results to real world industries, many promising results are being obtained in industrial applications, taking industrial robotics to a next development stage.

Acknowledgments

The works presented in this paper have been performed within the scope of the following European Funded projects:

"X-ACT: Expert cooperative robots for highly skilled operations for the factory of the future". X-ACT is a Small or Medium-scale focused research project supported by the European Commission in the 7th Framework Programme (FP7-314355). For further information, see http://www.xact-project.eu/

"Robo-partner: Seamless Human-Robot Cooperation for Intelligent, Flexible and Safe Operations in the Assembly Factories of the Future" is also a research project funded by the European Commission in the 7th Framework Programme (FP7-608855). Further information can be found in http://www.robo-partner.eu/

"EuRoC: European Robotics Challenges" is, as the two previous projects, also founded by the European Commission in the same 7th Framework Programme (FP7-608849). Its web site for further information is http://www.euroc-project.eu/.

References

[1] M.A. Goodrich, A.C. Shultz, "Human-Robot Interaction: A Survey" Foundations and Trends in Human-Computer Interaction, vol. 1, no 3, pp. 203-275, 2007.

[2] T. Wojtara, M. Uchihara, H. Murayama, S. Shimoda, S. Sakai, H. Fujimoto, H. Kimura, "Human-robot collaboration in precise positioning of a three-dimensional object", Automatica, vol. 45, no. 2, pp. 333-342, 2009

[3] A. Bicchi, M. Bavaro, G. Boccadamo, D. De Carli, R. Filippini, G. Grioli, et al. "Physical Human-Robot Interaction: Dependability, Safety, and Performance", Proc. 10th Intl. Workshop Advanced Motion Control, pp. 9-14, 2008

[4] A. Bannat, J. Gast, T. Rehrl, W Rösel, G. Rigoll, F. Wallhoff. "A Multimodal Human-Robot-Interaction Scenario: Working Together with an Industrial Robot", Proc. 13th International Conference on Human-Computer Interaction, July 2009.

[5] P.A. Lasota, G.F. Rossano, J.A. Shah. "Toward Safe Close-Proximity Human-Robot Interaction with Standard Industrial Robots". Proc. IEEE International Conference on Automation Science and Engineering, pp. 339-344, August 2014.

[6] S. Puls, P.Betz, M. Wyden, H. Wörn. "Path Planning for Industrial Robots in Human-Robot Interaction". Proc. IEEE International Conference on Intelligent Robots and Systems. 2012.

[7] IEEE International Conference on Human Robot Interaction http://humanrobotinteraction.org/2015/

[8] Journal on Human Robot Interaction. http://humanrobotinteraction.org/journal/index.php/HRI.

[9] X-Act project. http://www.xact-project.eu/

[10] Robo-Partner project. http://www.robo-partner.eu/

[11] EuRoC project. http://www.euroc-project.eu/

[12] MvTec Halcon Library. http://www.halcon.com/halcon/version12/

[13] Point Cloud Library. http://pointclouds.org/

[14] A.Tellaeche, I.Maurtua. "6DOF Pose estimation of objects for robotic manipulation. A review of different options." Proc. IEEE Emerging Technologies and Factory Automation, pp. 1-8, September 2014.

[15] Pilz safety Eye. https://www.pilz.com/en-CO/eshop/00014000337042/SafetyEYE-Safe-camera-system.

[16] OpenCv machine vision library. http://opencv.org/

[17] Kalman Filter. http://en.wikipedia.org/wiki/Kalman_filter.

7.17. 6DOF pose estimation of objects for robotic manipulation. A review of different options

Alberto Tellaeche, Iñaki Maurtua: 6DOF pose estimation of objects for robotic manipulation. A review of different options. ETFA 2014: 1-8 [97]

6DOF pose estimation of objects for robotic manipulation. A review of different options.

Alberto Tellaeche
IK4-Tekniker
Calle Iñaki Goenaga 5
20600 Eibar, Gipuzkoa, Spain
Alberto.tellaeche@tekniker.es

Iñaki Maurtua
IK4-Tekniker
Calle Iñaki Goenaga 5
20600 Eibar, Gipuzkoa, Spain
Iñaki.maurtua@tekniker.es

Abstract

6DOF pose estimation of objects in robotics is nowadays an active field of research under different approaches.

The solution of this problem depends directly on many aspects, such as the geometry of the object to be located, the flexibility of the solution needed, or the 3D sensors used.

Reaching a robust and reliable solution to this problem is essential, being this the first step in present research in industrial robotics, where advanced manipulation and identification of complex objects is fundamental.

This paper presents different results obtained solving the 6DOF pose estimation problem, using different software libraries and the Microsoft Kinect as 3D acquisition system to capture the scene.

1. Introduction

6DOF pose estimation of objects is the task of estimating the coordinates (*X,Y,Z*) and rotation angles (Yaw, Pitch and Roll) of an object with respect to a previously established reference coordinate system.

In advanced industrial robotics solutions, where robots collaborate with human operators, this is one of the main tasks to solve, because the object to manipulate is not always placed in a fixed position.

The pose estimation problem can present many different challenges, depending on the area to inspect, the geometry of the object used, time constraints, precision required, etc. It is also fundamental in human-robot collaborative environments where an exhaustive control of people within the robotic cell guarantees the cell safety.

Some approaches already exist in this field of application, each of them having different restrictions due to the particular applications for which they have been developed. This work intends to present a general solution to this problem, within the framework of the X-ACT European project, where geometrically complex objects must be located and positioned to be later manipulated by a dual arm robot from Comau.

The principal advantages that the system tested in this research presents over other previous industrial approaches are:

1. It is a flexible system, and it is not specifically designed for a certain type of object. Object models can be obtained from CAD designs. The system is adaptable to any kind of object, provided an CAD / STL model exists.
2. The hardware and software used for this development are affordable for industrial applications. Microsoft Kinect is used as 3D sensor to capture de scene. From the software point of view, two libraries have been selected for comparison of results: PCL (Point Cloud Library), an open source project to process 3D point clouds, and the most widely used library in industrial machine vision, MvTec Halcon framework.
3. The point clouds obtained are processed in real time, without delays.

This paper is organized as follows: In section 2, there is a review of the state of the art related with pose estimation and manipulation of objects in industrial environments.

Section 3 explains the problem to solve within the X-ACT project, the object used for testing, and the proposed approach using both previously cited alternatives.

Section 4 discusses the development of the solution using PCL and Section 5 explains the same process using MVTec Halcon.

Section 6 describes the test methods for performance assessment and the results obtained, and finally, in

2014 IEEE Emerging Technology and Factory Automation (ETFA)
978-1-4799-4845-1/14/$31.00 ©2014 IEEE

section 7, the conclusions of this research work are exposed.

2. State of the art

The 6DOF problem appears continuously in many advanced robotic environments. Research in robotics is focused on trying to solve problems with uncertainties, just as human beings do. One of these fundamental problems is manipulating objects in a 3D space. This problem, in its generic form, has been an active field of research for the last years [1].

In reverse engineering, the problem of pose estimation and 2D and 3D reconstruction of point datasets is of vital importance. Good examples of these techniques applied to reconstruction can be found in [2] and [3].

In industrial systems, the precise identification of objects for grasping and manipulation has been always an active field of research, where advances have been obtained lately.

In [4], they use color concurrence histograms and geometric modeling to identify objects for manipulation and grasping, then using a classical learning framework for decision making.

In [5], they focus in the real time processing of the 3D point clouds acquired in real time. This is one of the fundamental problems to solve, because the processing of pure 3D point clouds is very computationally intensive.

CAD models are used also to represent the knowledge of the world, giving information of the environment for object recognition and matching [6].

Apart from these cited examples, may other approaches use different techniques for 3D object recognition and pose estimation, for example, in [7], 3D view-based eigen spaces are used, Features histograms are used in [8], basic research has been carried out using single colored objects in [9], keypoint features for robotic manipulation are used in [10]. Active appearance models are another approach used in [11].

3. Problem description

In the proposed experiment scenario, A Comau dual arm robot is going to be used to manipulate and dismount a sewing machine, so the first main task in this layout is to detect precisely the position of this object, taking into account the displacements and rotations present after the object has been placed by a human operator.

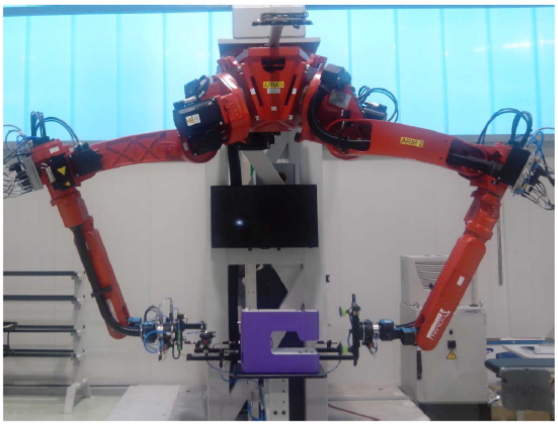


Figure 1. Layout of the experiment

To solve this problem, a Microsoft Kinect has been selected as 3D sensor, and has been placed in an azimutal position over the working area of the robot, obtaining a 3D point cloud of the area under inspection. This point cloud is already calibrated to real world coordinates by the Kinect. The aspect of this point cloud can be observed in Figure 2.

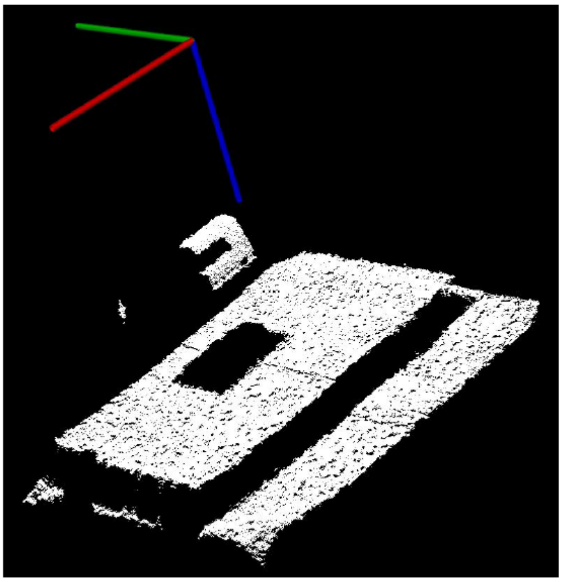


Figure 2. Calibrated point cloud of the scene

Also, in order to be able to perform the matching operation, it is necessary to have an STL model of the object under inspection. Figure 3 shows the STL file generated for the sewing machine used in this application.

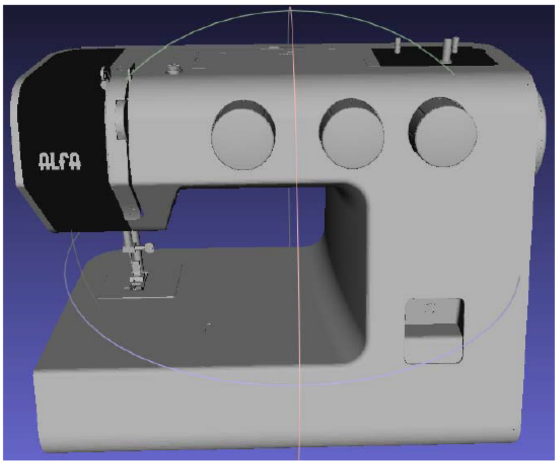


Figure 3. STL model of the sewing machine for 6DOF matching.

Taking this STL model and the 3D scene as a starting point, the following sections explain in detail the two different approaches (Point Cloud Library and MVTec Halcon) to solve the pose estimation problem.

4. Pose estimation problem solution using Point Cloud Library

The Point Cloud Library or PCL [12], is an open source, large scale software library for 2D/3D image and point cloud processing. This framework contains numerous state-of-the-art algorithms including filtering, feature estimation, surface reconstruction, registration, model fitting or segmentation. All these algorithms can be used to process and extract information of the 3D point clouds representing the environment.

The following subsections explain the process carried out to obtain the pose estimation results for the sewing machine in the robot working area.

4.1. Offline process. STL to Point Cloud model.

PCL library requires a point cloud model of the object under search. Using the library, the STL model is processed and it is later scaled to real world dimensions. In the next figure, the resultant point cloud 3D model is shown.

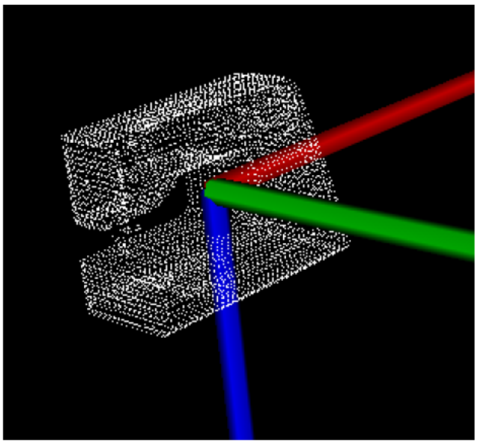


Figure 4. Point cloud model generated from the STL model.

4.2. Online process. Processing the point cloud of the scene

The original point cloud obtained by the Microsoft Kinect sensor, and presented in Figure 2, is processed to identify the object under inspection, performing the following operations:

1. Noise removal.
2. Pass through filtering.
3. Clustering of objects.
4. Approximate pose estimation.
5. Best fit matching using ICP.

4.2.1. Noise removal.

When dealing with 3D point clouds, filtering is one of the basic operations, and new methods are continuously being developed [13][14].

Datasets usually present outliers that can origin errors in the calculus of local point 3D features needed for further processing.

PCL filters outliers using statistical operations in each point neighborhood. For each point the mean distance from it to its neighbors is computed. Assuming the resulting distribution is Gaussian, all points whose mean distances are outside an interval defined by the global distances mean and standard deviation can be considered as outliers and trimmed from the dataset.

4.2.2. Pass through filtering.

Knowing the relative position of the Microsoft Kinect sensor, filtering the resulting point cloud by the *z* coordinate, allows to remove floor and background, leaving the object and platform where it is placed as the remaining elements in the resulting point cloud.

4.2.3. Clustering of objects.

With the point cloud obtained after the filtering, a clustering operation is performed to identify the remaining objects present in the scene. For this application, an Euclidean Cluster Extraction using a Kd-tree has been used. This algorithm steps are explained in [15].

For the application under study, the empirical parameters set are:

1. Cluster Tolerance = 2cm.
2. Minimum cluster size = 1000 points.
3. Maximum cluster size = 10e5 points.

With this setup, the sewing machine under inspection is the biggest cluster remaining in the image.

4.2.4. Approximate pose estimation

With the object isolated in the point cloud. The approximate pose estimation can be divided in the calculus of the translation matrix and rotation angles.

The translation matrix can be directly obtained from the coordinate limits in the cluster of the object:

$$V_T = \begin{bmatrix} x_{\min} + \dfrac{x_{\max} - x_{\min}}{2} \\ y_{\min} + \dfrac{y_{\max} - y_{\min}}{2} \\ z_{\min} + \dfrac{Obj\_Z\_size}{2} \end{bmatrix} \quad [1]$$

The Z translation is calculated using the Z dimension of the object, to avoid erroneous calculations due to possible perspective occlusions in the point cloud acquisition.

Rotation angles are directly obtained from the projection of the object over the *z* axis. Given the setup of the application, only the *Yaw* angle has a value, , which is not 0. Hence, the final transformation for the approximate pose estimation has the following form:

$$T_a = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 & x_{\min} + \dfrac{x_{\max} - x_{\min}}{2} \\ \sin(\alpha) & \cos(\alpha) & 0 & y_{\min} + \dfrac{y_{\max} - y_{\min}}{2} \\ 0 & 0 & 1 & z_{\min} + \dfrac{Obj\_Z\_size}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[2]

4.2.5. Best fit matching using ICP

The Iterative Closest Point algorithm is employed to minimize the difference between two clouds of points.

For best fit matching between the point cloud model of the sewing machine and the object already detected in the scene corresponding to it, an ICP variant is used for more robust calculations.

The ICP used in this work is a nonlinear approximation of the original algorithm, using the Levenberg-Marquardt optimization backend [16].

The final result of the ICP algorithm is a *4 x 4* dimension transformation matrix, containing the slight rotation and translation transformations for a best fit match between the model and de cloud corresponding to the sewing machine.

4.2.6. Best fit matching using ICP

The final pose estimation of the sewing machine is obtained as a combination of the approximate pose obtained and the best fit matching obtained from the ICP.

$$\begin{aligned} R_{final} &= R_{ICP} \cdot R_{aprox} \\ T_{final} &= T_{ICP} + T_{aprox} \end{aligned} \quad [4]$$

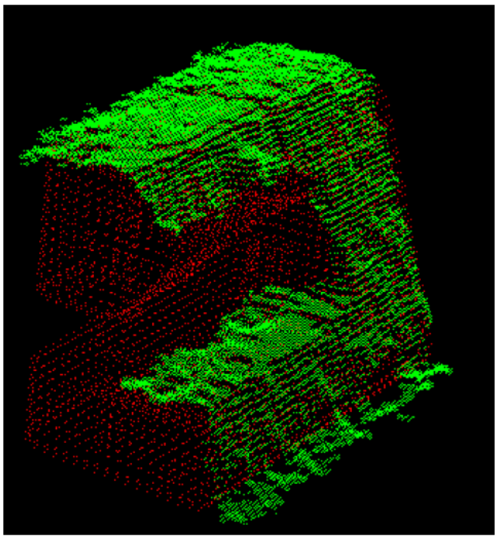


Figure 5. Final matching and 6DOF pose estimation using PCL

5. Pose estimation problem solution using MvTec Halcon.

Among all the commercial frameworks or libraries for image processing, Halcon is the most powerful and used.

This library has modules for matching, blob analysis, morphology and 3D vision, among others. This last module is the one used for solving the problem of pose estimation, being the position recognition of known objects one of its main applications.

If a model of a known 3D object is available, 3D matching can be applied to locate the object. The available 3D matching approaches perform a full 3D object recognition, locating the object and giving the pose in the search data. To do this there are two main approaches:

Shape-based 3D matching can be used to locate a complex 3D object in a single 2D image. The model of the 3D object must be available as a CAD model (DXF, OFF, PLY, etc)

Surface-based 3D matching is used to locate a 3D object in a 3D scene. In this case, a 3D object model must be available, and can be obtained either from a CAD model or from a reference 3D scene.

5.1. Offline process. STL loading.

Halcon library directly offers the possibility of loading an STL model of the object under inspection, to be used as the 3D object model in 3D matching.

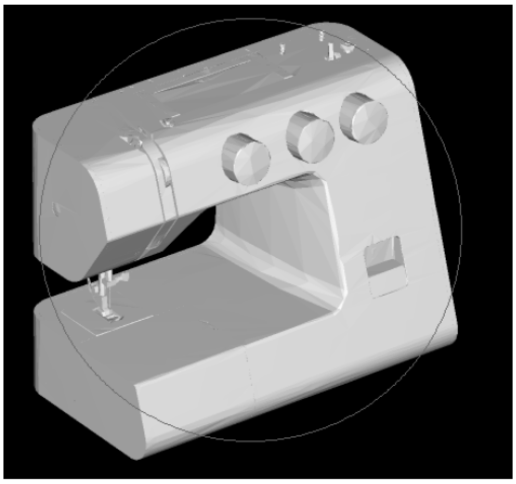


Figure 6. STL model loaded in Halcon

With the STL file loaded, Halcon internally creates a model, calculating also the normal vector for each 3D point in the model.

5.2. Online process.

Once the STL model is loaded, it is possible to perform a surface matching operation over the 3D scene.

This surface matching is performed in three main steps:

1. Approximate matching: The approximate poses of the instances of the surface model in the scene are searched. The main parameters to configure in this step are: number of matches to find, minimum difference of translation between two poses in absolute and relative values.
2. Sparse pose refinement: In this second step, the approximate poses found in the previous step are further refined. This increases the accuracy of the poses and the significance of the score value. In this step it is possible to configure the type of score to obtain in return, or the use of point normals in calculations.
3. Dense pose refinement: Accurately refines the poses found in the previous steps. This step works similar to the sparse pose refinement and minimizes the distances between the scene points and the planes of the closest model points. Parameters: number of iterations, number of points of the scene to use, type of score, or use of points with normals.

As a result of this process, Halcon gives the 3D pose estimation of the object within the scene referred to the original coordinate system given by the Kinect.

A 3D pose is a easier-to-understand representation of a rigid transformation. Instead of the 12 elements of a homogeneous transformation matrix, a pose describes the same transformation with 6 parameters, 3 translations and 3 rotations, one for each axis.

$$Pose = (T_x \quad T_y \quad T_z \quad R_x \quad R_y \quad R_z) \qquad [3]$$

With the pose, obtaining the homogeneous transformation matrix in case it is necesary, is a direct operation.

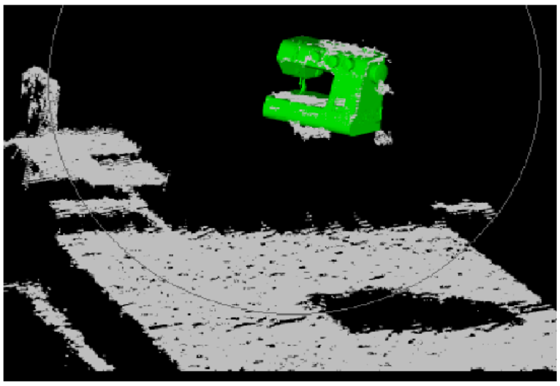


Figure 7. Successful matching operation of the object in the scene, using MVTec Halcon.

6. Obtained results

Numerical results have been obtained independently for both approaches using the same experiment setup, and recording controlled movements of the object under inspection. This approach allows comparing both approaches to assess precision and viability.

6.1. Results obtained with PCL.

The tests for numerical results have been carried out modifying *X* and *Y* coordinates of the sewing machine incrementally, with respect to a previously fixed position taken as reference. For each new position, the error of position estimation in mm, yaw angle in radians, and the overall error in the matching are taken as measurements to evaluate the precision of the proposed approach.

The performed tests follow a systematic order, considering different typical situations:

1. Displacement in X axis: three movements, positive, negative and large distance.
2. Displacement in Y axis three movements, positive, negative and large distance.
3. Rotations over Z axis (yaw angle) : Positive and negative rotations, small, medium and large variations.

Each of the previously cited test cases, have been repeated 10 times.

The results of the PCL matching process using the approximate positioning and the ICP for exact matching are given as the mean positioning error for each axis, and the mean error for yaw angle. As a final parameter the ICP estimation error is also given, as the sum of the distance errors of corresponding matching points.

The following table presents the results obtained. Linear measurements are in mm and angle measurements in radians.

TABLE I. TEST RESULTS FOR PCL

Increment Vector (X,Y,Yaw)	***Mean positioning error (X,Y,Yaw)***	***ICP estimation error***
(-35, 0, 0)	(4.67, < 1, < 0.02)	0.0032
(+35, 0, 0)	(2.27, < 1, < 0.02)	0.0041
(+70, 0, 0)	(2.86, < 1, < 0.02)	0.0037
(0, -35, 0)	(< 1, 3.68, < 0.02)	0.0027
(0, 35, 0)	(< 1, 4.02, < 0.02)	0.0033
(0, 70, 0)	(< 1, 2.97, < 0.02)	0.0037
(0, 0, 0.1745)	(< 1, < 1, 0.05)	0.0030
(0, 0, 0.3490)	(< 1, < 1, 0.07)	0.0038
(0, 0, 0.8726)	(< 1, < 1, 0.11)	0.044

The principal drawbacks of this approach are:

1. The method is very sensitive to algorithm parameters. ICP algorithm convergence is critical.
2. The algorithm gives convergence errors due to the object perspective if the displacements in *X* and *Y* are big enough.

6.2. Results obtained with MVTec Halcon.

As explained in section 5, MVTec Halcon solves the matching problem in three consecutive steps:

1. Approximate Matching
2. Sparse pose refinement
3. Dense pose refinement

As a result, the poses for the different test positions can be obtained. The coordinates are referred to the axis defined by the Microsoft Kinect sensor. In this case, the distances are given in mm and the angles in degrees.

TABLE II. POSES OBTAINED FOR THE DIFFERENT TEST POSITIONS

Position (X,Y,Yaw)	***X***	***Y***	***Z***	***Roll angle***	***Pitch angle***	***Yaw angle***
(0,0,0)	4.6	285.6	1390	9.17	359.3	182.2
(-35,-30,0)	34	317.9	1395	9.52	359.5	181.3
(0,35,0)	11	251.6	1380	8.3	358.2	182.4
(35,30,0)	-19	249	1380	8.76	358.3	182.4
(-35,0,0)	13	314.4	1397	8.48	358.1	181.9
(0,0,180)	14	281	1395	10.06	359.3	2.2
(0,0,90)	15	295.8	1394	8.1	357.6	91.4
(0,0,45)	8	282.8	1389	8.82	359.9	143.1

Taking as a reference position the (0, 0, 0) the errors in the Euclidean distances between the reference and the different test positions centers indicate the precision in results of this approach. In the last three test positions, where only rotations of the *yaw* angle are performed, the error is indicated as the estimation error of the new position angle. Table 3 shows the empirical results obtained. Positions and errors are expressed in mm, angle errors are expressed in degrees.

TABLE III. TEST RESUTLS FOR MVTEC HALCON

Position (X,Y,Z=cte)	***Distance Euclidean Error (mm), Angle error (deg)***
(-35,-30,0)	2.16
(0,35,0)	1.27
(35,30,0)	1.19
(-35,0,0)	4.3
(0,0,180)	2.2º
(0,0,90)	1.4º
(0,0,45)	1.88º

7. Conclusions

Both methods have proven to be effective obtaining a precise pose estimation of the object under study, but after testing both methods several conclusions can be taken out:

1. PCL library is free of charge and intended to be used in C/C++. It demands also a deep knowledge of the scientific basis of the algorithms implemented to adjust their parameters in an optimum way.
2. MVTec Halcon is a proprietary library, and it is necessary to purchase a license, but, on the other hand, it encapsulates powerful algorithms to perform matching operations, easier to configure. It also allows more flexibility defining object models and surfaces, and its results are not affected by perspective problems.
3. The solution provided by MVTec Halcon has proven to be more robust and flexible when the object presents big displacements from its original position, or partial occlusions. In situations where PCL solution fails to converge or give a correct solution, MVTec Halcon gives a correct pose.

To summarize, PCL library is a very powerful software, completely open and intended for research environments, where time to market is not a problem and people deals with problems related to algorithm setup and configuration. This library is mainly used in research environments.

MVTec Halcon is currently the most powerful library in industrial machine vision, being also valid for advanced and research applications. It has many different modules divided by different applications, among them the 3D Vision and Matching module. Although it is necessary a license to use it, this expense can be easily compensated by the savings in development time of complex applications

8. Acknowledgements

This work has been performed within the scope of the project "X-ACT: Expert cooperative robots for highly skilled operations for the factory of the future". X-ACT is a Small or Medium-scale focused research project supported by the European Commission in the 7th Framework Programme (FP7-314355). For further information, see http://www.xact-project.eu/

References

[1] Andreopoulos A, Tsotsos J K, "50 Years of object recognition: Directions forward.", *Computer Vision and Image Understanding*. Vol 117, Issue 8, pp 827-891, 2013.

[2] Eggert D W, Fitzgibbon A W, Fisher R B, "Simultaneous registration of multiple range views for use in reverse engineering of CAD models.", *Computer Vision and Image Understanding*, j.69(3):253-272, 1998.

[3] Fitzgibbon A W, "Robust registration of 2D and 3D point sets.", *Image and Vision Computing*. J. 21(13-14):1145-1153, 2003.

[4] Ekvall S, Kragic D, Hoffmann F,. "Object recognition and pose estimation using color cooccurrence histograms and geometric modeling." *Image and Vision Computing*. j :23(11) : 943-955, 2005.

[5] Beserra R, Marques B, Karin de Medeiros L, Vidal R, Pacheco L C, Garcia L M, "Efficient 3D object recognition using foveated point clouds.", *Computers & Graphics*. J :37(5) :496-508, 2013.

[6] Aldoma A, Vincze M, Blodow N, Gossow D, Gedikli S, Rusu R B, Bradski G. "CAD-model recognition and 6DOF pose estimation using 3D cues". *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. pp. 585-592, 2011.

[7] Morency L, Sundberg P, Darrell T, "Pose estimation using 3D view-based eigenspaces.", *IEEE International Workshop on Analysisn and Modeling of Faces and Gestures*. pp. 45-52, 2003

[8] Rusu R B, Bradski G, Thibaux R, Hsu J., "Fast 3D recognition and pose using the Viewpoint Feature Histogram", *IEEE International Conference on Intelligent Robots and systems (IROS)*. pp .2155-2162, 2010.

[9] Azad P, Asfour T, Dillmann R, "Accurate shape-based 6-DoF pose estimation of single-colored objects.", *IEEE International Conference on Intelligent Robots and Systems (IROS)*. pp.2690-2695, 2009

[10] Changhyun C, Christensen H I., "Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation", *IEEE International Conference on Robotics and Automation (ICRA)*. pp.4048-4055, 2010

[11] Mittrapiyanumic P, DeSouza G N, Kak A C, "Calculating the 3d-pose of rigid-objects using active appearance models.", *IEEE International Conference on Robotics and Automation (ICRA)*. pp.5147-5152, 2004

[12] The Point Cloud Library (PCL). Available: http://pointclouds.org/

[13] Schall O, Belyaev A, Seidel H P, "Robust filtering of noisy scattered point data.", *Proceedings of the Second Eurographics / IEEE conference on Point-Based Graphics*. pp. 71-77,2005

[14] Weyrich T, Pauly M, Keiser R, Heinzle S, Scandella S, Gross M, "Post-processing of scanned 3D surface data.", *Proceedings of the first Eurographics conference on Point-Based Graphics*. pp. 85-94, 2004

7.18. MAINBOT – Mobile Robots for Inspection and Maintenance in Extensive Industrial Plants

Iñaki Maurtua, Loreto Susperregi, Ane Fernández, Carlos Tubio, Torsten Felsch, Carmen Pérez, Jorge R. Rodriguez, and Meftah Ghrissi: MAINBOT – Mobile Robots for Inspection and Maintenance in Extensive Industrial Plants. Energy Procedia, Volume 49, Pages 1-2532 (2014), Proceedings of the SolarPACES 2013 International Conference, Las Vegas [103]

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 49 (2014) 1810 – 1819

Energy Procedia

SolarPACES 2013

MAINBOT – mobile robots for inspection and maintenance in extensive industrial plants

I. Maurtua[a] , L. Susperregi[a], A. Fernández[a], C. Tubío[a] , C. Perez[b], J. Rodríguez[c], T. Felsch[d], M. Ghrissi[e]

[a]IK4-Tekniker, Iñaki Goenaga 5, Eibar 20600, Spain
[b]TECNATOM, Avda. de Montes de Oca 1, San Sebastián de los Reyes (MADRID) 28703, Spain
[c]TORRESOL Energy, Avda. Zugazarte 61, Getxo 48930, Spain
[d] FRAUNHOFER IFF, Postfach 14 53, Magdeburg 39004, Germany
[e] ROBOSOFT, Technopole Izarbel, Bidart F-64210, France

Abstract

MAINBOT project is developing service robots applications to autonomously execute inspection tasks in extensive industrial plants in equipment that is arranged horizontally (using ground robots) or vertically (climbing robots). MAINBOT aims at using already available robotic solutions to deploy innovative systems in order to fulfill project industrial objectives: to provide a means to help measuring several physical parameters in multiple points by autonomous robots, able to navigate and climb structures, handling sensors or special non destructive testing equipment.
MAINBOT will validate the proposed solutions in two solar plants (cylindrical-parabolic collectors and central tower), that are very demanding from mobile manipulation point of view mainly due to the extension (e.g. a thermal solar plant of 50Mw, seven hours of storage, with 400 hectares, 400.000 mirrors, 180 km of absorber tubes, 140m tower height), the variability of conditions (outdoor, day-night), safety requirements, etc.. The objective is to increase the efficiency of the installation by improving the inspection procedures and technologies. Robot capabilities are developed at different levels: (1) Simulation: realistic testing environments are created in order to validate the algorithms developed for the project using available robot, sensors and application environments. (2) Autonomous navigation: Hybrid (topological-metric) localization and planning algorithms are integrated in order to manage the huge extensions. (3) Manipulation: Robot arm movement planning and control algorithms are developed for positioning sensing equipment with accuracy and collision avoidance. (4) Interoperability: Mechanisms to integrate the heterogeneous systems taking part in the robot operation, from third party inspection equipments to the end user maintenance planning. (5) Non-Destructive Inspection: based on eddy current and thermography, detection algorithms are developed in order to provide automatic inspection abilities to the robots.

Selection and peer review by the scientific conference committee of SolarPACES 2013 under responsibility of PSE AG.
Final manuscript published as received without editorial corrections.
Keywords: Mobile robotics, Maintenance, Non-destructive inspection, Thermosolar

Selection and peer review by the scientific conference committee of SolarPACES 2013 under responsibility of PSE AG.
Final manuscript published as received without editorial corrections.
doi:10.1016/j.egypro.2014.03.192

1. Introduction

MAINBOT project is developing service robots applications to autonomously execute inspection tasks in extensive industrial plants on equipment that is arranged horizontally (using ground robots) or vertically (climbing robots). MAINBOT aims at using already available robots to deploy innovative solutions in order to fulfil project industrial objectives: to provide a means to help measuring several physical parameters in multiple points by autonomous robots able to navigate and climb structures, handling sensors or special non destructive testing equipment.

Nomenclature

PT	Parabolic Through collector
CR	Central Receiver
SCA	Solar Collector Assembly
SCE	Solar Collector Element
NDT	Non Destructive Test
HTF	Heat Transfer Fluid
FA	Functional Analysis
FMEA	Failure Modes an Effects Analysis
ROS	Robot Operating System
GPS	Global Positioning System
INU	Inertial Navigation Unit

To define the requirements of this type of industries two validation scenarios have been selected, a Parabolic Through collector technology (PT) solar plant (50Mw, seven hours of storage) and a Central Receiver technology (CR) solar plant (19.9 Mw, fifteen hours of storage) shown in Fig. 1. Both plants pose strong challenges in terms of the number of elements to be inspected, the size of the elements, the working conditions, etc. Some figures can present an idea of the magnitude of the problem in extensive plants:

- 400.000 mirrors, with a total of 1.200.000 m2 of surface in PT.
- 2.650 heliostats (10 meters high and 11 meters width) with 35 mirrors in CR.
- About 90km of absorber tubes to be inspected (180 km) in PT.
- A tower of 140 m, at 120m receiver tubes area of 11m height

a b

Fig. 1. Solar plants used for validation; (a) parabolic through, (b) tower

Based on a set of selection criteria (positive impact in plant, novelty, feasibility, risk), several operations, to be performed autonomously by the robots, are selected:

- **Ubiquitous sensing.** The reflectivity index of the plant is a parameter of paramount importance in order to decide the cleaning and maintenance activities. Measurement of reflectivity is taken by a special purpose sensor, the reflectometer. A global field reflectivity index is obtained statistically using specific measurements in chosen mirrors from the solar field. The ground robot places a reflectometer on the specific points of the SCE, touching the mirrors and recording data.

- **Leakage detection**. In PT plants, Heat Transfer Fluid (HTF) circulates at high temperature (around 390ºC) inside the absorber tubes. HTF leakages are no desirables because oil must be replaced and this operation needs to put the SCE's out of service during two or three days. Robots using thermography inspection techniques are performing this detection.
- **Surface defects detection in vertical structures**. In CR plants a receiver located at the top of a tower heats molten salts. The receiver is a polyhedral structure composed of several panels of pipes. Receiver pipes have an external coating in order to improve radiation absorption. This coating has a thickness of microns. The climbing robot moves on top of those panels performing eddy current inspection, to assess the status of the coating by measuring its thickness. Moreover, a visual camera records external surface to detect loss of coating.
- **Surface defects detection in horizontal structures**. It is estimated that 2% of the mirrors must be replaced every year, and 0,83% mirrors are permanently broken in the plant. Ground robots in the plant look for broken mirrors since early detection can contribute to improve this efficiency. In addition, the ground robot patrolling at night and using thermography inspection is used to identify any kind of loss of vacuum in absorber tubes.
- **Internal defects detection**. Detection of corrosion and internal defects in general (cracks, etc.) is required in many components in a power plant. The climbing robot will test the presence of this kind of possible defects in the collector tubes.

This paper is structured as follows, section 2 shows the design requirements and a description of the robots used in the project, section 3 explains in detail all the basic technologies developed to endow MAINBOT robots with the capabilities to autonomously perform maintenance activities, section 4 explains the different NDT approaches selected to detect degradation problems in industrial plants. Finally section 5 provides the main conclusions and future work.

2. Robot design

In MAINBOT new robotic platforms are re-designed considering all the requirements defined in the application scenarios and using previously existing platforms as a starting point. Table 1 shows a summary of the requirements considered.

Table 1. Requirements summary.

Task	Applies to	Robot Requirements	Inspection requirements
Ubiquitous sensing	Ground Robot	(1) Precise positioning to reach specific points (2) Localization (15 points in a SCE, 33 SCA) (3) Obstacle avoidance strategies	(1) On line inspection (2) Accurate sensor placement on top of a surface (3) Trajectories (reach points, Linear in Loops, Terrace)
Leakages	Ground Robot	(1) Localization (joints, sensor field of view) (2) Trajectories (reach joints, Linear in Loops, Terrace) (3) Obstacle avoidance strategies	(1) On line inspection (2) Simultaneous inspection (visual servoing)
Surface defects	Climbing Robot & Ground Robot	(1) Access to vertical areas and autonomous movement and securing at object. (2) Non contact sensor manipulation (3) Contact sensor manipulation	(1) On line inspection (2) Accurate guiding of visual camera along surface (e.g. tubes of receiver) without contact (3) On line eddy current inspection with constant contact forces
Internal defects	Climbing robot	(1) Contact based sensor manipulation (2) Localization (sensor field of view)	(1) On line inspection (2) Accurate guiding of NDT sensors along surface (e.g. tubes of receiver) with constant contact

	(3) Trajectories (continuous, according to scanning, Terrace)	forces

In addition, a Reliability, Availability, Maintainability and Safety (RAMS) methodology, applied to the robots re-design, has been followed. Based on the validation scenarios several analysis have been performed and both, hardware and software levels have been considered.

- **Functional Analysis.** The Functional Analysis (FA) is a top-down structured and systematic evaluation of both robot types. It is a qualitative method to identify and analyze all the functions related to the systems and subsystems integrated into each robot. The purpose is to assure that the robot does not cause or contribute in a significant way to personal injuries and/or material damages. This approach is combined with the design FMEA approach to obtain a list of potential Failure Modes, with their consequences and the existing / proposed mitigations.
- **Reliability and Maintainability Analysis.** The objective is to calculate or predict the reliability of a robot at different stages during their design. Once the distributions for the Reliability and the Maintainability of each component have been calculated, simulation is used to calculate the Availability (A) of the whole robot. To evaluate system availability Mean Time Between Failures (MTBF) and Mean Time To Repair (MTTR) are considered. Combining Reliability Block Diagram (RBD) representation with Monte-Carlo simulation allows evaluating Availability when there are complex configurations (based on time dependant distributions).

Based on all these requirements ground robot and vertical robot have been re-designed.

2.1. Ground robot re-design

As illustrated in Fig. 2, the ground robot is built around a rigid mechanical structure adapted to the off-road navigation. The ground robot is an electrically driven, 4 wheel-drive / 4 wheel-steer mobile robot base and has a very good clearing capacity, offering a real solution for reconnaissance, monitoring and safety operations while minimizing human risks. It uses a hydropneumatic suspension capable of absorbing high and low frequency vibrations induced by the ground.

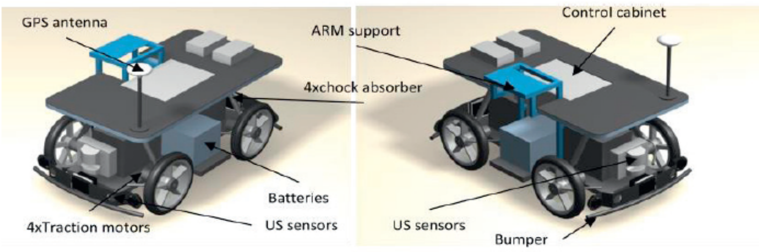


Fig. 2. Ground robot description

2.2. Climbing/vertical robot re-design

The MAINBOT approach for the is to design the climbing robot for maintenance tasks of large plants not from the scratch but using and adapting a existing robot design [1]. The physical structure of the climbing robot is shown in Fig. 3.

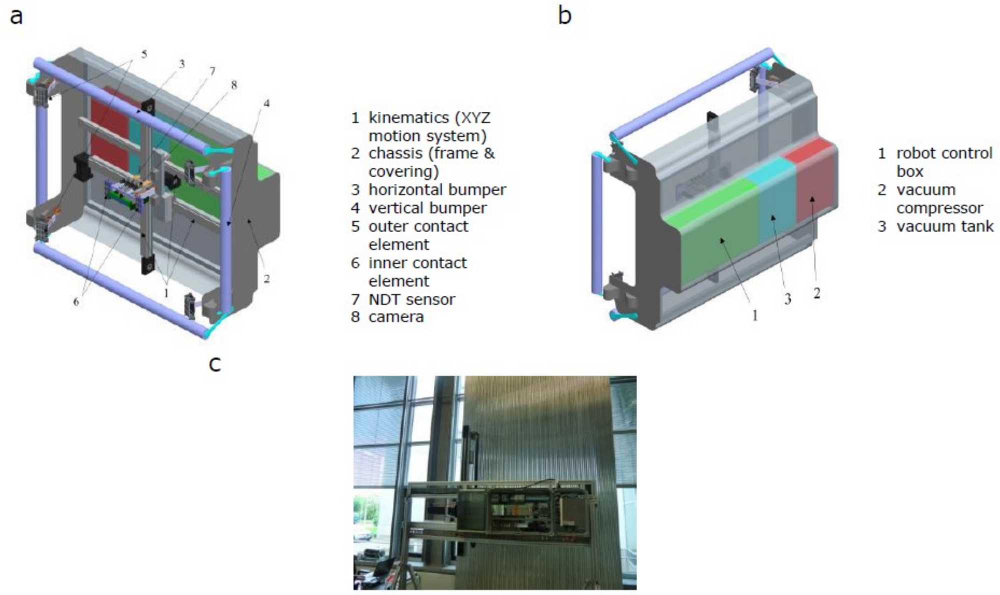


Fig. 3. Climbing robot; (a) front view, (b) back view, (c) prototype in a mockup

3. Autonomous navigation and manipulation

The aim is to endow the robots with the capability to autonomously navigate and manipulate in unstructured environments.

3.1. Simulation

3D simulation is a powerful tool for exploring what-if scenarios and for providing valuable information before developing real prototypes. Benefits of using 3D simulation in robotics come from: the ability to develop robotic applications without hardware dependency and the cost reduction.

In MAINBOT realistic testing environments have been created in order to validate the algorithms developed using available robot, sensors and application environments. A 3D simulation is used for evaluating the use and implementation of the ground mobile base and the manipulator in maintenance activities. After the analysis of state of the art simulation environments such as MORSE [2], Webots [3], UsarSim [4], finally Gazebo [5]has been selected. Using different CAD files representing the environment (the validation scenario) the ground robot provided by the robot manufacturer and the arm, a Gazebo model has been built for the complete ground robot system as shown in Fig. 4.

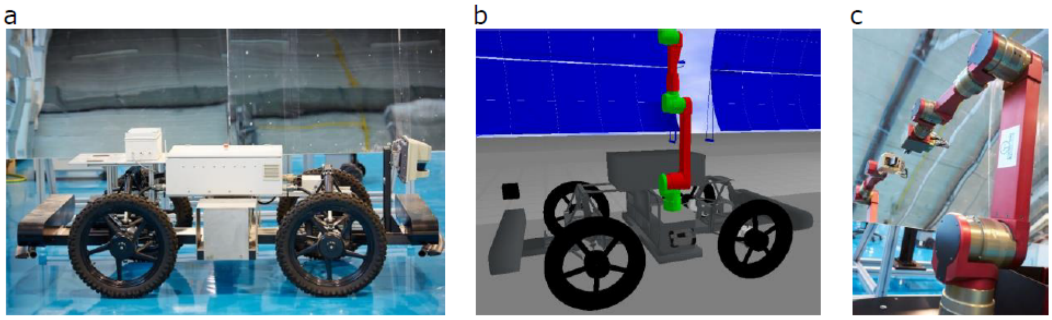


Fig. 4. Simulation environment; (a) Ground robot base, (b) Simulated ground robot, (c) Ground arm

3.2. Navigation

The navigation algorithms integrated in MAINBOT are designed to build a representation of the environment and generate robot trajectory plans considering the requirements of maintenance activities. Thus, several navigation strategies are considered, by the one hand, for ground navigation and on the other hand for vertical navigation.

In the case of ground robot, to navigate, the robot needs a representation of the environment given as a map. Navigation techniques heavily depend on the type of maps used. Most current localization systems use global, metric maps of the workspace. While convenient for small areas, global metric maps have inefficiencies of scale and in addition, are sensible to inaccuracies in both map-making and odometry performance of the robot. Because of the abovementioned reasons, it is becoming common in the field of mobile robotics to use hybrid maps that integrate different representations. Several authors [6] [7] propose combining the topological and the metric paradigm and they have shown that characteristics of both can be integrated. In MAINBOT the approach is to use a hybrid map consisting of a topological graph overlaid with local occupancy grids. As MAINBOT scenario is a completely outdoor scenario the localization problem is solved integrating a dGPS system, an RTK 2 (0.02 m accuracy) with a GPS/INS (inertial navigation system) to overcome the lack of information when GPS signal is lost.

The general localization approach in the vertical robot was adapted to the polygonal shape of the receiver scenario. Hereby the robot position relative to the tube panels is important for conducting accurate measurements with the onboard sensor equipment. The sensor localization depends on the position of the crane, the cable winches and the robot movement influenced by external facts such as wind forces and (contact) forces due to interaction with the panels. For the operation of the climbing robot at vertical structures global and local localization functions are considered:

- Global positioning at the receiver above the selected panel and at the right tower height for docking the robot to the panel and for climbing at the panel.
- Local positioning of the NDT sensors and cameras at the receiver tubes relative to the robot at the tubes generatrix and maintaining distance to the panel constant.

3.3. Manipulation

Robots working in unstructured environments have to be aware of their surroundings, avoiding collisions with any kind of obstacles. The manipulation algorithms integrated in MAINBOT are designed to build a representation of the environment using a set of sensors (laser, ultrasound, vision) and generate robot trajectory plans considering the requirements of maintenance activities. Thus, several planning strategies are considered:

- Planning arm movements with collision avoidance.
- Relative arm movements guided by sensory input.

These strategies allow providing mechanisms in order to perform inspection activities such as positioning of a inspection elements on a surface or tracking an element based on input coming from the inspection system.

3.4. Interoperability

The interoperability of the heterogeneous elements in the robotic solution is guaranteed by a general architecture on top of ROS [8] middleware that facilitates the organization, the maintainability and the efficiency of the software.

Fig. 5 shows the general overview of the robotic architecture: the robot receives as input maintenance tasks called missions either from the user interface (GUI) or an external application (End User). Afterwards, the Manager is in charge of decomposing the missions into tasks that can be performed by the Robotic Components. As explained before, the maintenance robotic system consists of several subcomponents like a mobile base, a manipulator and inspection sensors and systems, etc. During mission execution the Inspection Systems provide feedback about the status of the plant facilities.

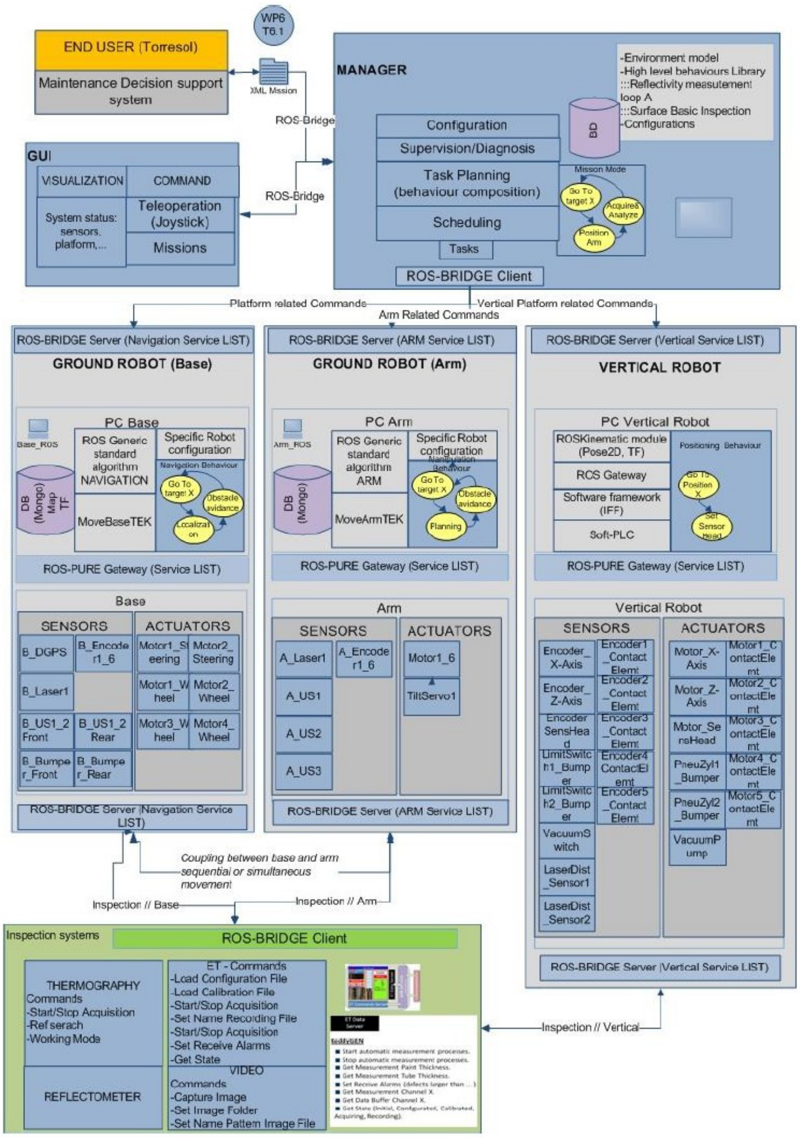


Fig. 5. System architecture

4. Non-destructive inspection techniques

Non Destructive Test (NDT) techniques are used to assess the different degradation problems to be tackled in an industrial plant: surface defects, leakages and internal defects.

4.1. Eddy current technology

Eddy current inspection is one of several NDT methods that uses "electromagnetism" as the basis for conducting examinations. Eddy current technique allows measuring coating thickness or tube thickness. Existing NDT instrumentation has been selected and adapted to MAINBOT requirements. Two types of ET sensors have been designed and manufactured: low frequency coils (to operate around 1 KHz) and high frequency coils (to operate around 1.5 MHz) shown in Fig. 6. The sensors are protected with a sapphire layer, to protect both tube paint layer and sensor surface. Raw eddy current data is processed on line, and compared with previously calibrated data.

Visual inspection is simultaneously carried out to detect external degradation in the tubes. This information is combined with eddy current data (data fusion).

Fig. 6. Eddy current sensor coils

4.2. Thermography technology

When a tube is degraded, and there is a vacuum loss, a gradient of temperature can be detected. The algorithm performs several operations on the thermal images acquired with a thermographic camera (FLIR ThermoVision® A20) that is mounted on the ground robot manipulator in an eye-in-hand configuration. Fig. 7 shows an example of the thermal image when vacuum lost is detected. The detection algorithm works in real time coordinated with the ground robot manipulator movements. The tracking algorithm calculates the arm movements in order to hold the tube in the field of view of the thermal camera.

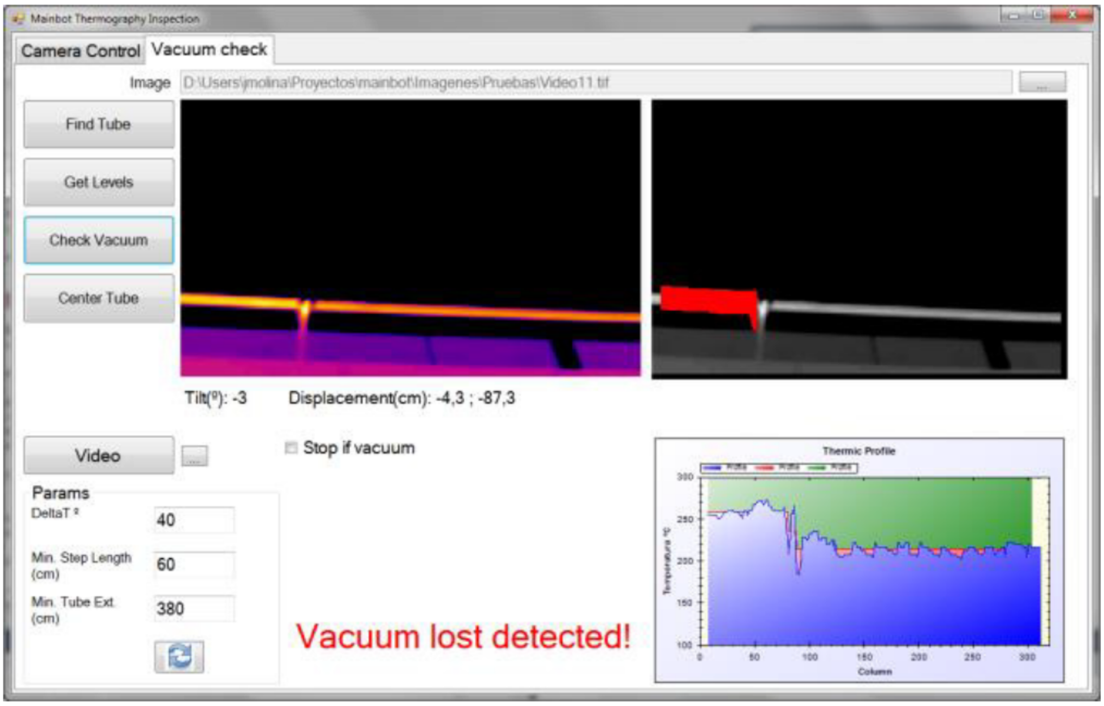


Fig. 7. Results of thermography inspection system, detecting vacuum lost

5. Conclusions

Efficient and effective maintenance is crucial for all kind of industries. In the case of capital intensive investment industries it is even more relevant and has an important impact in the operation costs during the long life cycle of their production means.

MAINBOT proposes using service robots to autonomously execute inspection tasks. A set of application scenarios that cover the general requirements of the maintenance activities in large industries have been selected.

Two kind of robotic solutions are developed in MAINBOT. Ground robot, a mobile manipulator composed of a mobile base and a 6DOF manipulator. The ground robot has to move in a large area, the solar field, and it has to reach different inspection areas in the plant and stop at pre-established points.

The vertical robot consists of a mobile base and an internal arm for inspection system positioning. The climbing robot has to move in a vertical structure, a tower.

MAINBOT is developing technologies for ground autonomous navigation and manipulation. Eddy current and thermography based algorithms have been developed and integrated in robotic platforms.

In the near future the validation of the development will be performed in two solar plants in the south of Spain.

Acknowledgements

This work has been partially funded by the European Community"s Seventh Framework Programme (FoF.NMP.2011-3) under grant agreement no 285245.

References

[1] M. Sack, N. Elkmann, T. Felsch and T. Bohme, "Intelligent control of modular kinematics - the robot platform SIRIUS," 2002.

[2] "Morse," [online] Available at: http://www.openrobots.org/wiki/morse/.

[3] "Webots," [online] Available at: http://www.cyberbotics.com/overview.

[4] "USARSim," [online] Available at: http://sourceforge.net/apps/mediawiki/usarsim/index.php?title=Main_Page.

[5] "Gazebo," [online] Available at: http://playerstage.sourceforge.net/gazebo/gazebo.html.

[6] E. M.-E. ,. B. M. Kurt Konolige, "Navigation in Hybrid Metric–Topological Maps," IEEE International Conference on Robotics and Automation,ICRA 2011, Shanghai, China, 9-13 May 2011, pp. 3041-3047, 2011.

[7] I. N. Nicola Tomatis, "Hybrid simultaneous localization and map building: closing the loop with multi-hypothesis tracking," vol. 3, 2002.

[8] ROS, "ROS," [Online]. Available at: http://www.ros.org/wiki/.

7.19. Robotic solutions for Footwear Industry

Iñaki Maurtua, Aitor Ibarguren, Alberto Tellaeche: Robotic solutions for Footwear Industry. ETFA 2012: 1-4, Cracovia [115]

Robotic Solutions for Footwear Industry

Iñaki Maurtua
IK4-Tekniker
c/Iñaki Goenaga,5
20600 Eibar, Spain
imaurtua@tekniker.es

Aitor Ibarguren
IK4-Tekniker
c/Iñaki Goenaga,5
20600 Eibar, Spain
aibarguren@tekniker.es

Alberto Tellaeche
IK4-Tekniker
c/Iñaki Goenaga,5
20600 Eibar, Spain
atellaeche@tekniker.es

Abstract

Since September 2010, the ROBOFOOT consortium, a group of 10 partners, including Footwear Industry, Research institutes and Robotic solution providers is working together to promote the introduction of robotics in the European Footwear Manufacturing Industry. This paper presents the initial results achieved, in particular they are described the user requirements and operations selected and the technical achievements reached so far.

The approach followed allows the coexistence of current working practices and facilities with robotic solutions. Due to the nature of the industry, either in size and financial capability, it has been identified as one of the requirements by end-users taking part in the project.

1. Introduction

Even if less known than other sectors, footwear industry is still relevant in Europe [1], in fact it includes more than 26.000 companies and almost 400.000 employees. This relevance is higher in some Mediterranean countries, in particular Spain, Italy and Portugal. However the trend shows a clear decline on business figures; low cost countries are becoming an obvious threat for the future of the sector.

Fig. 1: Manual cutting of leather parts

Fashion Footwear production is currently mainly **handcrafted**. Some manufacturing processes are assisted by specialized machinery (last manufacture, cementing, and cutting) and there exist highly automated lines in mass production of technical shoes (i.e. safety footwear). But most production is still handmade, being especially true in the case of **high added value shoes** production, where Europe maintains its leadership. This kind of production accounts for some of the shortest production runs to be found (eight pairs of shoes is the average order size).

The use of robots is limited to the above mentioned mass production of technical shoes. The main reasons that justify the extensive labour demand and this lack of robotic applications and automation are:

- The high number of products ***variants.***

On the one hand every year a minimum of two different collections (summer & winter) of shoes, sandals, boots, etc. are developed to be presented to the customers. As an average, more than 200 different **models** are manufactured for the two seasons.

On the other hand, it is necessary to adapt each model to at least six different **sizes** and two sides (left and right).

Finally, we have to take into account that each model can be manufactured in different **leather qualities** and for each quality in different **colours**.

- ***Complex manufacturing*** process.

For each model it is necessary to develop and manufacture the last (the rough form of a human foot used in shoemaking to provide the fit and style of the shoe), to produce the list of components (sole, heel, sock, strap, inner parts, etc.), to cut all the components of the uppers, to prepare the leather by means of several operations such as skiving, folding o perforating, to stitch the different parts that form the upper.

- ***Complex assembly*** process.

The assembly process is very laborious (up to 25 different operations) and especially complex in fitting operations due to the non uniformity and the different elasticity of the natural leather as well as the non-rigid nature of the components that difficult their manipulation. Finally each pair of shoes requires a final inspection and packaging.

Although some companies in this sector tried in the past to incorporate robotic solutions, they did not succeed in the objective except for specific operations related to the injection process, as stated before. The EU co-funded **ROBOFOOT** project is developing different solutions to facilitate the introduction of robotics in traditional footwear industry**.**

This paper presents some initial results, starting with the user requirements presented in Section 2 and the Operation selection in Section 3. First technical results are described in Section 4 with more detailed information on visual servoing achievements. Finally Section 5 summarizes the re-design of the manufacturing process.

2. User requirements

The project consortium has done a deep analysis in order to understand current practices and identify the main needs of the sector. The process has included reviewing several studies, the internal analysis of the two end-user partners ROTTA and PIKOLINOS and the contribution of the rest of partners that visited both manufacturing plants and participated in the decision making process providing their technical background.

- **Quality**: currently, as an average, 80% of the shoes need some kind of retouching at the end of the line. As most of these small faults are due to the low stability of some processes it is expected that the use of robots will contribute to reduce these retouching operations.
- **Impact in current production process**: The current state of the art in robotics and the complexity of some operations make unviable a full robotic automation. A basic requirement is the possibility to combine current production procedures with the robotized solutions proposed by ROBOFOOT. This includes the coexistence of manual operations with robotized ones and the reuse of current production means.
- **Efficiency**: reduction of manufacturing time. It should be taken into account that the robotized production has to be integrated in the production line where 'traditional' production means coexist. So, reduction of individual operation time cannot be considered an objective unless we consider this reduction the foundation for combining two operations.
- **Production flexibility**. ROBOFOOT has to guarantee the production flexibility, handling a wide variety of models/sizes coexisting in the production line and allowing frequent model changes
- **Reduction of costs**. Although it is not the main reason for introduction of robotics in this sector, it will allow some workers to do tasks with higher added value and overcome the lack of skilled workers for some operations.
- **Working conditions**. Currently there are several operations that involve potential risk for workers (dust, solvents, rotating parts, effort...). Introducing robots has to help in reducing the potential risk of those operations to the minimum.
- **Usability and maintainability**. The system has to be easy to use and maintain by no specialists.

3. Operation selection

According to the criteria established and the analysis of operations, it has been established a ranking of operations. They have been scheduled in the timeframe of the project and grouped in 3 prototypes:

- **Basic prototype**. It includes individual operations: Roughing, Gluing, Inking, Polishing, Last Manufacturing.
- **Intermediate prototype**. They correspond to some operations that can be combined in the same robotic cell: Roughing+ Gluing; Inking+ polishing+ last removal. Inspection is included as well, in order to detect the presence of nails, assess gluing and roughing processes and identify surface defects on the shoe.
- **Final prototype**: It corresponds to the most challenging operation, i.e. packaging. A complete robotized cell has been designed, although only some of the sub-operations will be implemented in the context of ROBOFOOT.

4. Main technologies develop in ROBOFOOT

4.1. Robot programming and controlling

The general approach for robot programming and controlling is as follows.

The starting point is the 3D information of the shoe. It can be obtained from the CAD or from a digitalization of the shoe mounted on the Last. Based on that, technicians define the trajectories for the different operations, including the technological features.

The resulting file is the input for the postprocessor that generates the robot program in COMAU's PDL2 language.

To overcome the problem of minor misalignments between the Grasping Device and the last it is necessary the correction of the resulting program. Finally in several operations, such as roughing, real time adjustment of trajectories is needed.

- *Automatic robot part program generation from digital data.*

  INESCOP has developed a CAM system to define the tool-path that has to be followed by the robot as well as the process parameters.

  The geometry is imported from the virtual model or from a digitalized model. Based on that, the user can define the area to be processed and the parameters to be used. After the post-processing of the resulting file, the program executable by the robot is generated.

However even a small assembly error between the GD and the last may lead to a significant difference between the theoretical and real tool trajectory – especially at shoe tip.

To overcome this problem it has been developed a system for online adaptation of off-line robot program.

- *On-Line smart adaptation of off-line automatic robot part program*

  CNR-ITIA has developed an innovative solution for on-line smart adaptation of off-line automatic robot part program generated from digital data (CAD/CAM) through the use of on-line sensors namely a new generation of modular, scalable and reconfigurable 3D laser scanners.

  To this aim, a 3D laser scanning device has been conceived and developed to fine-tune the program generated in the previous step by adding the real positioning of the GD with respect the Last. This operation is done only once for each pair of last and GD.

- *Manual Guidance Device*

  This wireless device can be place anywhere after the sixth axe of the robot. Operators can control the movements of the robot in a very intuitive way using the MGD. The main application in the context of the project is the correction of some points and the maintenance activities.

Fig. 2: Manual guidance device (Developed by CNR-ITIA for COMAU)

- *Sensor based robot controlling*

  Finally, we have to consider that for some operations it is needed to correct or generate the trajectory of the robot in real time. This is the case of operations like roughing, where it is important to guarantee that the leather is not damaged.

  CNR-ITIA is developing a real time robot trajectory adaptation mechanism based on force control and using the C4G Open feature that allows the possibility to change several robot parameters on-line.

On the other hand, visual servoing control has been developed to identify the pose of the shoes in the manovia and generate the robot trajectory to pick them up. To this aim, a vision system in an eye-in-hand configuration has been introduced.

- *Visual servoing*

Shoes go from one working station to another on trolleys that are placed in the Manovia. For each shoe on the trolley, the robot has to take it, manipulate it in the workstation (roughing, gluing,...) and finally leave it back on the trolley. Due to the fact that we are combining manual and robotized operations, it is not guaranteed an accurate positioning of the shoe on the trolley.

To overcome this problem and achieve a precise grasping of the shoe, a visual servoing system has been developed. The system uses images to estimate the pose of the Grasping Device attached to the shoe and based on those estimations the robot corrects its pose until the desired position is reached. It is necessary to perform the grasping task with a precision of around 1 millimetre and 1-2 degrees in each axis to avoid damages in the shoe. The manoeuvre should not take more than 5-6 seconds to maintain the production time cycle.

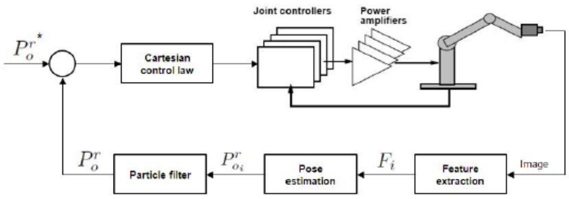


Fig. 3: Visual servoing schema

The pose identification is difficult due to the industrial nature of the scenario; problems like the poor illumination of those environments, the metallic nature of the Grasping Device which makes difficult a proper illumination, as well as the wax and ink used during the shoe making process that can be adhered to the Grasping Device. All the above facts make it difficult to acquire good quality images to estimate the pose of the shoe.

TEKNIKER has developed a 6DOF visual servoing system using a dynamic look-and-move approach. This system includes a particle filter to deal with the uncertainties (illumination, dirt in the Grasping Device…) of the image acquisition process. The process followed by the system can be described as:

- Feature extraction from the image: the image is analyzed, trying to find features of the Grasping Device, features as holes and edges. Due to the uncertainties of the images it is difficult to determine exactly the position and dimensions of those features. To overcome this problem, different hypothesis have been extracted for each feature, determining possible positions and dimensions of the holes and edges.

- **Pose estimation** of all possible hypotheses. Based on the features extracted in the previous step, different poses are estimated, each of them related with a hypothesis.
- **Fusion of those hypotheses by means of the particle filter** to get the final pose estimation. The particle filter uses the estimated poses to determine which of the particles (hypothesis) are the most suitable for the given state and uses this distribution to improve the estimations in further steps.
- **Robot movement** based on the estimated pose. The robot moves in the workspace, trying to correct the orientation and distances.

Based on the described loop, the system corrects iteratively the robot position until the desired position is achieved, as shown in Fig. 3. When the final position is reached, the robot will be able to grasp the lasted shoe in a safe and accurate way. The performed experiments show good performance of the system, reaching high success rate and fitting in the time constraints exposed previously.

4.2. Manipulation

ROBOFOOT considers two different problems: manipulation with the Last and manipulation without Last, just the shoe.

In the first case it has been necessary to modify the current lasts by introducing an external element, the Grasping Device (GD) that allows grasping the last with the required rigidity and repeatability. This modification has been done in such a way that Lasts are compatible with existing manufacturing machines at end-users facilities.

The vision system makes possible the identification of the pose of the lasts both on the manovia and when they lie on the exit of a chiller.

In the second scenario (shoes without lasts) the deployment of a simple parallel gripper would damage the shoe and special grippers (for instance suction) would only be helpful for some shoe types and materials. DFKI is developing a bimanual multifingered robotic approach to achieve this objective, based on the AILA [3] robot and the iCub hands [4].

5. Manufacturing process re-design

Three manufacturing robotic cells have been designed and implemented:

- **Roughing, gluing and last milling**: A multifunctional robotized cell for bottom and side roughing, gluing and last milling has been conceived by QDESIGN.
- **Polishing, inking and last removal:** A robotic cell has been designed by AYCN. This cell integrates a robot that takes the shoes from the ex**it of the chiller, does the inking** and polishing processes and, finally, opens the last for manual removal of the shoe by the operator.
- **Packaging:** a robotized packaging cell is being developed.

5.1. Quality assessment

The project is developing visual inspection techniques for different operation assessment: roughing (to control the boundaries of the roughed area and to verify that the leather has not been damaged due to over-roughing), gluing (detect the excess/lack of glue in the shoe) and nail removing steps (to detect that there are not nails left in the sole of the shoe once the operator has done this operation manually). Surface defects (cuts, scars, colour irregularities, etc.) are tackled as well.

6. Future work and acknowledgments

The project will last until February 2013. The last months will be devoted to experimental validation of the robotic solutions for each operation and the implementation of the packaging system.

This work has been performed within the scope of the project “ROBOFOOT: Smart robotics for high added value footwear industry “. ROBOFOOT is a Small or Medium- scale focused research project supported by the European Commission in the 7th Framework Programme (260159). For further information see http://www.robofoot.eu.

References

[1] “EUROPEAN INDUSTRY IN A CHANGING WORLD- UPDATED SECTORAL OVERVIEW 2009”, EU Commission, 2009.

[2] Bessey, E. et al., “Research, Technology and Development for Manufacturing”. *The ManuFuture Road: Towards Competitive and Sustainable High-Adding-Value Manufacturing*. Springer, 2009.

[3] Lemburg, J., de Gea Fernandez, J. et al., “AILA - design of an autonomous mobile dual-arm robot“,2011 IEEE International Conference on Robotics and Automation (ICRA), pp 5147-5153, 2011.

[4] Schmitz, A. et al., “Design, realization and sensorization of the dexterous iCub hand”, 2010 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp 186 – 191, 2010

7.20. Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments

Aitor Ibarguren, José María Martínez-Otzeta, Iñaki Maurtua: Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments. ICINCO, 9th International Conference on Informatics in Control, Automation and Robotics (2) 2012: 161-166. Roma [96]

Particle Filtering for Position Based 6DOF Visual Servoing in Industrial Environments

Aitor Ibarguren, José María Martínez-Otzeta and Iñaki Maurtua

***Abstract*— Visual Servoing allows the introduction of robotic manipulation in dynamic and uncontrolled environments. This paper presents a Position Based Visual Servoing algorithm using Particle Filtering. The objective is the grasping of objects using the 6 degrees of freedom of the robot manipulator (position and orientation) in non-automated industrial environments using monocular vision. A Particle Filter has been added to the Position Based Visual Servoing algorithm to deal with the different noise sources of those industrial environments (metallic nature of the objects, dirt, illumination problems). This addition allows dealing with those uncertainties and being able to recover from errors in the grasping process. Experiments performed in the real industrial scenario of ROBOFOOT[1] project showed accurate grasping and high level of stability in the Visual Servoing process.**

I. INTRODUCTION

Traditional industrial robotic applications, like part placement or spot welding, require precise information about the position of the objects to perform their task. Visual Servoing [1], [2] can enhance those industrial applications allowing corrections on the robot trajectories. This technique would help to introduce robots in applications where the workpiece is moving or is placed in the working area in an unknown pose.

Even so, industrial environments raise their own challenges in the inclusion of Visual Servoing techniques, especially when the production line is not completely automated. Dirt, imprecision in the workpiece placement, changing lighting conditions are some of the problems that must be tackled in this kind of environments, introducing uncertainties in the trajectory correction process.

Particle Filtering [3], [4], a sequential Monte Carlo algorithm, is a suitable choice to deal with uncertainties in the observed data. This technique offers a reliable way to estimate unknown states based on an observation sequence, as they are able to deal with multiple hypothesis in a simple and effective way. Due to those features Particle Filters are a suitable technique that can be efficiently introduced to deal with the uncertainties identified in industrial environments.

This paper presents a Position Based Visual Servoing algorithm using Particle Filtering. Based on the real industrial scenario of ROBOFOOT project, the paper proposes an algorithm to grasp a workpiece (a shoe last specifically) from a not constrained workshop, correcting the 6 degrees of freedom of the robot during the Visual Servoing process.

A. Ibarguren, J.M. Martínez-Otzeta and I. Maurtua are with Fundación Tekniker, Apdo. 44 Otaola 20, 20600 Eibar, Gipuzkoa, Spain aibarguren@tekniker.es

[1] http://www.robofoot.eu/

The paper is organized as follows. In Section II the related work is presented. Section III exposes briefly Particle Filters. Task specification and configuration is shown in Section IV. Section V is devoted to the proposed approach, while in Section VI the experimental results are shown. Finally, Section VII presents the conclusions as well as the future work to be done.

II. RELATED WORK

Several approaches tackle the use of Visual Servoing in industrial environments, posing different industrial scenarios and approaches.

Sung-Hyun et al. [5] propose an image-based Visual Servoing based on stereo vision. The use of stereo vision allows guiding the robot manipulator to the desired location without giving such prior knowledge about the relative distance to the desired location or the model of the object.

Nomura et al. [6] describe a Visual Servoing system able to track and grasp industrial parts moving on a conveyor using a 6DOF robot arm. A hybrid Kalman Filter is also incorporated to track a moving object stably against visual data noise. Experiments are also presented, performing both 3DOF and 6DOF Visual Servoing.

Finally, Lippiello et al. [7], [8] presented Visual Servoing applications on Industrial Robotic cells. On their setup, composed of two industrial robot manipulators equipped with pneumatic grippers, vision systems and a belt conveyor, a position-based Visual Servoing is proposed. The system also uses Extended Kalman Filters (EKF) [9] to manage the occlusions during the multi-arm manipulation.

III. PARTICLE FILTER

Particle filters, also known as sequential Monte Carlo methods (SMC), are sequential estimation techniques that allow estimating unknown states x_t from a collection of observations $z_{1:t} = \{z_1, ..., z_t\}$. The state-space model is usually described by state transition and measurement equations

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{1}$$

$$z_t = g_t(x_t, u_t) \tag{2}$$

where f and g are the state evolution and observation model functions respectively and v_t and u_t denote the process and observation noise respectively.

Based on the previous equations, particle filters allow approximating the posterior density (PDF) by means of a set of particles $\{x_t^{(i)}\}_{i=1,\dots,n}$ using equation

$$p(x_t|z_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{3}$$

where each particle $x_t^{(i)}$ has an importance weight $\omega_t^{(i)}$ associated and δ is the Kronecker delta. These weights are computed following equation

$$\omega_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(z_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})} \tag{4}$$

where $p(z_t|x_t^{(i)})$ is the likehood function of the measurements z_t and $q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})$ is the proposal density function.

Based on the previously presented equations the particle set evolves along time, changing the weights of the particles and resampling them in terms of the observations.

Particle filtering provides a robust tracking framework when dealing with non-linear and non-gaussian state and observation functions as it considers multiple state hypotheses simultaneously.

IV. TASK SPECIFICATION AND CONFIGURATION

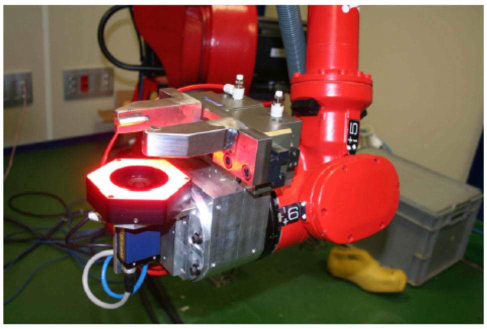


Fig. 1. Gripper, camera and lighting system

Based on the needs of ROBOFOOT project, whose aim is the introduction of robots in the footwear industry, an object grasping task has been designed using real specifications of footwear workshops. One of the principal aspects of the project is the focus on minimizing the impact of the introduction of the robots in the existing production means, nowadays basically handcrafted in high added value shoe production. Taking it into account, the grasping scenario has been specified as:

- A 6DOF robot arm with a gripper and a camera and lighting system mounted on the end-effector with an eye-in-hand configuration, as seen in Fig. 1.
- Lasts, material with the shape of a foot used to build shoes, are the object to be grasped. An iron piece (*grasping device*) has been added to lasts to allow a precise and stiff grasping, see Fig. 2. Those *grasping devices* will be the objects to be identified during the Visual Servoing process.

Fig. 2. Lasts with the *grasping device* on the trolley

- Lasts are carried in the manovia, specific trolley mounted on a conveyor. The trolleys are designed to allow the placement of lasts of different shapes and sizes and in the same way lasts are placed in the trolley by human operators. Due to those previous facts it is not possible to know the pose of the last in the trolley, as seen in Fig. 2. Therefore, it forces the Visual Servoing system to correct its 6DOF to grasp the workpiece.
- Based on the design of the gripper and the *grasping device*, the grasping process requires a precision of around a millimeter and 1-2 degrees on each axis. In the same way the maneuver should take no more than 4-5 seconds.

Based on this scenario, the initial set-up of the system has raised some problems related with the pose estimation of the *grasping device*:

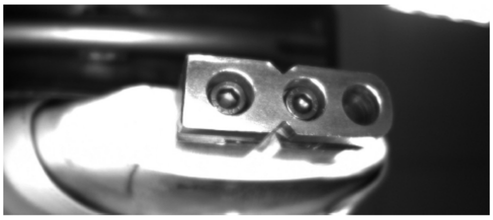


Fig. 3. Image of the *grasping devices* taken in real conditions

- Illumination is a key aspect in a vision system. In this industrial scenario is complicated to place a suitable general illumination, that is why it was decided to put a specific lighting system on the gripper. Even so, the metallic nature of the *grasping device* makes it difficult to get a good image due to the brightness, reflection and the impossibility of lighting all the image properly, as shown in Fig. 3.
- Some of the tasks to be performed by both the human operators and robots include ink, wax and other kind of possible dirt that will be adhered to the *grasping device*, as seen on Fig. 4.

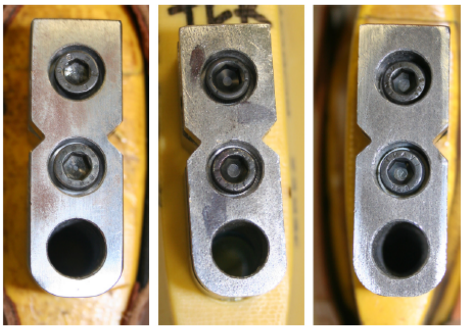


Fig. 4. State of the *grasping devices*

Those previous points will make it difficult to acquire clear images of the *grasping device*, adding uncertainties to the 6DOF pose estimation that will be the base of the Visual Servoing process.

V. PROPOSED APPROACH

Based on the previous scenario and its restrictions, this paper proposes a position-based Visual Servoing, using a dynamic look-and-move structure. Taking into account the needed precision, which implies a high resolution camera, and the demanding image processing due to the unstable conditions of the image, a look-and-move approach has been adopted.

A particle filter has also been added to the system to manage the uncertainties of the vision system. Problems on the illumination, the metallic nature of the *grasping device* and the possible dirt of the environment could be managed using a multiple hypothesis approach like a particle filter.

Next lines will describe the general structure of the system, as well as the vision module, pose estimation, the particle filter and the grasping algorithm.

A. *System Modeling and Architecture*

In the described scenario, the space can be represented by $P \in \Re^6$, a set of three positions and three orientations, where $P = [x, y, z, \alpha, \beta, \gamma]^T$. In the same way, this scenario will be composed of two different frames, the robot frame r and the camera frame c. Given those two frames, the homogeneous transformation matrix, denoted by ${}^rT_c$, transforms poses from frame c to frame r as:

$$P^r = {}^rT_c P^c \tag{5}$$

The error of the positioning task involved in the grasping process is represented by vector $E \in \Re^6$ which represents the difference between the pose of the object P_o^r in the robot frame and the pose of the end-effector P_e^r in the robot frame (6). The grasping process can be seen as a minimization of this error that will be fulfilled when $|E| = 0$.

$$E = P_e^r - P_o^r = \begin{bmatrix} x_e^r - x_o^r \\ y_e^r - y_o^r \\ z_e^r - z_o^r \\ \alpha_e^r - \alpha_o^r \\ \beta_e^r - \beta_o^r \\ \gamma_e^r - \gamma_o^r \end{bmatrix} \tag{6}$$

For pose estimation, position-based Visual Servoing systems extract features from the acquired images and estimate the pose of the object P_o^r and perform the corrections. Even so, the described scenario introduces uncertainties in the feature extraction step (illumination, metallic workpiece...), introducing errors in the pose estimation. To deal with this problem, the use of a particle filter is proposed. From each image, a set of n feature vectors $F_i = \{f_1, f_2, ..., f_m\}_{i=1...n}$ will be extracted for the pose estimation, each of them related with a specific image analysis procedure. Each of those n vectors will be a hypothesis of the values of the m features used for the pose estimation, as it will not be possible to have a unique feature vector extracted from each image due to the uncertainties in the image.

From each feature vector F_i, $P_{o_i}^c$ and $P_{o_i}^r$ will be calculated,

$$P_{o_i}^c = \mathrm{PE}(F_i) \tag{7}$$

$$P_{o_i}^r = {}^rT_c P_{o_i}^c \tag{8}$$

where $P_{o_i}^c$ is the i-th hypothesis of the pose of the object in the camera frame, $P_{o_i}^r$ is the i-th hypothesis of the pose of the object in the robot frame and pose estimation function PE is the function that relates a set of features with a pose of the object in the camera frame.

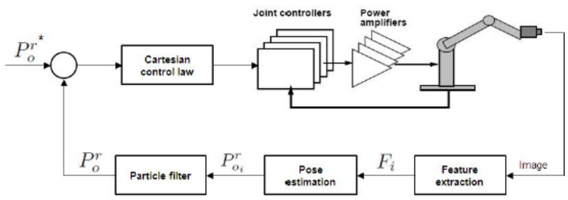


Fig. 5. Dynamic position based look-and-move structure with particle filter

Those n poses, $P_{o_{i=1..n}}^r$ will be the observations of the particle filter, which will output the final pose estimation of the object in the robot frame P_o^r. This final pose will be used to calculate the error E between the object and the end-effector, used to calculate the next robot movement. Fig. 5 shows the structure of the proposed Visual Servoing system.

Next lines will describe the feature extraction, pose estimation, particle filtering and grasping algorithm of the grasping process.

B. Feature Extraction

As stated before, one of the challenges of the presented scenario is the feature extraction for pose estimation. The metallic nature of the *grasping device* and the illumination problems make it difficult to detect the different features (edges, corners, holes) precisely. Taking also into account the perspective of the camera through the grasping process, the image features used for pose estimation, shown in Fig. 6, are:

- The center of the three holes (1, 2, 3) of the *grasping device*. Only the pixels of the center of the holes are included, excluding the size and dimensions of the holes, due to the difficulties of extracting their contour precisely.
- The inclination of the left edge (4) of the *grasping device*.

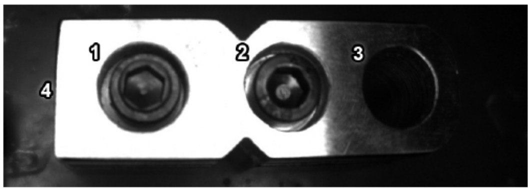


Fig. 6. Visual features for pose estimation

To detect those image features different thresholds, edge detection algorithms and filters are used. Even so, in some images it is not possible to determine the exact position of the three holes' centres as there are various possible circular shapes in each position (ex. the inner screw, outer circle and dirt around it). In those cases it is not possible to define a universal rule to determine which the real contour of the holes is. To overcome this problem, this approach proposes to use all those possible centers of the three holes (left, central and right), creating a set of hypothesis that will be used for pose estimation.

Once the centers of the holes and the left edge are detected, a feature vector will be calculated for each hypothesis as:

$$F_i = \{c_2, d_{12}, d_{13}, d_{23}, \phi_{12}, \phi_{13}, \phi_{23}, \phi_{edge}, \lambda\} \tag{9}$$

where c_2 is the coordinate in pixels of the central hole, p_i is the perimeter of the ith hole, d_{ij} is the distance in pixels between the holes i and j, ϕ_{ij} is the angle between the holes i and j, ϕ_{edge} is the angle of the left side of the *grasping device* and λ is a coefficient that measures the noise (quality) of the hypothesis based on the similitude of the circular shapes and their alignment and calculated as

$$\lambda = \frac{C_v(p_1, p_2, p_3) + \frac{|\phi_{12} - \phi_{13}| + 1}{|\phi_{13}| + 1}}{\mathrm{Min}(r_1^{xy}, r_2^{xy}, r_3^{xy})} \tag{10}$$

where p_i is the perimeter of the ith hole, C_v is the *coefficient of variation* of the perimeters and r_i^{xy} is the xy axis ratio of the bounding box of the ith hole.

Those are the features that will be used to estimate the pose of the workpiece.

C. Pose Estimation

Once the image is analyzed and the features are extracted, the pose of the object in the camera frame for each of the possible hypothesis are calculated as

$$P_{o_i}^c = [x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i]^T = \mathrm{PE}(F_i) \tag{11}$$

where each position and orientation is a quadratic function based on some of the features. The coefficients of the quadratic functions are omitted from the paper as they are related to the size of the *grasping device* and the aberration of the lens.

$$\mathrm{PE}(F_i) = \begin{cases} x_i = f(c_2, d_{12}, d_{13}, d_{23}) \\ y_i = g(c_2, d_{12}, d_{13}, d_{23}) \\ z_i = h(d_{12}, d_{13}, d_{23}) \\ \alpha_i = j(\phi_{12}, \phi_{13}, \phi_{23}) \\ \beta_i = k(d_{12}, d_{23}) \\ \gamma_i = l(\phi_{edge}) \end{cases} \tag{12}$$

Once the hypothetical poses of the object in the camera frame $P_{o_i}^c$ are estimated, the poses in the robot frame $P_{o_i}^r$ are calculated using the homogeneous transformation matrix ${}^rT_c$. Those hypothesis will be the observations of the particle filter.

D. Particle Filter

Once the possible hypothesis are calculated it is necessary to merge and fuse this information to perform the grasping process. To this end a particle filter is proposed, as it fits in this kind of non-gaussian problem.

Focusing on the posed problem, the state in time t will be defined as a pose of the object in the robot frame

$$X_t = [x_t, y_t, z_t, \alpha_t, \beta_t, \gamma_t]^T \tag{13}$$

As it is not possible to model the pose estimation error a priori, the state transition is defined as

$$X_t = X_{t-1} + V_{t-1} \tag{14}$$

where X_{t-1} is the previous state vector and V_{t-1} is the process noise.

The observation, on the other hand, is defined by a set of hypothetical poses of the object in the robot frame

$$Z_t = P_{o_{i=1..n}}^r \tag{15}$$

So based on this information source each particle will be defined by a probability $P(X_t|Z_t)$.

1) Probability: To calculate the probability of a state given an observation, initially the distance between the poses is calculated as

$$distPos_i = \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2 + (z_t - z_i)^2} \quad (16)$$

$$distAng_i = \sqrt{(\alpha_t - \alpha_i)^2 + (\beta_t - \beta_i)^2 + (\gamma_t - \gamma_i)^2} \quad (17)$$

where $distPos_i$ is the Euclidean distance between x, y and z positions of the state t and the ith hypothesis and $distAng_i$ is the Euclidean distance between the α, β and γ orientations of the state t and the ith hypothesis.

$$P(X_t|Z_t) = \prod_{i=1}^{n} e^{-distPos_i \cdot distAng_i \cdot (1+\lambda_i)} \quad (18)$$

Based on this distance, the probability of the state is calculated as the product of the exponential of the distances of all the hypothesis ponderated by the λ_i noise coefficient.

2) Particle Filtering Procedure: Finally the procedure of the particle filter is given as:

- 1: Find the *grasping device* in the initial image and initialise N particles $X_0^{(i)}$ with the different hypothesis randomly, where $w_0^{(i)} = 1/N$
- 2: if $ESS < threshold$ (*Effective Sample Size*), draw N samples with *selection with replacement*
- 3: Predict $x_t^{(i)} = x_{t-1}^{(i)} + v_{t-1}$
- 4: Replace the particles with the lowest weight by the new hypothesis found in the image
- 5: Update importance weights $w_t^{(i)} = w_{t-1}^{(i)} P(X_t|Z_t)$
- 6: Normalize weights $w_t^{(i')} = w_t^{(i)} / \sum_{j=1}^{N} w_t^{(j)}$
- 7: Set $t = t + 1$, goto Step 2

In this procedure EES [10] (Effective Sample Size) is calculated as

$$cv_t^2 = \frac{var(w_t^{(i)})}{E^2(w_t^{(i)})} = \frac{1}{N} \sum_{i=1}^{N} (Nw_t^{(i)} - 1)^2 \quad (19)$$

$$ESS_t = \frac{N}{1 + cv_t^2} \quad (20)$$

where N is the number of particles and $w_t^{(i)}$ is the weight of particle i in time t.

Based on this discrete approximation of the posterior probability, the object is tracked along the grasping procedure. The fourth step has been added to allow a fast convergence.

E. Grasping Algorithm

The feature extraction step has shown that the best images are acquired when camera is perpendicular to the *grasping device* and the end-effector to a 30mm distance, $E = [0, 0, 30, 0, 0, 0]^T$, as it solves in part the illumination problems. Taking it into account the grasping algorithm will try to minimize the error until this value is reached, adding an small tolerance of ± 1 mm in position and $\pm 1.5°$ in orientation to avoid an infinite loop. Once this error is reached the robot will make a final approach in just one axis and perform the grasping.

VI. Experimental Results

To be done.

VII. Conclusions and Future Work

To be done.

References

[1] L. Weiss, A. Sanderson, and C. Neuman, Dynamic sensor-based control of robots with visual feedback, *IEEE J. Robot. Automat.*, vol. 3, pp. 404-417, Oct. 1987.

[2] S. Hutchinson, G. Hager, and P. Corke, A tutorial on visual servo control, *IEEE Trans. Robot. Automat.*, vol. 12, pp. 651-670, Oct. 1996.

[3] A. Doucet, N. De Freitas, N and N. Gordon, Sequential Monte Carlo methods in practice, *Springer-Verlag*, 2001.

[4] J.H. Kotecha and P.M. Djuric, Gaussian particle filtering, *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, pp. 429-432, 2001.

[5] H. Sung-Hyun Han; W.H. Seo, K.S. Yoon and L. Man-Hyung, Real-time control of an industrial robot using image-based visual servoing, *Proceedings IEEE/RSJ International Conference on*, vol. 3, pp. 1762–1767, 1999.

[6] H. Nomura and T. Naito, T., Integrated visual servoing system to grasp industrial parts moving on conveyer by controlling 6DOF arm, *Systems, Man, and Cybernetics, 2000 IEEE International Conference on* , vol. 3, pp. 1768–1775, 2000.

[7] V. Lippiello, B. Siciliano and L. Villani, An experimental setup for visual servoing applications on an industrial robotic cell, *Advanced Intelligent Mechatronics. Proceedings, 2005 IEEE/ASME International Conference on*, pp.1431–1436, 2005.

[8] V. Lippiello, B. Siciliano and L. Villani, Position-Based Visual Servoing in Industrial Multirobot Cells Using a Hybrid Camera Configuration, *Robotics, IEEE Transactions on*, vol.23, no.1, pp.73–86, 2007.

[9] S.J. Julier,J.K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[10] J. Liu, R. Chen, and T. Logvinenko, *A theoretical framework for sequential importance sampling and resampling*, Technical report, Stanford University, Department of Statistics, 2000.

7.21. Robotics for the Benefit of Footwear Industry

Iñaki Maurtua, Aitor Ibarguren, Alberto Tellaeche: Robotics for the Benefit of Footwear Industry. ICIRA, 5th International Conference on Intelligent Robotics and Applications (2) 2012: 235-244, Montreal [116]

Robotics for the Benefit of Footwear Industry

Iñaki Maurtua, Aitor Ibarguren, and Alberto Tellaeche

IK4-TEKNIKER, Av. Otaola 20
20600 Eibar, Spain
{imaurtua,aibarguren,atellaeche}@tekniker.es

Abstract. This paper presents the initial results achieved by the ROBOFOOT project aimed at contributing to the introduction of robotics in the Footwear Manufacturing Industry. In particular, user requirements, operations selected and technical achievements reached so far are described. Visual servoing solution developed for shoe pose identification is described with deeper detail. The introduction of this technology allows the coexistence of current working practices and robotic solutions, with minor changes in the production means already existing in most companies. This has been identified as one of the requirements by the end-users taking part in the project.

Keywords: Footwear, Force Control, Visual Servoing, Programming.

1 Introduction

With more than 26.000 companies and almost 400.000 employees footwear industry is still relevant in Europe [1]. However trend shows a clear decline on business figures; low cost countries are becoming an obvious threat for the future of the sector.

Fashion Footwear production is currently mainly *handcrafted.* Some manufacturing processes are assisted by specialized machinery (last manufacture, cementing and cutting) and there exist highly automated lines in mass production of technical shoes (i.e. safety footwear). But most production is still handmade, being especially true in the case of *high added value shoes* production, where Europe maintains its leadership.

The introduction of intelligent *robotics may contribute to overcome the complexity in the automation of the processes of this industry* that accounts for some of the shortest production runs to be found (eight pairs of shoes is the average order size). The main reasons that justify this lack of automation and extensive labor demand are:

- High number of products *variants*. On the one hand, a minimum of two different collections (summer & winter) of shoes are developed to be presented to the customers every year. As an average, more than 200 different *models* are manufactured for the two seasons. On the other hand, it is necessary to adapt each model to at least six different *sizes* and two sides (left and right).

  Finally, we have to take into account that each model can be manufactured in different *leather qualities and colors*.

C.-Y. Su, S. Rakheja, H. Liu (Eds.): ICIRA 2012, Part II, LNAI 7507, pp. 235–244, 2012.

- *Complex manufacturing process*. For each model it is necessary to develop and manufacture the last (the rough form of a human foot used in shoemaking to provide the fit and style of the shoe), to produce the list of components (sole, heel, sock, strap, inner parts, etc.), to cut the inner and outside parts, to stitch different part to form the upper.
- *Complex assembly process*. The assembly process is very laborious (up to 25 different operations) and especially complex in fitting operations due to the non uniformity and the different elasticity of the natural leather as well as the non-rigid nature of the components. Finally each pair of shoes requires a final inspection (small spots or color differences in the leather, correct alignment of parts, etc.) and they are packaged.

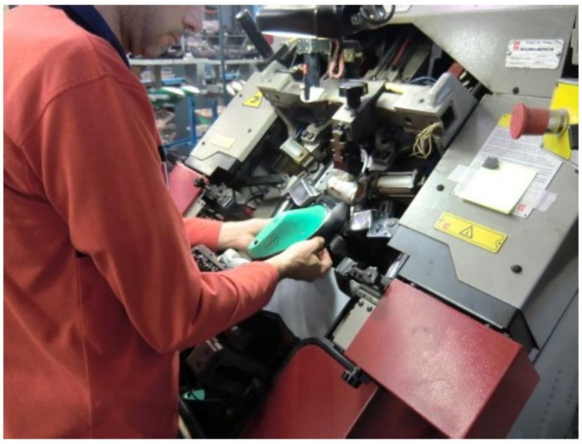


Fig. 1. Semi-automatic assembly operation

Although some companies in this sector tried to incorporate robotic solutions, they did not succeed in the objective except for the injection process. The EU co-funded *ROBOFOOT* project is developing different solutions to facilitate the introduction of robotics in traditional footwear industry**.**

This paper presents some initial results, starting with the user requirements presented in Section 2 and the Operation selection in Section 3. First technical results are described in Section 4 with more detailed information on visual servoing achievements. Finally Section 5 summarizes the re-design of the manufacturing process.

2 Robots in Footwear Industry: User Requirements

The project consortium has done a deep analysis of current practices and main needs of Footwear Industry. The process has included reviewing several studies, the internal analysis of the two end-user partners, ROTTA and PIKOLINOS, and the contribution of the rest of partners that visited both manufacturing plants and participated in the decision making process providing their technical background. In summary, the main requirements identified are the following:

- *Quality*: introduction of robots shouldn´t increase the number of shoes that need some kind of retouching at the end of the line (currently an average of 80%). On the contrary, as most of these small faults are due to the low stability of some processes, it is expected that the use of robots will contribute to reduce these retouching operations.
- *Impact in current production process*: A basic requirement is the possibility to combine current production procedures with the robotized solutions proposed by ROBOFOOT. This includes the coexistence of manual operations with robotized ones and the reuse of existing production means.
- *Efficiency*: reduction of manufacturing time. It should be taken into account that the robotized production has to be integrated in current production. So, reduction of individual operation time cannot be considered an objective unless we consider combining two operations.
- *Production flexibility*. Two business trends demand new and more flexible manufacturing technologies [2]: on the one hand, higher flexibility and adaptability at a process level: Due to changes in the habits and behaviours of (both industrial and private) consumers, product life cycles are getting shorter and more product variants have to be offered. On the other hand, demand for higher flexibility and adaptability at plant and supply chain level: Shoe manufacturing companies have to be able to shortly react on changes on the market.
- ROBOFOOT has to guarantee the production flexibility, handling a wide variety of models/sizes coexisting in the production line and allowing frequent model changes
- *Reduction of costs*. Although it is not the main reason for introduction of robotics in this sector, it will allow some workers to do tasks with higher added value and overcome the lack of skilled workers for some operations.
- *Working conditions*. Currently there are several operations that involve potential risk for workers (dust, use of solvents, rotating parts, effort...). Introducing robots has to help in reducing the potential risk of those operations to the minimum.
- *Usability and maintainability*. The system has to be easy to use and maintain by no specialists.

3 Operation Selection

The criteria used for operation selection have been:

- Has it a positive impact on initial requirements?
- Does it mean an innovation in the process?
- Is the operation applied in most shoe types?
- Are there many variants in the way of doing the process? Can we cope with most of them?
- Is the solution proposed suitable to be used in other operations?
- Does it seem feasible to be done in the timeframe of the project?
- Is it suitable to be introduced in a demonstrator?

According to the criteria established and the analysis of operations, it has been established a ranking of operations, grouped in three prototypes:

- *Basic prototype*. It includes individual operations: Roughing, Gluing, Inking, Polishing, Last Manufacturing.
- *Intermediate prototype*. They correspond to some operations that can be combined in the same robotic cell: Roughing+ Gluing; Inking+ polishing+ last removal. Inspection will be included as well, in order to detect the presence of nails, assess gluing and roughing processes and to identify defects on the shoe.
- *Final prototype*: It corresponds to the most challenging operation, i.e. packaging. A complete robotized cell has been designed, although only some of the sub-operations will be implemented in the context of ROBOFOOT.

4 Technologies for Robotic Footwear Production

4.1 Robot Programming and Controlling

The general approach for robot programming and controlling is presented in the picture below. The starting point is the 3D information of the shoe. It can be obtained from the CAD or from a digitalization of the shoe mounted on the Last (not all companies use 3D CAD system for shoe designing).

Technicians define the trajectories for the different operations, including technological features (1).

The resulting file is the input for the postprocessor that generates the robot program in PDL2 language, specific for the COMAU robots used in the project (2).

To overcome the problem of minor misalignments between the Grasping Device and the last it is necessary the correction of the resulting program (3). Finally, real time adjustment of trajectories is needed for several operations, such as roughing (4).

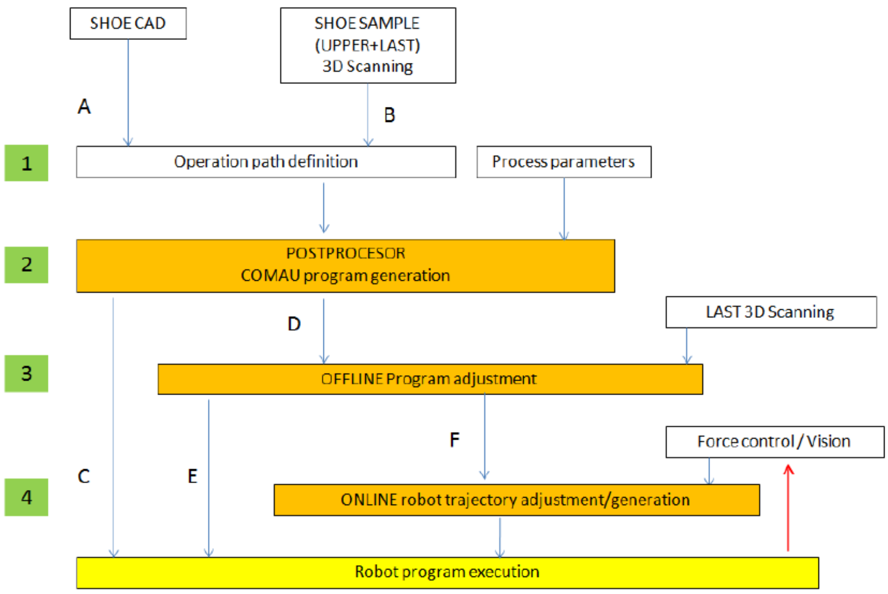


Fig. 2. General scheme for trajectory generation/control

Automatic Robot Part Program Generation from Digital Data

A CAM system has been developed by INESCOP to define the tool-path that has to be followed by the robot as well as process parameters.

The geometry is imported from the virtual model or from a digitalized model. Based on that, the user can define the area to be processed and the parameters to be used. After post-processing the resulting file it is generated the executable program.

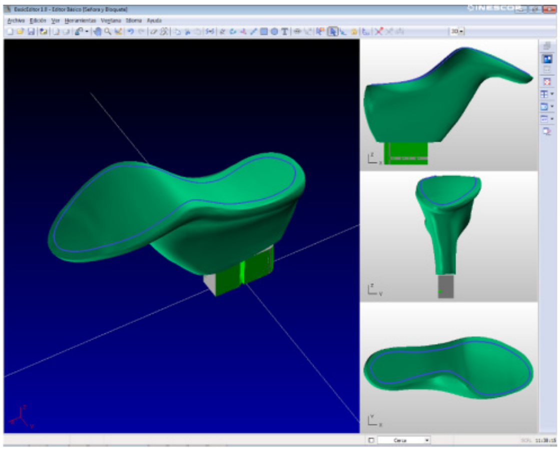


Fig. 3. Off-line robot programming (INESCOP)

However even a small assembly error between the GD and the last may lead to a significant difference between the theoretical and real tool trajectory – especially at shoe tip.

To overcome this problem it has been developed a system for online adaptation of off-line robot program.

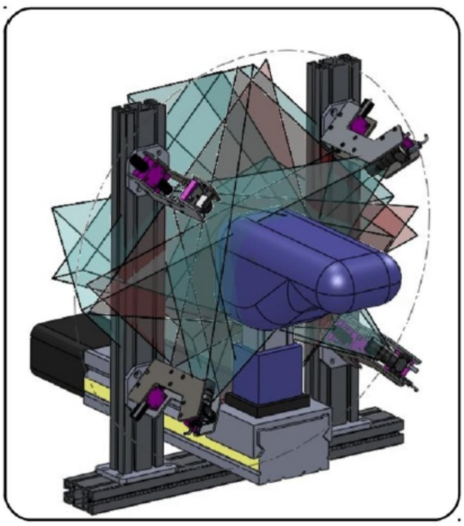


Fig. 4. 3D scanning for off-line program adjustment (CNR-ITIA)

On-Line Smart Adaptation of Off-Line Automatic Robot Part Program

CNR-ITIA has developed an innovative solution for on-line smart adaptation of off-line automatic robot part program generated from digital data (CAD/CAM) through the use of on-line sensors, namely a new generation of modular, scalable and reconfigurable 3D laser scanner. It allows fine-tuning the program generated in the previous step by adding the real positioning of the GD with respect the Last.

Manual Guidance Device

This wireless device can be place anywhere after the sixth axe of the robot. Operators can control the movements of the robot in a very intuitive way using the MGD. It offers the following benefits to end-users:

- It represents an innovative and smart way to program robot movements in an intuitive way at a very low cost.
- The design of the device is such that it is made out of low-cost parts, without losing focus on the performance of the system.
- The device will allow operators (even unskilled ones) to save time during the robot movement programming, without losing the quality of the trajectory definition.
- The device is extremely flexible as it can be positioned on the robot tool (or mounted on the robot flange) regardless of the robot type.

Fig. 5. Manual guidance device (Developed by CNR-ITIA for COMAU)

Sensor Based Robot Controlling

Finally, we have to consider that it is needed to correct or to generate the robot trajectory in real time for some operations. This is the case of operations like roughing, where it is important to guarantee that leather is not damaged.

CNR-ITIA is developing a real time robot trajectory adaptation mechanism based on force control and using the C4G Open feature that allows real-time parameters change (robot system variables).

On the other hand, visual servoing control has been developed to identify the pose of the shoes in the manovia and generate the robot trajectory to pick them up. To this aim, a vision system in an eye-in-hand configuration has been introduced.

Visual Servoing

Shoes go from one working station to the next one on trolleys that are placed in the Manovia. For each shoe on the trolley, the robot has to take it, manipulate it in the workstation (roughing, gluing,...) and leave it back on the trolley. Due to the fact that we are combining manual and robotized operations, it is not guaranteed an accurate positioning of the shoe on the trolley by workers.
Some colors show up very poorly when printed in black and white.

To overcome this problem and achieve a precise shoe grasping, a visual servoing system has been developed. The system uses images to estimate the pose of the Grasping Device attached to the shoe (external metallic element that allows a reliable grasping) and based on those estimations the robot corrects its pose until the desired position is reached. It is necessary to perform the grasping task with a precision of around 1 millimetre and 1-2 degrees in each axis to avoid damages in the shoe. The manoeuvre should not take more than 5-6 seconds to maintain the production time cycle.

The pose identification is difficult due to the industrial nature of the scenario; problems like the poor illumination of those environments, the metallic nature of the Grasping Device which makes difficult a proper illumination, as well as the wax and ink used during the shoe making process that can be adhered to the Grasping Device. All the above facts make it difficult to acquire good quality images to estimate the pose of the shoe.

TEKNIKER has developed a 6DOF visual servoing system using a dynamic look-and-move approach, including a particle filter to deal with the uncertainties (illumination, dirt in the Grasping Device…) of the image acquisition process. The process is as follows:

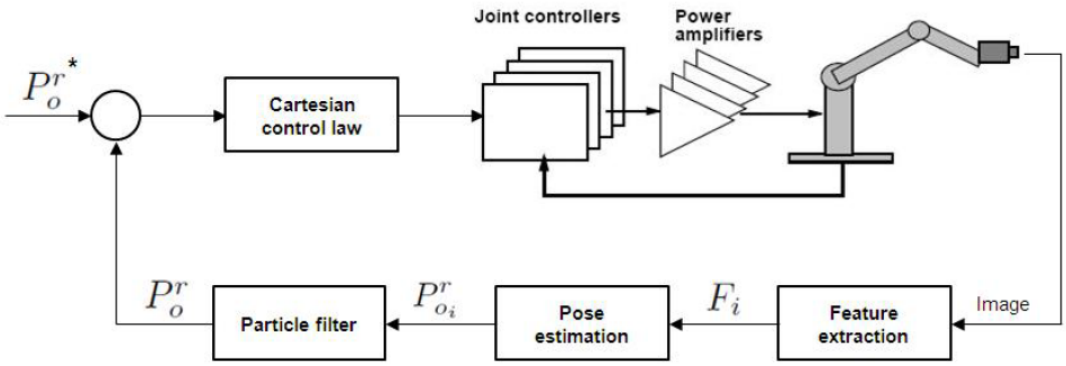


Fig. 6. Visual servoing schema

- *Feature extraction* from the image. In this step the image is analyzed, trying to find features of the Grasping Device, such as holes and edges. Even so, due to the uncertainties of the images it is difficult to determine exactly the position and dimensions of those features. To overcome this problem, different hypothesis have been extracted for each feature, determining possible positions and dimensions of the holes and edges.

- *Pose estimation* of all possible hypotheses. Based on the features extracted in the previous step, different poses are estimated, each of them related with a hypothesis.
- *Fusion of those hypotheses by means of the particle filter* to get the final pose estimation. The particle filter uses the estimated poses to determine which of the particles (hypothesis) are the most suitable for the given state and uses this distribution to improve the estimations in further steps.
- *Robot movement* based on the estimated pose. The robot moves in the workspace, trying to correct orientation and distance errors.

Based on the described loop, the system corrects iteratively the robot position until the desired one is achieved, as shown in Fig. 5. The experiments show good performance of the system, reaching high success rate and fitting in the time constraints previously explained.

4.2 Manipulation

In ROBOFOOT two different problems are considered: manipulation with Last and manipulation without Last, just the shoe.

In the first case it has been necessary to modify the current Lasts by introducing an external element, the Grasping Device (GD) that allows grasping the last with the required rigidity and repeatability. This modification has been done in such a way that Lasts are compatible with existing manufacturing machines at end-users facilities.
Some colors show up very poorly when printed in black and white.

The vision system makes possible the identification of the pose of the lasts both on the manovia and when they lie on the exit of a chiller.

In the second scenario (shoes without Lasts) the deployment of a simple parallel gripper would damage the shoe and special grippers (for instance suction) would only be helpful for some shoe types and materials. To achieve this objective DFKI is developing a bimanual multifingered robotic approach based on the AILA [3] robot and the iCub hands [4].

5 Manufacturing Process Redesign

5.1 Manufacturing Cells

Three manufacturing robotic cells have been designed and implemented:

- *Roughing, gluing and last milling*: A multifunctional robotized cell for bottom and side roughing, gluing and last milling has been conceived by QDESIGN. Attention has been paid on how the roughing and gluing activities are actually performed by ROTTA.
- *Polishing, inking and last removal*: A robotic cell has been designed by AYCN. This cell integrates a robot that takes the shoes from the exit of the chiller, does the inking and polishing processes and, finally, opens the last for manual removal of the shoe by the operator. Inking is carried out in a conventional cabinet where painting guns are placed at fixed positions. Polishing operations are carried out

using a roller similar to those used in the conventional process, but the frame has been adapted to allow force monitoring.

- *Packaging*: it includes all the phases needed to pack a pair the shoes.

5.2 Quality Assessment

Quality inspection is currently done by workers at the end of the manufacturing process, just before packaging. These operators are in charge of doing manual reworking of shoes to repair small defects (80% of shoes require this kind of task). Due to this fact, it was decided that quality assessment supported by robots had to be implemented to verify the goodness of intermediate operations to ensure the final quality of the shoe. Specifically it was decided implementing visual inspection techniques in roughing, gluing and nail removing steps. Surface defects (cuts, scars, colour irregularities, etc.) will be tackled as well.

The robot manipulates the shoe inside inspection cabinet, where the following operations are verified:

- *Roughing*: There are two different objectives: to control the boundaries of the roughed area and to verify that the leather has not been damaged due to over-roughing. Different machine vision approaches, including thermography, are under dvelopment.
- *Nail removing*: the aim is to detect that there are not nails left on the sole (their removal is done manually by an operator). The proposed approach is to search the position of nails initially (once they are hammered in the last, although not necessarily just after this operation) and restrict the search space to those locations after the nail removal operation to guarantee that all of them have been removed.
- *Gluing*: To detect the excess (or lack) of glue in the shoe an additive sensible to ultraviolet light is added to the glue, making it easier to detect the presence of such excess or lack of glue. Using the correct lighting a simple threshold algorithm allows glue detection.
- *Texture analysis*: 2D vision will be used to detect marks, patterns and color changes on the leather surface.

6 Future Work and Acknowledgments

The project will last until February 2013. The last months will be devoted to experimental validation of the robotic solutions for each operation and the implementation of the packaging system.

Acknowledgments. This work has been performed within the scope of the project "ROBOFOOT: Smart robotics for high added value footwear industry ". ROBOFOOT is a Small or Medium- scale focused research project supported by the European Commission in the 7th Framework Programme (260159). For further information see http://www.robofoot.eu.

References

1. EU Commission, European Industry in a Changing World-Updated Sectoral Overview 2009 (2009)
2. Bessey, E., et al.: Research, Technology and Development for Manufacturing. In: The ManuFuture Road: Towards Competitive and Sustainable High-Adding-Value Manufacturing. Springer (2009)
3. Lemburg, J., de Gea Fernandez, J., et al.: AILA - design of an autonomous mobile dual-arm robot. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5147–5153 (2011)
4. Schmitz, A., et al.: Design, realization and sensorization of the dexterous iCub hand. In: 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp. 186–191 (2010)

7.22. Accurate Correction of Robot Trajectories Generated by Teaching Using 3D Vision by Laser Triangulation

Alberto Tellaeche, Ramón Arana, Iñaki Maurtua: Accurate Correction of Robot Trajectories Generated by Teaching Using 3D Vision by Laser Triangulation. ICIRA, 5th International Conference on Intelligent Robotics and Applications (3) 2012: 385-394, Montreal [117]

Accurate Correction of Robot Trajectories Generated by Teaching Using 3D Vision by Laser Triangulation

Alberto Tellaeche, Ramón Arana, and Iñaki Maurtua

IK4 – Tekniker. Avda. Otaola 20, 20600 Eibar, Spain
{atellaeche,rarana,imaurtua}@tekniker.es

Abstract. Normal utilization of robot manipulators of anthropomorphic type does not reach beyond the reiteration of preprogrammed trajectories. While static robot programs may be sufficient for high volume manufacturers, they are not adequate in one-off or small-batch manufacturing, where programs must be adapted and modified in a dynamic way to fulfill the changing requirements in this type of production. Among the different techniques for robot programming, teaching is one of the fastest when changes have to be applied in complex trajectories. The main drawback of this technique is that a lot of time is lost defining the robot points very precisely. The objective of the work presented in this paper is to facilitate robot programming by combining teaching programming techniques and a 3D machine vision based accurate trajectory following.

Keywords: Teaching robot programming, 3D vision, laser triangulation, trajectory correction.

1 Introduction

There are multiple manufacturing processes where robots play (or might play) an important role. It is well known their use in welding, deburring and other un-safe or risky operations. Despite the intrinsic usefulness of robot manipulators of anthropomorphic type, their normal utilization is a repetition of established trajectories.

Nowadays, one of the main bounds for the growth and widespread of robotized cells in the context of small and medium enterprises is the complexity in the programming of robots. In the industry, the training level required for that kind of operation represents one of the biggest obstacles in order to prefer other automation solutions, intrinsically easier to setup.

This is particularly true for SMEs that cannot afford for big investments required for robot introduction and use, and cannot make expensive efforts in personnel robot training.

The objective of this work is to integrate a 3D visual servoing approach to adjust the rough trajectories generated by teaching robot programming using a Manual Guidance Device (MGD) and to allow accurate end-effector positioning by automatic correction of Tool Central Point (TCP) path. As a result, it will be possible to program

C.-Y. Su, S. Rakheja, H. Liu (Eds.): ICIRA 2012, Part III, LNAI 7508, pp. 385–394, 2012.

robotic applications in an easy and fast way, focusing the attention on the process (laser cladding, deburring, etc) regardless programmatic problems.

2 State of the Art and Related Work

The common robot programming techniques, namely offline programming and lead-through programming with teach pendant, lead to several problems for SME applications. A highly skilled operator is necessary and the teaching takes a long time.

With teaching techniques, several programming approximations are referred, such as Programming by Demonstration, Programming by Example, Walk-Through Approach, Computer Assisted Teach and Play, or Programming by Manual Guidance. Also, Virtual Reality plays an important role in robot teaching techniques.

Programming by Demonstration is an intuitive programming method that can be done using the manual guidance of the robot. The operator moves the robot by applying forces to the robot tip, or other parts of the robot. The robot moves with deactivated actuators or with active/passive compliance. Using this programming method the operator has not to program directly in an explicit way, but the robot program is composed of the result of an interaction between the robot and the human. The robot is moved to each desired position, the controller records the internal joint coordinates corresponding to that position. This information is collected to process a program composed by a sequence of vectors with joint coordinates and some activation signals for external equipment such as the gripper aperture. When the program is executed, the robot moves through the specified sequence of joint coordinates, reproducing the indicated signals. In [1] the programming of industrial robots with the methods of 'Programming by Manual Guidance' is described.

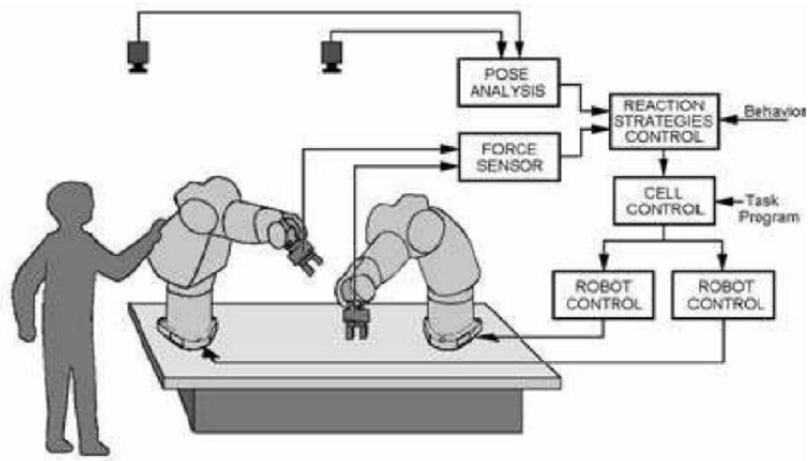


Fig. 1. Typical system configuration for programming by manual guidance [1]

In [2] a programming system for welding in shipyards is described. Robots can be programmed in a very fast pace by a 'Walk-Through Approach', a custom man-machine interface is implemented and the welding parameters are optimized by the system.

A further programming system called 'Computer-Assisted Teach and Play' is implemented for the whole arm manipulator (Barrett Technology Inc.) by Leeser et al. [3]. The operator can get help from the system by virtual surfaces, a gravity compensation for the arm and also for the tool and the payload are integrated.

Another interesting application, the insertion of pistons in a motor block is presented by Albu-Schaeffer et al. for 'Programming by Manual Guidance' [4]. The Light-Weight Arm developed at DLR can be guided manually on the desired trajectory.

Recently an innovative control strategy has been developed by Grunwald et al. [5]. The strategy is applied to a lightweight robot with a distributed system of sensors. In this case, it is possible to program the robot in an intuitive way.

Asada and Izumi [6] developed a method to generate a program for hybrid position/force control using a back-drivable robot guided by the human operator's hand directly.

In [7] an approach for real-time robot programming by human demonstration for 6D force controlled actions has been done. A human operator uses a joystick to guide a robot with a force sensor to execute a task including continuous contact between a manipulated object and an unmodeled environment.

Sato et al. present an alternative method of robot programming that does not need the use of force/torque sensors [8]. This implementation has been done on a high-speed parallel robot to carry out fast and complicated tasks.

Virtual reality is another approach to robot teaching. In [9], Takahashi et al. the authors propose a robot teaching interface which uses virtual reality. Takahashi et al. also propose robot teaching methods using VPL Data Gloves in a virtual workspace [10].

In [11] Kawasaki et al. Explain a virtual robot teaching system, consisting of human demonstration and motion-intention analysis in a virtual reality environment.

Finally, in [12] it is presented a new development method for event-driven robot teaching in a virtual environment.

3 Experiment Definition

This research has as main objective the development of a rapid robot programming system, combining teaching techniques for definition of robot points with a 3D system composed of a SICK Ranger E55 camera and a Class 2 red laser line.

The teaching points will be defined manually as approximate points by which the final trajectory will have to pass through. In this approach there is no need to define all the points of the trajectory, because the points will be recorded by the robot in real time and also will be corrected by the data obtained from the calibrated 3D image.

Based on the needs of the experiment, a prototype board has been constructed in Aluminum. This prototype has different trajectories defined, such as: straight line, curve, etc. The track defined in the prototype has a width of 60 mm and a depth of 7 mm, with a central nerve of 10 mm. Detail of this board can be observed in fig.2

Fig. 2. Detail of the prototype board with different trajectories

3.1 Robot Position Tracking

For this research a NM45 robot from COMAU Robotics has been used. The points defining the trajectory in a rough way have been defined using the provided COMAU MGD.

Once the trajectory has been defined, parallel to trajectory program, the robot executes a parallel task to register the TCP positions in a text file. These positions are recorded whenever an increment of 0.5 mm is detected in the Euclidean distance from the previous recorded point or when an increment of 1º is detected in the α angle (rotation of the camera). The registered points have the following formal definition:

$$< x, y, z, \alpha, \beta, \gamma > . \tag{1}$$

The robot also synchronizes the recording of each point with a 3D line scan of the camera by a trigger pulse.

Every 500 pulses, the file containing the 500 coordinates and the 1536 * 500 pixel 3D image acquired are processed for trajectory correction. In parallel, the next capture of points is being performed.

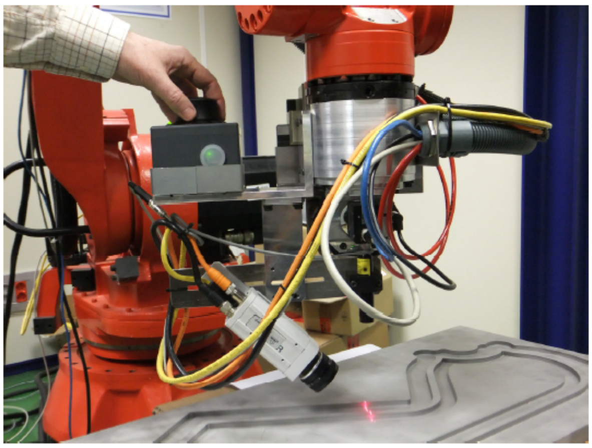


Fig. 3. COMAU robot with laser triangulation system scanning the prototype

3.2 3D Image Acquisition

For 3D visual trajectory correction, a laser triangulation system has been used. This system is composed by a SICK Ranger E55 Camera and a Class 2, red laser line. The camera has a 1536 * 512 pixel sensor and is capable of obtaining up to 35K profiles per second.

3D vision by means of laser triangulation has been widely used in industrial applications, like surface control or coordinate extraction. Works related to this can be found in [13] and [14].

There are four main principles for mounting the camera and the laser line in laser triangulation systems. These are:

1. *Ordinary setup*: The camera is mounted right above the object – perpendicular to the direction of movement – and the laser is illuminating the object from the side. This geometry gives the highest resolution when measuring range, but also results in miss-register – that is, a high range value in a profile corresponds to a different y coordinate than a low range value.
2. *Reversed ordinary setup*: As the ordinary setup, but the placement of the laser and the camera has been switched so that the lighting is placed above the object. When measuring range, the reversed ordinary geometry does not result in miss-register, but gives slightly lower resolution than the ordinary geometry.
3. *Specular:* The Ranger and the lighting are mounted on opposite sides of the normal. Specular geometries are useful for measuring dark or matte objects.
4. *Look-away:* The Ranger and the lighting are mounted on the same side of the normal. This geometry can be useful for avoiding unwanted reflexes but requires more light than the other methods and gives lower resolution.

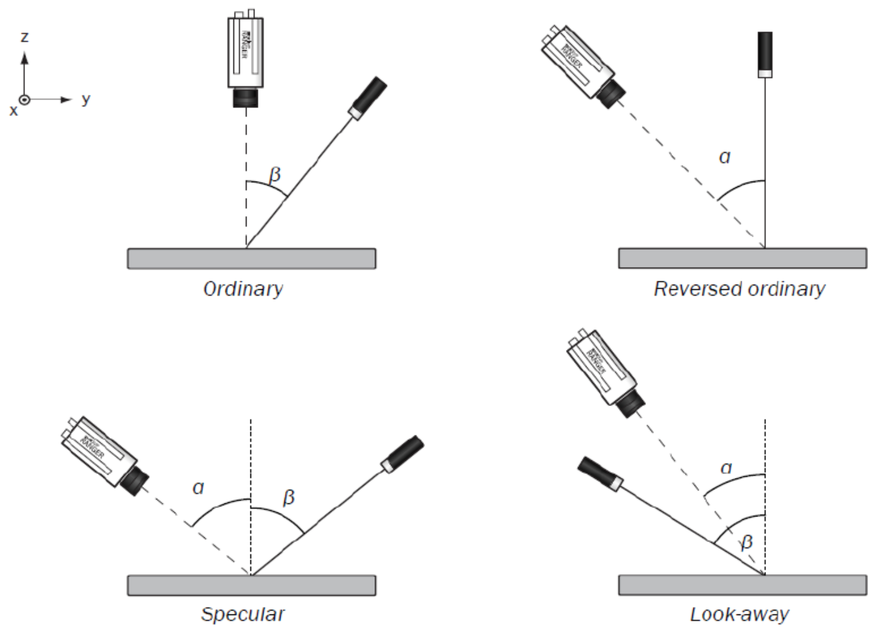


Fig. 4. Four different setup for laser triangulation

Fig. 4 shows a schematic representation of the geometrical setup for laser triangulation. According to [15], the *reversed ordinary setup* provides a good height resolution avoiding miss-register problems. The height resolution can be expressed as follows:

$$\Delta Z \approx \Delta X / \sin(\alpha). \tag{2}$$

The Ranger E55 camera internally calibrates the 3D images obtained, giving the *d* and *z* coordinates in real world mm. *d* is the Euclidean distance from the target (borders) to the center of the camera sensor, and *z* the point height. Also, during system set up, the center of the sensor of the Ranger E55 camera has been adjusted and calibrated to coincide with the robot Tool Center Point (TCP).

Using the 3D calibrated images acquired with the recorded coordinate points of the robot it is possible to correct the rough trajectory recorded by teaching, obtaining the very accurate path necessary in many robotic applications.

4 Trajectory Correction and Adjustment

Prior to robot movement, the user must select what type of edge must be adjusted in the subsequent operation. Attending to the shape of the track constructed in the prototype, options are: left border of the track, left border or the central nerve, right border of the central nerve, right border of the track. This setup will be used in the image processing to detect the correct border. Also, the reference *z* coordinate must be adjusted, to place in the height range where the camera provides correctly calibrated measures.

4.1 Image Processing

As explained in previous sections every 500 recorded points, two images are obtained from the ranger camera, *d* calibration and *z* calibration.

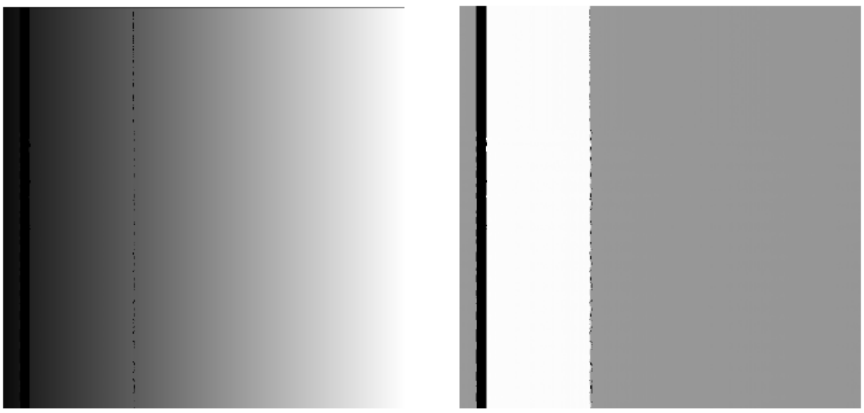


Fig. 5. d and *z* calibrated images from the central nerve of the prototype

After performing image filtering for noise reduction, the line profiles composing the 3D z calibrated image are processed, calculating the derivate signal for each one. With the derivative signal, it is possible to detect in a very precise manner the level transitions (borders) present in the line profile. When there is a high to low transition, corresponding to the left track border or to the right border of the central nerve, a minimum peak appears in that point in the derivative signal. On the other hand, when there is a low to high transition corresponding to left border of the central nerve, or to right track border, a maximum appears in the derivative signal. The rest of the values in the derivative signal tend to 0.

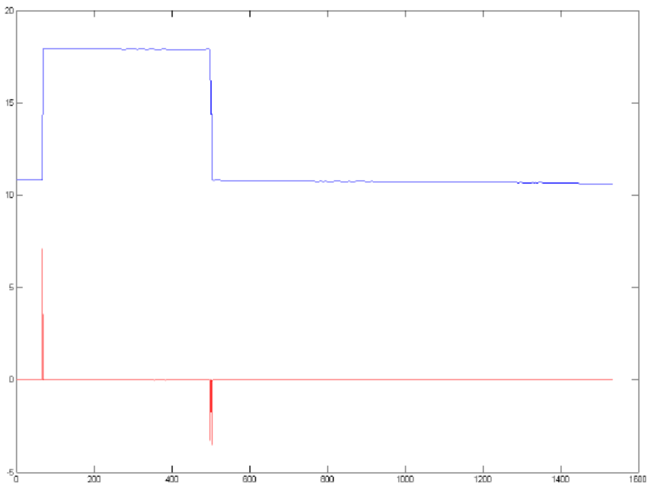


Fig. 6. line profile (above) and its derivative (below)

Taking into account the maximums and minimums present in the derivative signal and their relative position, it is possible to estimate the pixel in the line profile where a certain border is. Using that pixel value, the z calibrated value is obtained from the z calibrated image, and the d calibrated value is obtained from the d calibrated image. These values are converted to robot coordinates in the format presented in (1). This conversion is:

$$x = d\cos(\alpha)\ . \tag{3}$$

$$y = d\,\mathrm{sen}(\alpha)\ . \tag{4}$$

4.2 Estimation of the Trajectory from the 3D Image.

By the process explained in the previous section, it is possible to estimate the real points of the border under inspection and thus, obtain the trajectory. Although the image is processed to avoid errors due to noise or lack of data in certain points, it has been implemented a restriction in the position of the points in the trajectory, taking into account the previous (x,y) positions of the two previous correct points obtained.

Let (x_0,y_0) and (x_1,y_1) be the two previous correct points calculated in the trajectory. The maximum increment in x and y coordinates is defined by a maximum Euclidean distance of 0.5 mm between two consecutive points.

$$(x_{2max}, y_{2max}) = (x_1+\Delta x_{max}, y_1+\Delta y_{max}). \tag{5}$$

If $x_2 > x_1+\Delta x_{max}$ or $y_2 > y_1+\Delta y_{max}$, then (x_2, y_2) is estimated as follows:

$$y_2 = y_1+\Delta y_{max} \quad . \tag{6}$$

$$x_2=(x_1-x_0)(y_2-y_0)/(y_1-y_0) + x_0. \tag{7}$$

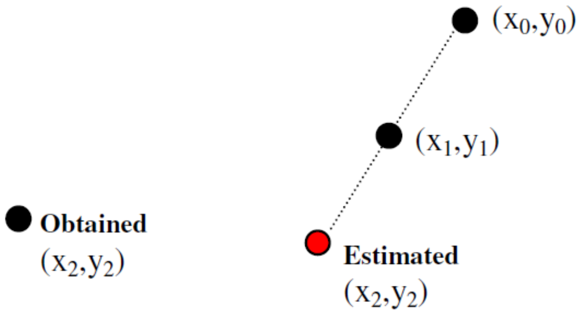


Fig. 7. Graphical representation of point estimation

4.3 Final Trajectory Correction

The center of the sensor of the Ranger E55 coincides with the robot TCP. Thus, the x_c and y_c point coordinates of the border detected in the 3D image correspond to the Δx and Δy final corrections of each point recorded by the robot.

Given a recorded point of the robot:

$$X_r = < x_r, y_r, z_r, \alpha_r, \beta_r, \gamma_r > \quad . \tag{8}$$

and the coordinates of the border obtained from the camera:

$$X_c = < x_c, y_c, z_c, \alpha_c, \beta_c, \gamma_c > \quad . \tag{9}$$

The final correction is:

$$X_r–X_c= < x_r - x_c, y_r - y_c, z_r, \alpha_r, \beta_r, \gamma_r > \quad . \tag{10}$$

The α angle variation is implicitly taken into account in (3) and (4), so it is not necessary to change the recorded value from the robot. *z* coordinate is calibrated and the real height value of the point of the trajectory being followed is obtained from the camera.

5 Experimental Results

To assess the performance of the correction method proposed, several tests have been carried out, using three different types of trajectories and the four different borders in the prototype.

The three trajectories used for validation are: straight, curve and combined straight-curve trajectories.

The borders are: left border of the track, left border or the central nerve, right border of the central nerve and right border of the track.

Table 1 shows the results obtained. The precision of the trajectory corrected is expressed in terms of mean and standard deviation of the accumulated error in trajectory estimation. These measurements are expressed in mm.

Table 1. Results obtained for trajectory estimation

		Predefined Trajectories		
		Straight	Curve	Straight-Curve
left border of the track	Mean	0.020	0.078	0.065
	Std. Dev	0.041	0.1	0.071
left border or the c. nerve	Mean	0.013	0.068	0.057
	Std. Dev	0.043	0.091	0.077
right border of the c. nerve	Mean	0.012	0.063	0.060
	Std. Dev	0.039	0.087	0.071
right border of the track	Mean	0.014	0.072	0.051
	Std. Dev	0.037	0.102	0.069

Analyzing the results obtained, it can clearly be seen that the system has a better performance in correcting straight trajectories. This is caused because in curve trajectories, there are more variables that can increment the overall error. While in straight trajectories all the inaccuracies depend purely on linear movements, in curve trajectories, apart from these linear movements, there are also flange rotations. However, the errors obtained can be assumed in the majority of robotic applications.

6 Conclusions and Future Work

With this research, it has been developed a new system for quick robot programming. The rough trajectory generated with teaching points in the COMAU robot, can afterwards be precisely corrected with the calibrated data obtained from the 3D images provided by the Ranger E55 camera.

The work done until now corrects the trajectories but does not take into account the robot tool orientation (β and γ angles). Tool correction and orientation will be the next steps for this research to obtain total robot trajectory correction.

Acknowledgments. This work has been carried out under the experiment "EASYPRO – Accurate Manual Guided robot programming". EASYPRO experiment belongs to the ECHORD platform, an EU-funded project within the Seventh Framework Program.

In the EASYPRO experiment, the consortium is composed by an Spanish research center, Tekniker-IK4, and two Italian companies, CNR-ITIA (Research center) and COMAU Robotics.

References

1. Frigola, M., Poyatos, J., Casals, A., Amat, J.: Improving Robot Programming Flexibility through Physical Human - Robot Interaction. In: IROS Workshop on Robot Programming by Demonstration, Las Vegas (October 2003)
2. Ang, H., Lin, W., Lim, S.-Y.: A Walk-Through Programmed Robot for Welding in Shipyards. Industrial Robots 26(5) (1999)
3. Leeser, K., Donoghue, J., Townsend, W.: Computer-Assisted Teach and Play: Novel User-Friendly Robot Teach Mode Using Gravity Compensation and Backdrivability. In: Proceedings of S ME Fifth World Conference on Robotics Research, Cambridge, USA (1994)
4. Albu-Schaffer, A., Hirzinger, G.: Cartesian Impedance Control Techniques for Torque Controlled Light-Weight Robots. In: Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington DC (May 2002)
5. Grunwald, G., Schreiber, G., Albu-Schaffer, A., Hirzinger, G.: Programming by touch: The different way of Human-Robot Interaction. IEEE Transactions on Industrial Electronics 50(4) (2003)
6. Asada, H., Izumi, S.: Direct Teaching and Automatic Program Generation for the Hybrid Control of Robot Manipulators. In: IEEE International Conference on Robotics and Automation, Raleigh (1987)
7. Bruyninckx, H., De Schutter, J.: Specification of force controlled actions in the task frame formalism — a synthesis. IEEE Trans. on Robotics and Automation 12(4), 581–589
8. Sato, D., Shitashimizu, T., Uchyama, M.: Task Teaching to a Force Controlled High Speed Parallel Robot. In: Proceedings of IEEE International Conference on Robotics & Automation, Taipei, Taiwan (2003)
9. Takahashi, T., Ogata, H.: Robotic assembly operation based on task-level teaching in virtual reality. In: Proceedings of IEEE International Conference on Robotics & Automation, Nice, France (1992)
10. Takahashi, T., Sakai, T.: Teaching robot´s movement in virtual reality. In: IEEE/RJS International Workshop on Intelligent Robots and Systems. Proceedings 'Intelligence for Mechanical Systems' IROS, Osaka, Japan (1991)
11. Kawasaki, H., Furukawa, T., Ueki, S., Mouri, T.: Virtual Robot Teaching Based on Motion Analysis and Hand Manipulability for Multi-fingered Robot. Journal of Advanced Mechanical Design, Systems, and Manufacturing 3(1), 1–12 (2009)
12. Cui, M., Dong, Z., Tian, Y.: Simulation and execution of event-driven robot teaching in a virtual environment. In: Fifth World Congress on Intelligent Control and Automation, WCICA, Hangzhou (2004)
13. Leopold, J., Gunther, H., Leopold, R.: New developments in fast 3D-surface quality control. Measurement 33(2), 179–187 (2003)
14. Picón, A., Bereciartua, M.A., Gutiérrez, J.A., Pérez, J.: Machine vision in quality control. Development of 3D robotized laser-scanner. Dyna. 84(9), 733–742 (2010)
15. Boehnke, K.E.: Hierarchical Object Localization for Robotic Bin Picking. Ph.D. dissertation. Faculty of Electronics and Telecommunications. Politehnica University of Timisoara (September 2008)

7.23. A Wearable Computing Prototype for supporting training activities in Automotive Production

Iñaki Maurtua, Pierre T. Kirisci, Thomas Stiefmeier, Marco Luca Sbodio, Hendrik Witt. A Wearable Computing Prototype for supporting training activities in Automotive Production. 4th International Forum on Applied Wearable Computing 2007. Tel Aviv [121]

A Wearable Computing Prototype for supporting training activities in Automotive Production

Iñaki Maurtua[1], Pierre T. Kirisci[2] Thomas Stiefmeier[3], Marco Luca Sbodio[4], Hendrik Witt[5]

[1]Production Engineering Department Fundación Tekniker, Spain, {imaurtua@tekniker.es}

[2] Bremen Institute for Industrial Technology and Applied Work Science at the University of Bremen (BIBA), Germany, {kir@biba.uni-bremen.de}

[3]ETH Zürich, Institut für Elektronik, Switzerland, {stiefmeier@ife.ee.ethz.ch}

[4]Hewlett-Packard, Italy Innovation Center, Italy, {marco.sbodio@hp.com}

5University of Bremen, TZI Wearable Computing Lab., Germany, {hwitt@tzi.de}

Abstract. This paper presents the results of the wearable computing prototype supporting training- and qualification activities at the SKODA production facilities in Czech Republic. The emerged prototype is based upon the first of the 2 main “Variant Production Showcases” (training and assembly-line) which are to be implemented in the WearIT@work project (EC IP 004216). As an introduction, the authors of this paper investigate current training processes at Skoda, and derive the potential benefits and risks of applying wearable computing technology. Accordingly, the approach of creating the wearable prototypes, via usability experiments at the Skoda production site, is explained in detail. As a preliminary result, the first functional prototypes, including a task recognition prototype, based upon the components of the European Wearable Computing Platform, are described. The paper is rounded up by providing a short outlook regarding the second envisaged test case, which is focussed upon selected assembly line operations of blue collar workers.

Keywords: wearable computing, automotive production, training, usability, task recognition

1 Introduction

Since the late 90’s, applying Wearable Computing in industrial work environments has become an attractive approach to efficiently support mobile work processes [Siew96]. Automotive production inhibits an industrial environment where variant products of high complexity and short life-cycles are manufactured [Kir06]. Hence, many situations exist where work is performed in a mobile manner (e.g. maintenance, work at the assembly line, etc.). In order to cope with these conditions, concerted training of personnel is a very crucial issue in automotive production. Accordingly, training requirements are usually high. The European research project WearIT@work investigates, amongst other scenarios, the impact of wearable computing in automotive production. Since the beginning of 2004, a thorough analysis was carried out, involving interviews, field studies and comprehensive process analysis at the Skoda production facilities in *Mlada Boleslav* and *Vrchlabi* (Czech Republic). The aim was to implement a wearable computing solution which is capable of supporting the training procedures of Skoda blue collar assembly line workers. The wearable prototype which derived from these field-studies, offers semi-autonomous training by mobile- and context-sensitive support of trainee personnel. The trainees are provided with all necessary (digital) information in order to successfully perform individual production tasks. At the same time the performed tasks are tracked via mobile sensors mounted on a data glove and a car body. Particularly, the wearable system supports the trainees by detecting errors when tasks were not performed correctly, and by providing appropriate help and suggestions [Ma06]. The trainee can interact with the system through voice, and head-mounted display (HMD). What is even more important is that he/she does not need to interact at all, because the system recognizes the tasks performed and presents the information automatically.

2 The Training Process at Skoda

The two Skoda production facilities involved in the variant production showcase are signed by the fact that the personnel must receive theoretical and practical training before they are authorized to work at the assembly line.

The current training process is carried out at the so-called “E-factory” in Vrchlabi, and comprises two separate phases:

- Theoretical training is provided at the E-Learning Institute of Skoda using didactic material (Fig. 1) created by the Skoda training department, combining text, images and videos. The material is provided in paper- or in electronic format. The students must pass various tests before going on to the practical stage.
- The Learning Island. Using a real vehicle chassis, the students are required to put the theory into practice under the observation of a supervisor. The supervisor evaluates the individual actions of the trainee, while pointing out errors and providing suggestions for improvement. At the end the supervisor decides whether the trainees are ready to start working at the assembly line.

Fig. 2 illustrates how training is currently performed under the presence of the supervisor. Thus, a fully autonomous learning, without the presence of a supervisor, is not yet practiced at Skoda.

Fig. 1: Didactic Material used during the Skoda training procedure.

Fig. 2: Demonstration of the installation of the front headlights by the supervisor.

2.1 Training with Wearable Computing – Benefits and Risks

Compared to stationary computer systems, mobile- and wearable computing technology have seriously caught up in performance, functionality, scalability. This makes training solutions based on mobile- and wearable computing an attractive consideration for industrial organisations. In this sense, one of the objectives of WearIT@work was to supplement the training procedures at Skoda with a context-sensitive wearable computing solution. The idea was that the trainees gain mobile access to the information (e.g. instructions and required tools) to carry out their assembly tasks. In fact, the wearable system was used to recognize the context of performed work, and as a result provide the trainee with the required information to adequately perform individual assembly tasks. Concurrently the wearable computing system tracks the trainees' activities and analyses them. While the workers perform their training, the supervisor is connected to all active wearable systems via his PC, and can monitor all activities. The nature of the assembly activity itself made it necessary to design a system that does not restrict workers' freedom of movement, while allowing them to handle all necessary components and tools. It was especially crucial to take into account that workers had to adopt many different postures during the assembly process: crouching, standing, seated, inside- and outside of the car. It can be assumed that one of the main advantages of a wearable training solution is that the constant direct presence of the supervisor is no longer required. Thus, the supervisor has the opportunity to observe a number of trainees at the same time via his PC. Since the supervisor has a continuous overview of real-time information such as performed activity, number of mistakes, and number of repetitions, he may interact with the trainees in difficult situations. Eventually, an immediate benefit of using wearable solutions in automotive production is that the time of training procedures may be reduced. On the other hand this only applies when the threshold for getting acquainted with the wearable system is low enough. In order to ensure this, the wearable prototype is designed in line with the real requirements of the user (user-centred design). Additionally the chosen setup does not introduce much additional effort on the user, because accessing required information is done without or only with minimized explicit interaction with the system. In this manner there is no need for the user to be distracted e.g. by a stationary interaction device, such as keyboard or mouse that would impair the way the trainees do their job.

3 Initial Work

3.1 Usability Experiments at Vrchlabi

In the first phase of the project the aim was to create a wearable system that supported the training process in the learning island. This first prototype formed the basis to evaluate the different modalities of interaction with the assembly line workers under real conditions in the SKODA production site at Vrchlabi, Czech Republic. These were voice recognition, textile keyboard, non-explicit- or task recognition-based interaction, and Head Mounted Displays. The front headlight assembly process was selected as a test case since this specific process represents a complex enough task which justifies the use of wearable technologies during training. An additional aim was to acquire the users' feedback, regarding their preferences and attitudes towards several hardware-, software-, user interface- and remote support-related-features. Finally, the results obtained had to be useful for future analysis on the impact of wearable technologies in the training process itself. The detailed evaluation procedure has been described comprehensively in [Mau06].

The findings of the experiments were manifold. Generally, the usability test with the real end-users was rather complicated, as they had to be picked out from the running assembly line. Therefore the time schedule for usability testing was very tight, making it impossible to extend the tests when required. On the other hand it was not possible to recruit the amount of users needed in order to obtain statistically significant results. Dealing with real end-users, it was not feasible, trying to apply a user-centred approach through a human translator. In fact, all end users were unable to understand and speak English, which required the services of a professional Czech/English translator. These constraints prevented the usage of some effective techniques like 'thinking aloud', thus, interaction with the workers turned out to be un-natural and difficult. Additionally, it had to be dealt with the fact that the design-/developing team consisted of remotely located partners, what made it very complicated to guarantee a real iterative design process. Apart from these aspects, it was difficult to measure system performance. As a result it was not easy to quantitatively evaluate the training process. Due to the "learning effects", it was not legitimate to use the same worker to measure the same training process before and after the introduction of wearable technology. Furthermore it was difficult to compare two different users because of their different skills and learning capabilities. Nevertheless, the first results of the experiment confirmed that the wearable system was well accepted. Regarding all theoretical concerns about lack of privacy and loss in autonomy, it was a surprise that one of the favourite features of the workers was the ability of the wearable system to monitor the task completeness. In fact, when one of the workers made a mistake in the assembling process, the system detected it, and triggered an error message. Later on, during the post-questionnaire, it was one of the most valuables features the worker identified. It was also observed that one of the workers was very uncomfortable wearing the HMDs. This particular worker preferred using a large display in order to view the information. It has to be underlined as well, that video support was not requested by the majority of the workers. Generally they preferred pictures with aggregated information in comparison to simple text.

The main outcome of the Vrchlabi Usability Experiments was the decision to perform new usability experiments in Spain. The purpose was to involve a larger group of end-users located near the research team in order to overcome the constraints experienced at Skoda.

3.2 Usability Experiments at Tekniker

In order that the experiments were successful, an infrastructure was set-up to carry out assembly tasks. The prerequisites for such an infrastructure were rather simple:

- It should allow creating as many different tasks as we need,
- All tasks had to be of a similar complexity degree,
- The assembly task should involve the use of manually- or tool assisted part manipulation

Two main experiments on usability have been carried out using this platform. The aim of the first experiment was to extend the initial findings of the experiment made with the workers of Skoda at Vrchlabi. The intention was to measure the acceptance of the system. Besides, the performance in terms of memorability (how fast workers get trained), and in terms of task completion (time consumed and errors made). All in all 40 workers took part in the experiment.

In summary the main findings were:

- Users improved their performance when using the wearable system with implicit interaction: The assembly tasks were performed faster and with less error. In fact, it took 67 seconds less in average than when paper-based information was used which was actually the second best alternative.
- Users did not learn faster using a wearable system. In fact, people were able to learn faster through paper-based support. Although the difference was neglectable when compared to those using context based interaction.
- In the test performed the day after, paper-based learning performed the best, while context-based interaction performed the worst.
- Voice recognition-based interaction was the preferred interaction modality by workers.
- Workers preferred graphical information to text.
- Workers found the system very useful when doing a complex task, allowing hands free access to information, avoiding dispensable movements in order to check information

The second experiment was aimed to compare the benefits of using Head Mounted Displays (HMDs) to access information, versus the presentation of the same information on a large screen near the working place. 20 workers took part in the experiment and 3 different HMD evaluated..

The best performance was obtained when the information was presented on the large display, and the worst when accessing through the binocular HMD. However, when asked about the user's preferences, most of users chose the binocular HMD as the best choice.

3.3 The task recognition Prototype

In order to track the progress of the headlight assembly, sensors were integrated on distinctive parts of the car body, on the worker, and also on the tools. This was done to guarantee a detection of sub-tasks which are relevant for the different steps in the workflow.

As on-body sensor, a RFID reader has been attached on the back of the user's hand. With the information coming from this reader, required tools such as two cordless screwdrivers can be detected and uniquely identified. In addition, an inertial sensor package has been placed on the back of the user's hand. This module provides a three-axis accelerometer and gyroscope. It is used to pick up the incidence of the torque limiter of the cordless screwdrivers, which occurs when a screw is properly tightened according to the chosen torque.

As wiring up the worker would be too great an impediment for the user during his work, all data streams from the wearable sensors are transmitted using Bluetooth modules.

The correct positioning of the assembled car components is monitored by a set of stationary sensors mounted directly on the car body. Critical locations with permanent contact to the component, e.g. the contact surfaces behind screws, are monitored by measuring the force exerted on force sensitive resistors (FSR) on these surfaces. The FSRs' very low thickness (about 0.5 mm) allows them to be placed on such positions without modifications of the components or the car body. At locations where the assembled components do not touch the car body, we used magnetically triggered reed switches. They also measure the proximity of alignment checking tools used at specific places for quality control.

The current setup uses 5 FSR sensors and 4 reed switches around screw positions, the back of the lamp body, and the checkpoints for the alignment tools. The signals from stationary sensors are read and preprocessed by a microcontroller based data acquisition module, which is mounted inside the engine compartment.

The data streams coming from the wearable sensors on the user and the stationary sensors inside the car body are gathered and further processed by the Context Recognition Toolbox (CRN) [Ba06]. This software framework provides a library of data processing algorithm modules frequently used for context recognition, and allows to set up a process network using these modules. Fig. 3 illustrates the detailed network, which has been used for the task recognition prototype. The left part comprises the processing of FSRs and reed switches. After some signal conditioning, a threshold operator is applied to detect the status of the respective sensor. The middle thread in figure 1 shows the CRN tasks, which are dealing with the RFID reader. On the right side, the chain of tasks is depicted which detects the occurrence of the described torque limiter. A more detailed description of the data processing and the resulting event creation is given in [St05]. A merger task brings these three streams together and sends it to the JContextAPI using a TCP/IP connection.

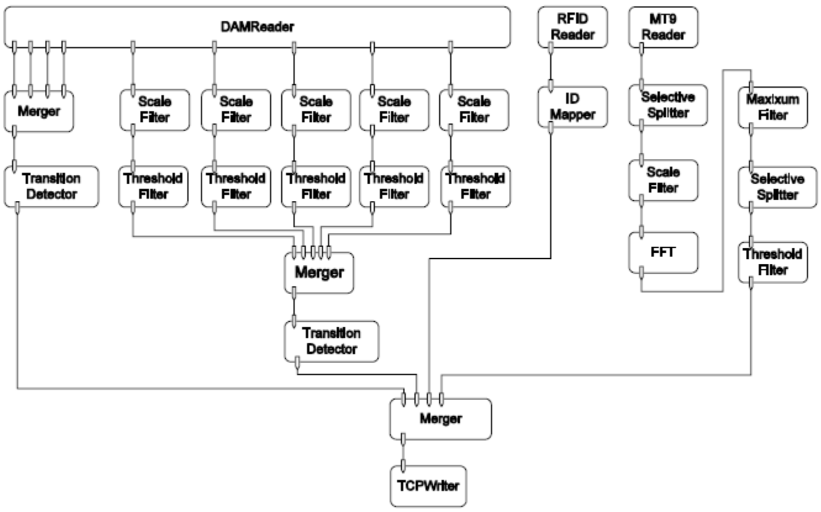


Fig. 3: Detailed Network used for the Task recognition Prototype

4. The final Prototype

4.1 Hardware Architecture

The current prototype is based on a distributed architecture using several components from the Wearit@work framework. The main device is an OQO Model

01+ [Oq01], which the user wears attached to his belt. The OQO technical characteristics offer enough computational power to fully support the application, and it also allows for appropriate connectivity: network, bluetooth, external VGA output. The user has a Carl Zeiss binocular look-around head mounted display [Ze01], and a Sony Ericson HBH-300 bluetooth headset to interact with the prototype application via voice commands.

The tracking of the user's actions is enabled by a special data glove that has been engineered by ETH Zurich. It comprises an inertial sensor package for motion detection and a RFID reader to identify utilized tools. In addition, a set of force resistive sensors and magnetic switches is attached to the car body for monitoring interactions between tools, assembly parts and the car body itself. The output of the sensors is collected by a stationary system (a laptop), which processes them and makes them available for the recognition of user's actions. Fig. 4 shows a user wearing the wearable hardware components.

Fig. 4: User wearing the hardware components

4.2 Software Architecture

The software architecture is shown in Fig. 5. The end-user application (henceforth also referred to as application, for brevity) is written in Java and it runs on the OQO. Further, it relies on the Open Wearable Computing Framework (OWCF), which has been developed within the Wearit@work project. Specifically, the application uses the following OWCF components: Wearable User Interface (WUI) and JContextAPI (JContextAPI is an implementation and an extension of the ideas presented in [Sb06]).

The application shown in Fig. 6 is modelled internally as a finite state machine: each state corresponds to a specific task of the assembly procedure for which the user is being trained; transitions are triggered by user actions, both explicit (for example voice commands) and implicit (i.e. actions that are performed as part of the assembly procedure, and that are detected and recognized automatically by the system). The application is capable of tracking the sequence of user's actions, and to monitor that such sequence corresponds to what is expected in the assembly procedure. Whenever the user performs an unexpected action, the application displays a warning message, and can contextually provide appropriate help to support the user.

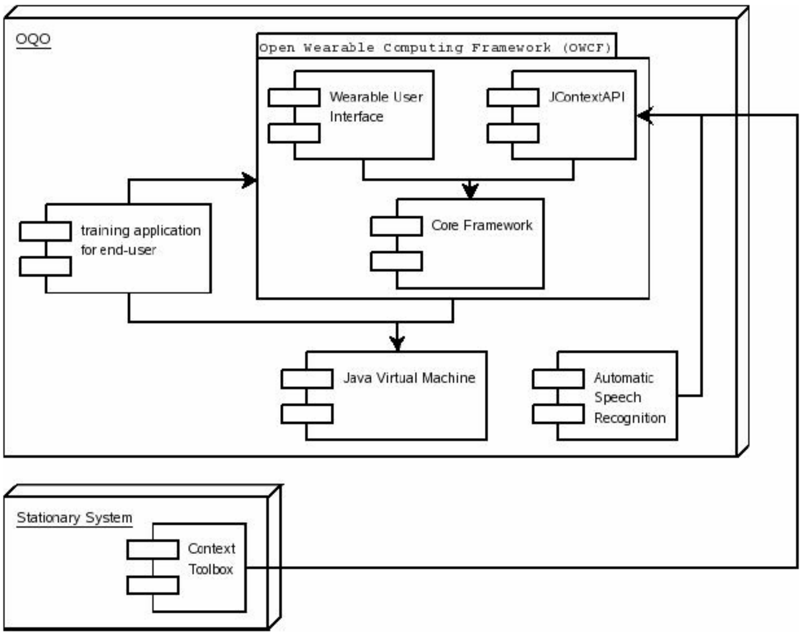


Fig. 5: Software Architecture

The application user interface is based on the WUI [Wi07,Wi06], which presents the required information to support the user during the steps of the assembly procedure in the most suitable way. The WUI is engineered to obtain the best result in presenting the output on "look around" head mounted displays. Further, it provides a very good support for the state-based architecture of the application. As such, states can be associated with abstract screen structures, each of which are rendered as the required graphical widgets (text boxes, pictures, menu items, etc.), and the navigation towards other screens. The WUI also takes care of building the best rendering of the user interface dependent upon the output device. For this, the envisioned interface capabilities are described with an abstract model independent of any specific interaction style or output medium; implementing a 'separation of concerns software' approach. Hence, application developers can focus on specifying required in- and output behaviour of the envisioned interface, without being in need to take care of its resulting rendering.

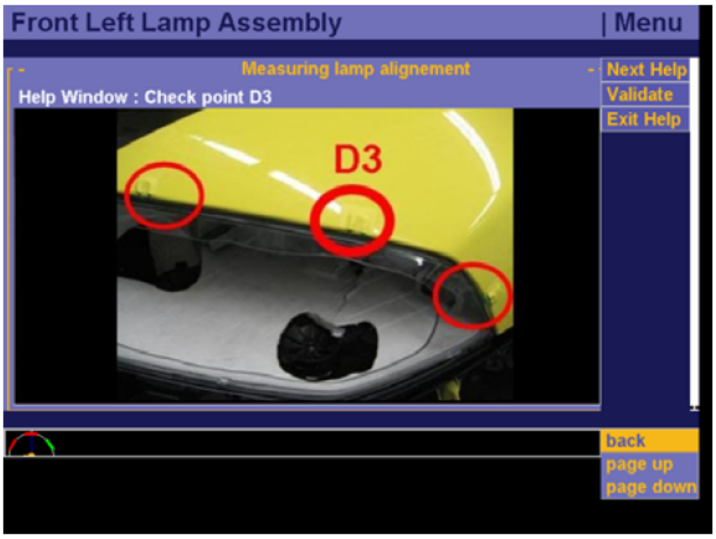


Fig. 6: The user-oriented application

Transitions of the application state machine are provided by JContextAPI in the form of events: The application simply registers listeners for the relevant events, and JContextAPI notifies the registered listeners whenever those events happen within the system. There are two major event sources in the current software architecture: the Context Toolbox [Ba06], which is running on the Stationary System, and the Automatic Speech Recognition (ASR) component, which is also running on the OQO. The *Context Toolbox* is used to process the signals coming from the special data glove worn by the user, and from the sensors attached to the car body. Such data are transferred to the OQO via TCP socket, and are processed by JContextAPI, which transform them in events, and eventually notifies registered listeners. The same approach is used with the *Automatic Speech Recognition (ASR)* component. The current version of the prototype uses a commercial ASR, Dragon Naturally Speaking from Scansoft [Sc01]. The ASR is trained to recognize some specific commands (words like “show help”, “next”, etc.), and publishes the recognized token on a TCP socket. JContextAPI transforms this data into events that are forwarded to registered listeners.

Notice that JContextAPI performs also some additional elaboration of the data received by the Context Toolbox, in order to generate events of a higher abstraction level. This feature is useful and meaningful to the application, and can be used to trigger transitions in its internal finite state machine.

The usage of the OWCF components has simplified the engineering of the application providing several benefits:

- the WUI supports the rendering of a simple user interface structure specifically targeting head mounted displays
- the JContextAPI enables context awareness through the simple and straightforward mechanism of event subscription/notification.

In general OWCF has provided a set of reusable components that shielded the application developers from the complexities related to the interfacing and handling of lower level system resources and data sources.

5. Conclusions and Further Work

The results of the study have confirmed that effective training of personnel in automotive production is one of the most crucial factors which are responsible for securing production flexibility. The developed wearable computing prototype enables a context-sensitive provision of necessary information to the training personnel. The wearable solution was able to track and analyse the trainee's actions, while providing the end user with actions for error handling. As a result, semi-autonomous training of trainees in automotive production was realised.

In the usage of wearable computing solutions for supporting training procedures, high benefits can be expected. However, at current stage there is not yet enough experimental data to draw clear conclusions on further benefits and issues of the proposed solution. Nevertheless the consortium will continue to refine the solution according to end-users feedback, and to conduct further tests and field studies within Skoda in order to gather enough knowledge to evaluate the prototype more comprehensively. In this respect there are some interesting features that have to be tested before the final prototype can be deployed in the E-Learning Institute, in 2007:

- There is a new wristband designed and developed that has to be tested, both in terms of performance and -user acceptance,
- Collaboration between trainer and trainee has to be implemented,
- Different methods for event notification have to be tested.

Besides the refinement of the final training prototype, WearIt@work envisages a second wearable prototype which will empower blue collar workers in selected stations of the assembly -line. In this context an extensive field study was recently initiated, and will continue until the second quarter of 2007. The studies will include an end user- and process study, which shall directly impact the features of the second prototype According to the most recent requirements analysis with the Skoda end users, WearIT@work has identified the quality assurance operations at the assembly-line as a promising test scenario for employing wearable computing. Regarding this particular scenario, blue collar workers manually and visually check the functionalities of specific features of the almost finished car (e.g. doors, windows, and trunk). It is the aim of the workers to find errors and malfunctions which are then to be documented (for the time being: paper-based). Thus, the total number of errors is to be reduced before the car is audited in its final station (namely before being shipped to the customer). The objective of the scenario shall be to empower the workers with wearable computing, in order that they may be supported in error detection and –documentation. As such, we believe that as an outcome, a substantial reduction of total errors can be achieved. For the technical realisation of the wearable solution, components (hardware as well as software) developed for the training scenario will be reused and adapted to the specific requirements of the end users. Additionally, we will make use of synergies existing among other WearIT@work scenarios such as inspection procedures for aircraft maintenance.

Regarding the application of the User Centred Design approach, it came out that within an international integrated project, guaranteeing this kind of approach is not at all an easy task. In fact, the cultural and geographical distance makes it quite difficult to apply an orthodox approach. In WearIT@work, the production pilot team is following their own sophisticated methodology: namely an initial requirement elicitation process with real end-users, usability tests with local users close to the research team, and final validation with final end users. It must however be mentioned that it is nevertheless very difficult to get hold of the real end users in a global company where production is highly dependent upon human resources. Moreover, a critical selection of the 'right' end users is crucial due to diverse cultural- and educational backgrounds. Conclusively not any end user, e.g. of the assembly-line, is suitable for interviewing and evaluation purposes. WearITwork is currently in the middle of this process, and is planning to have preliminary results in early 2007, which will naturally be subject to publishment.

Acknowledgements

This work was partly funded under grant 004216 by the European integrated project "WearIT@work - Empowering the mobile worker by wearable computing"

References

[Ba06] Bannach, D., Kunze, K., Lukowicz, P., Amft, O.: Distributed modular toolbox for multi-modal context recognition, Proceedings of the Architecture of Computing Systems Conference, (2006) 99-113

[Kir06] Kirisci, P., Morales, E.: Der Einsatz von Wearable Computing im industriellen Kontext, Mobilfunk:Technologien und Anwendungen, ITG-Fachbericht Nr. 194, 17.-18. Mai, 2006, VDE Verlag, Osnabrück, (2006)115-120

[Ma06] Matysczok, C.: Einsatzmöglichkeiten mobiler IT-Systeme in der Automobilindustrie im Bereich der Aus- und Weiterbildung, Zeitschrift für den wirtschaftlichen Fabrikbetrieb (ZWF), 3/2006, Hanser, (2006)

[Mau06] Maurtua, I., Pérez, M. A., Unceta, M., Garmendia, I.: Experimenting Wearable Technology at Production, Deliverable 22 (D22) of the WearIt@work project, (2006) 45-52

[St06] Stiefmeier, T., Lombriser, C., Roggen, D., Junker, H., Ogris, G. Tröster G.: Event-Based Activity Tracking in Work Environments, 3rd International Forum on Applied Wearable Computing (IFAWC), Bremen, Germany, March 15 - 16, (2006)

[Siew96] Siewiorek, D. P, Finger, S. Terk, M., Subrahmanian, E., Kasabach, C., Prinz, F., Smailagic, A., Stivoric, J., Weiss, L.: Rapid Design and Manufacture of Wearable Computers, Communications of the ACM, Vol. 39, No. 2, (1996) 63-70

[Wi05] Witt, H.: A Toolkit for Context-aware User Interface Development for Wearable Computers. Doctoral Colloquium at the 9th International Symposium on Wearable Computers (ISWC), Osaka, Japan, October 18-21, (2005)

7.24. Ambient Intelligence in Manufacturing: Organizational Implications

Irene Lopez de Vallejo, Iñaki Maurtua, Miren Unceta. Ambient Intelligence in Manufacturing: Organizational Implications. Workshop on 'Ubiquitous Computing and effects on Social Issues' Portland, Oregon. 2005 [124]

Ambient Intelligence in Manufacturing: Organizational Implications

Inaki Maurtua, Miren Unceta
Engineering Production Department,
Fundacion TEKNIKER,
Avenida Otaola, 20
20600 Eibar, Gipuzkoa, Spain
{ imaurtua;munceta }@tekniker.es

Irene Lopez de Vallejo
The Bartlett, Faculty of the Built Environment,
University College London,
Gower Street, London,
WC1E 6BT, UK
{ i.vallejo }@ucl.ac.uk

ABSTRACT
There are several references to research projects involving the Ambient Intelligence concept (AmI) in technical publications. These are mainly focused on applications linked to the everyday life of individuals - at home, in the street, in the car, in public spaces – and largely related to leisure and to a specific concept of quality of life. But, we wonder, how does the AmI concept apply to the manufacturing field? What's going on in the industrial environment? In this article we introduce a research project, Tekniker Foundation[1] is currently conducting on organizational implications of the application of the AmI vision in a Manufacturing environment.

Author Keywords
Ambient Intelligence, Manufacturing, Organizational implications.

ACM Classification Keywords
H5.3 [Information interfaces and presentation (e.g., HCI)]: Group and Organization Interfaces---Computer-supported cooperative work, Organizational design.

H5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces---User-centered design.

INTRODUCTION

There are several references to projects linked to the AmI concept in technical publications. These are mostly centered on applications linked to everyday situations concerning human beings: at home, in the street, in the car, in public places and relating to leisure. What happens in the Industrial environment? The opportunities for its application in the field of work are frequently referred to, although it is mainly the office environment that is ever mentioned [1].

In two of the most important studies relating to manufacturing recently completed [2] [3], where some of the key technologies in the future of manufacturing have been identified such as flexible manufacturing and control systems, decision support systems, improved people-machine interfaces, equipment, re-configurable and adaptable processes and systems, distributed computation, etc., the link to AmI related technologies has been made clear.

Potentially, the adoption of this vision could affect all stages in the manufacturing process: the design of the plant or the product, engineering, the organization and management of production, process control, quality control, maintenance, logistics and the management of the product throughout its life cycle, including its re-cycling.

In this perspective:

- AmI approach affects the products that are manufactured: the products designed for this type of manufacture will also be intelligent, and able to communicate and report on advantages throughout the life cycle. The products will have greater added value, since they will incorporate "intelligence" and "connectivity" capabilities which will make it possible to provide new services.
- Systems will be easier to use, not because they are aimed at less well trained operators, but because operators with the right skills can find the most relevant information, at the moment they need it and use it to make more appropriate decisions. A consequence of this process will be freeing the worker from most routine tasks.
- Integrating human and technical resources to improve satisfaction and efficiency among workers. Given that the vision is human – centered, the implementation of all technologies will be thought from the workers perspective, to accomplish their needs and help to solve their potential problems.

How does this affect the work processes, the physical and mental health of the operators, motivation, stress, satisfaction, team work, the method of learning? For AmI to be socially accepted in the workshop, a great effort will have to be made to provide a human - centered design, in order to prevent it from becoming an environment where people feel surrounded by a unit of electronic parts. The present research also issues aspects such as privacy and

[1] http://www.fundaciontekniker.com

social impact of the implementation of AmI in a manufacturing environment.

AMI LABORATORY AT TEKNIKER

AmI in Manufacturing: a scenario

"At the Hirurak manufacturing plant the daily routine goes on. Mikel is going to the office area to have a meeting with Jon. As he passes by the transfer machine being operated by María, he receives a warning on the optical device fitted to his glasses. The monitoring agent has detected that the motorized spindle performing the second drilling operation is vibrating at slightly above the nominal levels. Mikel displays the RMS signal for the vibration on his glasses, together with possible causes for this, suggested by the agent. He has time before the meeting with the manager and decides to solve the problem.

Mikel goes to the rear part of the machine and opens the gate leading to the access to the motorized spindle. The tracking and localization system has detected him. The maintenance agent uses this information to help him carry out the maintenance. He can see the steps to be followed, estimated time and materials needed on his glasses An alarm message is displayed on his glasses, Mikel is not wearing the safety gloves demanded by internal Health and Safety at Work procedures. Mikel reluctantly puts them on. Now he can continue with his work.

There is very little space in which to carry out the maintenance. Fortunately, he has all the necessary information within his reach. Mikel does not remember what the X21 connector is referred to in the instructions. He asks in Basque, his first language, for a detailed plan, and it is made available immediately on his glasses.

Meanwhile, María is preparing the programs for the next manufacturing order which is starting this afternoon. María is not a great expert in machining, but it does not matter. The Hirurak programming system is able to interpret what María wants when she expresses it in a semi-colloquial language that it can translate to the ISO programming language understood by the CNC on the machine. María, who is slightly disabled in one of her hands, often uses her voice as an interface device with the machine. This has made her able to find a job much more easily."

Mikel has finished the maintenance and is going to the office to have the meeting. He is arriving there a little late, but his personal agent had been in contact with Jon's personal agent, and they had re-arranged their agendas. There was no problem in starting half an hour later."

Description of the AmILAB Laboratory at Tekniker

Tekniker is a research foundation centered on subject matters related to manufacture and micro-technology. It has set up a laboratory to spread the model of AmI to the area of manufacturing.

In order to achieve this, there must be convergence in areas traditionally linked with manufacture (quality, production, maintenance, design, safety, etc.) and the technology made possible by AmI (multi-modal interfaces, communications, learning systems, embedded systems, etc.)

The main research objectives of the laboratory are:

- to support complex tasks with a minimum of human-machine interaction
- to enable mobile professionals to keep their attention focused on the interaction with the work environment.
- to investigate the user acceptance of wearables, methods for user interaction
- to identify processes suited to wearables in industry

Figure 1. General vision of the Laboratory. The Milling Machine in the center of the image

The central unit in the laboratory is a high-speed milling machine with linear motors and a magnetically positioned head, and applications/functions normal in a manufacturing environment, such as monitoring and operating the machine, production control and maintenance.

Its functions have been augmented by accessorizing to it various devices such as a voice recognition system, 2 head mounted displays, 1 tablet PC, 2 PDAs, 1 Xybernaut wearable computer, 1 smart phone (Nokia 6600), and a localization system based on radio frequency identification tags (RFID).

Maintenance and production control applications have been re-designed and implemented in the form of Agents on the JADE platform [4], incorporating the concept of intelligent interfaces, context sensitivity and automatic learning.

2

From the point of view of the interfaces, we are currently conducting experiments to incorporate the voice recognition system and the Head Mounted Display to control the machine and provide access to the information needed by the worker.

Fig.2. A worker wearing HMD connected to the Xybernaut and using the Milling Machine's CNC keyboard to interface the wearable computer

In the next months new equipment will be introduced in the laboratory to explore new forms of interface with the machine: a robot and new interface technologies based on haptic systems.

Even assuming that the AmI vision is technically possible, we do not know if it will be economically, but what we believe is that social acceptance will be the critical factor that will enable it to come into being. To help construct this vision of the future whether in manufacturing or in any other area, aspects that go beyond the purely technological must be confronted: Sociological, legal, organizational and ethical aspects [5], such as the perception of loss of independence, control and privacy, the augmentation of the physical surroundings, accepting and using new interfaces, new learning opportunities, new ways of interacting, have to be taken into account whenever AmI vision is tried to be accomplished. All these aspects centered on people at work are also tackled in the laboratory.

The AmILAB is a research funded by the Basque and Spanish Science and Technology Research programmes, starting in the late 2003. Tekniker is also the only Spanish partner to take part in the largest world-wide European Commission funded IP (Integrated Project) on the area of wearable computing: "wearit@work: Empowering the mobile worker by wearable computing"[2].

INTRODUCING AMI IN INDUSTRIAL COMPANIES

When planning the introduction of AmI vision in industrial companies, we are talking about supporting complex tasks with a minimum of human – machine interaction to enable the workers to keep their attention focused on the interaction with the work environment, co-workers and tasks. [6] We are talking here about an organizational change process, where the focus relies on identifying suitable processes for the implementation of AmI technologies in industrial companies.

In this paper we will summarize a preliminary vision on organizational issues that will entail the introduction of AmI, exploring three main research questions: do organizations have to fulfill any previous conditions? Which role plays the General Management in this process? How should the process be implemented? [7] [8]

There are different criteria to define the characteristics and type of company suitable for the introduction of this vision:

- Management Style. AmI implementation asks for a participative process, present in management styles such as self-management models, cooperatives models. But it is also possible to implement it in a more autocratic managed company.
- Type of activity. One could think that knowledge based companies are more suitable, but in some purely manufacturing companies AmI related concepts might be easier to introduce, i.e. workshop workers find very useful interface technologies that allow them an easy command of the machine and a better acquisition of information
- Type of knowledge predominant, tacit knowledge vs. explicit knowledge companies. Experience "stored" in people vs. experience formalized in forms and procedures. Which type of company is more adequate to implement AmI?
- Human resources policy, here aspects such as transparency and an open communication policy are definitely positive. But the key is the agreement or social contract between workers and management about the use of the information handled by the AmI systems, potentially highly disruptive with large influence on privacy and control issues.

The conclusion is that there is no right combination of criteria that guarantees a successful adoption of AmI. Some are more adequate than others – a company with a

[2] http://www.wearitwork.com

3

participative management style and a transparent human resources and IT resources policy, but regarding types of activity or knowledge predominant we can't take a position.

Regarding criteria to apply in the process of implementing this type of vision, we can say that participation, a step-by-step process, voluntary where people can opt in or out of it when desired, and where a social contract or agreement between the different stakeholders are key issues to be taken into account:

- Participation. All stakeholders involved have to take active part, from the very beginning, on the implementation, providing their unique expertise and perspective. This will smooth the process and will help to create a common and tailored vision that answers to the real needs of the people working on the company.
- Step-by-step process. Start with a pilot Project inside the company to put in practice the vision. In order to demonstrate benefits and minimize inherent risks to the implementation of new technologies in any company.
- Voluntary, the possibility of opting in and/or out of the technology is key in this process. People have to feel in control of the technology.
- Social contract. A multilateral commitment between the different stakeholders in the company – management, workers, trade unions – has to be achieved. The negotiation and agreement on the managerial and social rules of use of AmI technology has to be explicit and revisited when possible new unsuspected aspects arise:
    - Understanding by the workers that they are working in an AmI environment, and that some of their activities, movements and interactions are registered by the system.
    - Allow opt-in/opt-out options of use of the system
    - Allow access to the personal information registered and stored by the system. And the possibility of modify or delete it.
    - Assure the absolute confidenciality of the data stored.
    - Clearly define which personal data are going to be used by the management.
    - Creation of "AmI free" spaces
    - Guarantee not to use the data for punitive processes (rewards. Salary policy)
    - Consider the possibility of a total rejection of AmI by the workers. Contingency plan.

During the implementation process a strong leadership is needed in order to motivate the workers and to lead the change process. The leader, recognized by the staff and who transmits authority, has to facilitate a context in which the human mistake is acceptable while using these new devices.

CONCLUSION

In the area of manufacturing, AmI is not only going to affect the way in which processes develop, but will also provide new ways of working and doing business. The development of new products and services and the shift in the focus of attention of the worker from the machine to their immediate working environment will be the immediate consequences of the adoption of AmI vision. In no case this is proposed as a new manufacturing paradigm, but, whatever models are followed, these will have to take into account these changes and feed organizational processes and communication and collaboration interactions.

Not only several technological challenges such as miniaturization, inter-operability and energy management have to be addressed by research teams across the world in the present decade, but also a strong focus on the social and organizational aspects of AmI has to be taken to overcome barriers to its realization.

The introduction of this new range of information and communication technologies in industrial companies entails changes in working procedures and the way people interact and relate to each other, having a potentially significant influence in the way these organizations will be managed and experienced in the near future.

ACKNOWLEDGMENTS

We want to thank to the different research funding bodies that allow us to dedicate part of our efforts to try to make the world a better place through a more human understanding and use of technology in our workplaces, the Basque Government Technology and Industry Ministry, the Spanish Science and Technology Ministry, and in particular to the EU Commission and the project wearIT@work.

REFERENCES

1. ISTAG: Scenarios for Ambient Intelligence in 2010. Final Report, Feb. 2001. IPTS
2. Expert Group on Competitive & Sustainable Production and Related service Industries in Europe in the Period to 2020, EU Commission, "Sustainable Production: Challenges & objectives for EU Research Policy", 2001
3. Visionary Manufacturing Challenges for 2020. Committee on Visionary Manufacturing Challenges, National Research Council. NATIONAL ACADEMY PRESS,Washington, D.C. 1998
4. JADE: Java Agent Development Framework, http://jade.cselt.it/index.html

5. López de Vallejo, Irene "Ambient Intelligence Vision: exploring the social risks of its construction", October 2004, published on the proceedings of eChallenges 2004 International Conference.
6. Mark Weiser, "The Computer of the 21st Century", Scientific American, September 1991, pg. 78
7. C. Fernández, F. Sánchez, J.I. Yánez. "Rumbo, camino a la empresa innovadora"
8. López de Vallejo, Irene "Soft Factors in the New ICT powered workspace", October 2003, published on the proceedings of eChallenges 2003 International Conference.

The columns on the last page should be of approximately equal length.

8. PATENTES

8.1. MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD

MINISTERIO
DE INDUSTRIA, ENERGÍA
Y TURISMO

Oficina Española
de Patentes y Marcas

Justificante de presentación electrónica de solicitud de patente

Este documento es un justificante de que se ha recibido una solicitud española de patente por vía electrónica, utilizando la conexión segura de la O.E.P.M. Asimismo, se le ha asignado de forma automática un número de solicitud y una fecha de recepción, conforme al artículo 14.3 del Reglamento para la ejecución de la Ley 11/1986, de 20 de marzo, de Patentes. La fecha de presentación de la solicitud de acuerdo con el art. 22 de la Ley de Patentes, le será comunicada posteriormente.

Número de solicitud:	P201730325	
Fecha de recepción:	10 marzo 2017, 17:22 (CET)	
Oficina receptora:	OEPM Madrid	
Su referencia:	P170295ES	
Solicitante:	FUNDACIÓN TEKNIKER	
Número de solicitantes:	1	
País:	ES	
Título:	MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD	
Documentos enviados:	Descripcion.pdf (17 p.)	package-data.xml
	Reivindicaciones.pdf (3 p.)	es-request.xml
	Dibujos.pdf (3 p.)	application-body.xml
	Resumen.pdf (1 p.)	es-fee-sheet.xml
	OLF-ARCHIVE.zip	feesheet.pdf
	FEERCPT-1.pdf (1 p.)	request.pdf
Enviados por:	CN=VALLEJO LOPEZ JUAN PEDRO - 33501382L,givenName=JUAN PEDRO,SN=VALLEJO LOPEZ,serialNumber=33501382L,C=ES	
Fecha y hora de recepción:	10 marzo 2017, 17:22 (CET)	
Codificación del envío:	1E:B5:06:4C:05:0E:2C:BC:6D:F5:5C:E8:81:B1:BA:67:44:AC:A4:53	

AVISO IMPORTANTE
Las tasas pagaderas al solicitar y durante la tramitación de una patente o un modelo de utilidad son las que se recogen en el Apartado "Tasas y precios públicos" de la página web de la OEPM (http://www.oepm.es/es/propiedad_industrial/tasas/). Consecuentemente, si recibe una comunicación informándole de la necesidad de hacer un pago por la inscripción de su patente o su modelo de utilidad en un "registro central" o en un "registro de internet" posiblemente se trate de un fraude.
La anotación en este tipo de autodenominados "registros" no despliega ningún tipo de eficacia jurídica ni tiene carácter oficial.
En estos casos le aconsejamos que se ponga en contacto con la Oficina Española de Patentes y Marcas en el correo electrónico informacion@oepm.es.

ADVERTENCIA: POR DISPOSICIÓN LEGAL LOS DATOS CONTENIDOS EN ESTA SOLICITUD PODRÁN SER PUBLICADOS EN EL BOLETÍN OFICIAL DE LA PROPIEDAD INDUSTRIAL E INSCRITOS EN EL REGISTRO DE PATENTES DE LA OEPM, SIENDO AMBAS BASES DE DATOS DE CARÁCTER PÚBLICO Y ACCESIBLES VÍA REDES MUNDIALES DE INFORMÁTICA.
Para cualquier aclaración puede contactar con la O.E.P.M.

/Madrid, Oficina Receptora/

MINISTERIO
DE INDUSTRIA, ENERGÍA
Y TURISMO

Oficina Española
de Patentes y Marcas

(1) MODALIDAD:	
PATENTE DE INVENCIÓN	[✓]
MODELO DE UTILIDAD	[ ]
(2) TIPO DE SOLICITUD:	
PRIMERA PRESENTACIÓN	[✓]
ADICIÓN A LA PATENTE EUROPEA	[ ]
ADICIÓN A LA PATENTE ESPAÑOLA	[ ]
SOLICITUD DIVISIONAL	[ ]
CAMBIO DE MODALIDAD	[ ]
TRANSFORMACIÓN SOLICITUD PATENTE EUROPEA	[ ]
PCT: ENTRADA FASE NACIONAL	[ ]
(3) EXP. PRINCIPAL O DE ORIGEN:	
MODALIDAD:	
N.º SOLICITUD:	
FECHA SOLICITUD:	
4) LUGAR DE PRESENTACIÓN:	OEPM, Presentación Electrónica
(5-1) SOLICITANTE 1:	
DENOMINACIÓN SOCIAL:	FUNDACIÓN TEKNIKER
UNIVERSIDAD PÚBLICA	[ ]
NACIONALIDAD:	España
CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
CNAE:	
PYME:	
DOMICILIO:	Polo tecnológico de Eibar, c/ Iñaki Goenaga 5
LOCALIDAD:	Eibar
PROVINCIA:	20 Guipúzcoa
CÓDIGO POSTAL:	20600
PAÍS RESIDENCIA:	España
CÓDIGO PAÍS:	ES
TELÉFONO:	
FAX:	
CORREO ELECTRÓNICO:	
PERSONA DE CONTACTO:	
MODO DE OBTENCIÓN DEL DERECHO:	
INVENCIÓN LABORAL:	[✓]
CONTRATO:	[ ]
SUCESIÓN:	[ ]
PORCENTAJE DE TITULARIDAD:	100,00 %
(6-1) INVENTOR 1:	
APELLIDOS:	TUBIO OTERO
NOMBRE:	Carlos
NACIONALIDAD:	España
CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
(6-2) INVENTOR 2:	
APELLIDOS:	SUSPERREGI ZABALO
NOMBRE:	Loreto
NACIONALIDAD:	España
CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
(6-3) INVENTOR 3:	
APELLIDOS:	FERNANDEZ GONZALEZ
NOMBRE:	Ane
NACIONALIDAD:	España

CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
(6-4) INVENTOR 4:	
APELLIDOS:	ANSUATEGI COBO
NOMBRE:	Ander
NACIONALIDAD:	España
CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
(6-5) INVENTOR 5:	
APELLIDOS:	MAURTUA ORMAECHEA
NOMBRE:	Iñaki
NACIONALIDAD:	España
CÓDIGO PAÍS:	ES
NIF/NIE/PASAPORTE:	
(7) TÍTULO DE LA INVENCIÓN:	MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD
(8) PETICIÓN DE INFORME SOBRE EL ESTADO DE LA TÉCNICA:	
SI	[ ]
NO	[✓]
(9) SOLICITA LA INCLUSIÓN EN EL PROCEDIMIENTO ACELERADO DE CONCESIÓN	
SI	[ ]
NO	[✓]
(10) EFECTUADO DEPÓSITO DE MATERÍA BIOLÓGICA:	
SI	[ ]
NO	[✓]
(11) DEPÓSITO:	
REFERENCIA DE IDENTIFICACIÓN:	
INSTITUCIÓN DE DEPÓSITO:	
NÚMERO DE DEPÓSITO:	
ACCESIBILIDAD RESTRINGIDA A UN EXPERTO (ART. 45.1. B):	
(12) DECLARACIONES RELATIVAS A LA LISTA DE SECUENCIAS:	
LA LISTA DE SECUENCIAS NO VA MÁS ALLÁ DEL CONTENIDO DE LA SOLICITUD	[ ]
LA LISTA DE SECUENCIAS EN FORMATO PDF Y ASCII SON IDENTICOS	[ ]
(13) EXPOSICIONES OFICIALES:	
LUGAR:	
FECHA:	
(14) DECLARACIONES DE PRIORIDAD:	
PAÍS DE ORIGEN:	
CÓDIGO PAÍS:	
NÚMERO:	
FECHA:	
(15) AGENTE DE PROPIEDAD INDUSTRIAL:	
APELLIDOS:	VALLEJO LÓPEZ
NOMBRE:	Juan Pedro
CÓDIGO DE AGENTE:	0994/6
NÚMERO DE PODER:	
(16) RELACIÓN DE DOCUMENTOS QUE SE ACOMPAÑAN:	
DESCRIPCIÓN:	[✓] N.º de páginas: 17
REIVINDICACIONES:	[✓] N.º de reivindicaciones: 16
DIBUJOS:	[✓] N.º de dibujos: 4
RESUMEN:	[✓] N.º de páginas: 1
FIGURA(S) A PUBLICAR CON EL RESUMEN:	[✓] N.º de figura(s): 1A
ARCHIVO DE PRECONVERSION:	[✓]
DOCUMENTO DE REPRESENTACIÓN:	[ ] N.º de páginas:
JUSTIFICANTE DE PAGO (1):	[✓] N.º de páginas: 1
LISTA DE SECUENCIAS PDF:	[ ] N.º de páginas:

ARCHIVO PARA LA BUSQUEDA DE LS: OTROS (Aparecerán detallados):	[ ]
(17) EL SOLICITANTE SE ACOGE AL APLAZAMIENTO DE PAGO DE TASA PREVISTO EN EL ART. 162 DE LA LEY 11/1986 DE PATENTES, DECLARA: BAJO JURAMIENTO O PROMESA SER CIERTOS TODOS LOS DATOS QUE FIGURAN EN LA DOCUMENTACIÓN ADJUNTA:	[ ]
DOC COPIA DNI:	[ ] N.º de páginas:
DOC COPIA DECLARACIÓN DE CARENCIA DE MEDIOS:	[ ] N.º de páginas:
DOC COPIA CERTIFICACIÓN DE HABERES:	[ ] N.º de páginas:
DOC COPIA ÚLTIMA DECLARACIÓN DE LA RENTA:	[ ] N.º de páginas:
DOC COPIA LIBRO DE FAMILIA:	[ ] N.º de páginas:
DOC COPIA OTROS:	[ ] N.º de páginas:
(18) NOTAS:	
(19) FIRMA:	
FIRMA DEL SOLICITANTE O REPRESENTANTE:	VALLEJO LOPEZ JUAN PEDRO - 33501382L
LUGAR DE FIRMA:	MADRID
FECHA DE FIRMA:	10 Marzo 2017

MINISTERIO
DE INDUSTRIA, ENERGÍA
Y TURISMO

Oficina Española
de Patentes y Marcas

OFICINA ESPAÑOLA DE PATENTES Y MARCAS		
Hoja informativa sobre pago de tasas de una solicitud de patente o modelo de utilidad		
1. REFERENCIA DE SOLICITUD	**P170295ES**	
2. TASAS	**Importe (en euros)**	
Concepto	**Código de barras asignado**	**Importe**
IE01 Solicitud de demanda de depósito o de rehabilitación.	88153437746	63,68
IE02 Solicitud de cambio de modalidad en la protección		0,00
IE04 Petición IET		0,00
IE06 Prioridad extranjera (0)		0,00
El solicitante se acoge al aplazamiento de tasas previsto en el art. 162 de la Ley 11/1986 de Patentes	❑	
El solicitante es una Universidad pública	❑	
	Importe total	63,68
	Importe abonado	63,68

Se ha aplicado el 15% de descuento sobre la tasa de solicitud de acuerdo con la D. Adic. 8.2 Ley de Marcas.

MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO

Oficina Española de Patentes y Marcas

TASA en materia de Propiedad Industrial
CODIGO 511

Modelo 791

Identificación

Ejercicio: 2017
Nro Justificante: 7915112871680

Sujeto Pasivo

NIF:

Apellidos y Nombre o Razón Social:

Agente o Representante legal (1):

NIF: 33501382L

Apellidos y Nombre o Razón Social: JUAN PEDRO VALLEJO LÓPEZ

Código de Agente o Representante (2): 0994

Autoliquidación

Titular del expediente si es distinto del pagador: FUNDACIÓN TEKNIKER

Modalidad Expediente: P   Número Expediente: Tipo (3):

Clave: IE01   Año: 2017   Concepto: SOL. DE INVENCION O REAHABILITACIÓN POR INTERNET

Unidades: 1   Importe: 63,68

Referencia OEPM: 88153437746

909992100200188153437746

Declarante

Fecha: 10/03/17 13:12

Firma: JUAN PEDRO VALLEJO LÓPEZ

Ingreso

Importe en euros: 63,68   Adeudo en cuenta: ☒

NRC Asignado: 7915112871680000000001

Modelo 791

(1) Solo cuando el pago se realice con cargo a la cuenta corriente del representante o agente.
(2) En el caso de que tenga asignado un número por la OEPM.
(3) En el caso de patentes europeas, se pondrá una P si es el número de publicación o una S si es el número de solicitud.

OEPM, Paseo de la Castellana, 75, 28071 Madrid   NIF: Q2820005C

TBP

DESCRIPCIÓN

MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD

CAMPO DE LA INVENCIÓN

La presente invención pertenece al campo de los métodos y sistemas de seguridad. Más concretamente, la invención pertenece al campo de los métodos y sistemas de detección de proximidad para evitar colisiones entre personas u objetos.

ANTECEDENTES DE LA INVENCIÓN

Evitar la colisión entre un objeto en movimiento y cualquier elemento (incluyendo personas o animales) que se puedan cruzar en su camino es una necesidad que actualmente se aborda de distintas formas.

En el caso de vehículos terrestres, que normalmente circulan a una velocidad relativamente elevada, estos vehículos a menudo incorporan sistemas embarcados de prevención de colisiones basados en sistemas GPS y sensores, tales como radares o cámaras, que permiten la localización tanto del propio vehículo como la de potenciales obstáculos. Por ejemplo, la patente US 8924139 B2 describe uno de estos sistemas de prevención de colisiones.

En otros tipos de objetos en movimiento no es viable o, debido por ejemplo a la relativamente baja velocidad que pueden alcanzar estos objetos, no merece la pena, incorporar sistemas de detección como el descrito en US 8924139 B2. Tal es el caso, por ejemplo, de pequeños vehículos para el transporte de personas o cosas en áreas concretas, como almacenes, aeropuertos, estaciones de tren, rutas turísticas, etc.; o de equipos de transporte de cargas, como manipuladores, elementos mecatrónicos, grúas, normalmente remotamente operados o programados para seguir una trayectoria, como puede ser el caso de robots manipuladores; o de otros vehículos dedicados a sectores específicos como la agricultura (tractores, etc.) o la construcción (excavadoras, etc.). La solicitud de patente US 2006/0287829 A1 describe un método para detectar colisiones potenciales entre un vehículo de manejo de carga y objetos que se encuentren a su alrededor, basado en disponer en el vehículo un sistema de sensores para detectar objetos en su entorno y un sistema de alerta para detectar señales de alarma si se detecta un objeto a cierta distancia. Sin embargo, este método requiere la instalación de sensores de distancia (ultrasonidos, radar, laser, y/o detectores electromagnéticos) en la

1

periferia del vehículo a vigilar, lo cual implica seleccionar a priori los puntos de riesgo del vehículo (por ejemplo, avión), cuando en realidad los puntos de riesgo dependen del tipo de situación que puede acontecer.

A su vez, la solicitud de patente US 2015/0254985 A1 describe un sistema para evitar colisiones en minas subterráneas. El sistema está basado en un sistema de visión térmica para la captura de imágenes en tiempo real para la identificación y posición de objetivos, de los que se evalúa también su dirección y velocidad. Se incluye también un software para evitar colisiones, configurado para detectar la proximidad de objetivos, determinar su amenaza y evitar la colisión basándose en la determinación de la amenaza. Sin embargo, este sistema necesita generar una infraestructura basada en dispositivos fuera del elemento móvil. Además, el sistema de visión térmica requiere de entrenamiento para la identificación de objetos predefinidos en base a la huella térmica. También se conocen métodos y dispositivos para proporcionar información a su portador sobre la presencia de objetos cercanos y la distancia a la que se encuentran. Por ejemplo, la solicitud de patente US 2011/0025492 A1 describe un aparato que informa a su portador sobre la presencia de un objeto próximo mediante vibraciones táctiles tras la detección del objeto próximo mediante un detector de proximidad.

DESCRIPCIÓN DE LA INVENCIÓN

La presente invención proporciona un método, sistema y programa informático de detección de proximidad que resuelve los inconvenientes de propuestas anteriores. El método, sistema y programa informático se basan en monitorizar o supervisar un volumen alrededor de un dispositivo bajo control, mediante el cálculo de una distancia entre el dispositivo bajo control y un obstáculo. Preferentemente se calcula la distancia mínima entre ambos elementos. Se puede determinar así el riesgo de colisión entre el dispositivo bajo control y un obstáculo situado dentro de ese volumen monitorizado.

El método de detección de proximidad se implemente mediante un módulo de software (programa informático) que se puede configurar e integrar en cualquier aplicación en la que exista un interés en controlar un volumen alrededor de un dispositivo bajo control, ya sea un dispositivo móvil o un dispositivo estático.

En el contexto de la presente divulgación, se entiende por “dispositivo bajo control” cualquier dispositivo o equipo, ya sea estático o en movimiento, incluyendo dispositivos que pueden moverse tanto por sí mismos como transportados en otro equipo o empujados o impulsados por medios externos al dispositivo. La invención tiene especial aplicación en relación con dispositivos en movimiento o con capacidad para moverse, ya sea por sí mismos o transportados o empujados por otros dispositivos o equipos. El dispositivo (o un equipo en el que el dispositivo está situado o en el que el dispositivo es transportado) puede estar conducido por un operador que lo maneja localmente, u operado remotamente, o se puede mover de forma autónoma, por ejemplo pero de forma no limitativa, porque esté programado para seguir una trayectoria o para buscar de forma autónoma una trayectoria hasta un determinado objetivo. El operador es una persona que controla el movimiento del dispositivo (o del equipo que lo transporta, si éste fuera el caso), ya sea controlando el dispositivo (o equipo que lo transporta) como tal o controlando medios externos que impulsan o mueven el dispositivo (o equipo que lo transporta). Entre los equipos que pueden transportar un dispositivo bajo control se incluyen, de forma no limitativa, equipos que pueden transportar una carga, ya sean local o remotamente operados, o programados para seguir o buscar una trayectoria, tales como un manipulador, un elemento mecatrónico, una grúa, un robot, etc.; vehículos para el transporte de personas o cosas, por ejemplo en áreas concretas, como almacenes, aeropuertos, estaciones de tren, rutas turísticas, etc.; o vehículos dedicados a sectores específicos como la agricultura (tractores, etc.) o la construcción (excavadoras, etc.). Estos equipos pueden ser en sí mismos el dispositivo bajo control. En el caso de que el método de la invención se utilice para monitorizar el riesgo de colisión entre una persona o animal y un potencial obstáculo, el término “dispositivo bajo control” engloba también a la persona o animal cuyo riesgo de colisión se esté monitorizando (por ejemplo, en el caso de personas invidentes, niños de corta edad o de personas con un trastorno que les impida evaluar riesgos de colisión).

En el contexto de la presente divulgación, se entiende por “obstáculo” cualquier objeto, persona o animal que se encuentre dentro de un volumen que rodea al dispositivo bajo control. Es decir, por “obstáculo” se incluyen no solo aquellos objetos, personas o animales que, estando dentro de dicho volumen, se interpongan en la trayectoria del dispositivo bajo control, sino cualquier objeto, persona o animal que, por estar dentro de dicho volumen, se consideran amenazas potenciales de colisión. El obstáculo puede estar estático o en movimiento.

En el contexto de la presente divulgación, se entiende por “sensor 3D” cualquier sensor capaz de ofrecer información tridimensional del entorno que se puede representar en forma de nube de puntos. Los sensores 3D se basan en distintas tecnologías, tales como visión, láser, radar, entre otras. La nube de puntos representa la información obtenida por

3

el sensor en volumen, asociando a cada punto X,Y,Z un valor, que permite representar los obstáculos detectados por el sensor. A modo de ejemplo, sin carácter limitativo, se puede utilizar el Velodyne's PUCK™ (VLP-16) LiDAR, que ofrece información de distancia a los objetos en tiempo real con un alcance de 100 m, 360º de radio de campo de visión en horizontal, y +-15º en vertical.

El sensor 3D (en adelante, el sensor) puede situarse o bien en el entorno o alrededores del dispositivo bajo control, o bien situado sobre o en el propio dispositivo bajo control. Cuando el sensor está situado en el propio dispositivo bajo control, si el dispositivo bajo control se está desplazando, el sensor se desplaza a la vez que lo hace el dispositivo bajo control, de forma que siempre se vigila un volumen previamente definido alrededor del dispositivo bajo control. Es decir, el volumen vigilado se desplaza con el dispositivo bajo control. Alternativamente, el sensor puede situarse fuera del dispositivo bajo control. En este caso, el volumen a vigilar puede asociarse al sensor, en cuyo caso se vigila un volumen fijo, o el volumen vigilado se puede asociar y desplazar con el dispositivo bajo control.

El método y programa informático computan el riesgo de colisión que existe entre el dispositivo bajo control y un obstáculo, a partir de varias fuentes de información. Estas fuentes de información son, al menos: la geometría del dispositivo bajo control en un instante determinado A(t); y la geometría del o de los obstáculos detectados por el sensor dentro de un volumen en un instante determinado B(t). Nótese que tanto la geometría del dispositivo bajo control como la del o los obstáculos detectados dentro del volumen monitorizado varían o pueden variar con el tiempo, ya que tanto el dispositivo bajo control como el o los obstáculos pueden variar su posición con el tiempo (es decir, pueden estar en movimiento). La geometría del dispositivo bajo control puede ser fija o variable; si, por ejemplo, se trata de un manipulador robótico de 7 grados de libertad, la geometría cambia en función del estado de cada uno de los grados de libertad, cambiando la configuración geométrica al realizar un movimiento de un punto a otro. Lo mismo ocurre, por ejemplo, en el caso de un vehículo que lleva un manipulador que a su vez transporta una carga: en este caso, el conjunto formado por el manipulador y la carga puede considerarse el "dispositivo bajo control" (el conjunto de manipulador y carga determinan la configuración geométrica del dispositivo. O en otro ejemplo, una grúa que puede transportar cargas diferentes: la carga determina la configuración geométrica del dispositivo. Para evaluar el riesgo de colisión, el método, sistema y programa informático computan una distancia, preferentemente mínima, entre dichas geometrías A(t) y B(t).

4

La descripción de la geometría del dispositivo bajo control A(t) se define como un conjunto de formas primitivas, tales como, pero de forma no limitativa, cajas, cilindros, etc. Esta representación simplificada de la geometría del dispositivo bajo control reduce significativamente el tiempo de cálculo de la distancia mínima. Debido a esto, el método, sistema y programa informático son capaces de reaccionar a entornos altamente dinámicos.

Con respecto a la geometría del o de los obstáculos detectados por el sensor dentro de un volumen V B(t), la geometría dentro del volumen V se representa como un conjunto de vóxeles ocupados por los elementos (obstáculos) dentro de este volumen. Cuando un elemento se encuentra dentro del volumen y por tanto se considera un obstáculo, los vóxeles correspondientes al obstáculo están marcados como ocupados. Es decir, cada vóxel lleva asociado un bit que indica si el vóxel está marcado o no marcado. Como un experto sabe, un vóxel (del inglés *volumetric pixel*) es la unidad cúbica que compone un objeto tridimensional. El vóxel constituye la unidad mínima procesable de una matriz tridimensional, siendo por tanto el equivalente del píxel en un objeto 2D.

En un primer aspecto de la invención, se proporciona un método de detección de proximidad entre un dispositivo y un obstáculo, que comprende las siguientes etapas: establecer una descripción geométrica inicial de un dispositivo bajo control como un conjunto de formas primitivas; definir un volumen a monitorizar en torno a dicho dispositivo bajo control; obtener un conjunto de puntos en el espacio mediante un sensor 3D; realizar un primer filtrado de dicho conjunto de puntos en el espacio para eliminar de dicho conjunto de puntos los puntos que quedan fuera de dicho volumen a monitorizar, obteniéndose un subconjunto de puntos; monitorizar la geometría del dispositivo bajo control, actualizando la descripción geométrica del dispositivo bajo control representada como un conjunto de formas primitivas; realizar un segundo filtrado de dicho subconjunto de puntos para eliminar los puntos que quedan dentro de dicha descripción geométrica actualizada del dispositivo bajo control, obteniéndose un conjunto de puntos en el espacio que representan un obstáculo dentro de dicho volumen a monitorizar; calcular una distancia entre dicha descripción geométrica actualizada del dispositivo bajo control y dicho conjunto de puntos en el espacio que representan dicho obstáculo.

En realizaciones de la invención, el sensor 3D está situado en dicho dispositivo bajo control o cercano al mismo.

En realizaciones de la invención, el sensor 3D se desplaza a la vez que se desplaza el dispositivo bajo control o a la vez que se desplaza un equipo que porta o transporta a

5

dicho dispositivo bajo control.

En realizaciones de la invención, el volumen a monitorizar se desplaza a la vez que se desplaza dicho dispositivo bajo control o a la vez que se desplaza un equipo que porta o transporta a dicho dispositivo bajo control.

En realizaciones de la invención, el primer filtrado de dicho conjunto de puntos en el espacio obtenidos mediante un sensor 3D se eliminan también los puntos que representan ruido.

En realizaciones de la invención, la distancia calculada entre dicha descripción geométrica actualizada del dispositivo bajo control y dicho conjunto de puntos en el espacio que representan dicho obstáculo es una distancia mínima.

En realizaciones de la invención, en dicha etapa de monitorización se obtiene una descripción geométrica simplificada del dispositivo bajo control. En algunas de estas realizaciones, para el cálculo de dicha distancia se utiliza la descripción geométrica simplificada del dispositivo bajo control.

En realizaciones de la invención, se realiza además una etapa de, a partir de dicha distancia, establecer un riesgo de colisión entre dicho dispositivo bajo control y dicho obstáculo, o seleccionar un conjunto de puntos de riesgo de colisión situados en dicho dispositivo bajo control y/o en dicho obstáculo.

En realizaciones de la invención, las etapas anteriores se repiten con una determinada frecuencia, actualizándose con dicha frecuencia la descripción geométrica actualizada del dispositivo bajo control y el conjunto de puntos en el espacio que representan dicho obstáculo, recalculándose así dicha distancia.

En realizaciones de la invención, se realiza además la etapa de calibrar al inicio la posición relativa entre el sensor, el dispositivo bajo control y el volumen a monitorizar.

En un segundo aspecto de la invención, se proporciona un sistema de detección de proximidad entre un dispositivo y un obstáculo, que comprende: medios para establecer una descripción geométrica inicial de un dispositivo bajo control como un conjunto de formas primitivas; medios para definir un volumen a monitorizar en torno a dicho dispositivo bajo control; medios para obtener un conjunto de puntos en el espacio mediante un sensor 3D; medios para realizar un primer filtrado de dicho conjunto de puntos en el espacio para eliminar de dicho conjunto de puntos los puntos que quedan fuera de dicho volumen a monitorizar, obteniéndose un subconjunto de puntos; medios para monitorizar la geometría del dispositivo bajo control, actualizando la descripción

6

geométrica del dispositivo bajo control representada como un conjunto de formas primitivas; medios para realizar un segundo filtrado de dicho subconjunto de puntos para eliminar los puntos que quedan dentro de dicha descripción geométrica actualizada del dispositivo bajo control, obteniéndose un conjunto de puntos en el espacio que representan un obstáculo dentro de dicho volumen a monitorizar; medios para calcular una distancia entre dicha descripción geométrica actualizada del dispositivo bajo control y dicho conjunto de puntos en el espacio que representan dicho obstáculo.

En realizaciones de la invención, el sistema comprende además un módulo de configuración configurado para proporcionar una interfaz en la que establecer dicho volumen a monitorizar y dicha descripción geométrica inicial del dispositivo bajo control.

En realizaciones de la invención, el sistema comprende además medios de almacenamiento de memoria.

En realizaciones de la invención, el sistema comprende además una interfaz de control configurada para visualizar los posibles puntos de colisión entre el dispositivo bajo control y el obstáculo.

En un tercer aspecto de la invención, se proporciona un programa informático que comprende instrucciones de código de programa de ordenador para realizar el método anteriormente descrito.

Como puede observarse, el método, sistema y programa informático proporcionan ventajas apreciables con respecto a métodos conocidos. Por ejemplo, a diferencia de US 2006/0287829 A1, cuyo método requiere seleccionar a priori los puntos de riesgo del vehículo a monitorizar, en la presente invención basta con un sensor 3D que cubre una zona a vigilar alrededor del dispositivo bajo control, sin necesidad de desplegar una instalación de sensores. Con respecto a US 2015/0254985 A1, que necesita generar una infraestructura basada en dispositivos fuera del elemento móvil y requiere de entrenamiento para la identificación de objetos predefinidos en base a una huella térmica, el presente método y sistema establecen un riesgo de colisión analizando y comparando nubes de puntos procesadas, sin necesidad de generar infraestructura en base a dispositivos fuera del elemento móvil.

Ventajas y características adicionales de la invención serán evidentes a partir de la descripción en detalle que sigue y se señalarán en particular en las reivindicaciones adjuntas.

BREVE DESCRIPCIÓN DE LAS FIGURAS

Para complementar la descripción y con objeto de ayudar a una mejor comprensión de las características de la invención, de acuerdo con un ejemplo de realización práctica de la misma, se acompaña como parte integrante de la descripción, un juego de figuras en el que con carácter ilustrativo y no limitativo, se ha representado lo siguiente:

Las figuras 1A y 1B ilustran esquemáticamente dos posibles escenarios de aplicación del método y programa informático de la invención.

La figura 2 ilustra un diagrama de bloques del método de la invención.

La figura 3 ilustra un diagrama de bloques del método de la invención que incluye un módulo de configuración.

La figura 4 ilustra un ejemplo de aplicación en el que el método, sistema y programa informático de la invención se ha utilizado para vigilar una zona alrededor de un brazo robótico colgado de una grúa. Una interfaz de control va mostrando en pantalla los potenciales puntos de colisión en un determinado instante.

DESCRIPCIÓN DE UN MODO DE REALIZACIÓN DE LA INVENCIÓN

La figura 1A ilustra un esquema de un posible escenario de aplicación del método, sistema y programa informático de detección de proximidad de acuerdo con una posible realización de la invención. En ella se muestra de forma esquemática un dispositivo bajo control 12. El dispositivo bajo control 12 está colgado o situado en un equipo 11, concretamente se ha representado una guía 11 de la que puede colgar el dispositivo 12, aumentando su radio de acción, ya que el dispositivo 12 se mueve (o potencialmente puede moverse), cambiando su configuración, y además todo él puede trasladarse en el eje o guía. Nótese que éste es un caso de uso típico (pero no limitativo) en la industria de automoción para el pintado o soldadura de piezas. Se muestra también un sensor 3D 13, en este caso situado en el propio dispositivo bajo control 12. Alternativamente, el sensor puede situarse en el entorno o alrededores del dispositivo bajo control, por ejemplo acoplado a un equipo que porta o transporta al dispositivo bajo control, de forma que el sensor se desplaza a la vez que se desplaza el dispositivo bajo control. El sensor 3D queda fuera del alcance de la presente invención. Se muestra también un volumen V1 bajo monitorización. El volumen V1 representa el espacio dentro del cual se desea monitorizar la presencia de obstáculos que puedan suponer un riesgo de colisión con el dispositivo bajo control 12. En esta realización, el volumen V1 se ha definido como un

8

ortoedro, pero la forma tridimensional del volumen V1 puede ser otra (por ejemplo, pero de forma no limitativa, un cilindro, una esfera, un prisma, etc.). Los potenciales obstáculos en diferentes instantes de tiempo dentro del volumen V1 se detectar mediante el sensor 13. El volumen vigilado V1 se desplaza con el dispositivo bajo control 12. Para llevar a cabo el cálculo de distancia (y riesgo de colisiones asociado) por el método y programa informático, que se explica más adelante, el dispositivo bajo control 12 se representa en dicho programa y método mediante su geometría en un instante determinado A(t). Esta geometría se modela o representa en dicho programa informático como un conjunto de formas primitivas, por ejemplo como un conjunto de prismas.

En el esquema de la figura 1A, se ha representado un obstáculo 14 como una persona que se aproxima al volumen V1. La geometría B(t) del obstáculo 14 detectada por el sensor se representa o modela por el método y programa informático, como un conjunto de vóxeles ocupados por el obstáculo 14 dentro del volumen V1 bajo inspección. Es decir, cuando el obstáculo 14 se encuentra dentro del volumen V1, los vóxeles correspondientes al obstáculo 14 están marcados como ocupados. Cuando el obstáculo 14, de forma general en movimiento, está parcialmente dentro del volumen V1 y parcialmente fuera del mismo, como es el caso del ejemplo mostrado en la figura 1A, en cada instante de tiempo solo los vóxeles de dentro del volumen V1 están marcados como ocupados (de hecho, el método y programa excluye los vóxeles fuera del volumen V1 aunque pertenezcan a un mismo obstáculo, como se explica en relación con el diagrama de la figura 2). Por eso la geometría B(t) varía con el tiempo, a medida que varía la percepción que el sensor tiene del obstáculo según el sensor se desplaza (y opcionalmente el obstáculo también se desplaza, según sea el caso), es decir, a medida que varían los datos relativos a un obstáculo dentro del volumen V1 que acompaña siempre al dispositivo bajo control.

La figura 1A muestra también una flecha que representa una distancia D1 que a su vez indica un riesgo de colisión R1 entre el dispositivo bajo control 12 y el obstáculo 14. Es decir, en el método y programa informático, el riesgo de colisión R1 se calcula como una distancia D1 entre la geometría A(t) del dispositivo bajo control 12 en un instante determinado y la geometría B(t) del obstáculo 14 detectada por el sensor 13 dentro del volumen seleccionado V1 en un instante determinado. Como puede observarse, tanto el dispositivo bajo control 12 como el obstáculo 14 son potencialmente móviles, es decir, no están necesariamente estáticos.

La figura 1B ilustra un esquema de otro posible escenario de aplicación del método, sistema y programa informático de detección de proximidad de acuerdo con otra realización de la invención. En ella se muestra un dispositivo bajo control 22. En este caso, el dispositivo bajo control 22 está siendo transportado por un carrito o pequeño vehículo de transporte 21, concretamente el dispositivo bajo control 22 está colocado en un mástil de dicho vehículo 21. En la parte inferior del mástil se encuentra un sensor 3D 23, que queda fuera del alcance de la invención. El sensor 23 podría situarse en otra parte del vehículo 21 o del propio dispositivo bajo control 22. Se muestra también un volumen V2 bajo monitorización, que representa el espacio dentro del cual se desea monitorizar la presencia de obstáculos que puedan suponer un riesgo de colisión con el dispositivo bajo control 22. En este caso, tanto el volumen V2 como el dispositivo bajo control 22 se desplazan cuando el vehículo 21 se desplaza. En esta realización, el volumen V2 se ha definido como un cilindro, pero la forma tridimensional del volumen V2 puede ser otra (por ejemplo, pero de forma no limitativa, un ortoedro, una esfera, un prisma, etc.). Los potenciales obstáculos en diferentes instantes de tiempo se detectar mediante el sensor 23. Como en el caso descrito en la figura 1A, para el cálculo de colisiones llevado a cabo por el método y programa informático, que se explica más adelante, el dispositivo 22 se representa mediante su geometría en un instante determinado A(t). Esta geometría se representa como un conjunto de formas primitivas (no ilustradas en las figuras 1A o 1B).

En el esquema de la figura 1B, se ha representado un obstáculo 24 como una columna. Por tanto, se ha ilustrado en este caso un obstáculo estático. Nótese que a pesar de que en este caso el obstáculo es estático, la geometría del obstáculo B(t) detectada por el sensor 23 varía en el tiempo debido a que el sensor 23 se desplaza junto con el vehículo 21 que porta al dispositivo bajo control 22. Como en el caso anterior, B(t) representa un conjunto de vóxeles ocupados por el obstáculo dentro de un volumen V2 y detectados por el sensor 23 dentro del volumen V2 que acompaña siempre al dispositivo bajo control 22. Es decir, a medida que el dispositivo bajo control 22 se desplaza, lo que detecta el sensor 23 que se desplaza con el dispositivo bajo control 22 va cambiando. Por ejemplo, si el dispositivo bajo control 22, al desplazarse sobre el vehículo 21, se aproxima al obstáculo 24, en cada instante de tiempo B(t) tendrá más vóxeles ocupados (marcados como ocupados). La geometría B(t) del obstáculo 24 captada por el sensor 23 se representa como un conjunto de vóxeles ocupados por el obstáculo 24 dentro del volumen V2 bajo inspección. Es decir, cuando al aproximarse el dispositivo bajo control 22 al obstáculo 24, y por tanto, al desplazarse el volumen V2 junto con el sensor 23, el obstáculo 24 puede pasar a estar, total o parcialmente, dentro del volumen V2. Por el contrario, si el movimiento es diferente, por ejemplo alejándose del obstáculo 24, éste puede dejar de estar dentro del volumen V2. Los vóxeles correspondientes al obstáculo 24 que quedan dentro del volumen V2 en cada instante se marcan como ocupados y representan la geometría de lo que detecta el sensor 3D 23.

La figura 1B muestra también una flecha que representa una distancia D2 que a su vez indica el riesgo de colisión R2 entre el dispositivo bajo control 22 y el obstáculo 24. El riesgo de colisión R2 se calcula como una distancia D2 entre la geometría A(t) del dispositivo bajo control 22 en un instante determinado y la geometría B(t) del obstáculo 24 en un instante determinado (vista o captada por el sensor 23).

Tanto en la realización de la figura 1A como en la de la figura 1B, se utiliza preferentemente la distancia mínima entre ambas geometrías para calcular el riesgo de colisión R1, R2 en cada instante de tiempo. Gracias a que la descripción de la geometría del dispositivo bajo control 21, 22 se ha definido como un conjunto de formas primitivas (es decir, de forma simplificada), se reduce significativamente el tiempo de cálculo de la distancia, preferentemente mínima. Se consigue así reaccionar a entornos altamente dinámicos.

Para ejecutar el método, sistema y programa informático, se calibra previamente la posición relativa entre el sensor 13 (o 23), el dispositivo bajo control 12 (o 22) y el volumen V1 (o V2). La calibración se basa en establecer un marco de referencia común para el dispositivo 12, 22 y el sensor 13, 23, de forma que se conozca la relación geométrica entre la información proveniente del sensor 13, 23 respecto a un obstáculo 14, 24 y su posición en el espacio en el volumen vigilado V1, V2. Es decir, durante la calibración se establecen los parámetros de transformación necesarios para establecer cuál es la posición respecto al dispositivo bajo control 12, 22 de un vóxel indicado por el sensor 13, 23 (que puede estar en el dispositivo 12, 22 o fuera del mismo). De esta manera, se pueden calcular la distancias entre A(t) y B(t) con respecto a un mismo origen de referencia. El volumen V1, V2 es fijo y se configura al inicio. Es decir, se vigila un volumen fijo V1, V2 alrededor del dispositivo 12, 22. El volumen V1, V2 puede cambiarse para diferentes sesiones o aplicaciones, pero de cambiarse, debe hacerse al inicio, por ejemplo durante una etapa de configuración, y se queda fijado para toda la sesión o, por ejemplo, hasta una nueva configuración. Es decir, el volumen definido no cambia dinámicamente. La frecuencia de muestreo viene preferentemente dada por el sensor 13. 23. Es decir, A(t) y B(t) se actualizan en el tiempo cada 1/f segundos, siendo f la frecuencia de muestreo del sensor 13, 23. En ejemplos no limitativos, la frecuencia de

11

muestreo puede ser de 5 Hz (Hercios), 10 Hz o 20 Hz.

La figura 2 ilustra un diagrama de bloques del método, sistema y programa informático de cálculo de la distancia (y riesgo de colisión derivado de la distancia), de acuerdo con una posible realización de la invención. El diagrama de bloques parte, como datos de entrada, de:

-un conjunto de puntos o nube de puntos 31 obtenida por el sensor 3D (13, 23 en las figuras 1A y 1B, respectivamente) en cada instante de tiempo, determinado preferentemente por la frecuencia de muestreo del sensor; este conjunto de puntos representa el o los elementos captados por el sensor. Es decir, representa lo que el sensor captura dentro de su campo de visión y rango. Pueden ser obstáculos o no. El sensor capta datos, que forman una nube de puntos. Los datos que da el sensor se representan en forma de vóxel. El sensor no sabe si los datos captados pertenecen a un obstáculo, o al dispositivo bajo control, o a un vehículo que porta al dispositivo bajo control, o a cualquier otra cosa. El sensor 3D capta una nube de puntos periódicamente El periodo viene dado por la frecuencia de muestreo del sensor. En general, la nube de puntos en cada instante de tiempo es diferente de la nube en el instante de tiempo anterior. El sensor puede desplazarse a la vez que lo hace el dispositivo bajo control, o no hacerlo.

-un volumen monitorizado 32 (V1, V2 en las figuras 1A y 1B, respectivamente); este volumen representa el espacio tridimensional dentro del cual se ha decidido buscar posibles obstáculos. Este volumen 32 se parametriza inicialmente por un experto (por ejemplo, en una aplicación informática para configuración). El volumen monitorizado se calibra al inicio, referenciándose con respecto a la geometría inicial del dispositivo bajo control A(t=0) que se explica a continuación.

-una descripción geométrica tridimensional inicial del dispositivo bajo control A(t=0) (12, 22 en las figuras 1A y 1B, respectivamente); Esta descripción geométrica es una representación geométrica en base a primitivas (por ejemplo, pero de forma no limitativa, prismas, cilindros, esferas, etc. o combinaciones de las mismas) del dispositivo bajo control. Esta representación la realiza un experto, por ejemplo mediante una aplicación de configuración desarrollada a tal efecto. Alternativamente, puede realizarse mediante un programa de diseño, como por ejemplo CAD. Es decir, en la descripción geométrica tridimensional inicial se define el dispositivo bajo control como un conjunto de volúmenes simplificados 33. Nótese que esta representación es la inicial porque estos volúmenes pueden variar en posición a medida que el dispositivo se mueve (cambia la geometría

12

A(t)).

A partir de la nube de puntos 31 obtenida por el sensor se realizan distintas operaciones o etapas:

(1) Filtrado (bloque 35) de la nube de puntos 31 obtenida por el sensor 3D, a partir del volumen a monitorizar (V1, V2) establecido y representado en el bloque 32. El filtrado 35 se realiza considerando aquellos datos que caen dentro del volumen a vigilar y que es de interés. Es decir, en esta etapa se excluyen los puntos de la nube de puntos N1 que caigan fuera del volumen seleccionado 32, obteniéndose una nube de puntos N2. Este filtrado 35 para obtener la nube de puntos N2 se actualiza a medida que el sensor va tomando nuevos datos.

(2) Preferentemente, un filtrado de ruido (no ilustrado específicamente en la figura 2, aunque puede entenderse comprendido en el bloque 35): El sensor tiene un cierto ruido de base que genera falsas detecciones. Para minimizar el efecto de ruido asociado al sensor, preferentemente se establece un umbral de número mínimo vóxeles consecutivos ocupados para considerarlo como algo relevante y ser considerado como obstáculo.

(3) Auto-filtrado (bloque 36) del dispositivo bajo control en la nube de puntos N2, para no ser considerado como un obstáculo dentro del volumen vigilado por el sensor (ya que en algunos casos el sensor puede estar “viendo” elementos del propio dispositivo bajo control). Durante la ejecución del programa informático y método, la descripción geométrica tridimensional del dispositivo bajo control (que inicialmente se ha considerado como A(t=0)), se va actualizando A(t) (bloque 34) durante la ejecución de movimientos. Además, se obtiene también una descripción geométrica tridimensional simplificada As(t) (bloque 33). Tanto el dato actualizado de la geometría del equipo monitorizada en el tiempo A(t) (bloque 34) como el dato actualizado de la geometría simplificada del equipo monitorizada en el tiempo As(t) (bloque 33) se obtiene de un controlador del dispositivo (dicho controlador queda fuera del alcance de la presente invención). El dato actualizado A(t) alimenta preferentemente el bloque de auto-filtrado del volumen (bloque 36), pues A(t) es una representación más fidedigna del dispositivo bajo control que As(t). Imagínese, por ejemplo, un brazo manipulador que se desplaza sobre un vehículo. En un instante inicial está en una posición fija, pero si va a agarrar algo, su geometría va a cambiar: de brazo recogido a brazo extendido. Esto se representa como A(t). Este auto-filtrado (bloque 36) tiene en cuenta la representación geométrica monitorizada (en el instante de tiempo t) A(t), que puede ir cambiando en cada instante, cada vez que se mueve el dispositivo bajo control. Preferentemente, este auto-filtrado 36 tiene también en

13

cuenta, por seguridad, un conjunto de parámetros de relleno (o padding) (como un engorde de la geometría) alrededor de los volúmenes geométricos que representan A(t). La geometría simplificada As(t) se utiliza preferentemente para el cálculo posterior de distancias, como se explica más adelante, pues simplifica el cálculo. Alternativamente, puede usarse la geometría no simplificado A(t) para el cálculo de las distancias. A modo de ejemplo, el controlador del dispositivo bajo control del que se obtiene el dato de la geometría actualizada (y también geometría actualizada simplificada) del dispositivo bajo control puede ser el controlador de un robot (si se está controlando un robot) o, si se está controlando una grúa, el armario de control de la misma, etc.

En una posible realización, se utilizan los tres tipos de operaciones (1), (2) y (3).

A partir de los citados datos de entrada (conjunto de puntos o nube de puntos N1 obtenida por el sensor 3D en el bloque 31, volumen monitorizado (bloque 32) y descripción geométrica A(t) y/o As(t) del dispositivo bajo control), el cálculo de la distancia que representa un riesgo de colisión se realiza como sigue:

Se realiza la etapa de filtrado (bloque 35) para, a partir del conjunto tridimensional de puntos de datos o nube de puntos N1 obtenidos mediante el sensor en el bloque 31, y del volumen monitorizado 32, el conjunto de puntos N1 se filtra (bloque 35) para eliminar: los puntos que queden fuera del volumen monitorizado 32, y opcionalmente, los puntos que se consideran ruido.

Tras este bloque de filtrado 35 que da como resultado un conjunto reducido de puntos o nube reducida de puntos N2, se realiza un nuevo filtrado 36 de los puntos que quedan dentro de la geometría del dispositivo bajo control, teniendo en cuenta la geometría monitorizada del dispositivo bajo control (preferentemente la geometría monitorizada no simplificada A(t) obtenida en el bloque 34). Se selecciona así solamente los puntos que representan uno o más obstáculos (bloque 37). Estos puntos representan la geometría del o de los obstáculos detectados por el sensor dentro de un volumen en un instante determinado B(t). Al igual que la geometría monitorizada del dispositivo bajo control A(t) obtenida en el bloque 34 (o, en su caso, As(t) obtenida en el bloque 33), la geometría de los obstáculos B(t) detectados por el sensor también se va actualizando a medida que se desplaza al menos el dispositivo bajo control (y opcionalmente el obstáculo, que también puede estar en movimiento). Es decir, se consigue un subconjunto de puntos en el espacio B(t) que representan un obstáculo 14, 24 situado dentro del volumen a monitorizar V1, V2, 32. Con otras palabras, se crea una representación 3D B(t) de los elementos (obstáculos) que hay dentro del volumen monitorizado 32, a partir de la nube

14

de puntos filtrados N2 (puntos capturados por el sensor que caen dentro del volumen seleccionado) y de la geometría tridimensional monitorizada del dispositivo bajo control (preferentemente geometría no simplificada A(t)) en cada instante de tiempo.

Por último, a partir de la representación 3D B(t) de los obstáculos que hay dentro del volumen monitorizado y de la representación geométrica del dispositivo bajo control monitorizada, se realiza (bloque 38) una consulta sobre la proximidad de ambas representaciones, es decir, se calcula una distancia entre la geometría del dispositivo bajo control y la geometría del obstáculo B(t). Para este cálculo, se utiliza preferentemente la geometría simplificada As(t), aunque alternativamente pude usarse la geometría A(t). Preferentemente se calcula la distancia mínima entre ambas geometrías. Dependiendo de la distancia obtenida, se puede establecer un riesgo de colisión basado en unos determinados umbrales de riesgo definidos previamente. Estos umbrales pueden variar en función de las aplicaciones o sectores industriales en que se use el presente método y programa informático.

Con otras palabras, todo el tiempo se manejan dos nubes de puntos: A(t) (y/o As(t)), que representa la geometría del dispositivo en cada instante de tiempo, y un conjunto de puntos (que varía en el tiempo) que representa lo que capta el sensor. De este conjunto de puntos se eliminan los puntos que caen fuera del volumen monitorizado y los puntos que corresponden al dispositivo bajo control (si el sensor está viendo parte del dispositivo), obteniéndose B(t). Luego se calculan, en cada instante de tiempo, distancias entre As(t) y B(t) (o, menos preferentemente, entre A(t) y B(t)), y en cada instante de tiempo la distancia más pequeña es la que determina el riesgo de colisión (estableciendo un umbral de riesgo). También así se puede obtener el punto del dispositivo bajo control que tiene más riesgo (porque conocemos la geometría), y se puede representar dicho punto de mayor riesgo, por ejemplo en una interfaz de control.

Este proceso se repite con la frecuencia que se estime oportuna en función de la aplicación, sector, velocidad de desplazamiento del dispositivo bajo control, etc. En realizaciones de la invención, este proceso se repite a intervalos menores de 5 minutos, por ejemplo pero no limitativamente, menores de 1 minuto, menores de 10 segundos, menores de 1 segundo o menores de 100 milisegundos.

El método y programa informático puede incluir también un bloque, módulo o sistema de configuración. La figura 3 muestra un esquema completo del método y programa informático, en la que al diagrama de bloques de la figura 2 se ha añadido el citado módulo de configuración 41. El módulo de configuración 41 proporciona una interfaz

15

donde establecer los parámetros asociados al volumen monitorizado 32 y la descripción geométrica A(t=0) del dispositivo bajo control usados como entradas por el método o programa informático de la figura 2 (o parte inferior de la figura 3). El módulo de configuración 41 permite definir dichos parámetros a través de una interfaz que luego utiliza el método y programa de ordenador.

El método y programa informático se implementan en un dispositivo que comprende medios de procesado (tales como un procesador o microprocesador) y medios de almacenamiento de memoria configurados para almacenar los datos que se van adquiriendo y calculando, tales como las distintas nubes de puntos y las distancias entre dispositivo bajo control y obstáculo. El método, sistema y programa informático pueden conectarse a una interfaz de control. Por ejmplo, puede instalarse un PC conectado con un cable (o de forma inalámbrica) con el armario de control del dispositivo bajo control; este PC también se conecta por cable (o de forma inalámbrica) al sensor. El PC puede estar solidario al dispositivo bajo control, es decir, si éste se desplaza en un carrito, el PC puede estar en el carrito.

Se ha realizado un experimento en el que el método, sistema y programa informático de la invención se ha utilizado para vigilar una zona alrededor de un brazo robótico colgado de una grúa. La figura 4 muestra la grúa y brazo robótico del experimento. El dispositivo que implementa el método y programa informático es un procesador PC conectado a un sensor 3D y al armario de control del brazo robótico bajo control. El sensor 3D es un detector de proximidad LIDAR (Velodyne) y se han colocado en el brazo robótico. Una interfaz de control, va mostrando los potenciales puntos de colisión en un determinado instante. La figura 4 muestra una representación de dos posibles puntos de colisión (los dos puntos negros de la imagen) en la interfaz de control en un instante determinado. Como puede observarse, se muestra un punto de riesgo de colisión detectado en el brazo robótico y otro punto de riesgo de colisión detectado en un obstáculo detectado (en la parte inferior izquierda de le figura). Las líneas punteadas formando un ortoedro representan el volumen vigilado.

En este texto, la palabra “comprende” y sus variantes (como “comprendiendo”, etc.) no deben interpretarse de forma excluyente, es decir, no excluyen la posibilidad de que lo descrito incluya otros elementos, pasos etc.

Por otra parte, la invención no está limitada a las realizaciones concretas que se han descrito sino abarca también, por ejemplo, las variantes que pueden ser realizadas por el experto medio en la materia (por ejemplo, en cuanto a la elección de materiales, dimensiones, componentes, configuración, etc.), dentro de lo que se desprende de las reivindicaciones.

REIVINDICACIONES

1.- Un método de detección de proximidad entre un dispositivo y un obstáculo, caracterizado por:

establecer una descripción geométrica inicial (A(t=0)) de un dispositivo bajo control (12, 22) como un conjunto de formas primitivas;

definir (32) un volumen a monitorizar (V1, V2) en torno a dicho dispositivo bajo control (12, 22);

obtener (31) un conjunto de puntos en el espacio (N1) mediante un sensor 3D (13, 23);

realizar un primer filtrado (35) de dicho conjunto de puntos en el espacio (N1) para eliminar de dicho conjunto de puntos (N1) los puntos que quedan fuera de dicho volumen a monitorizar (V1, V2), obteniéndose un subconjunto de puntos (N2);

monitorizar (33, 34) la geometría del dispositivo bajo control (12, 22), actualizando la descripción geométrica del dispositivo bajo control (As(t), A(t)) representada como un conjunto de formas primitivas;

realizar un segundo filtrado (36) de dicho subconjunto de puntos (N2) para eliminar los puntos que quedan dentro de dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (11, 22), obteniéndose un conjunto de puntos en el espacio B(t) que representan un obstáculo (14, 24) dentro de dicho volumen a monitorizar (V1, V2);

calcular una distancia (D1, D2) entre dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (12, 22) y dicho conjunto de puntos en el espacio B(t) que representan dicho obstáculo (14, 24).

2.- El método de la reivindicación 1, en el que dicho sensor 3D (13, 23) está situado en dicho dispositivo bajo control (12, 22) o cercano al mismo.

3.- El método de cualquiera de las reivindicaciones anteriores, en el que dicho sensor 3D (13, 23) se desplaza a la vez que se desplaza el dispositivo bajo control (12, 22) o a la vez que se desplaza un equipo (11, 21) que porta o transporta a dicho dispositivo bajo control (12, 22).

4.- El método de cualquiera de las reivindicaciones anteriores, en el que dicho volumen a monitorizar (V1, V2) se desplaza a la vez que se desplaza dicho dispositivo bajo control (12, 22) o a la vez que se desplaza un equipo (11, 21) que porta o transporta a dicho dispositivo bajo control (12, 22).

5.- El método de cualquiera de las reivindicaciones anteriores, en el que en dicho primer filtrado (35) de dicho conjunto de puntos en el espacio (N1) obtenidos mediante un sensor 3D (13, 23) se eliminan también los puntos que representan ruido.

6.- El método de cualquiera de las reivindicaciones anteriores, en el que dicha distancia (D1, D2) calculada entre dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (12, 22) y dicho conjunto de puntos en el espacio B(t) que representan dicho obstáculo (14, 24) es una distancia mínima.

7.- El método de cualquiera de las reivindicaciones anteriores, en el que en dicha etapa de monitorización (33) se obtiene una descripción geométrica simplificada (As(t)) del dispositivo bajo control (12, 22).

8.- El método de la reivindicación 7, en el que para el cálculo de dicha distancia (D1, D2) se utiliza dicha descripción geométrica simplificada (As(t)) del dispositivo bajo control (11, 12).

9.- El método de cualquiera de las reivindicaciones anteriores, que comprende además una etapa de, a partir de dicha distancia (D1, D2), establecer un riesgo de colisión (R1, R2) entre dicho dispositivo bajo control (12, 22) y dicho obstáculo (14, 24) o seleccionar un conjunto de puntos de riesgo de colisión situados en dicho dispositivo bajo control (12, 22) y/o en dicho obstáculo (14, 24).

10.- El método de cualquiera de las reivindicaciones anteriores, en el que las etapas anteriores se repiten con una determinada frecuencia, actualizándose con dicha frecuencia la descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (12, 22) y el conjunto de puntos en el espacio B(t) que representan dicho obstáculo (14, 24), recalculándose así dicha distancia (D1, D2).

11.- El método de cualquiera de las reivindicaciones anteriores, que comprende además la etapa de calibrar al inicio la posición relativa entre el sensor (13, 23), el dispositivo bajo control (12, 22) y el volumen a monitorizar (V1, V2).

12.- Un sistema de detección de proximidad entre un dispositivo y un obstáculo, caracterizado por:

19

medios para establecer una descripción geométrica inicial (A(t=0) de un dispositivo bajo control (12, 22) como un conjunto de formas primitivas;

medios para definir (32) un volumen a monitorizar (V1, V2) en torno a dicho dispositivo bajo control (12, 22);

medios para obtener (31) un conjunto de puntos en el espacio (N1) mediante un sensor 3D (13, 23);

medios para realizar un primer filtrado (35) de dicho conjunto de puntos en el espacio (N1) para eliminar de dicho conjunto de puntos (N1) los puntos que quedan fuera de dicho volumen a monitorizar (V1, V2), obteniéndose un subconjunto de puntos (N2);

medios para monitorizar (34) la geometría del dispositivo bajo control (12, 22), actualizando la descripción geométrica del dispositivo bajo control (A(t), As(t)) representada como un conjunto de formas primitivas;

medios para realizar un segundo filtrado (36) de dicho subconjunto de puntos (N2) para eliminar los puntos que quedan dentro de dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (11, 22), obteniéndose un conjunto de puntos en el espacio B(t) que representan un obstáculo (14, 24) dentro de dicho volumen a monitorizar (V1, V2);

medios para calcular una distancia (D1, D2) entre dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (12, 22) y dicho conjunto de puntos en el espacio B(t) que representan dicho obstáculo (14, 24).

13.- El sistema de la reivindicación 12, que comprende además un módulo de configuración (41) configurado para proporcionar una interfaz en la que establecer dicho volumen a monitorizar (V1, V2) y dicha descripción geométrica inicial (A(t=0)) del dispositivo bajo control (12, 22).

14.- El sistema de cualquiera de las reivindicaciones 12 o 13, que comprende además medios de almacenamiento de memoria.

15.- El sistema de cualquiera de las reivindicaciones 12 a 14, que comprende además una interfaz de control configurada para visualizar los posibles puntos de colisión entre el dispositivo bajo control (12, 22) y el obstáculo (14, 24).

16.- Un programa informático que comprende instrucciones de código de programa de ordenador para realizar el método de acuerdo con cualquiera de las reivindicaciones 1 a 11.

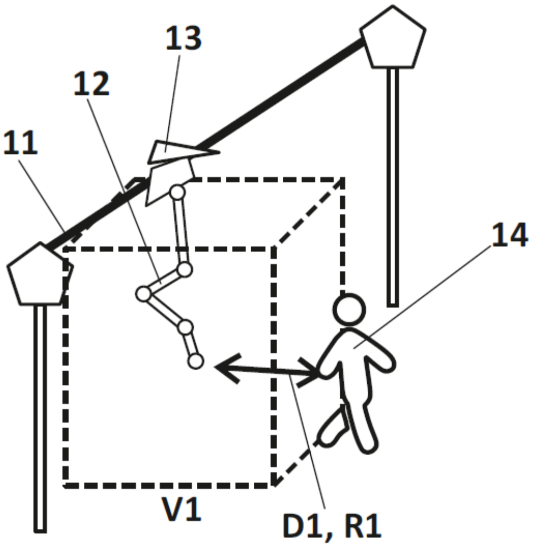


FIG. 1A

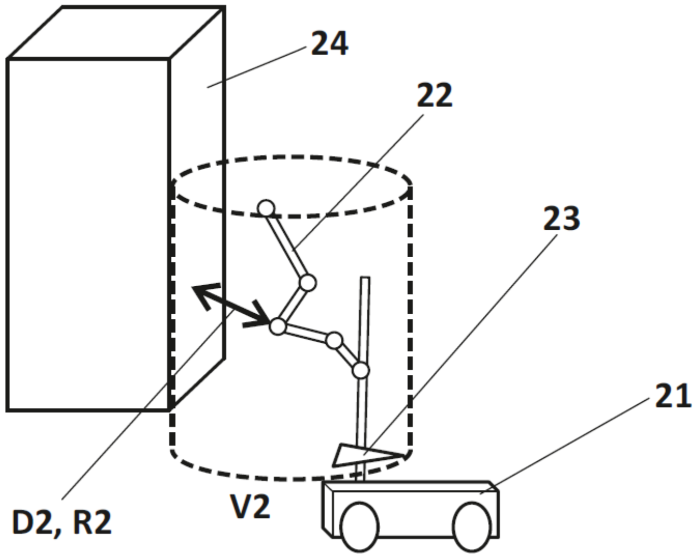


FIG. 1B

21

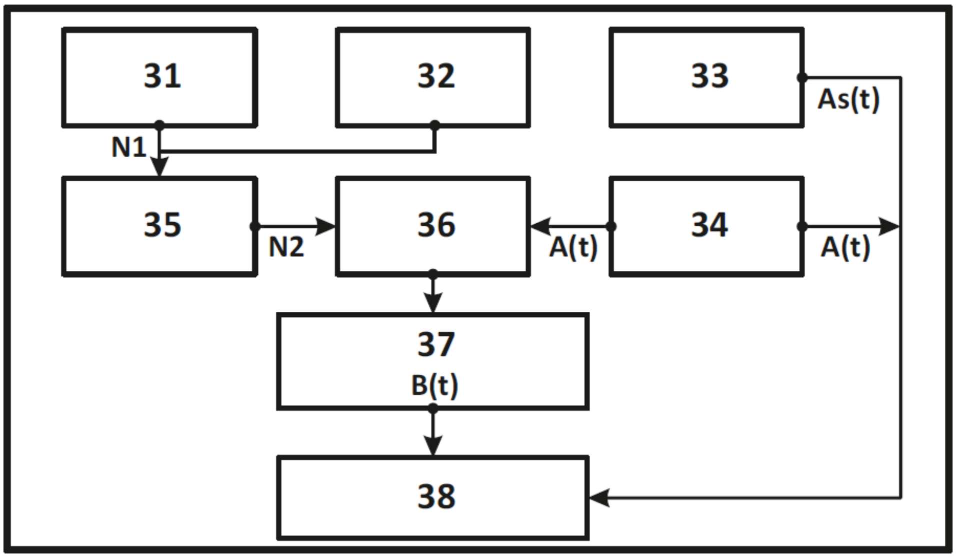

31
32
33
As(t)
N1
35
N2
36
A(t)
34
A(t)
37
B(t)
38

FIG. 2

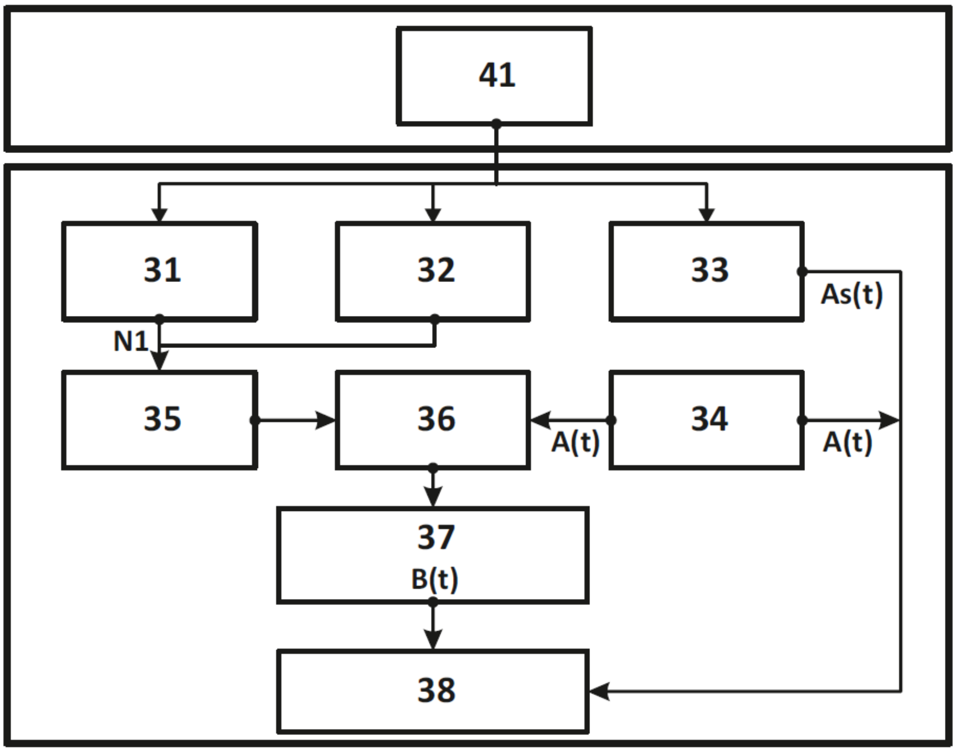

41
31
32
33
As(t)
N1
35
36
A(t)
34
A(t)
37
B(t)
38

FIG. 3

22

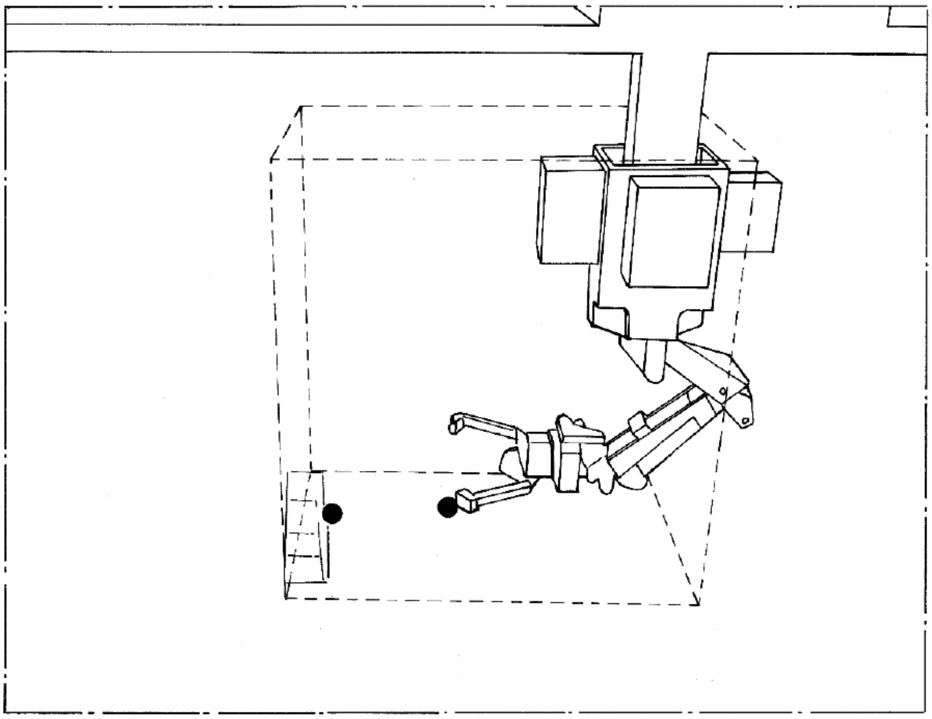


FIG. 4

23

RESUMEN

MÉTODO, SISTEMA Y PROGRAMA INFORMÁTICO DE DETECCIÓN DE PROXIMIDAD

Un método de detección de proximidad entre un dispositivo y un obstáculo, que comprende: establecer una descripción geométrica inicial (A(t=0) de un dispositivo bajo control (12, 22) como un conjunto de formas primitivas; definir (32) un volumen a monitorizar (V1, V2) en torno a dicho dispositivo bajo control (12, 22); obtener (31) un conjunto de puntos en el espacio (N1) mediante un sensor 3D (13, 23); realizar un primer filtrado (35) de dicho conjunto de puntos en el espacio (N1) para eliminar de dicho conjunto de puntos (N1) los puntos que quedan fuera de dicho volumen a monitorizar (V1, V2), obteniéndose un subconjunto de puntos (N2); monitorizar (33, 34) la geometría del dispositivo bajo control (12, 22), actualizando la descripción geométrica del dispositivo bajo control (A(t), As(t)) representada como un conjunto de formas primitivas; realizar un segundo filtrado (36) de dicho subconjunto de puntos (N2) para eliminar los puntos que quedan dentro de dicha descripción geométrica actualizada (A(t)) del dispositivo bajo control (11, 22), obteniéndose un conjunto de puntos en el espacio B(t) que representan un obstáculo (14, 24) dentro de dicho volumen a monitorizar (V1, V2); calcular una distancia (D1, D2) entre dicha descripción geométrica actualizada (A(t), As(t)) del dispositivo bajo control (12, 22) y dicho conjunto de puntos en el espacio B(t) que representan dicho obstáculo (14, 24). Sistema y programa informático.

[FIG. 1A]

Sección III Bibliografía

[1] ISTAG, «Scenarios for Ambient Intelligence 2010,» Sevilla, 2001.

[2] X-ACT Consortium, «WEB del proyecto,» [En línea]. Available: http://www.xact-project.eu/. [Último acceso: 23 Mayo 2017].

[3] ROBOPARTNER Consortium, «WEB del proyecto,» [En línea]. Available: http://www.robo-partner.eu/. [Último acceso: 22 05 2017].

[4] ECHORD, «WEB del experimento EASYPRO,» ECHORD, [En línea]. Available: http://www.echord.info/wikis/website/easypro.html. [Último acceso: 22 Mayo 2017].

[5] ECHORD++ and IK4-TEKNIKER, «WEB del experimento en el proyecto,» [En línea]. Available: http://echord.eu/debur/. [Último acceso: 22 Mayo 2017].

[6] AUTOWARE Consortium, «WEB del proyecto,» [En línea]. Available: http://www.autoware-eu.org/. [Último acceso: 22 Mayo 2017].

[7] MANUWORK Consortium, «WEB del proyecto,» [En línea]. Available: http://www.manuwork.eu/. [Último acceso: 22 Mayo 2017].

[8] EUROC Consortium, «WEB del proyecto,» [En línea]. Available: http://www.euroc-project.eu/. [Último acceso: 22 Mayo 2017].

[9] ROBOFOOT Consortium, «WEB del proyecto,» [En línea]. Available: http://www.robofoot.eu/. [Último acceso: 22 Mayo 2017].

[10] MAINBOT Consortium, «WEB del proyecto,» [En línea]. Available: http://www.mainbot.eu/. [Último acceso: 22 Mayo 2017].

[11] FOURBYTHREE, «WEB del proyecto,» [En línea]. Available: http://www.fourbythree.eu/. [Último acceso: 23 Mayo 2017].

[12] CRO-INSPECT Consortium, «WEB del proyecto,» [En línea]. Available: http://cro-inspect.eu/. [Último acceso: 22 Mayo 2017].

[13] International Organization for Standardization (ISO), «ISO 10218-1:2011, Robots and robotic devices -- Safety requirements for industrial robots -- Part 1: Robots,» International Organization for Standardization, Ginebra, 2011.

[14] International Organization for Standardization (ISO), «ISO-10218-2:2011, Robots and robotic devices -- Safety requirements for industrial robots -- Part 2: Robot systems and integration,» ISO, Ginebra, 2011.

[15] International Organization for Standardization (ISO), «ISO/TS 15066:2016, Robots and robotic devices -- Collaborative robots,» ISO, Ginebra, 2016.

[16] SMERobotics, «WEB del demostrador en el proyecto,» SMERobotics, [En línea]. Available:

http://www.smerobotics.org/demonstrations/d7.html. [Último acceso: 22 5 2017].

[17] EUROC-PIROS, «WEB del equipo PIROS,» EUROC, 2015. [En línea]. Available: http://www.euroc-project.eu/index.php?id=piros. [Último acceso: 21 5 2017].

[18] EUROC-RSAII, «WEB del equipo RSAII,» EUROC, 2015. [En línea]. Available: http://www.euroc-project.eu/index.php?id=rsaii. [Último acceso: 21 5 2017].

[19] A4BLUE, «WEB del proyeccto,» A4BLUE, 2017. [En línea]. Available: http://a4blue.eu/. [Último acceso: 21 5 2017].

[20] MANUWORK, «WEB del proyecto,» MANUWORK, 2017. [En línea]. Available: http://www.manuwork.eu/. [Último acceso: 21 5 2017].

[21] WEARIT@WORK, «WEB del proyecto,» WEARIT@WORK, [En línea]. Available: http://www.wearitatwork.com/. [Último acceso: 22 5 2017].

[22] D. Acemoglu y P. Restrepo, «Robots and Jobs: Evidence from US Labor Markets,» National Bureau of Economic Research (NBER), Cambridge, 2017.

[23] Word Economic Forum, «The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution,» 2016.

[24] A. Berg, E. F. Buffie y L.-F. Zanna, «Robots, crecimiento y desigualdad,» *Finanzas y Desarrollo,* vol. 53, nº 3, pp. 10-13, 2016.

[25] E. Mundo, «El Mundo online,» 3 Abril 2017. [En línea]. Available: http://www.elmundo.es/tecnologia/2017/04/03/58e202e5e5fdea79758b45cf.html. [Último acceso: 26 05 2017].

[26] Expansión, «www.expansion.com,» Unidad Editorial Información General, S.L.U, 18 2 2017. [En línea]. Available: http://www.expansion.com/economia-digital/protagonistas/2017/02/18/58a89ca3e5fdeafa0c8b4587.html. [Último acceso: 26 05 2017].

[27] A. van Dam, «Post-WIMP user interfaces,» *Communications of the ACM* , vol. 40, nº 2, p. 63–67, 1997.

[28] J. Hirschberg y C. Manning, «Advances in natural language processing,» *Science,* nº 349, p. 261–266, 2015.

[29] H. Ishii y B. Ullmer, «Tangible bits: towards seamless interfaces between people, bits and atoms,» de *CHI'97*, Los Angeles, 1997.

[30] C. Schwesig, I. Poupyrev y E. Mori, «Gummi: a bendable computer,» de *CHI '04*, Viena, 2004.

[31] J. Kildal, S. Paasovaara y V. Aaltonen, «Kinetic Device: Designing Interactions with a Deformable Mobile Interface,» de *CHI '12 CHI Conference on Human Factors in Computing Systems*, Austin, 2012.

[32] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian y K. Hornbæk, «Opportunities and Challenges for Data Physicalization,» de *CHI '15*, Seul, 2015.

[33] J. Fryman y B. Matthias, «Safety of Industrial Robots: From Conventional to Collaborative Applications,» de *7th German Conference on robotics,ROBOTIK 2012*, Munich, 2012.

[34] M. Endsley, «Toward a theory of situation awareness in dynamic systems,» *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, nº 1, pp. 32-64, 1995.

[35] M. Vasic y A. Billard, «Safety issues in human-robot interactions,» de *IEEE International Conference on Robotics and Automation* , Karlsruhe, 2013.

[36] T. Malm, J. Viitaniemi, J. Latokartano, S. Lind, O. Venho-Ahonen y J. Schabel, «Safety of Interactive Robotics—Learning from Accidents,» *International Journal of Social Robotics,* vol. 2, nº 3, p. 221–227, 2010.

[37] T. Carlson y Y. Demiris, «Using visual attention to evaluate collaborative control architectures for human robot interaction,» de *Proceedings of New Frontiers in Human-Robot Interaction: A symposium at the AISB 2009 Convention*, Edinburgh, 2009.

[38] M. MacMahon, B. Stankiewicz y B. Kuipers, «Walk the talk: Connecting language, knowledge, and action in route instructions,» de *AAAI'06 proceedings of the 21st national conference on Artificial intelligence*, Austin.

[39] P. Stone, M. Sridharan, D. Stronger, G. Kuhlmann, N. Kohl, P. Fidelman y N. Jong, «From pixels to multi-robot decision-making: A study in uncertainty,» de *Ninth International Conference onIntelligent Systems Design and Applications, ISDA '09*, Pisa, 2009.

[40] N. Juristo, A. Moreno y M.-I. Sanchez-Segura, «Guidelines for eliciting usability functionalities,» *IEEE Transactions on Software Engineering* , vol. 33, nº 11, pp. 744-758 , 2007.

[41] XSENS, «Página del producto XSENS MVN,» XSENS, [En línea]. Available: https://www.xsens.com/products/xsens-mvn/. [Último acceso: 27 5 2017].

[42] MyO, «WEB oficial de la empresa,» Thalmic Labs Inc, [En línea]. Available: https://www.myo.com/. [Último acceso: 21 5 2017].

[43] B. Fang, F. Sun, H. Liu y D. Guo, «A novel data glove using inertial and magnetic sensors for motion capture and robotic arm-hand teleoperation,» *Industrial Robot,* vol. 44, nº 2, p. 155–165, 2017.

[44] A. Ibarguren, I. Maurtua y B. Sierra, «Layered architecture for real time sign recognition: Hand gesture and movement,» *Engineering Applications of Artificial Intelligence,* vol. 23, nº 7, pp. 1216-1228, 2010.

[45] A. Ibarguren, I. Maurtua y B. Sierra, «Layered Architecture for Real-Time Sign Recognition,» *he Computer Journal,* vol. 53, nº 8, pp. 1169-1183, 2010.

[46] J. Suarez y R. Murphy, «Hand gesture recognition with depth images: A review,» de *Proceedings of RO-MAN, IEEE*, Paris, 2012.

[47] Y. Wang, C. Yang, X. Wu, S. Xu y H. Li, «Kinect Based Dynamic Hand Gesture Recognition Algorithm Research,» de *4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Nanchang, 2012.

[48] B. Burger, I. Ferrane y F. Lerasle, «Towards multimodal interface for interactive robots: challenges and robotic systems,» de *Robotics 2010 Current and Future Challenges*, Intech, 2010.

[49] S. Thrun y Otros, «Experiences with two deployed interactive tour-guide robots,» *Artificial Intelligence,* vol. 114, nº 1-2, pp. 3-55, 1999.

[50] A. Bannat, G. J., R. T, W. Rösel, G. Rigoll y F. Wallhof, «A multimodal human-robot-interaction scenario: Working together with an industrial robot,» de *HCI PartII: Novel Interaction Methods and Techniques*, 2009.

[51] S. Dobrisek, R. Gajsek, F. Mihelic, N. Pavesic y V. Struc, «Towards efficient multi-modal emotion recognition. International,» *International journal of Advanced Robotic Systems,* vol. 10, nº 1, 2013.

[52] Zheng, B. Chen, X. Wang, Y. Huang y Q. Wang, «On the design of a wearable multi-sensor system for recognizing motion modes and sit-to-stand transition,» *International Journal of Advanced Robotic Systems,* vol. 11, nº 30, 2014..

[53] S. Rossi, E. Leone, M. Fiore, A. Finzi y F. Cutugno, «An extensible architecture for robust multimodal human-robot communication,» de *International Conference on Intellegent Robots and Systems (IROS)*, Tokyo, 2013.

[54] International Organization for Standardization (ISO), «ISO 13855:2010: Safety of machinery -- Positioning of safeguards with respect to the approach speeds of parts of the human body,» ISO, Ginebra, 2010.

[55] K. Ikuta, H. Ishii y M. Nokata, «Safety Evaluation Method of Design and Control for Human-Care Robots,» *The International Journal of Robotics Research,* vol. 22, nº 5, p. 281–297, 2003.

[56] J. A. Marvel y R. Bostelman, «A Cross-domain Survey of Metrics for Modelling and Evaluating Collisions,» de *International Journal of Advanced Robotic Systems*, InTECH, 2014, p. 142.

[57] K. Ousama, «Real-Time Obstacle Avoidance for Manipulators and Mobile Robots,» *The International Journal of Robotics Research,* vol. 5, nº 1, pp. 90-98, 1986.

[58] S. Wen, W. Zheng., J. Zhu y otros, «Elman Fuzzy Adaptive Control for Obstacle Avoidance of Mobile Robots Using Hybrid Force/Position Incorporation,» *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 42, nº 4, pp. 603-608, 2012.

[59] D. Park, H. Hoffmann, P. Pastor y otros, «Movement Reproduction and Obstacle Avoidance with Dynamic Movement Primitives and Potential Fields,» *IEEE International Conference on Humanoid Robotics*, pp. 91-98, 2008.

[60] M. Fischer y D. Henrich, «3D Collision Detection for Industrial Robots and Unknown Obstacles using Multiple Depth Images,» de *erman Workshop on Robotics – GWR*, Braunschweig,, 2009.

[61] C. Walter, C. Vogel y E. Norbert, «Stationary Sensor-System supporting Manipulators at Safe Human-Robot interaction,» de *roceedings for the joint conference of ISR 2010 und ROBOTIK 2010*, Munich, 2010.

[62] D. Henrich y T. Gecks, «Multi-camera collision detection between known and unknwon objects,» de *2nd ACM/IEEE International Conference on Distributed Smart Cameras*, Stanford,, 2008.

[63] Pilz GmbH, «Homepage of Pilz GmbH and Co. KG,» Pilz GmbH, 2017. [En línea]. Available: https://www.pilz.com/en-GB/products-solutions/sensor-technology/safe-camera-systems. [Último acceso: 22 5 2017].

[64] C. Vogel, C. Walter y N. Elkmann, «A Projection-based Sensor System for Safe Physical Human-Robot Collaboration,» de *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, 2013.

[65] S. Phan, Z. F. Quek, P. Shah, D. Shin, O. K. Zubair Ahmed y M. Cutkosky, «Capacitive Skin Sensors for Robot Impact Monitoring,» de *EEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, 2011..

[66] Robert Bosch GmbH, «bosch-apas,» 2015. [En línea]. Available: http://bosch-apas.com/media/apas/bosch_apas/newsroom/downloads/Datenblatt_APASassistant_EN_web.pdf. [Último acceso: 20 5 2017].

[67] KUKA, «KUKA IIWA Robot,» KUKA Aktiengesellschaft , [En línea]. Available: https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa. [Último acceso: 23 5 2017].

[68] Universal Robot, «Universal Robots,» Universal Robots A/S, [En línea]. Available: https://www.universal-robots.com/. [Último acceso: 23 5 2017].

[69] Rethink Robotics, «Rethink robots,» Rethink Robotics, 2017. [En línea]. Available: http://www.rethinkrobotics.com. [Último acceso: 23 5 2017].

[70] S. Haddadin , A. Albu-Schäffer y G. Hirzinger, «Soft-Tissue Injury in Robotics,» de *IEEE International Conference on Robotics and Automation*, Anchorage, 2010.

[71] S. Haddadin, Towards Safe Robots: Approaching Asimov's 1st Law, Springer, 2014.

[72] S. Haddadin, A. Albu-Schäffer y G. Hirzinger, «Safety Evaluation of Physical Human-Robot Interaction via Crash-Testing,» *Proceedings of Robotics: Science and Systems,* vol. 3, pp. 217-224, 2007.

[73] S. Oberer y R. Schraft, «Robot-dummy crash tests for robot safety assessment,» de *IEEE International Conference on Robotics and Automation*, Rome, 2007.

[74] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser y R. Parasuraman, « Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction,» *he Journal of the Human Factors and Ergonomics Society,* vol. 53, nº 5, pp. 517-527 , 2011.

[75] W. A. Bainbridge, J. Hart, E. S. Kim y B. Scassellati, «The effect of presence on human-robot interaction,» de *Proceedings of the 17th IEEE Symposium on Robot and Human Interactive Community*, Munich, 2008.

[76] K. M. Tsui, M. Desai y H. A. Yanco, «Considering the bystander's perspective for indirect human-robot interaction,» de *Proceedings of the 5th ACM/IEEE International Conference on Human Robot Interaction*, New York, 2010.

[77] E. Park, Q. Jenkins y X. & Jiang, «Measuring trust of human operators in new generation rescue robots,» de *7th JFPS International Symposium on Fluid Power*, Toyama, 2008.

[78] B. Sadrfaridpour, J. Burke y Y. Wang, «Human and Robot Collaborative Assembly Manufacturing: Trust Dynamics and Control,» de *RSS 2014 conference*, Berkeley , 2014.

[79] B. ADRFARIDPOUR, H. SAEIDI, Y. WANG y J. BURKE, «Modeling and Control of Trust in Human and Robot Collaborative Manufacturing,» de *Robust Intelligence and Trust in Autonomous Systems*, Springer International Publishing AG, 2014, pp. 115-141.

[80] A. Andreopoulos y T. J K, «50 Years of object recognition: Directions forward,» *Computer Vision and Image Understanding,* vol. 117, nº 8, pp. 827-891, 2013.

[81] S. Ekvall, D. Kragic y F. Hoffmann, «Object recognition and pose estimation using color cooccurrence histograms and geometric modeling,» *mage and Vision Computing,* vol. 23, nº 11, pp. 943-955, 2005.

[82] R. Beserra, B. Marques, L. Karin de Medeiros, R. Vidal, L. C. Pacheco y L. M. Garcia, «Efficient 3D object recognition using foveated point clouds,» *Computers & Graphics,* vol. 37, nº 5, pp. 496-508, 2013.

[83] H. Sung-Hyun, H. Seo, K. Yoon y L. Man-Hyung, «Real-time control of an industrial robot using image-based visual servoing,» de *EEE/RSJ International Conference on Intelligent Robots and Systems, IROS '99*, 1999.

[84] H. Nomura y T. Naito, «Integrated visual servoing system to grasp industrial parts moving on conveyer by controlling 6DOF arm,» de *2000 IEEE International Conference on Systems, Man, and Cybernetics*, 2000.

[85] V. Lippiello, B. Siciliano y L. Villani, «Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration,» *IEEE Trans. Robot,* vol. 13, nº 21, p. 73–86, 2007.

[86] G. Klein y D. Murray, «Full-3D edge tracking with a particle filter,» de *British Machine Vision Conference (BMVC'06)*, 2016.

[87] C. Teuliere, E. Marchand y L. Eck, «Using multiple hypothesis in model-based tracking,» de *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[88] Google, «Google Cloud Speech Api,» [En línea]. Available: https://cloud.google.com/speech/. [Último acceso: 2017 5 22].

[89] U.-G. d. i. d. p. d. l. natural, «WEB oficial de Freeling,» [En línea]. Available: http://nlp.lsi.upc.edu/freeling/node/1. [Último acceso: 2017 5 22].

[90] S. Harris, A. Seaborne y E. Prud, «SPARQL 1.1 query language,» *W3C Recommendation,* vol. 21, 2013.

[91] I. Maurtua, I. Fernández, A. Tellaeche, J. Kildal, A. Ibarguren y B. Sierra, «Natural Multimodal Communication for Human-Robot Collaboration,» *International Journal of Advanced Robotic Systems,* 2017.

[92] J. Nielsen, Usability Engineering, Morgan Kaufman, 1993.

[93] I. Maurtua, A. Ibarguren, J. Kildal, L. Susperregi y B. Sierra, «Human Robot collaboration in Industrial applications: safety, interaction and trust,» *International Journal of Advanced Robotic Systems,* 2017.

[94] A. Ibarguren, I. Maurtua, M. A. Pérez y B. Sierra, «Multiple target tracking based on particle filtering for safety in industrial robotic cells,» *Robotics and Autonomous Systems,* vol. 72, nº C, pp. 105-113, 2015.

[95] «Página oficial de la libreria,» [En línea]. Available: https://github.com/flexible-collision-library.

[96] A. Ibarguren, o. M. JMartínez-Otzeta y I. Maurtua, «Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments,» de *9th International Conference on Informatics in Control, Automation and Robotics*, Roma, 2012.

[97] A. Tellaeche y I. Maurtua, «6DOF pose estimation of objects for robotic manipulation. A review of different options,» de *20th IEEE International Conference on Emerging Technologies And Factory Automation*, Barcelona, 2014.

[98] A. Ibarguren, J. Molina, L. Susperregi y I. Maurtua, «Thermal Tracking in Mobile Robots for Leak Inspection Activities,» *Sensors,* vol. 13, nº 10, pp. 13560-13574, 2013.

[99] PointClouds.org , «Tutorial uso de Kdtree,» PointClouds.org , [En línea]. Available: http://pointclouds.org/documentation/tutorials/kdtree_search.php. [Último acceso: 26 5 2017].

[100] A. Tellaeche, I. Maurtua y A. Ibarguren, «Use of machine vision in collaborative robotics: An industrial case,» de *22nd IEEE International Conference on Emerging Technologies And Factory Automation*, Berlin, 2016.

[101] I. Maurtua, L. Susperregi, A. Ansuategui, A. Fernández, A. Ibarguren, J. Molina, C. Tubio, C. Villasante, T. Felsch, C. Pérez, J. R. Rodriguez y M. Ghrissi, «Non-destructive inspection in industrial equipment using robotic mobile manipulation,» de *AIP Conference Proceedings* , 2016.

[102] F. Torsten, G. Strauss, C. Perez, J. M. Rego, I. Maurtua, L. Susperregi y J. R. JRodríguez, «Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques,» *Robotics,* vol. 4, nº 2, pp. 103-119, 2015.

[103] I. Maurtua, L. Susperregi, A. Fernández, C. Tubio, T. Felsch, C. Pérez, J. R. Rodriguez y M. Ghrissi, «MAINBOT – Mobile Robots for Inspection and Maintenance in Extensive Industrial Plants,» de *nergy Procedia, Volume 49, Pages 1-2532 (2014), Proceedings of the SolarPACES 2013 International Conference*, Las Vegas, 2013.

[104] A. Ibarguren, J. M. Martínez-Otzeta y I. Maurtua, «Particle Filtering for Industrial 6DOF Visual Servoing,» *Journal of Intelligent and Robotic Systems,* vol. 74, nº 3-4, pp. 689-696, 2014.

[105] I. Maurtua, «Wearable Technology in Automotive Industry: from Training to Real Production, Human-Computer Interaction,» de *Human-Computer Interaction*, Rijeka, InTech, 2009.

[106] I. Maurtua, M. Unceta y M. A. Pérez, «Experimenting Wearable Solutions for Workers' Training in Manufacturing,» de *Lecture Notes in Computer Science*, vol. 4553, Berlin, Springer, 2007, pp. 663-671.

[107] I. Maurtua y (Editor), Human Machine Interaction - Getting Closer, Rijeka: InTech, 2012, p. 270.

[108] I. Maurtua y (editorr), Human-Computer Interaction, Rijeka: InTech, 2009.

[109] J. Kildal, K. Tahiroglu, J. C. Vasquez y I. Maurtua, «Studying Human-Robot Collaboration in an Artistic Creative Process,» de *2017 Conference on Human-Robot Interaction (HRI2017), ReHRI'17 – International Workshop on reproducible HRI experiments: scientific endeavors, benchmarking and standardization*, Viena, 2017.

[110] I. Maurtua, I. Fernández, J. Kildal, L. Susperregi, A. Tellaeche y A. Ibarguren, «Enhancing safe human-robot collaboration through natural multimodal communication,» de *22nd IEEE International Conference on Emerging Technologies And Factory Automation*, Berlin, 2016.

[111] I. Maurtua, N. Pedrocchi, A. Orlandini, F. de Gea, C. Vogel, A. Geenen, K. Althoefer y A. Shafti, «FourByThree: Imagine humans and robots working hand in hand,» de *22nd IEEE International Conference on Emerging Technologies And Factory Automation*, Berlin, 2016.

[112] I. Maurtua, I. Fernandez, J. Kildal, L. Susperregi, A. Tellaeche y A. Ibarguren, «Interacting with collaborative robots in industrial environments: A semantic approach,» de *ICAPS 2016 Workshop on "Planning, Scheduling and Dependability in Safe Human-Robot Interactions*, Londres, 2016.

[113] J. Kildal y I. Maurtua, «Revisiting the end user's perspective in collaborative human-robot interaction,» de *19th International Conference on CLAWAR 9th International Conference on Climbing and Walking Robots and Support Technologies for Mobile Machines*, Londres, 2016.

[114] A. Tellaeche, I. Maurtua y A. Ibarguren, «Human robot interaction in industrial robotics. Examples from research centers to industry,» de *21st IEEE International Conference on Emerging Technologies And Factory Automation*, Luxemburgo, 2015.

[115] I. Maurtua, A. Ibarguren y A. Tellaeche, «Robotic solutions for Footwear Industry,» de *17th IEEE International Conference on Emerging Technologies And Factory Automation*, Cracovia, 2012.

[116] I. Maurtua, A. Ibarguren y A. Tellaeche, «Robotics for the Benefit of Footwear Industry,» de *5th International Conference on Intelligent Robotics and Applications*, Montreal, 2012.

[117] A. Tellaeche, R. Arana y I. Maurtua, «Accurate Correction of Robot Trajectories Generated by Teaching Using 3D Vision by Laser Triangulation,» de *5th International Conference on Intelligent Robotics and Applications*, Montreal, 2012.

[118] L. Susperregi, I. Fernández, A. Fernandez, S. Fernandez, I. Maurtua y I. Lopez de Vallejo, «Interacting with a Robot: A Guide Robot Understanding Natural Language Instructions,» de *6th International Conference on Ubiquitous Computing and Ambient Intelligence*, Vitoria-Gasteiz, 2012.

[119] M. I. de la Fuente, J. Echanobe, I. del Campo, L. Susperregui y I. Maurtua, «Hardware Implementation of a Neural-Network Recognition Module for Visual Servoing in a Mobile Robot,» de *DEXA 21st International Conference on Database and Expert Systems Applications*, Bilbao, 2010.

[120] M. I. de la Fuente, J. Echanobe, I. del Campo, L. Susperregui y I. Maurtua, «Development of an Embedded System for Visual Servoing in an Industrial Scenario,» de *2010 International Symposium on Industrial Embedded Systems*, Trento, 2010.

[121] I. Maurtua, P. T. S. Kirisci, Thomas, M. L. Sbodio y H. Witt, « Wearable Computing Prototype for supporting training activities in Automotive Production,» de *4th International Forum on Applied Wearable Computing* , Tel Aviv, 2007.

[122] M. Á. Pérez, L. Susperregi, I. Maurtua, A. S. Ibarguren y Basilio, «Software Agents for Ambient Intelligence based Manufacturing,» de *IEEE Workshop on Distributed Intelligent Systems*, Praga, 2006.

[123] L. Susperregi, I. Maurtua, C. S. Tubío, Inigo, M. Á. Pérez y B. Sierra, «Context aware agents for Ambient Intelligence in Manufacturing at Tekniker,» de *AgentLink*, 2005.

[124] I. Lopez de Vallejo, I. Maurtua y M. Unceta, «Ambient Intelligence in Manufacturing:Organizational Implications,» de *Workshop on 'Ubiquitous Computing and effects on Social Issues' Portland, Oregon. April 2005*, Portland, 2005.

[125] L. Susperregi, I. Maurtua, C. S. Tubío, Inigo, M. Á. Pérez y B. Sierra, «An agent based ambient intelligence experience in manufacturing,» de *Ambient Intelligence and (Everyday) Life International Workshop*, Donostia-San Sebastian, 2005.

[126] «3D Collision Detection for Industrial Robots and Unknown Obstacles using Multiple Depth Images».

[127] AUTOWARE, «WEB del proyeccto,» AUTOWARE, 2017. [En línea]. Available: http://www.autoware-eu.org/. [Último acceso: 21 5 2017].

[128] H. Liu, Y. Yu, F. Sun y J. Gu, «Visual-tactile fusion for object recognition,» *IEEE Transactions on Automation Science and,* vol. 14, nº 2, p. 996 – 1008, 2017.