eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Departamento de Ciencia de la Computación e Inteligencia Artificial
Konputazio Zientzia eta Adimen Artifiziala Saila
Department of Computer Sciences and Artificial Intelligence

Técnicas de Visión e Inteligencia Artificial para Aplicaciones Industriales en Producción

Mikel Labayen Esnaola

septiembre, 2023

Supervisores / Gainbegiraleak / Supervisors:
Ph.D. Naiara Aginako Bengoa
Ph.D. Basilio Sierra Araujo

Mikel Labayen Esnaola

Técnicas de Visión e Inteligencia Artificial para Aplicaciones Industriales en Producción
Supervisores: Ph.D. Naiara Aginako Bengoa y Ph.D. Basilio Sierra Araujo
Universidad del País Vasco
Departamento de Ciencia de la Computación e Inteligencia Artificial
CAF Signalling
Área de Vehículo Autónomo
Donostia

Ikusmen eta Adimen Artifizialeko Teknikak Ekoizpeneko Industria Aplikazioetarako
Gainbegiraleak: Ph.D. Naiara Aginako Bengoa eta Ph.D. Basilio Sierra Araujo
Euskal Herriko Unibertsitatea
Konputazio Zientzia eta Adimen Artifiziala Saila
CAF Signalling
Ibilgailu Autonomoen Saila
Donostia

Computer Vision and Artificial Intelligence Techniques for Industrial Applications in Production
Supervisors: Ph.D. Naiara Aginako Bengoa and Ph.D. Basilio Sierra Araujo
University of the Basque Country
Department of Computer Sciences and Artificial Intelligence
CAF Signalling
Autonomous Vehicle Area
Donostia

Resumen

La Visión Artificial (VA) o Computer Vision (CV) está muy presente tanto en aplicaciones personales cotidianas que se utilizan en smartphones, ordenadores o incluso vehículos, como en la industria productiva, donde es una de las tecnologías fundacionales de la automatización de muchos de los procesos visuales (fabricación y control automatizado, análisis de contenido audiovisual, videovigilancia, conducción autónoma...). Esta línea de investigación (VA), que pertenece a la denominada Inteligencia Artificial (IA) o Artificial Intelligence (AI), ha ayudado a añadir valor y a mejorar la calidad de los productos, así como gestionar la fabricación y facilitar la comercialización y la distribución de bienes. En general, ha permitido crear nuevas funcionalidades de valor añadido y optimizar numerosos procesos, bajando sus costes y haciéndolos más eficientes.

La visión artificial empezó su andadura en la industria a finales de la primera década del siglo XXI (2005-2012). Por aquel entonces, las técnicas clásicas de procesamiento de imagen afrontaban el reto de reducir complejidad en sus algoritmos SW y hacerlos ejecutar en tiempo real en un set-up HW aceptable en costes y tamaño. Con la revolución del Aprendizaje Profundo (AP) o Deep Learning (DL) que se inició en la segunda década de este siglo y que se ha desarrollado especialmente en los últimos años (2012-actualidad), se ha dotado de IA a los sistemas basados en VA, introduciendo el uso de técnicas basadas en Redes Neuronales Profundas (RNP) o Deep Neuronal Networks (DNN). Estas redes neuronales se han convertido rápidamente en parte esencial del campo de VA, dado su potencial y aplicabilidad, demostrada en muchos sectores de la industria. Sin embargo, la introducción de las RNP en productos/servicios/aplicaciones industriales acarrea una serie de retos que han de ser resueltos poco a poco. A día de hoy, algunos de ellos están todavía pendientes de una solución real y factible.

La elección de las arquitecturas SW y HW, así como su configuración/modificación para los casos de uso concretos, requieren nuevas líneas y tareas de investigación para poder garantizar la Fiabilidad, Disponibilidad, Mantenibilidad, Escalabilidad (FDME), y en último caso, y si la funcionalidad así lo requiere, la Seguridad de

los productos/servicios para que al final sistemas basados en estas tecnologías puedan ser comercializados. Este proyecto de tesis tiene como objetivo proporcionar conocimientos y herramientas basados en sistema de Visión e Inteligencia Artificial (VA&IA) para el desarrollo de:

- Prototipos funcionales y accesibles para cualquier tipo de empresa.
- Productos comerciales en producción.
- Productos para sistemas embebidos.
- Funcionalidades de tipo Nivel de Integridad de Seguridad (NIS) o Safety Integrity Level (SIL).

Este trabajo también expone los resultados de las investigaciones realizadas en diferentes casos de uso y en diferentes sectores industriales a lo largo de la vida profesional del autor de esta memoria. Dichas investigaciones han sido publicadas en revistas y congresos internacionales, en aras a obtener el respaldo de las personas que han aceptado publicar los artículos, y de las personas que tengan a bien leerlos.

En esta memoria, que se realiza bajo la modalidad de compendio de artículos, se presenta un amplio resumen de la labor investigadora realizada, así como los artículos y las patentes que la avalan.

Laburpena

Ikusmen Artifiziala (IkA) edo Computer Vision-a (CV) oso presente dago bai telefono adimendunetan, ordenagailuetan edo baita ibilgailuetan erabiltzen diren eguneroko aplikazio pertsonaletan, baita ekoizpen-industrian ere, non begizko prozesu asko automatizatzeko oinarrizko teknologietako bat den (fabrikazio eta kontrol automatizatua, ikus-entzunezko edukien azterketa, bideo-zaintza, gidatze autonomoa...). Adimen Artifiziala (AA) edo Artificial Intelligence (AI) izenekoari dagokion ikerketa-lerro honek (IkA) balio erantsia ematen eta produktuen kalitatea hobetzen lagundu du, baita ekoizpena kudeatzen eta salgaien banaketa eta merkaturatze prozesua errazten ere. Oro har, balio erantsiko funtzionaltasun berriak sortzea eta prozesu ugari optimizatzea ahalbidetu du, haien kostuak murriztuz eta eraginkorragoak eginez.

Ikusmen artifizialak XXI. mendeko lehen hamarkadaren amaieran (2005-2012) hasi zuen bere ibilbidea industrian. Garai hartan, irudiak prozesatzeko teknika klasikoek euren SW algoritmoen konplexutasuna murrizteko eta kostu eta tamaina aldetik onargarria zen HW konfigurazio batean denbora errealean aplikazioak exekutatzeko erronkei aurre egin zioten. Mende honetako bigarren hamarkadan hasi eta bereziki azken urteotan (2012-gaur egun) garatu den Ikaskuntza Sakonaren (IS) edo Deep Learning-aren (DL) iraultzarekin, IkA-en oinarritutako sistemak AA-az hornitu dira, Sare Neuronal Sakonetan (SNS) edo Deep Neuronal Networks-etan (DNN) oinarritutako tekniken erabilera integratuz. Sare neuronal hauek IkA eremuan ezinbesteko zati bihurtu dira, industria-sektore askotan dituzten potentziala eta aplikagarritasuna egiaztatu bait dira. Hala ere, SNS-ak produktu/zerbitzu/aplikazio industrialetan sartzeak hainbat erronka dakartza berekin, pixkanaka konpondu beharrekoak. Gaur egun, horietako batzuk konponbide erreal eta bideragarri baten zain daude oraindik.

SW eta HW arkitekturak aukeratzeak, bai eta erabilera kasu zehatzetarako beraien konfigurazioak/aldaketak burutzeak, ikerketa-lan eta ildo berriak eskatzen ditu Fidagarritasuna, Erabilgarritasuna, Mantenigarritasuna, Eskalagarritasuna (FEME), eta azken kasuan, eta funtzionaltasunak hala eskatzen badu, teknologia horietan

oinarritutako sistemak merkaturatzea ahalbideratuko duen produktu/zerbitzuen Segurtasuna bermatzeko. Tesi proiektu honek ondorengo garapen hauek aurrera eramateko Ikusmen eta Adimen Artifizialean (IkA&AA) oinarritutako sistemen ezagutzak eta tresnak eskaintzea du helburu:

- Prototipo funtzionalak eta eskuragarriak edozein motatako enpresarentzat.
- Produktu komertzialak ekoizpenean.
- Sistema txertatuetarako produktuak.
- Segurtasun Osotasun Maila (SOM) edo Safety Integrity Level (SIL) motako funtzionalitateak.

Lan honek txosten honen egilearen lanbide-bizitzan zehar erabilera-kasu ezberdinetan eta industria-sektore ezberdinetan egindako ikerketen emaitzak azaltzen ditu. Ikerketa hauek nazioarteko aldizkari eta kongresuetan argitaratu dira, artikuluak argitaratzea adostu duten pertsonen eta irakurtzeko prest dauden pertsonen aitortza lortzeko asmoz.

Artikuluen bildumaren bidez egiten den memoria honetan, egindako ikerketalanaren laburpen zabala aurkezten da, baita berau eusten duten artikuluak eta patenteak ere.

Abstract

Computer Vision (CV) is very present both in everyday personal applications used in smartphones, computers or even vehicles, and in the production industry, where it is one of the foundational technologies for the automation of many visual processes (automated manufacturing and control, audiovisual content analysis, video surveillance, autonomous driving...). This research line (CV), which belongs to the so-called Artificial Intelligence (AI), has helped to add value and improve the quality of products, as well as managing manufacturing and facilitating the marketing and distribution of goods. In general, it has enabled the creation of new value-added functionalities and the optimisation of numerous processes, decreasing their costs and making them more efficient.

Machine vision began its journey in industry at the end of the first decade of the 21st century (2005-2012). At that time, classical image processing techniques faced the challenge of reducing complexity in their SW algorithms and making them run in real time on an acceptable HW set-up in terms of cost and size. With the Deep Learning (DL) revolution that started in the second decade of this century and has developed especially in the last few years (2012-present), CV-based systems have been endowed with AI, introducing the use of techniques based on Deep Neural Networks (DNN). These neural networks have quickly become an essential part of the CV field, given their potential and applicability, demonstrated in many industry sectors. However, the introduction of DNNs in industrial products/services/applications brings with it a number of challenges that need to be solved gradually. Today, some of them are still awaiting a real and feasible solution.

The choice of SW and HW architectures, as well as their configuration/modification for specific use cases, require new research lines and tasks in order to guarantee the Reliability, Availability, Maintainability, Scalability (RAMS), and ultimately, if the functionality requires it, the Security of the products/services so that systems based on these technologies can be marketed. This thesis project aims to provide knowledge and tools based on Computer Vision and Artificial Intelligence (CV&AI) for the development of:

- Functional and accessible prototypes for any type of company.
- Commercial products in production.
- Products for embedded systems.
- Safety Integrity Level (SIL) type functionalities.

This work also presents the results of research carried out in different use cases and in different industrial sectors throughout the author's professional life. This research has been published in international journals and conferences, in order to obtain the support of the people who have accepted to publish the articles, and of the people who are willing to read them.

This report, which is a compilation of articles, presents a broad summary of the research work carried out, as well as the articles and patents that support it.

Eskertzak

Nahiz eta inguruko askorentzat lan hau isilpean eta oharkabean pasa den,

Etxekoei

Egunero zaintzen nauten nire hiru sorginei, momentu baxuetan aurrera egiteko indarrak eman dizkidaten irribarre goxo horiengatik.

Lankideei

Nere alboan lan egin nahi edo beste erremediorik izan ez duten guzti horiei. Zuekin guztiokin egindako elkarlan eta parrandengatik. Horiek gabe, ez zen posible izango.

Zuzendari eta Mentoreei

Momentu txarretan, eta proiektu honekin jarraitzeko nire borondate eza aurka izanda ere, lan guzti hau aurrera eramatera animatu izanagatik. Baita hasiera haietan mentore gisara lan honi ekitera intsistentziaz bultzatu izanagatik ere.

Familiari

Lan honetan nenbilela jakin ez bazekiten ere, txikitatik hona, hurrengo orri hauetako lana burutzeko emandako erreminta, laguntza eta maitasun guztiagatik.

Lagunei

Lan hau burutu bitartean biziarazitako ekintza *'ludiko-festibo'* guztiengatik.

Eskerrik
Asko!

Índice general

1 Introducción

La visión artificial es un campo de la inteligencia artificial que enseña a los ordenadores a *'ver'*, observar, inspeccionar y comprender el contenido de las imágenes digitales. Por tanto, un sistema de visión e inteligencia artificial implica la capacidad de una máquina de obtener información significativa a partir de entradas visuales como imágenes y vídeos, y de realizar acciones o recomendaciones basadas en esa información. Para ello requiere el empleo de sensores visuales como cámaras de fotos o vídeos y técnicas de procesamiento digital de señales. La visión artificial funciona en tres pasos básicos:

- **Adquisición de la imagen:** las imágenes, se pueden adquirir en tiempo real a través de cámaras de rangos espectrales diferentes (entre ellos el visual) o a través de representaciones gráficas de otros sensores como el LiDAR (light detection and ranging).
- **Procesamiento de la imagen:** se puede realizar mediante filtros digitales, técnicas clásicas de segmentación o tracking, o por modelos de aprendizaje profundo que automatizan gran parte de este proceso con el fin de identificar y clasificar los objetos que aparecen en ellas. Estos modelos se entrenan con miles de imágenes etiquetadas previamente.
- **Entendimiento de la imagen:** el paso final es el paso interpretativo, donde se identifica o clasifica un objeto.

Se puede considerar que la visión artificial es un campo de la inteligencia artificial y si la IA permite a los ordenadores *'pensar'*, la VA les permite ver, *'observar y comprender'*.

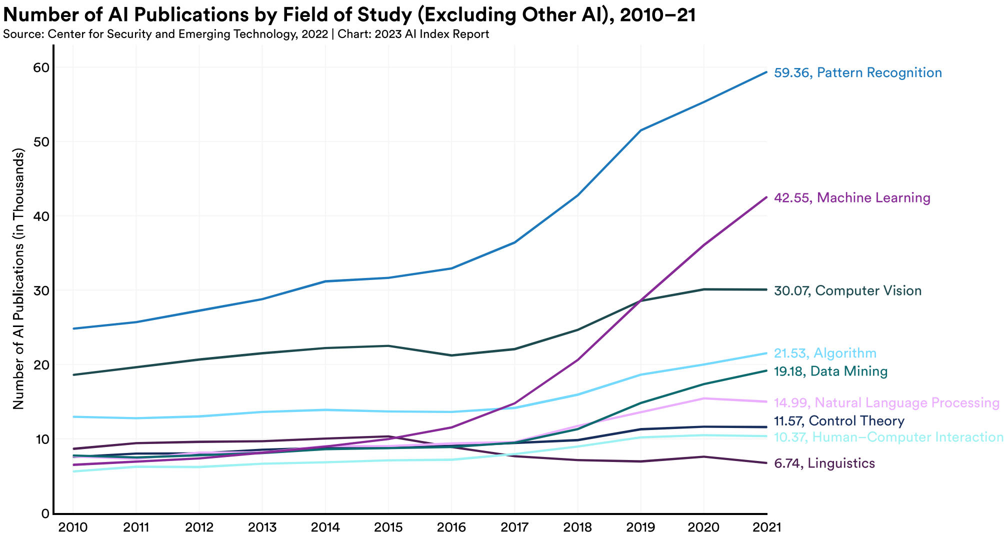
Actualmente, las aplicaciones de visión artificial están presentes desde en pequeñas aplicaciones en la vida cotidiana de cada persona, hasta en grandes y complejas tareas industriales donde se utiliza para automatizar procesos. Las aplicaciones son numerosas (ver Fig. 1.1): detección de defectos, control de calidad, metrología, detección de intrusos alimenticios, lectura de códigos, verificación de montajes, videovigilancia, autenticación de identidad, análisis de contenido audiovisual, conducción autónoma... Y los sectores que se benefician de ellos también: automoción, sector agroalimentario, envases, packaging y embalajes, electrónica, logística, sector audiovisual, seguridad, movilidad...

(a) Robótica. (b) Autenticación. (c) Conducción autónoma.

(d) Alimentación. (e) Contenido audiovisual. (f) Control de calidad.

Fig. 1.1: Ejemplos de uso de técnicas de visión e inteligencia artificial en producción en distintos sectores de la industria.

El tamaño del mercado mundial que mueven las tecnologías de visión e inteligencia artificial se valoró en 12,78 mil millones de dólares en 2021 y se espera que se expanda a una tasa de crecimiento anual compuesta (CAGR) del 12,5 % hasta alcanzar los 24,19 mil millones de dólares en 2029 [DBM23]. Las crecientes necesidades en inspección de calidad, en la automatización en diferentes verticales industriales y en sistemas basados en robótica guiada por visión, impulsan con fuerza el crecimiento del mercado. Estas tecnologías también son de importancia para los sectores como el de la automoción, la aeronáutica y el transporte ferroviario, ya que están en plena fase de desarrollo de sus vehículos autónomos. Toda esta espiral de demanda tecnológica hace disparar la financiación de numerosos proyectos de investigación en el campo de VA&IA, y en consecuencia, el goteo de resultados publicables crece de forma considerable (ver Fig. 1.2).



Number of AI Publications by Field of Study (Excluding Other AI), 2010–21

Source: Center for Security and Emerging Technology, 2022 | Chart: 2023 AI Index Report

Fig. 1.2: Número de publicaciones por área de estudio dentro de IA [Sta23]).

Sin embargo, todavía hoy en día, la visión artificial se enfrenta a retos no resueltos en su objetivo de penetrar definitivamente en aplicaciones industriales y ofrecer servicios fiables, estables y de bajo coste, tanto en su implantación como en su mantenimiento:

- **Costes elevados de set-up:** aunque dependa mucho del tipo de aplicación, si ésta necesita de un set-up multi-sensor y el equipo necesario para procesar dicha tecnología en tiempo real y de forma continuada, los costes se disparan. La conducción autónoma es un buen ejemplo de ello, ya que requiere un procesamiento en tiempo real, en equipos embarcados y gestionando información proveniente de muchos sensores.
- **Necesidad de mucha información inicial:** los modelos de AP requieren un preentrenamiento que consume muchos datos iniciales. Capturarlas, procesarlas, etiquetarlas y entrenar los modelos tiene un coste logístico, temporal y de recursos elevado.
- **Costes elevados en mantenimiento:** muchas de las aplicaciones industriales actuales exigen que los sistemas cumplan con requisitos de fiabilidad, disponibilidad, mantenibilidad y escalabilidad. Dado que la tecnología es relativamente nueva y evoluciona rápido, las situaciones inesperadas en su correcto funcionamiento son frecuentes. Mantenerlos, escalarlos y ofrecer un servicio disponible en todo momento supone unos gastos elevados.
- **Legislación y procesos de certificación:** las industrias que desarrollan aplicaciones de seguridad que requieren altos niveles de integridad, todavía no pueden incorporar estas tecnologías ya que la legislación vigente no lo permite en muchos sectores. La tecnología de visión artificial y la forma de demostrar su fiabilidad todavía tendrán que desarrollarse antes de que estos algoritmos puedan ser utilizados para controlar sistemas críticos de seguridad.

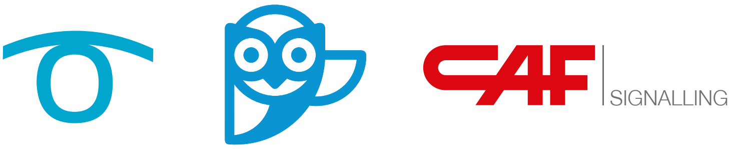
1.1 Contexto del trabajo de investigación

Los resultados presentados en este proyecto de tesis doctoral son fruto de diversos trabajos de investigación realizados durante la experiencia profesional del doctorando. Concretamente, se apoya en un compendio de artículos que han sido desarrollados gracias a la cooperación durante los últimos años entre grupos de investigación de diferentes empresas y el departamento de Ciencias de la Computación e Inteligencia Artificial (CCIA, ↗) de la Universidad del País Vasco (UPV, ↗). Este departamento trabaja en proyectos que comparten las mismas necesidades tecnológicas relativas al VA e IA que las empresas industriales para conseguir soluciones competitivas que ofrezcan resultados de vanguardia. Estas necesidades son el nexo de unión que han dado lugar a todos los artículos que componen este proyecto de tesis.

1.1.1 Currículum investigador y profesional

En su paso por varias empresas de diferentes sectores (Fig. 1.3), el doctorando ha trabajado más de 15 años en proyectos de automatización basados en VA y IA (Sec. 1.7). Primeramente, en el centro de investigación Vicomtech (2007-12), como investigador e ingeniero SW, diseñando y desarrollando sistemas, además de como líder de proyectos, gestionando proyectos cliente y equipos I+D. En una segunda etapa en la start-up de base científico-tecnológica Smowltech (2012-18), como máximo responsable en la fase fundacional de la empresa y como posterior director técnico, comercializando internacionalmente aplicaciones de autenticación biométrica. Por último, en CAF Signalling (2018-actualidad), como responsable de la división del vehículo autónomo, liderando proyectos cliente y equipos I+D en el campo de la automatización de operaciones ferroviarias basada en Sistemas de Transporte Inteligente (STI). Además, durante dos cursos académicos (2010/11 y 2017/18), compagino su trabajo con un puesto de profesor adjunto en el departamento de Ingeniería Electrónica de la UPV.

En cuanto a su currículum investigador son reseñables sus diferentes actuaciones como líder de tarea y paquete de trabajo en 6 proyectos de investigación de ámbito nacional e internacional, así como su papel de coordinador e investigador principal en otros 2 proyectos: en un Retos-Colaboración (↗) y un SME Instrument-H2020 (↗). Por otra parte, su trayectoria como investigador se ve reforzada por numerosas publicaciones como autor principal, 5 de ellas en revistas de renombre con cuartiles Q1 y Q2 (Sec. 4), y por 4 patentes internacionales (Sec. 5), que habiendo siendo reconocidas por los organismos pertinentes, protegen sus trabajos de investigación.



(a) Vicomtech. ↗ (b) Smowltech. ↗ (c) CAF Signalling. ↗

Fig. 1.3: Centro tecnológico y empresas de diferentes sectores industriales donde se han realizado los trabajos presentados en esta tesis.

1.1.2 Vicomtech

Vicomtech es un centro de investigación tecnológica especializado en VA y IA. Aunque también colabora en proyectos de investigación básica, el principal objetivo de Vicomtech es tender un puente entre la investigación básica y la industria, desarrollando soluciones reales para las empresas. Por este motivo, las innovaciones están muy centradas en aplicaciones industriales.

Los resultados presentados en este proyecto de tesis, y que corresponden al paso del tesitando por el centro de investigación, están muy centradas en aplicaciones concretas y prácticas del sector audiovisual. En concreto, se obtuvieron en proyectos desarrollados dentro de la subdivisión de Análisis Automatizado de Contenido Multimedia del departamento de Televisión Digital y Servicios Multimedia.

1.1.3 Smowltech

Smowltech es una start-up que nació como una spin-off del centro de investigación Vicomtech para el desarrollo de producto y la comercialización de un prototipo que autenticación de usuarios online mediante reconocimiento biométrico facial. Actualmente es una tecnología patentada (ver secciones 5.4 y 5.3), que aparte de autenticar la identidad de usuario, no solo mediante reconocimiento facial sino también por voz y forma de tecleo, incorpora una opción de supervisión de su actividad en la sesión online mediante vigilancia automatizada, captura de screenshots e incluso pudiendo bloquear periféricos y aplicativos en su dispositivo si así se requiere.

Desarrollar un producto basado en tecnologías de VA y IA, funcional, fiable, disponible, mantenible y escalable, ofreciendo un servicio en tiempo real supone una actividad investigadora que a día de hoy es una de las tareas que más recursos demanda en la empresa. Los resultados exportados al proyecto de tesis doctoral corresponden a la fase de investigación y desarrollo de este producto industrial por parte del departamento técnico liderado por el doctorando.

1.1.4 CAF Signalling

CAF Signalling es una filial tecnológica del grupo CAF. Sus actividades se orientan al diseño, desarrollo, fabricación, suministro y mantenimiento de sistemas de señalización ferroviaria. Proporciona soluciones integrales, cubriendo tanto el ámbito de infraestructuras como el de material embarcado. Para llevar a cabo dichas actividades, la empresa cuenta con capacidad de ingeniería propia, siendo reconocida como tal por diferentes operadores y gestores de infraestructura ferroviarios en los cinco continentes.

Innovar es una de las actitudes esenciales de la estrategia empresarial perseguida por CAF Signalling y actualmente uno de sus principales retos, ya que la siguiente revolución tecnológica en el sector supondrá la total automatización de la conducción basada en tecnologías de VA y IA. Así lo recoge el plan de innovación 2022-2026 del grupo CAF, siendo la conducción autónoma uno de los cuatro pilares de la estrategia de futuro. En este contexto, CAF Signalling se plantea dar una respuesta tecnológica sólida y de largo alcance a una demanda creciente en los niveles de automatización que garanticen niveles más eficientes de operación y mayores cotas de seguridad. Para ello está inmersa en una intensa actividad de investigación industrial que está liderando el doctorando. Es en este área de conducción autónoma donde se centran las aportaciones a este proyecto de tesis correspondientes al sector ferroviario.

1.2 Motivación

La inclusión de la inteligencia artificial en forma de aprendizaje profundo sobre RNPs ha revolucionado el campo de la visión artificial en los últimos años. Gracias a ello, continuamente se están sucediendo nuevos casos de éxito, prototipos y productos que incorporan dichas tecnologías y revolucionan la automatización de procesos industriales de todo tipo. Por ejemplo, en detección/identificación de objetos, una de las tareas más conocidas en visión artificial, los modelos basados en RNPs han superado ampliamente a los modelos tradicionales [Zou+23], y día a día siguen mejorando sus resultados.

Sin embargo, esto no siempre fue así. Los primeros prototipos y productos industrializables sólo se basaban en técnicas tradicionales de visión artificial como segmentación, detección, clasificación o tracking mediante procesamiento de imágenes o de vídeos. De hecho, actualmente, las aplicaciones de visión artificial sin módulos de inteligencia artificial son muy útiles en aquellos contextos en el que el problema no requiere de mayor complejidad tecnológica para ser resuelta y los criterios de seguridad requieren software determinista. Por aquel entonces y en las actuales aproximaciones estrictamente de visión artificial, las investigaciones eran y son motivadas por la necesidad de:

- Crear prototipos totalmente funcionales.
- Crear prototipos válidos para análisis de mercado y la viabilidad de la solución.
- Diseñar soluciones con algoritmos optimizados que rebajen el coste computacional y permitir su ejecución en tiempo real.
- Diseñar soluciones con set-up moderados para garantizar la accesibilidad de la tecnología por todo tipo y tamaño de empresas.

No fue hasta el año 2012 cuando la tecnología basada de aprendizaje profundo basado en RNPs dotó de inteligencia a las aplicaciones de visión artificial, pasándolas a llamar aplicaciones basadas en visión e inteligencia artificial. Los retos desde entonces han sido variados y las investigaciones han sido promovidas por la necesidad de:

- Insertar funcionalidades basadas en visión por computador y RNPs en productos/servicios en producción [Xia+23].
- Ejecutar funcionalidades de visión e inteligencia artificial en tiempo real y con baja latencia [ZL23].
- Obtención de resultados fiables mediante fusión de datos, métodos o diferentes algoritmos para aumentar la seguridad [Dua23] [Li+23].
- Diseño de set-up reducidos que hagan de los productos mantenibles, escalables y que garanticen su disponibilidad durante su ciclo de vida.

Actualmente, gracias a su éxito y auge, estas tecnologías han permitido imaginar todo tipo de aplicaciones. Uno de los ejemplos más notorios es el de la robótica autónoma, el cual requiere poder ejecutar las aplicaciones de forma fiable, en tiempo real y en sistemas embebidos de bajo consumo. A su vez, si planean operar funcionalidades de nivel de integridad de seguridad, deben conseguir poder ser certificables. Las nuevas motivaciones se pueden resumir en:

- Portar las soluciones de visión e inteligencia artificial a equipos embebidos, con recursos limitados y consumo mínimo de energía, tal y como requieren lo sistemas autónomos [BT23].
- Crear plataformas HW embebidas que aíslen las operaciones/ejecuciones cruciales y mantengan de forma eficaz y segura los recursos del sistema [MGL23].
- Mantener el rendimiento y las características de fiabilidad, mantenibilidad, disponibilidad (ejecuciones redundadas) y escalabilidad del sistema en entornos embebidos [SLS23].
- Hacer que las soluciones basadas en inteligencia artificial puedan ser explicables y certificables mediante procesos de verificación y validación estandarizados [Ali+23].

La motivación de transferir estas tecnologías a la industria (que se inició hace ya una década y media), ha suscitado diferentes líneas de investigación. Dependiendo de la madurez de la tecnología disponible en cada momento, han perseguido hitos diferentes pero progresivos hacia un mismo fin: conseguir ejecutar en tiempo real sistemas complejos de visión e inteligencia artificial en cualquier plataforma HW/SW subyacente, incluso lo más limitantes en recursos y consumos de energía, y con un grado de seguridad y fiabilidad certificables.

Las diferentes líneas de investigación descritas en este proyecto de tesis han sido motivadas por las diferentes necesidades de la industria provenientes del deseo de integrar estas tecnologías siempre dinámicas y cambiantes. Por tanto, siempre han tenido como objeto proporcionar conocimientos y herramientas para la comercialización de sistemas basados en visión e inteligencia artificial aplicados a diferentes casos de uso reales, cada uno con sus requisitos, y de diferentes sectores industriales.

1.3 Hipótesis

Teniendo en cuenta las motivaciones expuestas en la sección anterior, se formulan las siguientes hipótesis, que han servido de base para la investigación llevada a cabo en este proyecto de tesis:

1. Los sistemas de visión e inteligencia artificial basados en Deep Learning son parte esencial y muy demandada de las aplicaciones de automatización industrial.

2. Los sistemas de visión e inteligencia artificial han de ser diseñados con el mínimo set-up de sensores y equipamiento, así como con soluciones simples a nivel de SW para poder garantizar su accesibilidad por cualquier tipo y tamaño de empresa.

    a) La reducción en la complejidad en el diseño de la solución, la eficiencia en el desarrollo de algoritmos y la definición de mínimos en las arquitecturas de IA son esenciales para poder reducir el set-up de sensores y requisitos de cómputo de la solución.

    b) La reducción de costes en la solución final hará del producto una solución más accesible para todo tipo de empresas y se garantizará la transferencia tecnológica total a la industria productiva independientemente de su volumen de negocio.

3. Los sistemas de visión e inteligencia artificial han de demostrar fiabilidad, disponibilidad, mantenibilidad y escalabilidad para poder ser incluidos dentro de un producto/servicio en producción.

    a) Los modelos de IA son normalmente opacos y poco explicables y es difícil valorar su fiabilidad de forma clara y concisa. Por tanto, es de vital importancia ejecutar varios aplicativos que persigan un mismo fin basados en diferentes modelos IA para poder garantizar cierto grado de fiabilidad. La fusión de resultados y su gestión son necesarias para garantizar el grado de fiabilidad requerido por ciertas aplicaciones industriales.

    b) Las aplicaciones han de estar disponibles siempre que se requieran. Por tanto, deberán de existir módulos secundarios de redundancia para garantizar la disponibilidad ante un fallo del sistema.

    c) Las aplicaciones han de ser mantenibles. Será de importancia tener un buen sistema de mejora y re-entrenamiento.

    d) Las aplicaciones basadas en un modelo IA que necesita de procesos de etiquetado, entrenamiento y validación en bucle para garantizar su escalabilidad, tienen que tener sistemas ágiles y rápidos para ello.

4. Los sistemas de Visión e Inteligencia Artificial requeridos por la industria de la movilidad autónoma han de poder ser ejecutados en tiempo real en sistemas embebidos y de bajo consumo.

    a) Las inferencias de las distintas aplicaciones basadas en IA que puedan estar compartiendo recursos computacionales, han de poder hacerlas en entornos de sistema embebidos con recursos limitados y consumos reducidos.

5. Los sistemas de visión e inteligencia artificial han de poder ser validados y verificados de forma concisa y segura, así como certificados para poder utilizarlos en las funcionalidades de nivel de integridad de seguridad.

    a) El comportamiento de los modelos de IA no puede ser acotado y puede caer en respuestas no controladas. Es de obligado cumplimiento que se validen y verifiquen en entornos virtuales que simulen el mayor de los casos y escenarios posibles. Así podrán ser parte de una funcionalidad Nivel de Integridad de Seguridad, que requiere datos lo más exactos posibles sobre probabilidades de fallo de cada subsistema.

    b) El HW donde se ejecutan los algoritmos y su arquitectura deberá seguir diseños e implementaciones que garanticen el control de recursos, ejecuciones redundadas y sistemas de botadores en toma de decisiones para mayor seguridad.

1.4 Objetivos

Atendiendo a las hipótesis formuladas anteriormente, se definen tres líneas de investigación, con varios objetivos asociados a cada una de ellas. La Fig. 1.4 muestra un resumen gráfico de la relación entre contexto y líneas de investigación.

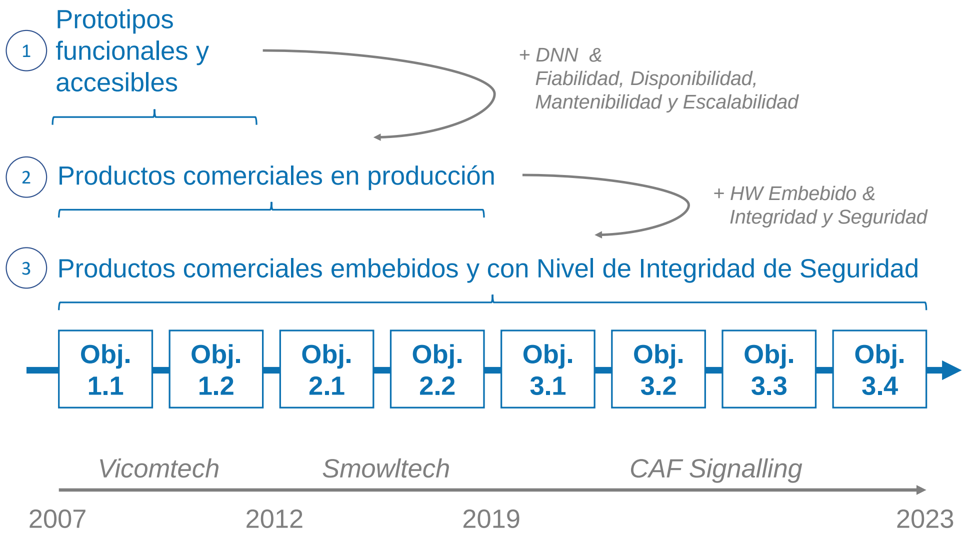

1
Prototipos funcionales y accesibles
+ DNN &
Fiabilidad, Disponibilidad,
Mantenibilidad y Escalabilidad
2
Productos comerciales en producción
+ HW Embebido &
Integridad y Seguridad
3
Productos comerciales embebidos y con Nivel de Integridad de Seguridad
Obj. 1.1
Obj. 1.2
Obj. 2.1
Obj. 2.2
Obj. 3.1
Obj. 3.2
Obj. 3.3
Obj. 3.4
Vicomtech
Smowltech
CAF Signalling
2007
2012
2019
2023

Fig. 1.4: Resumen de relación Líneas de investigación - Objetivos - Contexto de la investigación.

1.4.1 Prototipos funcionales y asequibles basados en visión artificial para la industria audiovisual

Esta línea de investigación se ha centrado en realizar prototipos totalmente funcionales, capaces de operar en tiempo real y basados en técnicas de visión artificial. El objetivo es lograr que las empresas puedan explorar el mercado en busca de necesidades de cliente y esclarecer la viabilidad de un producto más maduro de cara a una futura comercialización. Las soluciones, además de funcionar con precisión y con aceptable rendimiento, tienen que ser lo más simples posibles y operar con el mínimo set-up de equipamiento posible. Con ello se garantiza la accesibilidad a esta tecnología a cualquier tamaño de empresa. Los objetivos concretos son:

Obj 1.1 Crear prototipos funcionales basados en soluciones por visión artificial:

- Estudiar las últimas técnicas de visión artificial (segmentación, detección, clasificación y tracking), configuraciones (set-up) de sensores visuales (cámaras) y equipamiento de procesamiento para determinar la mejor solución ad-hoc para diferentes casos de uso del sector.

- Probar y desarrollar diferentes soluciones HW y SW para las aplicaciones automatizadas y ofrecer un diferencial de innovación optimizando y proponiendo nuevas técnicas basándose en las peculiaridades de los problemas a solucionar.

Obj 1.2 Crear diseños accesibles para todo tipo y tamaño de empresas, simplificando las soluciones y adaptando el HW y el SW a las características de cada caso de uso:

- Diseñar configuraciones simplificadas de set-up de sensores y equipamiento de procesamiento que permitan soluciones con costes acotados y simples en operación y mantenimiento.

- Simplificar el SW de procesamiento y optimizar algoritmia aprovechando las peculiaridades de cada caso de uso con los cuales poder reducir el alcance de la aplicación hasta el mínimo requerido.

Los resultados presentados en este proyecto de tesis respecto a estos objetivos se centran en un caso de uso concreto donde mediante análisis automatizado por visión artificial de contenido multimedia, se ofrece información virtual de valor añadido en retransmisiones de eventos deportivos por televisión.

1.4.2 Sistemas comerciales fiables con técnicas de visión e inteligencia artificial para la industria de aplicaciones y servicios online

Esta línea de investigación se centra en el diseño y desarrollo de sistemas basados en la tecnología del aprendizaje profundo o deep learning, siendo primeramente prototipos y posteriormente productos/servicios en producción que operan en tiempo real. En otras palabras, el objetivo es introducir nuevos productos/servicios innovadores en el mercado basándose en las nuevas bondades que ofrece el AP y hacerlo funcionar en producción, con garantías y en tiempo real.

Las soluciones, además de funcionar con precisión y con aceptable rendimiento (Fiabilidad), tienen que cumplir con características típicas de un producto/servicio maduro en producción; la Disponibilidad, la Mantenibilidad y la Escalabilidad (FDME). Además, todo ello tiene que ser diseñado con soluciones simples y set-up limitados, siendo accesibles para empresas de tamaño reducido, con bajas/limitadas capacidades financieras y operativas:

Obj 2.1 Crear sistemas/servicios de visión artificial que introduzcan un cambio de paradigma para la visión artificial con la inclusión de técnicas de AP:

- Estudiar las últimas técnicas de visión e inteligencia artificial, configuraciones (set-up) de sensores (tanto visuales, como de audio, teclados de escritura...) y equipamiento de procesamiento (local y/o online) para determinar la mejor solución ad-hoc para diferentes casos de uso del sector.

- Probar y desarrollar diferentes soluciones HW y SW para las aplicaciones automatizadas ofreciendo un diferencial de innovación que optimice y proponga nuevas técnicas gracias a las peculiaridades de los casos de uso y de los problemas a solucionar.

Obj 2.2 Crear productos/servicios listos para entrar en producción y dar servicio a clientes de forma duradera y continuada. Para ello, el sistema tiene que cumplir con las características FDME:

- Diseñar y desarrollar soluciones fiables con altos índices de precisión y *recall*. Dado que los sistemas basados en AP pueden arrojar resultados a priori impredecibles o incontrolables, se requieren soluciones AP ejecutadas por separado y fusionar sus resultados (i.e. identificación multi-biométrica con reconocimiento facial + reconocimiento de voz + reconocimiento de forma de escritura o typing).

- Diseñar y desarrollar soluciones estables que estén siempre disponibles y accesibles para los clientes independientemente de factores externos. La redundancia en la ejecución y mecanismos de balanceo de carga serán esenciales para lograr este objetivo.

- Diseñar y desarrollar soluciones mantenibles para que sean sostenibles en el tiempo.

- Diseñar y desarrollar soluciones escalables que permitan crecer de forma ordenada el negocio manteniendo la fiabilidad, disponibilidad y la mantenibilidad del sistema.

Los resultados presentados en este proyecto de tesis se centran en un caso de uso de autenticación y verificación multi-biométrica (facial, voz, typing...) de usuarios online basada en tecnologías de AP y desarrollada para la industria de aplicaciones y servicios online.

1.4.3 Sistemas embebidos de percepción del entorno basado en visión e inteligencia artificial con nivel de integridad de seguridad crítica para la industria de sistemas inteligentes para transporte autónomo

Esta línea de investigación se centra en el diseño y desarrollo de sistemas de percepción del entorno basados en visión e inteligencia artificial para la conducción autónoma. Dicho sistema, aparte de funcionar con precisión y con aceptable rendimiento bajo diferentes condiciones de iluminación y estados meteorológicos (día, noche, con lluvia, con nieve, con niebla...), debe garantizar la disponibilidad, la mantenibilidad y la escalabilidad (FDME). Además, han de poder ser ejecutados en sistema embebidos de bajo consumo y ser introducidos en cadenas funcionales con Nivel de Integridad de Seguridad. Los objetivos concretos son:

Obj 3.1 Definición y desarrollo funcional de aplicaciones embebidas de percepción del entorno multi-sensor para conducción autónoma como:

- Detección de obstáculos: detección de objetos que puedan obstaculizar la circulación del vehículo autónomo.
- Detección de señalización lateral: detección de límites de velocidad, semáforos o autoridad de movimiento.
- Odometría visual: posicionamiento y odometría del vehículo basado en información de sensores visuales.
- Estimación del camino: detección del futuro recorrido del vehículo para poder determinar la aplicabilidad de los objetos detectados como posibles obstáculos o señalización lateral.

Obj 3.2 Desarrollo de sistemas de Validación y Verificación (V&V) semi-automática que permitan la simulación de todos los escenarios posibles en entornos virtuales. Especialmente aquellos escenarios que no son usuales (y cuesta tener datos reales para entrenamiento), pero críticos desde el punto de vista de la seguridad (atropellos, condiciones climáticas extremas...). Este tipo de herramientas podrá reducir los costes de comercialización de los sistemas de percepción seguros basados en visión e inteligencia artificial, evitando importantes barreras iniciales.

Obj 3.3 Desarrollo de arquitecturas HW para sistemas embebidos NIS que permita la portabilidad de las aplicaciones con alto coste computacional (por los tiempos de inferencia de los modelos AP) a sistemas embebidos con aceleración HW, aislamiento entre las operaciones cruciales, gestión eficaz y segura los recursos del sistema y ejecuciones redundadas con botadores para toma de decisión.

Obj 3.4 Definición de estándares para la certificación de sistemas de percepción basados en aprendizaje profundo y que son inexistentes actualmente en sectores como el ferroviario.

Los resultados presentados en este proyecto de tesis se centran en el sector ferroviario y en concreto en el campo del futuro vehículo (tren, tranvía, alta velocidad...) autónomo, donde se pretende obtener un módulo de percepción exterior basado en técnicas de visión e inteligencia artificial. Dicho módulo tendrá que ser certificable y podrá ser parte de las funcionalidades críticas, con nivel de integridad de seguridad, acorde con la férrea normativa del sector que actualmente las prohíbe.

1.5 Publicaciones principales

Esta tesis se apoya en 6 publicaciones principales que contribuyen a las diferentes líneas de investigación y objetivos definidos en el apartado anterior (ver Tabla 1.1). En 5 de ellas, el autor de esta tesis se presenta como autor principal y en la restante como co-autor. 5 de ellas ya han sido editadas y publicadas en revistas de primer (Q1) o segundo cuartil (Q2). La restante ha sido enviada a una revista de primer cuartil y está en proceso de revisión a la espera de una respuesta positiva. Para información más detallada ver la Sección 4.

[Lab+14] **Mikel Labayen** and Igor G. Olaizola and Naiara Aginako and Julián Flórez (2014). *Accurate Object Tracking and 3D Visualization of Controversial Plays in Sports Events Broadcast*. Multimedia Tools and Applications. (ver Sección 4.1)

Abstract: The application of computer-aided controversial play resolution in sport events significantly benefits organizers, referees and audience. Nowadays, especially in ball sports, very accurate technological solutions can be found. The main drawback of these systems is the high rent expenses which makes them not affordable for less-known regional/traditional sports events. The lack of competitive systems with reduced hardware/software complexity and requirements motivates this research. Using Visual Analytics technologies the system detects the trajectory of balls, permitting to solve with precision possible controversial plays. Ball is extracted from the video scene using its colour and shape characteristics and velocity vector properties. Afterwards, its relative position to border line is computed based on polynomial approximations. Remark the necessity of a unique camera even for 3D information extraction. In order to enhance user visual experience, real-time rendering technologies are introduced to obtain virtual 3D reconstruction in quasi real-time. Testing of the system has been done in real scenarios, comparing the system output with referees' judgement. Results of the system have been broadcasted during Basque Pelota matches.

[Lab+21] **Mikel Labayen** and Ricardo Vea and Julián Flórez and Naiara Aginako and Basilio Sierra (2021). *Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology*. IEEE Access. (ver Sección 4.2)

Abstract: Identity verification and proctoring of online students are one of the key challenges to online learning today. Especially for online certification and accreditation, the training organizations need to verify that the online students who completed the learning process and received the academic credits are those who registered for the courses. Furthermore, they need to ensure that these students complete all the activities of online training without cheating or inappropriate behaviours. The COVID-19 pandemic has accelerated (abruptly in certain cases) the migration and implementation of online education strategies and consequently the need for safe mechanisms to authenticate and proctor online students. Nowadays, there are several technologies with different grades of automation. In this paper, we deeply describe a specific solution based on the authentication of different biometric technologies and an automatic proctoring system (system workflow as well as AI algorithms), which incorporates features to solve the main concerns in the market: highly scalable, automatic, affordable, with few hardware and software requirements for the user, reliable and passive for the student. Finally, the technological performance test of the large scale system, the usability-privacy perception survey of the user and their results are discussed in this work.

[Etx+22a] Mikel Etxeberria and Maider Zamalloa and Nestor Arana-Arexolaleiba and **Mikel Labayen** (2022). *Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case*. IEEE Access. (ver Sección 4.3)

Abstract: Localization is one of the most critical tasks for an autonomous vehicle, as position information is required to understand its surroundings and move accordingly. Visual Odometry (VO) has shown promising results in the last years. However, VO algorithms are usually evaluated in outdoor street scenarios and do not consider underground railway scenarios, with low lighting conditions in tunnels and significant lighting changes between tunnels and railway platforms. Besides, there is a lack of GPS, and it is not easy to access such infrastructures. This research proposes a method to create a ground truth of images and poses in underground railway scenarios. Second, the EnlightenGAN algorithm is proposed to face challenging lighting conditions, which can be coupled with any state-of-the-art VO techniques. Finally, the obtained ground truth and the EnlightenGAN have been tested in a real scenario. Two different VO approaches have been used: ORB-SLAM2 and DF-VO. The results show that the EnlightenGAN enhancement improves the performance of both approaches.

[Lab+23c] **Mikel Labayen** and Xabier Mendialdua and Naiara Aginako and Basilio Sierra (2023). *Semi-Automatic Validation and Verification Framework for CV&AI-enhanced Railway Signalling and Landmark Detector*. IEEE Transactions on Instrumentation and Measurement. (ver Sección 4.4)

Abstract: The automation of railway operations is an activity in constant growth. Different railway stakeholders are already developing their research activities for the future driverless autonomous driving based on Computer Vision (CV) and Artificial Intelligence (AI) enhanced perception technologies (e.g., obstacle detection). Unfortunately, the AI models are opaque in nature and here is no certification accepted rules for CV&AI-enhanced functionality certification. To meet the increasing needs of trusted CV&AI-based solutions, numerous Validation and Verification (V&V) approaches have been proposed in other sectors like automotive, most of the based on virtual simulators. Unfortunately, there is currently no virtual perception simulator for railway scenario. Capturing and labelling camera image in real environment is expensive in terms of time and resources. In addition, these data gathering sessions do not differ enough in meteorological or lighting conditions, which makes resulting database less valuable for the V&V processes, as they are very similar. This work aims to create an semi-automatic system based on virtual scenarios taking advantage of the scenario design parameterisation possibilities of train driving videogames (by exploiting their virtual cameras as perception sensors) and measuring the CV&AI-enhanced system performance based on the global accuracy metrics and detected potential safety and operation rules violations. This work also demonstrates the quantitative (in number of test carried out) and qualitative (as the diversity of the created scenarios, hard to replicate in real environment) improvements while reducing current V&V cost.

[Lab+23b] **Mikel Labayen** and Laura Medina and Fernando Eizaguirre and José Flich and Naiara Aginako (2023). *HPC Platform for Railway Safety-Critical Functionalities based on Artificial Intelligence*. Applied Sciences. (ver Sección 4.5)

Abstract: The automation of railroad operations is a rapidly growing industry. In 2023, a new European standard for the automated Grade of Automation (GoA) 2 over European Train Control System (ETCS) driving is anticipated. Meanwhile, other railway stakeholders are already planning their research initiatives for the following stage: unattended and driverless autonomous driving. As a result, the industry is particularly active in research and proofs of concept regarding perception technologies based on Computer Vision (CV) and Artificial Intelligence (AI), with outstanding results at the application level. However, executing high-performance and safety-critical applications on embedded systems and in real-time is a challenge for the railway industry, just like it is for the automotive industry. There aren't many commercially available solutions since High-Performance Computing (HPC) platforms are typically seen as being beyond the price range of the safety-critical systems business. This work proposes a novel safety-critical and high-performance computing platform for CV&AI-enhanced technology execution used for Automatic accurate stopping and Safe passenger transfer railway functionalities. The resulting computing platform is compatible with the majority of widely used AI inference methodologies, AI model architectures, and AI model formats thanks to its design, which enables process separation, redundant execution, and HW acceleration in a transparent manner. The innovation introduced in this work is related with the creation of a hardware accelerator module, an AI runtime, and a modified inference SW. The proposed technology increases the portability of railway applications into embedded systems, isolates crucial operations, and effectively and securely maintains system resources.

[Lab+23a] **Mikel Labayen** and Daniel Ochoa de Eribe and Ander Aramburu and Marcos Nieto and Naiara Aginako (2023). *European Common Data Management Platform Definition for Railway AI Function Development*. Transportation Research Part C: Emerging Technologies. ESTADO: bajo revisión, pendiente de aceptación. (ver Sección 4.6)

Abstract: Digitalisation and automation of operations in the railway industry includes the use of Automatic Train Operation systems that provide automated functions to reach different levels of automation, known as the Grade of Automation (GoA) levels. Artificial Intelligence has emerged as technology that can substitute humans in certain driving tasks, in GoA3 (driverless) and GoA4 (unattended) modes. AI capabilities include perception, decision-making, precise positioning, or optimization of communications. The success of AI models depends on the quality and diversity of the data used for training, along with the set-up of a data life-cycle framework that covers creation, training, testing, deployment and monitorisation. The management of training datasets implies both expensive and time-consuming data gathering, labelling, curation and formatting efforts, potentially hindering the development of reliable AI systems. This paper presents a Common Data Management Platform developed by a consortium of European railway stakeholders, devised to efficiently manage data for AI training, and which is demonstrated in two different Proofs of Concept.

Tab. 1.1: Lista de las principales publicaciones que apoyan la tesis junto con el tipo de publicación, la posición del autor y los objetivos abordados.

Publicación	Tipo	Pos. Autor	Objetivos
[Lab+14]	Journal Q2	1º Autor	Obj1.1, Obj1.2
[Lab+21]	Journal Q2	1º Autor	Obj2.1, Obj2.2
[Etx+22a]	Journal Q2	Co-Autor	Obj3.1
[Lab+23c]	Journal Q1	1º Autor	Obj3.2
[Lab+23b]	Journal Q2	1º Autor	Obj3.3
[Lab+23a]	Journal Q1	1º Autor	Obj3.4

1.6 Patentes principales

Este proyecto de tesis también se apoya en 3 patentes principales. Dichas patentes registran los inventos creados fruto de los resultados de las distintas investigaciones (ver Tabla 1.2). Dos de ellas, [Ric+17] y [MRM21], constituyen una herramienta esencial que garantiza la propiedad intelectual de las invenciones y protegen la actividad comercial de un servicio en más de 7 países internacionales. Para información más detallada ver la Sección 5.

[Igo+13] García Olaizola Igor and Flórez Esnal Julián and San Román Otegui Juan Carlos and Aginako Bengoa Naiara and **Labayen Esnaola Mikel** (2013). *Method for Detecting the Point of Impact of a Ball in Sports Events*

Description: The invention relates to a method for determining the point of impact of a ball in a playing field during a controversial piece of play in a sports event and comprises the steps of: recording the contentious area during the game by means of a single camera, extracting the images corresponding to the controversial piece of play, selecting the area corresponding to the ball, calculating the coordinates of the ball in pixels in each image, determining the point of intersection of the two straight lines joining the previous points and transforming the point of intersection into real coordinates. As a result of these steps, it is possible to resolve the controversial piece of play with a single camera.

[Ric+17] Vea Orte Ricardo and **Labayen Esnaola Mikel** and Flórez Esnal Julián and Marcos Ortego Gorka (2017). *Method and System for Verifying the Identity of a User of an Online Service*

Description: The invention relates to a method for verifying the identity of a user of an online service, with the steps of: when a user is connected to an online service, sending an IP address of an authentication server; connecting to said IP address and downloading one application for taking photos with the webcam of the user terminal; taking a photo; sending said photo and associated metadata to a management unit; storing it in a data base; automatically extracting one set of biometrical parameters per each face which appears in said photo; comparing said set of biometrical parameters with a reference biometrical model of the user to which said user ID belongs; if the result of said comparison does not unequivocally match the person in the photo with the user to which said user ID belongs, either informing the web service provider or sending said photo to a manual recognition unit for manual validation of the photo; continuously verifying the identity of the user connected to the online service through said user terminal.

[MRM21] Labayen Esnaola Mikel and Vea Orte Ricardo and Fraile Yarza Manuel (2021). *Method and System for Verifying the Identity of a User of An Online Service Using Multi-Biometric Data*

Description: The invention relates to a method for verifying the identity of a user of an online service, comprising: when a user is connected to an online service from a user terminal, establishing a connection with a biometric data collecting module; downloading one application for taking photos with the webcam of the user terminal and at least one application for capturing audio or for extracting keystroke pattern; while a session with the online service is active: taking a photo of the user terminal; sending each photo and associated metadata to a management module; automatically extracting features of each face which appears in said photo; comparing said features with a biometrical model of a reference photo of the user; repeating the former steps, thus continuously verifying the identity of the user connected to the online service; using a second biometrical technique for verifying the user identity, said second biometric technique being either sound recognition or keystroke pattern analysis. If voice is detected or a keystroke pattern is extracted, voice/keystroke features are extracted and compared with a reference model. If the result of a comparison does not unequivocally match the person in the photo/audio clip/keystroke pattern with the registered user, either informing the service provider or sending the photo/audio clip for manual validation. Besides, if sound or keystroke pattern recognition is used, another photo may be automatically taken when voice or keystroke pattern is detected and sent to the management module.

Tab. 1.2: Lista de las principales patentes que apoyan la tesis junto con el tipo de patente, la posición del inventor y los objetivos abordados.

Patente	Tipo	Pos. Autor	Objetivos
[Igo+13]	Internacional	Co-Inventor	Obj1.1, Obj1.2
[Ric+17]	Internacional	Co-Inventor	Obj2.1, Obj2.2
[MRM21]	Internacional	1º Inventor	Obj2.1, Obj2.2

1.7 Proyectos I+D

El trabajo de investigación presentado en este documento se sustenta en diferentes proyectos de investigación realizados en las distintas empresas y centro de investigación. La aplicabilidad en la industria de la mayoría de los proyectos ha influido en las decisiones tomadas durante el proceso de investigación, así como en los resultados obtenidos. En las siguientes líneas se describen los proyectos más importantes. Además de su descripción y objetivos, también se indican los resultados obtenidos y su relación con el trabajo de investigación (ver Tabla. 1.3). Para terminar, se enumeran todos los proyectos que han aportado resultados útiles (aunque menos significativos) que han ayudado a alcanzar los objetivos planteados en este trabajo.

1.7.1 FP2 - R2DATO

Este proyecto tiene como objeto aprovechar las ventajas de la digitalización y la automatización para desarrollar la próxima generación de Automatic Train Control (ATC) y ofrecer así servicios escalables de explotación digital y automática (hasta autónoma) de trenes mejorando la capacidad de las redes ferroviarias y satisfaciendo la creciente demanda de transporte tanto de pasajeros como de mercancías.

- **Título**: Europe's Rail Flagship Project 2 - Rail to Digital Automated up to autonomous Train Operation
- **Financiado por (contribución) / Código:** UE (53,9M€) / 101102001
- **Fecha:** 2022-2026

Objetivos
Se espera obtener resultados tangibles para 2025 en temas clave como: Automatic Train Operation (ATO), European Train Control System (ETCS) híbrido L3, bloque móvil L3, tecnologías digitales (conectividad 5G...), y directrices y métodos para un despliegue rápido y rentable de esta tecnología en toda Europa. La metodología general se articula en torno a tres etapas: desde el desarrollo de habilitadores técnicos, pasando por prototipos funcionales y terminando en demostradores dedicados con nivel TRL 6/7. A través de estas mejoras técnicas, FP2-R2DATO cumplirá los objetivos e impactos definidos en el Plan Director y el Programa de Trabajo Anual de la Europe's Rail Joint Undertaking (ERJU): contribuir a aumentar la puntualidad, la fiabilidad y la productividad del personal, el material rodante y la infraestructura.

Resultados obtenidos
Se trata de un proyecto en curso que comenzó en diciembre de 2022. Aún no hay resultados exportables.

1.7.2 THAVA

Actualmente, en entornos cerrados como pueden ser las líneas de metro, se están incluyendo funcionalidades de automatización ferroviaria. El uso de estos sistemas de automatización, debido a las estrictas normativas que existen, no pueden aplicarse del mismo modo a líneas de perímetro abierto, como los tranvías y las vías interurbanas, de alta velocidad y largo recorrido. En estos escenarios el número de factores que pueden afectar a la seguridad aumenta exponencialmente y los sistemas de automatización actuales no pueden garantizar la seguridad; por este motivo es necesario investigar en este área.

- **Título**: Tecnologías y Herramientas Avanzadas para Vehículos Autónomos del sector ferroviario
- **Financiado por (presupuesto):** EUS (HAZITEK) (5,2M€)
- **Fecha:** 2022-2023

Objetivos
El objetivo general de este proyecto es investigar tecnologías habilitadoras (hardware y software) de percepción, posicionamiento y toma de decisiones automatizada que sentarán las bases y serán pieza clave en el futuro desarrollo de sistemas de automatización con niveles Grade of Automation (GoA) 3 y GoA4 de vehículos ferroviarios en líneas de perímetro abierto y Mainlines.

Resultados obtenidos
A la espera de que termine el *retrofit* de dos unidades de trenes donde se probará el sistema en vía, el primer año de proyecto se ha desarrollado un sistema de percepción y toma de decisión para la conducción autónoma en perímetro exterior. Entre otras cosas, dicho sistema posibilita las funciones de Previsión de Colisiones y Parada de Precisión en Plataforma basados en un sistema multi-sensor (cámaras multi-espectrales y LiDAR). El sistema está siendo validado en laboratorio, tanto el comportamiento del tren incluyendo la toma de decisión, como la capacidad de detección del sistema de percepción en entorno virtual (simulado) garantizando su funcionamiento en diversas condiciones de visibilidad.

1.7.3 TAURO

El proyecto TAURO pretende desarrollar las tecnologías necesarias para que el transporte ferroviario autónomo se haga realidad en Europa. Pretende trabajar en sistemas para la percepción del entorno, la conducción remota, la supervisión y el diagnóstico automáticos. Para ello, TAURO se divide en cuatro grupos de trabajo técnicos: 1) Percepción del entorno para la automatización, 2) Conducción y control remoto, 3) Supervisión automática del estado y diagnóstico para trenes autónomos y 4) Tecnologías de apoyo a la migración a ATO sobre ETCS.

- **Título**: Technologies for Autonomous Rail Operation
- **Financiado por (contribución) / Código:** UE (1,9M€) / 101014984
- **Fecha:** 2020-2023

Objetivos
Desarrollar el futuro del transporte ferroviario autónomo europeo: identificar, analizar y, finalmente, proponer tecnologías de base adecuadas, que se seguirán desarrollando, certificando y desplegando a través de las actividades previstas para la Asociación Europea para la Transformación del Sistema Ferroviario Europeo.

Resultados obtenidos
Se han definido las bases (con requisitos concretos y propuestas de arquitectura) para una base de datos común en la industria ferroviaria con el cual poder entrenar y testear los futuros modelos de inteligencia artificial necesarias en el módulo de percepción interno/externo del tren autónomo. El proyecto también termina con una propuesta inicial sobre la futura normativa que regulará el uso de la IA en el sector. Por otro lado, después de identificar un amplio conjunto de requisitos funcionales y no funcionales, el proyecto propone una propuesta conjunta para la conducción remota de vehículos ferroviarios. Por último, el proyecto ha definido las funcionalidades de supervisión automática del estado y diagnóstico para trenes autónomos, y además, ha especificado el primer borrador de los módulos SW necesarios para que un ATO pueda operar sobre sistemas *legacy* (bajo la supervisión de un conductor).

1.7.4 FRACTAL

El valor de las redes fractales complejas radica principalmente en la complejidad que representan con una simplicidad única, al mismo tiempo que conservan las características principales del sistema mediante nodos y las interacciones entre ellos. El proyecto FRACTAL está desarrollando un nodo de computación sobre el cual basar una red fractal cognitiva capaz de aprender del entorno y de responder ante este. Ayudará a la interacción continua, rápida y fiable entre el mundo físico y la nube en aplicaciones que van desde los vehículos autónomos a la medicina remota.

- **Título**: A Cognitive Fractal and Secure EDGE based on an unique Open-Safe-Reliable-Low Power Hardware Platform Node
- **Financiado por (presupuesto) / Código:** UE (15,7M€)/ 877056
- **Fecha:** 2020-2023

Objetivos

El objetivo es crear un nodo informático fiable que cree un Edge Cognitivo bajo los estándares de la industria. Este nodo de computación será el bloque de construcción del internet de las cosas. La capacidad cognitiva vendrá dada por una arquitectura interna y externa que permita prever su rendimiento interno y el estado del mundo circundante. Por lo tanto, este nodo tendrá la capacidad de aprender a mejorar su rendimiento frente a la incertidumbre del entorno. Esta red compleja transferirá todas esas ventajas cognitivas al Edge, un paradigma informático que se sitúa entre el mundo físico y la nube.

Resultados obtenidos

Se ha llevado a cabo el estudio y análisis de trabajos relacionados en las áreas de consumo de energía en circuitos digitales, ordenadores de alto rendimiento y bajo consumo, técnicas embebidas de *power gating*, arquitecturas de bajo consumo, Dynamic Voltage and Frequency Scaling (DVFS) adaptativo para Multi-Procesador System on a Chip (MPSoC), técnicas de compresión y descompresión de datos, técnicas de re-mapeo estático y dinámico (metascheduling). Se ha diseñado el sistema FRACTAL que consta de múltiples nodos y se ha establecido una arquitectura de plataforma para sus servicios. Una arquitectura multi-núcleo jerárquica adaptativa activada por tiempo para proporcionar servicios de bajo consumo y seguridad. También se ha implementado un planificador estático para un sistema Network on a Chip (NoC) basado en un algoritmo genético y un meta-cronificador para un sistema NoC basado en el tiempo. Por último, se ha establecido la arquitectura y el concepto de servicio planificados. La plataforma será validad por los casos de uso a finales del 2023.

1.7.5 VALU3S

Los fallos en los sistemas altamente automatizados pueden ser catastróficos. Al ser cada vez más complejos y ofrecer más conectividad, pueden surgir propiedades desconocidas de los sistemas, por lo que es esencial realizar una Verificación y Validación (V&V) exhaustivas. La elevada complejidad del proceso de V&V hace que sea caro y requiera tiempo. El proyecto VALU3S evalúa los métodos y herramientas de V&V para los sistemas automatizados del sector automovilístico, agrícola, ferroviario, sanitario, aeroespacial, de la automatización industrial y la robótica.

- **Título**: Verification and Validation of Automated Systems' Safety and Security
- **Financiado por (presupuesto) / Código:** UE (25,7M€) / 876852
- **Fecha:** 2020-2023

Objetivos

El objetivo de VALU3S es diseñar, implantar y evaluar métodos y herramientas de V&V de última generación con el fin de reducir el tiempo y el coste necesarios para verificar y validar los sistemas automatizados con respecto a los requisitos de Seguridad, Ciberseguridad y Privacidad (SCP). Este proceso se lleva a cabo a través de la identificación y clasificación de los métodos de evaluación, herramientas, entornos y conceptos que son necesarios para verificar y validar los sistemas automatizados con respecto a los requisitos de SCP. En el proyecto también se evaluarán los métodos de verificación y verificación aplicados, así como los flujos de trabajo y herramientas de proceso mejorados, mediante un amplio conjunto de demostradores.

Resultados obtenidos

El proyecto ha definido en detalle 57 escenarios de evaluación, 192 casos de prueba y 239 requisitos que se han utilizado para evaluar los casos de uso. También ha recopilado la descripción detallada de 43 métodos de V&V, 41 mejoras de métodos de V&V y 7 nuevos métodos combinados. Sin embargo, los outputs más tangibles son la creación de una solución personalizada para el modelado de flujos de trabajo. Por otra parte, ha recopilado y presentado 15 criterios de evaluación para vulnerabilidades de ciberseguridad y 13 criterios para medir los procesos de V&V. En general, el proyecto ha aportado el diseño, implementación y mejora de un amplio conjunto de herramientas de V&V y sus correspondientes criterios de de validación.

1.7.6 SELENE

La inteligencia artificial impulsa la informática de alto rendimiento utilizada en sistemas de seguridad crítica. Estos sistemas utilizan componentes comerciales de venta al público que ofrecen una vía alternativa para aumentar la capacidad computacional de las aplicaciones de seguridad crítica. Sin embargo, a pesar de su potencial en varios ámbitos, el uso de estos sistemas es limitado debido a la falta de plataformas HW certificadas y fiables.

- **Título**: SELENE: Self-monitored Dependable platform for High-Performance Safety-Critical Systems
- **Financiado por (presupuesto) / Código:** UE (4,99M€) / 871467
- **Fecha:** 2019-2022

Objetivos

SELENE sigue un planteamiento radicalmente nuevo y propone una Plataforma de Computación Cognitiva (CCP) para seguridad crítica con capacidades de autoconocimiento, auto-adaptación y autonomía. La CCP de SELENE utiliza técnicas de IA para adaptar el sistema a las condiciones particulares internas y externas (del entorno) con el objetivo de maximizar la eficiencia del sistema siendo capaz al mismo tiempo de cumplir los requisitos de la aplicación. Las técnicas de IA se alimentan con la información proporcionada por los monitores en línea y los sensores externos y se aplican de forma transparente sin comprometer la seguridad del sistema. Para garantizar que se preservan los requisitos de seguridad, el CCP de SELENE se basa en las sólidas capacidades de aislamiento proporcionadas a nivel de HW y SW.

Resultados obtenidos

Respecto al HW, se ha configurado un SoC SELENE basado en RISC-V de referencia. Este SoC soporta actualmente seis núcleos NOEL-V RISC-V de 64 bits que se han mejorado para que admitan virtualización y extensiones de instrucciones comprimidas. El soporte inicial de HW para la seguridad se ha centrado en la supervisión de la contención multi-núcleo, el soporte de diversas ejecuciones redundantes y la adaptación de una metodología de verificación de inyección de fallos al SoC SELENE. Respecto al SW, su arquitectura se asemeja a la arquitectura de software del proyecto SIL2LinuxMP con objeto de asegurar la seguridad. Esta arquitectura de referencia se ha ampliado con interfaces adecuadas para el conjunto de elementos de monitorización incluidos en el HW de SELENE. La arquitectura básica también se ha ampliado con el hipervisor Jailhouse para permitir la arquitectura de sistemas de criticidad mixta en un multi-núcleo básico. La plataforma SELENE ha sido validad por 4 casos de uso después de asegurar la portabilidad de sus aplicaciones de software.

1.7.7 X2Rail-4

X2RAIL-4 es el cuarto proyecto de los miembros de Shift2Rail en el ámbito IP2 'Sistemas avanzados de gestión y control del tráfico' y su objetivo es continuar la investigación y el desarrollo de tecnologías para: a) aumento de la capacidad de las líneas ferroviarias (ATO hasta GoA4), b) integridad del tren. Definir, desarrollar y probar prototipos completos de sistemas de integridad del tren que puedan aplicarse a todas las tipologías de trenes, c) evolución de la gestión del tráfico, y d) objetos conectados por radio.

- **Título**: Advanced signalling and automation system - Completion of activities for enhanced automation systems, train integrity, traffic management evolution and smart object controllers
- **Financiado por (contribución) / Código:** UE (17,9M€) / 881806
- **Fecha:** 2019-2023

Objetivos

En este proyecto se desarrollarán prototipos embarcados que garanticen la integridad del tren (con detector de obstáculos integrado) y que serán aprovechables en trenes de pasajeros y mercancías. Se diseñará además una plataforma para la gestión de las comunicaciones y el intercambio de datos entre distintos servicios. Por último, se creará un prototipo inalámbrico de intercambio de datos entre los equipos en la vía y los instrumentos de señalización que presente menos costes de mantenimiento y precise menos cables.

Resultados obtenidos
Se ha aprobado la especificación del ATO GoA2 con especificación de requisitos funcionales y de sistema. Dicha especificación de los requisitos del sistema (excepto el SUBSET-140) ha formado parte de la TSI2022.También se ha terminado con especificación del ATO üp to GoA4"(GoA3 y GoA4) incluyendo el TMS para la gestión de tráfico. Algunos de los socios han iniciado el desarrollo de prototipos de 'Detección de obstáculos' y 'Sensores de entorno' dedicados a la primera experimentación. También se ha definido la arquitectura estándar para los bancos de pruebas de referencia de GoA3/4 que se llevaran a cabo a finales del 2023.

1.7.8 SMOWL

SMOWL (producto/servicio) es una solución práctica y fiable para la identificación y supervisión de usuarios en línea. Consiste en un nuevo servicio de ciberseguridad que cubre la necesidad de una autenticación continua, automática y escalable de la identidad y el seguimiento de los usuarios en línea. La tecnología, basada en el reconocimiento automático y continuo del rostro, la voz y las pulsaciones del teclado, puede aplicarse a numerosos sectores.

- **Título**: A continuous identity authentication monitoring system for online services using automatic face, voice and keystroke recognition ↗
- **Financiado por (presupuesto) / Código:** UE (840K€) / 811319
- **Fecha:** 2019-2023

Objetivos
El principal objetivo de SMOWL (proyecto) es ofrecer al mercado mundial del e-learning una solución fiable y barata que permita la autenticación continua de los usuarios/estudiantes en línea y evite las usurpaciones de identidad durante las sesiones de aprendizaje y los exámenes online. Sus objetivos principales son: a) implementar el plan de negocio y desarrollar una amplia introducción en el mercado, b) realizar grandes pruebas pilotos de demostración (más de 10.000 usuarios), c) optimizar, ampliar y personalizar las funciones de SMOWL (producto/servicio), y d) crear un departamento de soporte técnico continuo para la autenticación 100 % fiable de los alumnos.

Resultados obtenidos
Después de centrarse inicialmente en mercados clave, se ha terminado por desarrollar un plan de negocio con una amplia introducción en el mercado mundial. Se han realizado grandes demostraciones piloto con las universidades/instituciones/empresas de e-learning más relevantes de diferentes países internacionales de todo el mundo. Además, se ha mejorado el sistema al añadir algunas otras características como el reconocimiento de voz y pulsaciones de teclas. También se le ha añadido un sistema de control o bloqueo del ordenador (característica clave para la vigilancia) para aumentar la oferta y ser más competitivos en el servicio de proctoring de exámenes.

1.7.9 MULTIBIO

En los últimos años, los campus virtuales y plataformas e-learning están incrementando exponencialmente su número de alumnos/usuarios y uno de los mayores problemas por parte de la entidad que ofrece los cursos, es la incertidumbre de no saber con certeza quién está cursando la materia al otro lado del ordenador y poder así certificar o no el título que recibirá el alumno.

- **Título**: Sistema biométrico multimodal para autenticación continua de usuarios online ↗
- **Financiado por (presupuesto) / Código:** ES (Retos-Colaboración) (500K€) / RTC-2016-5711-7
- **Fecha:** 2019-2023

Objetivos
El objetivo principal del proyecto es por tanto crear el primer prototipo comercial MULTIBIO, un producto/servicio totalmente nuevo, independiente de SMOWL (su predecesor) y más completo. Este nuevo servicio MULTIBIO, será un servicio de autenticación automático, pasivo (el usuario no participa activamente, como por ejemplo posando ante la cámara) y continuado de usuarios de servicios online que requieren un mayor grado de seguridad en la identificación de sus usuarios. Dicha autenticación se realizará mediante una identificación facial y de voz automática, masiva, eficaz y en entornos no controlados sobre imágenes y clips de audio capturadas mediante la webcam y a través del micrófono de las terminales. También capturará *screenshots* para vigilar qué es lo que pasa en la sesión iniciada.

Resultados obtenidos
Se ha creado un servicio continuo de autenticación de identidad del estudiante online a través de un sistema de reconocimiento biométrico continuado (cara, voz, escritura) y un sistema continuo de monitorización online. Los resultados técnicos

muestran que las soluciones de autenticación biométrica y supervisión totalmente automatizadas, continuas (no programadas), pasivas (para los estudiantes), escalables, totalmente integradas en el Learning Management System (LMS), seguras y privadas son asequibles y fiables. También se ha realizado un estudio basado en encuestas a usuarios del sistema. Los resultados han aportado datos cualitativos y cuantitativos que apoyan el uso de este tipo de software en la educación a distancia.

1.7.10 BEGIRA

La extracción de información a partir del análisis de imágenes se está haciendo cada vez más importante dada la relevancia de los datos extraídos. En el caso de retransmisiones deportivas, el propio mundo del deporte demanda soluciones tecnológicas avanzadas para el análisis de jugadas dudosas. Además, los radiodifusores apuestan por estos datos al ser una de las maneras de atraer a los usuarios. Sin embargo, los costes de estos sistemas no son asumibles por eventos deportivos de menor envergadura. Por ello, se pretende desarrollar un sistema que ofrezca servicios al alcance por todos los radiodifusores de distintos eventos deportivos.

- **Título**: Diseño y desarrollo de un sistema seguimiento preciso de objetos en transmisiones deportivas.
- **Financiado por / Código:** EUS (G93 Telecomunicaciones - GAITEK)
- **Fecha:** 2009-2011

Objetivos
El objetivo es crear un sistema de una sola cámara que muestre virtualmente la trayectoria de la pelota y el punto donde ha botado durante los partidos de pelota a mano televisados en directo. Para ello se debe procesar el vídeo en tiempo real para poder determinar la ubicación de la pelota en cada imagen del vídeo. Los algoritmos que se encuentran en el procesamiento de imágenes se utilizan tanto para encontrar la pelota en el vídeo como para trackearla a través de los consecutivos frames.

Resultados obtenidos
Los resultados obtenidos en el proyecto son indicativos de las utilidades que se puede lograr con un simple procesamiento de imágenes a bajo nivel. Tras analizar cada imagen, se ha construido un sistema que es capaz de segmentar la pelota, seguir su trayectoria a través de las imágenes de vídeo, encontrar la línea de tierra e identificar el punto de bote de la pelota. Además, todo el procesamiento se realiza en tiempo real. Cabe señalar que los resultados obtenidos en el proyecto se emitieron durante las retransmisiones en directo de los partidos de pelota a mano de la Radio Televisión Vasca (EITB), que el sistema fue publicado en una revista de referencia y que también está protegido por una patente europea.

Tab. 1.3: Lista de los principales proyectos de investigación que apoyan la tesis junto con el tipo de financiación, el rol del tesitando y los objetivos abordados.

Proyecto	Financiación	Rol	Inicio	Objetivos
1.7.10 BEGIRA	EUS (GAITEK)	Invest. Principal	2009	Obj1.1, Obj1.2
1.7.9 MULTIBIO	ES (RETOS)	Invest. Principal	2016	Obj2.1, Obj2.2
1.7.8 SMOWL	EU	Main Coord.	2018	Obj2.1, Obj2.2
1.7.7 X2Rail-4	EU	Participant	2019	Obj3.1
1.7.6 SELENE	EU	Task leader	2019	Obj3.3
1.7.5 VALU3S	EU	Task leader	2020	Obj3.2
1.7.3 TAURO	EU	WP & Task Leader	2020	Obj3.1, Obj3.4
1.7.2 THAVA	EUS (HAZITEK)	Líder de Tarea	2022	Obj3.1
1.7.1 R2DATO	EU	Participant	2022	Obj3.1

1.7.11 Otros proyectos

A continuación, se listas aquellos proyectos de menor relevancia, pero que han contribuido poco a poco a la consecución de los objetivos que recoge esta tesis.

SMOWL+

Sistema automático de reconocimiento facial para la autenticación continua de usuarios en servicios Online y entornos no controlados
GAITEK, Gobierno Vasco, 2014-2015.

Relación con los objetivos de la tesis: Obj2.1, Obj2.2

MICROMETEO

Plataforma web de información de meteorológica a nivel de micro zonas
Programa Innpacto, Ministerio de Ciencia e Innovación, 2010-2013.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

SEMANTS

Distribución inteligente de contenidos semánticos a través de interfaces multimodales
Programa AVANZA I+D, Ministerio de Industria, Turismo y Comercio, 2009-2011.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2, Obj2.1

ITACA3D

Plataforma de creación, producción y distribución de vídeo estereoscópico de entretenimiento para la visualización de televisión en 3D a través de redes broadcast
Programa AVANZA I+D, Ministerio de Industria, Turismo y Comercio, 2009-2011.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

360TV

Diseño y desarrollo de un sistema de producción audiovisual basado en el concepto de visión 360º
GAITEK, Digital Mobiles S.L., 2009-2010.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

SIAM

Diseño y desarrollo de un sistema de análisis multimedia de contenido audiovisual en plataformas web colaborativas
GAITEK, Hispavista, 2008-2010.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

ELISA

Entorno de localización inteligente para servicios asistidos
Proyectos Singulares Estratégicos, Ministerio de Industria, Turismo y Comercio, 2007-2010.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

SKEYE

Sistema de análisis meteorológico basado en imágenes del cielo tomadas desde tierra
GAITEK, Dominion Tecnologías, 2006-2008.

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

1.8 Estructura de la memoria

Este documento se divide en 5 capítulos, que se describen a continuación:

- **Introducción**: proporciona contexto sobre el trabajo de investigación realizado y expone las motivaciones, las hipótesis y los objetivos perseguidos. Además, enumera las principales publicaciones y patentes. Por último, describe los proyectos industriales en los cuales se ha enmarcado la investigación.

- **Resultados de la investigación**: presenta un análisis de los resultados y las aportaciones de cada publicación principal a las diferentes líneas de investigación y objetivos definidos en la introducción.

- **Conclusiones**: recoge las conclusiones extraídas del trabajo de investigación y se proponen nuevas ideas que servirán para continuar en el futuro con las líneas de investigación todavía abiertas y no resueltas.

- **Publicaciones**: lista las publicaciones principales en las que se basa este trabajo, así como otras publicaciones secundarias que también son relevantes en su campo.

- **Patentes**: enumera todas las patentes industriales que han resultado de alguna de las líneas de investigación de la tesis.

Por último, al final del documento se incluye la bibliografía y la lista de figuras y tablas.

Conclusiones 3

3.1 Discusión

Este proyecto de tesis expone sendas investigaciones realizadas en diferentes sectores industriales con objetivo de dar solución a los requerimientos de las aplicaciones comerciales basadas en visión e inteligencia artificial.

Para dar cumplimiento a los objetivos Obj1.1 y Obj1.2 se ha generado un prototipo funcional y accesible basado en visión artificial para la industria de contenidos multimedia. El prototipo consiste en un detector/tracker de pelota para estimar los botes cercanos a la línea de juego y un generador de repeticiones virtuales en 3D de las jugadas polémicas. El enfoque central del diseño de la solución ha sido reducir la complejidad del SW para obtener un rendimiento en tiempo real, así como reducir los requisitos de HW y su configuración a una única cámara para una región de juego reducida. El algoritmo introduce una ventaja adicional que lo hace robusto en términos de posibles diferentes ubicaciones de la cámara. El sensor y el plano del terreno de juego pueden formar cualquier ángulo, ya que las transformaciones necesarias para calcular el punto de impacto en coordenadas reales son eficaces y precisas. Los resultados obtenidos muestran que los errores de medida están por debajo del rango exigido (inferior a 1 cm en todos los casos e inferior a 5 mm en el 80 % de ellos), mejorando la precisión de un ojo humano.

Los objetivos Obj2.1 y Obj2.2, relativos al desarrollo de productos comerciales fiables con técnicas de visión e inteligencia artificial, se han alcanzado en forma de una aplicación para la industria de servicios online. Esta aplicación ofrece un servicio de autenticación continua de la identidad del estudiante online a través de un sistema de reconocimiento multi-biométrico constante (facial, voz y tipo de escritura) y un sistema continuo de vigilancia y seguimiento online. Para ello se incluyen técnicas de AP en un producto listo para entrar en producción y dar servicio a clientes de forma continua y sin cortes. Los resultados técnicos muestran que las soluciones de autenticación multi-biométrica y supervisión totalmente automatizadas, continuas, pasivas (para los estudiantes), escalables, totalmente integradas en el LMS, seguras y privadas, son asequibles y fiables. El sistema desarrollado actualmente es un servicio en producción que cumple con las características de fiabilidad, disponibilidad, mantenibilidad y escalabilidad y da servicio 24h/7d en países de todo el mundo.

Respecto a los objetivos Obj3.1 Obj3.2 Obj3.3 Obj3.4, todos ellos encaminados al desarrollo de productos embebidos de percepción del entorno basado en visión e inteligencia artificial, y con nivel de integridad de seguridad para la industria de la movilidad autónoma, este trabajo reúne cuatro investigaciones independientes, cada una con una contribución destacable en la consecución de cada uno de los objetivos.

Se ha presentado el desarrollo funcional de una aplicación embebida de percepción del entorno multi-sensor para posicionamiento y odometría visual, tal y como perseguía el objetivo Obj3.1. Los resultados de la investigación concluyen que los algoritmos de VO seleccionados, DF-VO y ORB-SLAM2, se pueden aplicar en escenarios ferroviarios subterráneos. Sin embargo, la precisión de la localización puede mejorarse aún más. El uso de técnicas de mejora del dataset para reducir el efecto de los desafíos perceptuales de los escenarios subterráneos también se ha estudiado en esta investigación. Se puede observar que la aplicación del algoritmo EnlightenGAN reduce los errores de ambos algoritmos VO. Además, también reduce la dispersión de los resultados del algoritmo ORB-SLAM2. No obstante, sigue siendo un problema inconcluso y este trabajo debe considerarse como un proceso de mejora. La aplicación de otros enfoques (otros algoritmos, otros sensores o fusión de datos), serán determinantes en el futuro, ya que el rendimiento puede mejorarse aún más para alcanzar la precisión que requieren algunas operaciones de tren autónomo en un escenario de metro urbano.

Por otra parte, el cumplimiento del objetivo Obj3.2, relativo al desarrollo de sistemas de validación y verificación semi-automática que permitan la simulación de todos los escenarios posibles en entornos virtuales, se ha materializado en la definición y desarrollo del primer framework semi-automático de V&V para la señalización ferroviaria. Se ha detallado el diseño y la implementación del framework, así como el sistema basado en VA&IA que se ha evaluado con él. Además, se han identificado y definido las métricas para comparar la cantidad y variedad de pruebas realizadas, y sus costes asociados, tanto del proceso de validación actual como del propuesto en este trabajo. Los resultados obtenidos en diferentes pruebas realizadas muestran un aumento significativo en la cantidad de pruebas que se pueden realizar al mismo tiempo para una validación más robusta del sistema. Además, aunque la mejora cuantitativa es considerable, el cambio más importante se produce en la parte cualitativa. Está claro que la variedad en los escenarios de las pruebas aumenta sustancialmente, ya que el entorno de generación de datos simulados permite simular una gran variedad de condiciones de visibilidad que sólo podrían conseguirse en grabaciones reales tras mucho tiempo y altos costes. Por último, el ahorro de recursos es otro punto clave a tener en cuenta, con una reducción significativa del esfuerzo para la misma carga de trabajo requerida.

Este proyecto de tesis también presenta una solución HW para sistemas embebidos de nivel de integridad de seguridad que permita la portabilidad de las aplicaciones con alto coste computacional a sistemas embebidos, dando así cumplimiento al objetivo Obj3.3. Esta nueva contribución ha introducido una nueva plataforma de computación embebida, de seguridad y alto rendimiento para aplicaciones ferroviarias basadas en VA&IA en tiempo real. Su diseño permite el aislamiento de procesos, la ejecución redundante, la aceleración HW y la abstracción, haciendo que la plataforma sea compatible con las diferentes técnicas de inferencia, arquitectura y formatos de modelos de IA (incluyendo estándares abiertos como ONNX). Cabe destacar la implementación de los diferentes módulos HW y SW específicos para la plataforma SELENE. Se ha desplegado un módulo acelerador HW, que puede personalizarse para soportar formatos de datos y capas de red neuronal específicos. El acelerador HLSinf muestra un gran rendimiento en la evaluación agnóstica de frecuencias. Utilizando la cuantización y otros métodos de optimización de modelos de IA, el rendimiento mejora el SoA. Además, se ha llevado a cabo un runtime de IA personalizado y un SW de inferencia adaptado, que abstrae la capa de aplicación del usuario de la configuración HW específica de la plataforma. La plataforma integrada añade flexibilidad para portar la solución al sistema integrado, aísla las funciones críticas y gestiona los recursos del sistema de forma segura y eficiente. Con una mayor madurez, la implementación ASIC y la certificación ferroviaria, la plataforma SELENE podría adaptarse a los requisitos de la industria ferroviaria tanto para aplicaciones de nivel de no seguridad como de seguridad.

Para terminar, en el camino hacia la consecución del objetivo Obj3.4 de este trabajo, la futura definición de estándares para la certificación de sistemas de percepción basados en aprendizaje profundo (inexistentes actualmente en el sector ferroviario), se ha definido la primera plataforma de gestión de datos para el entrenamiento, testeo y certificado de modelos IA. Aunque enfoques como el descrito en este trabajo se están convirtiendo en un estándar de facto en el sector de la automoción, el sector ferroviario todavía está evolucionando para adoptar metodologías centradas en la IA. En el momento de escribir este trabajo, sólo existen unos pocos conjuntos de datos abiertos destacables relacionados con la IA para el ámbito ferroviario. Este trabajo da respuesta a esta necesidad, diseñando y definiendo la primera plataforma común europea de gestión de datos ferroviarios acordada entre los más importantes *stakeholders* del sector. Teniendo en cuenta los resultados obtenidos en los PoC, se puede concluir que la solución en la nube es una mejor alternativa en comparación con una solución *in situ* para la construcción de la plataforma de gestión de datos en términos de disponibilidad, compartibilidad, escalabilidad y mantenibilidad.

3.2 Trabajos futuros

Este apartado presenta los posibles futuros trabajos separados por cada objetivo planteando en este proyecto de tesis.

Una vez validada la aplicación de detección y tracking de pelota en el juego de la pelota vasca, esta tecnología se está extendiendo a otros deportes que necesitan una distribución multi-cámara como es el tenis, ampliando la lista de posibles proveedores de este servicio y reduciendo sus costes. Los módulos de captura de imágenes, análisis de imágenes y renderización en tiempo real de cada cámara son reutilizables en esta nueva configuración modular y escalable.

Para el sistema en producción de autenticación y vigilancia automática se necesitan modelos biométricos más robustos que eviten desviaciones indeseables debidas a la variación en pose facial y a las condiciones de luz y ruido. También hay que reducir las necesidades de verificación humana cruzada, limitándolos sólo a efectos de garantía de calidad y no para complementar las limitaciones de los sistemas automáticos.

Respecto al sistema de odometría visual, el trabajo futuro podría orientarse en dos direcciones. Un análisis más detallado de la dispersión de los resultados de ORB-SLAM2 podría llevar a encontrar las características cruciales del escenario que reducen la precisión del algoritmo VO. A continuación, podrían tomarse medidas específicas dirigidas a esas características del escenario. Al mismo tiempo, la aplicación de otro tipo de algoritmos, como los algoritmos de VO basados en la fusión de sensores, podría conducir a la mejora de los resultados, específicamente para las operaciones de trenes autónomos que requieren una estimación precisa de la localización.

Por otra parte, cabe destacar que el framework de V&V presentado en este proyecto de tesis es un trabajo intermedio que aporta una solución a un problema actual, pero no deja de ser una solución semi-automática y, por tanto, temporal. Como trabajo futuro, se pretende trasladar el entorno ferroviario a herramientas de código abierto como CARLA. Esto permitirá automatizar completamente todo el proceso de generación de datos, la manipulación de las fuentes de datos de los sensores y la automatización total del proceso de etiquetado sin tener que seguir el mismo trayecto de tren para cada lote de pruebas. Además, se evaluará la opción de dotar al marco de la capacidad de realizar pruebas de ablación. Puede ser interesante probar el rendimiento de las funcionalidades basadas en IA eliminando determinados componentes para comprender la contribución del componente al sistema global.

Los siguientes pasos de la investigación respecto al HW embebido para aplicaciones de VA&IA se centrarán en mejorar el rendimiento de la ejecución en tiempo real, alcanzando tiempos de inferencia más bajos a la vez que se mantiene/incrementa la precisión de la detección. Se tendrán en cuenta nuevas arquitecturas de NN como candidatas para portarlas a la placa SELENE. Por otro lado, también podrán realizar pruebas más profundas y validar las posibilidades de la plataforma para la ejecución redundante seguida de un sistema de votación diferente, un 2oo3 por ejemplo, con el fin de aumentar el nivel de seguridad.

Para terminar, la futura certificación requerirá que la plataforma diseñada disponga de futuras iteraciones que converjan finalmente en la implementación a gran escala. Ello propiciará un escenario en el que todas las partes interesadas de la UE compartan datos y se beneficien de los datos compartidos por otros, acelerando los procesos de certificación. También se reunirán todos los requisitos y se seguirá debatiendo quién gestionará y alojará la plataforma, quién contribuirá y quién podrá utilizarla y en qué condiciones. ’Data Factory’, el grupo de trabajo convocado en el marco de la Empresa Común Ferroviaria Europea (EU-Rail) [EUC23], seguirá esta tarea en los próximos años.

4 Publications

4.1 Accurate Ball Trajectory Tracking and 3D Visualization for Computer-Assisted Sports Broadcast

- **Autores: Mikel Labayen** and Igor García and Naiara Aginako and Julián Flórez
- **Revista:** Multimedia Tools and Applications
- **Volumen:** 73
- **Páginas:** 199-208
- **Año:** 2014
- **Editor:** Springer US

Multimed Tools Appl (2014) 73:1819–1842
DOI 10.1007/s11042-013-1558-x

Accurate ball trajectory tracking and 3D visualization for computer-assisted sports broadcast

Mikel Labayen · Igor G. Olaizola · Naiara Aginako · Julian Florez

Published online: 24 August 2013

Abstract The application of computer-aided controversial plays resolution in sport events significantly benefits organizers, referees and audience. Nowadays, especially in ball sports, very accurate technological solutions can be found. The main drawback of these systems is the need of complex and expensive hardware which makes them not affordable for less-known regional/traditional sports events. The lack of competitive systems with reduced hardware/software complexity and requirements motivates this research. Visual Analytics technologies permit system detecting the ball trajectory, solving with precision possible controversial plays. Ball is extracted from the video scene exploiting its shape features and velocity vector properties. Afterwards, its relative position to border line is calculated based on polynomial approximations. In order to enhance user visual experience, real-time rendering technologies are introduced to obtain virtual 3D reconstruction in quasi real-time. Comparing to other set ups, the main contribution of this work lays on the utilization of an unique camera per border line to extract 3D bounce point information. In addition, the system has no camera location/orientation limit, provided that line view is not occluded. Testing of the system has been done in real world scenarios, comparing the system output with referees' judgment. Visual results of the system have been broadcasted during Basque Pelota matches.

M. Labayen · I. G. Olaizola · N. Aginako (✉) · J. Florez
Department of Digital Television and Multimedia Services, Vicomtech - Ik4 Research Alliance,
San Sebastian-Donostia, Spain
e-mail: naginako@vicomtech.org

M. Labayen
e-mail: mlabayen@vicomtech.org

I. G. Olaizola
e-mail: iolaizola@vicomtech.org

J. Florez
e-mail: jflorez@vicomtech.org

Keywords Computer graphics · Camera calibration · Tracking · Segmentation · Sports events broadcast

1 Introduction

Object tracking has a prominent role within the field of computer vision. The proliferation of high performance computers, the availability of high quality video cameras at affordable prices, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. Detection of target moving objects frame by frame, tracking and analysis to recognize their behavior are the usual pipeline in video analysis [2].

From application domain point of view, tracking systems are being introduced in sport game broadcasts, providing spectators with additional information. Due to the high performance equipment requirements, the renting of this kind of systems is quite expensive, making them unaffordable for small producers or broadcasters. This is exactly the Basque Pelota case. This regional/traditional game is produced by small producers and broadcasted by regional broadcasters. Their low budget does not allow to contract current setups to support controversial plays.

In this work, a system to assist referees solving controversial plays in sport games is described. The game has to be played with a ball and its playground must be delimited by lines. The developed software allows to reduce the set-up requirements, creating an accurate system that is affordable for a wider range of clients.

The set-up design, which is able to cover all border lines, is a challenge. This border line number can be high (e.g. tennis playground) driving the solution to multi-camera set-up. The modularity and scalability are important approaches for required solution. In this work The Basque Pelota test-case is presented in order to simplify the explanation. It is an ideal scenario to test the system first prototype because of its technical peculiarities. Since the playground is delimited by walls on 3 of its borders, it has only one border line to be covered. Once the application is validated in this game, this technology is being extended to other sports (i.e. tennis) which need a multi-camera distribution. For each camera image capture, image analysis and real-time rendering modules are reusable in this new modular and scalable set-up.

In the following Section 1.1, this article carries out a short analysis of the state of the art in controversial play resolution. Afterwards, in Section 2, a system overview is presented in terms of its objectives, description and specifications. Section 3 details the hardware and software (HW/SW) implementation of the core system, including camera calibration, image analysis and real time virtual 3D reconstruction processes. Finally, in Section 4 the document shows the results obtained from tests carried out in real scenarios and it ends up with Section 5 summarizing the conclusions.

1.1 Related works

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in the different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object. Therefore, the use of object tracking is pertinent in the tasks of [2]:

- Motion-based recognition, that is, human identification based on gait, automatic object detection, etc.
- Automated surveillance, that is, monitoring a scene to detect suspicious activities or unlikely events.
- Video indexing, that is, automatic annotation and retrieval of the videos in multimedia databases.
- Human-computer interaction, that is, gesture recognition, eye gaze tracking for data input to computers, etc.
- Traffic monitoring, that is, real-time gathering of traffic statistics to direct traffic flow.
- Vehicle navigation, that is, video-based path planning and obstacle avoidance capabilities.

This work focuses on motion-based object recognition in sport broadcasting. Tracking systems in the TV broadcast domain are not a recent approach at all. Most of the researched systems in this field are based on prediction algorithms based on Kalman [4, 17] or particle filters [10]. Extend state of the art material is available about methodologies dedicated e.g. to player and ball tracking in soccer [14, 18, 19] or tennis [5, 7, 13].

Some companies such as *Sportvision*[1] and *Virtual eye*[2] market systems which provide data content and enhancements for sports broadcasts and applications:

- The FoxTrax hockey puck tracking system [3] based on an infrared sensor. The circuit board inside a puck contained a shock sensor and infrared emitters. The puck emitted infrared pulses that were detected by both the 20 pulse detectors and the 10 modified IR cameras that were located in the rafters. Each IR camera processes the video locally and transmits the coordinates of candidate targets to the "Puck Truck".
- *Strick zone* control by ball and player tracking. Three PCs connected to three video cameras track a pitched baseball's flight toward the strike zone. Two cameras observe the baseball, while the third observes the batter to provide proper sizing for the strike zone.
- Playground lines drawings of *1ST & TEN*[3] in American football. This application uses a number of cameras shooting the field. Recent implementations require around four computers, one computer per camera plus a shared computer for chroma-keying and other tasks that can be run by a single operator.
- Cricket[4] and golf ball tracking. Based on image computer graphics technology, 4 high-speed cameras (250 fps), two Infrared cameras and sophisticated computer rack are used to track the cricket ball. This set up needs at least a group of 4 operators to its management.
- Additional information for viewers as graphics and statistics in golf.

[1] www.sportvision.com

[2] http://virtualeye.tv/

[3] http://www.ieeeghn.org/wiki/index.php/The_Making_of_Football%27s_Yellow_First-and-Ten_Line

[4] https://www.youtube.com/watch?v=LjLe06H7EJg

Due to their closed system, the algorithms on which they are based are in most of cases unknown.

Hawk-Eye[5] markets the most important controversial play image-based analysis and 3D virtual replay reconstruction approach for situations in which a tennis ball sized object is used in the play. Although, it started as cricket ball tracker, it is well known because it is able to point the location of a ball bounce in a tennis court with high accuracy. *Hawk-Eye* uses 6 high speed specialized vision processing cameras which are positioned around the ground and calibrated. In addition the system uses two broadcast cameras and calibrates them so that the graphic is always overlaid in the right place. All cameras have anti-wobble software to deal with camera movement. According to information in its web [9], it is able to deliver a pinpoint accuracy of under 5 mm.

However, the complexity of the set-up, high-speed cameras are needed, and the equipment requirements make the system too expensive for less-known regional/traditional sports. Even more, it cost is around $60.000 for one court which increases by 100 the cost of the system presented in this approach.

All these approaches use at least two or more high speed specialized vision processing cameras to determine the bounce distance from border line. In addition, they need operator team to control them. In order to reduce the existing solutions requirements, this work presents an alternative set-up, robust in terms of different possible camera operating location/orientation, based on unique broadcast-type camera per border line. This solution can be managed by one operator, even playground has more than one line under control, changing camera views from the system. The challenge in reduction of hardware complexity and achieving market solutions' accuracy motivates this work.

2 System overview

The industrial project called *Begira*, in which this research has been carried out, establishes the technical specifications that the developed system has to fulfil. Although the state of the art can offer specific solutions for some of the technical requirements, it cannot afford the consecution of all the technical specifications. Even more, the economical limitations are a also the variables that constrains this research.

2.1 System objectives

The system must be able to pinpoint accurately the distance of ball bounce from the line. The output 3D virtual video will simulate the last ball trajectory and will be inserted in TV PAL broadcast signal. The solution must be deployed on top of a simple HW/SW system to make it affordable for any producer, sport event broadcaster or less-known sport event organizer.

A system set-up design driven by flexibility in terms of size and operating location, can significantly reduce the costs rising this challenge as a major aim. In addition,

[5] www.hawkeyeinnovations.co.uk

the system needs to operate in quasi real-time, at least faster than the estimated time for video replay which is about 30 sec. Respect to pinpoint accuracy, defined by Basque Pelota referee committee, the estimation error should be under 1 cm for all cases and under 5 mm for 80 % of them. This error has been set taking into account the typical human eye incertitude in the appreciation of bounce point (from a distance and with millisecond duration), which is also subjective. As this limit was considered achievable after the demonstration of our first version of the system, it was determined as a requirement. Referee committee is aware of the difficulty of approaching these accuracy values, but consider them necessary in order to standardize the system.

2.2 System description an requirements

In this section, the system general workflow, as well as module specifications and functionalities are described (Fig. 1).

The prototype has four independent modules: a) the camera, b) the capturer/encoder, c) a laptop for storage and all processing tasks and d) a video adapter for TV PAL broadcast or playout. The camera captures images and transfers them in RAW format using an standard professional TV interface to the capturer/encoder module. This second module encodes frames using the H.264 codec and transfers it to the laptop where it stores them into a Transport Stream (.ts) video file. The image processing and later 3D virtual replay generation tasks are executed in the laptop. In the last module, the 3D virtual representation output video is adapted to broadcast quality video .

- **Camera** The cameras used for image capture must fulfil some characteristics in order to make the ball detection easier for segmentation and identification operations carried out in the next steps:
  - **Frame rate** The broadcast camera must provide enough images per second to track and predict the ball trajectory in each frame. The choice of this factor is defined taking into account the trade-off between the data processing time (in capturing/coding/storing) and the necessity of the amount of real images to be able to approach accurately the ball trajectory. To avoid missing frames, the time elapsed in recording/storing each frame must be less than reciprocal of the frame rate.
  - **Shutter speed and diaphragm aperture** In the case under study, the accuracy of the tracker can be improved if a target with stable shape, without blur effect, and with stable color (grayscale intensity) is acquired. Therefore, a high shutter speed camera is required. The choice of this factor is to be

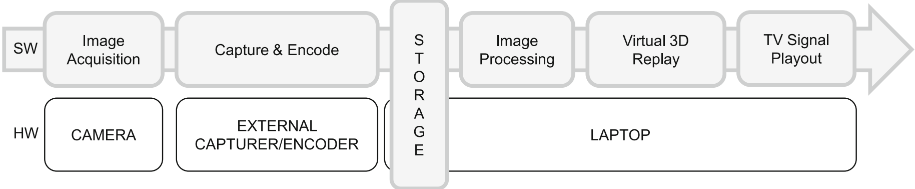
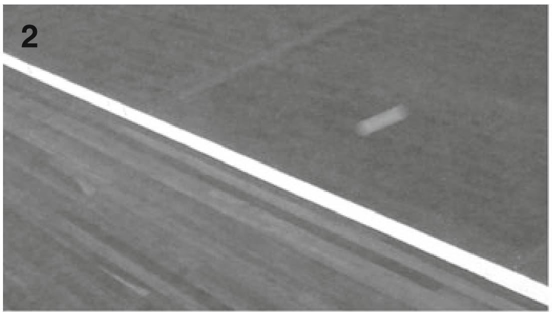


Fig. 1 System HW/SW workflow & modules



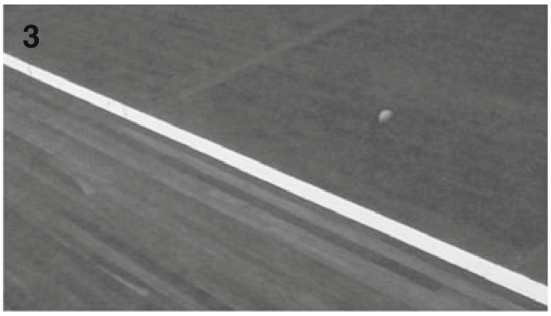


Fig. 2 Low shutter speed (auto). Electronic gain disabled. Blurry ball

Fig. 3 High shutter speed (1/500). Electronic gain enabled. Blur-free ball

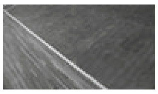
defined taking into account the trade-off between the minimum illumination required in the segmentation process and the necessity to keep the shape of the ball stable. It will be set to the minimum that allows the ball to appear as a clear round object. The maximum speed of the ball will be relevant on it (see Figs. 2 and 3).

The minimum shutter speed and consequently the diaphragm aperture are set according to the illumination of the sport events place and to the expectable maximum ball speed in each game. In this work, these values are set for Basque Pelota courts (indoor, illuminated for TV broadcast).

- **Image resolution** The accuracy in ball bounce pinpointing is also related to the resolution of the captured images. The higher the resolution, the lower the pixels/distance ratio. Once again, the system performance is based on a trade off between lower processing time and higher accuracy in measure (see Table 1).

  To approximately calculate Pixel/Distance ratios, we use as reference object the border line. Calculating the amount of pixels in the horizontal and vertical vertices of the image, we can compute border line width pixel amount and compare it to border line real width measurement.
- **Color space** The color space influences the ball segmentation process. Although multi-component color spaces can offer extra information in image object understanding, the bright white color of the ball and the dark color of the playground, offer high contrast which makes single component color spaces enough for segmentation purposes. This reduces the generated data amount for capture, encoding, storage and processing tasks.
- **Optical lens with fixed camera-to-playing field distance** The field of view of the camera must also be taken into account to determine lens distortion

Table 1 Pixel/Distance (pixel/cm) ratios calculated for different image resolutions and camera distance

Image resolutions	Distance from camera (m)		Reference image
	9 m	0.5 m	
1080p	0.45	0.065	
720p	0.7	0.09	
576p	0.85	0.15	

and system precision in terms of *image pixel/real distance*. The greater the field of view, the greater the covered scene area in which the ball trajectory can be analyzed. However, the greater the field of view, the greater the lens distortion and the lower the precision (pixel/distance). Even though this parameter must be taken into account, it is not as critical as others like velocity of the ball.

- **Capture & encoding module** The capturer/encoder module captures the RAW multi component video signal provided by the camera. After that, the signal is converted to a single component color space. Then, the signal is compressed and encoded in order to reduce the information data flow for the storage and processing tasks.
  - **Codec** The codec requirements have to solve the controversial relation between image quality and compression ratio. The goal is to obtain the maximum compression ratio, keeping the minimum image quality which ensures correct segmentation conditions after decoding.
  - **File container** The video file must be read and written at the same time. In addition, the read process must offer quasi random access capabilities for the retrieval of part of the whole recorded video starting from a specific frame. Moreover, most of multimedia container formats include timestamps and data just before file closing, becoming navigation more difficult. The chosen container must solve this problem.
- **Storage and processing laptop** The laptop and capture/encoding module are connected trough USB 2.0. The laptop stores the encoded images in its hard disc, it retrieves and analyzes them and finally generates a 3D virtual replay of the action. A multi-tasking approach for quasi real-time performance establishes the hardware characteristics of the laptop. The system core software is stored and executed in this module. The algorithm robustness is directly related to the system set-up flexibility.
- **Video adapter to TV broadcast signal** The broadcasted output video signal must comply with the broadcaster graphical requirements and signal quality specifications at its mobile units. This module adapts the rendered video signal into a TV broadcast signal.

3 Implementation

In the first step of the implementation, state of the art and market study has been carried out to identify the existing HW/SW developments which best fit the needs of the system based on the requirements outlined above. Two main issues have been encountered at this point: on one hand, no specific HW/SW solution exists for the established requirements. On the other, available HW solutions deal with independent tasks identified in the system workflow & modules figure (Fig. 1). This context pushes the development of our own algorithms, as the outcome of a research process. The unique existing Open Source algorithms used in the implementation are Camera Calibration (OpenCV) and Polynomial Approximation (GNU).

The system as a whole has been integrated using Qt:[6] a cross-platform application framework that is widely used for developing application software with graphical user interfaces (GUIs).

3.1 Image capture

From the beginning, this system was developed using conventional TV broadcaster equipment in order to reduce costs in later market adoption processes. The camera used in the tested prototype is a common professional HD handheld camera (Panasonic HVX200A[7]), widely used by many kinds of producers/broadcasters. It provides a RAW YUV(4:2:2) component signal at a maximum resolution of FULL-HD 1080i and a maximum frame rate of 50fps far away from the throughput and features of cameras required by other market solutions. The camera is set-up at HD 720p 50fps both providing enough resolution and frame rate for our approach.

The scene illumination conditions are then to be analyzed. The amount of available light is a combination of pelota court lights and of additional spotlights used in special competition broadcasting. Under these conditions, the scene often is not enough illuminated, providing resulting images (at 50fps and 1/500 shutter speed) which are low-contrast.

Taking into account the minimum illumination required to keep the color and constant shape characteristics of the ball in the segmentation process, the balanced compromise between shutter speed, which keeps constant the ball round shape, and diaphragm aperture and electronic light gain, which keep the scene contrast, is defined for each broadcasted event.

3.2 Image encoding and storage

The amount of data generated capturing HD 720p images at 50fps and described by the bit rate BR parameter makes necessary the use of compress/encode algorithms.

*Resolution:(1280 * 720) pixels/frame*
Frame Rate: 50 frames/sec
Bit Depth: 8 bits/pixel
Components: (1 (Y) + 0.5 (U) + 0.5 (V))
(Note: YUV 4:2:2 format)

$$BR = 1280 \times 720 \times 50 \times 8 \times 2 = 703.125 Gbps \tag{1}$$

The generated throughput would impose special storage, transmission bandwidth and equipment. This makes the system set-up more expensive and less compact. However, reduced system cost and dimensionality are central requirements from the beginning: to reduce the throughput the signal must be compressed. The dominant video codec today for web and mobile video (limited by the transmission channel bandwidth) is H.264 [6, 15]. H.264 compression preserves the video quality at high compression ratio better than other popular codecs widely available on the market [15, 16].

[6] www.qt.nokia.com

[7] www.panasonic.com/business/provideo/home.asp

Although the standard defines 17 sets of profiles, H.264 has three commonly-used: Baseline (lowest), Main, and High. Higher profiles (Main and High profiles) ensure the best signal quality-compression relation. Since the system needs high compression ratios with the best signal quality, the High profile is chosen.

H.264 is typically deployed into *.MP4* file containers. However, a wide range of different containers can be used. One of the main difficulties of working with open videos is the random access within the content. Most seek function implementations require closed video files to function properly. However, in our case, the positioning at specific frame is performed will the video file is open and the encoder is appending information on it. To achieve this purpose, it is necessary to have time marks periodically embedded in the video file. Nevertheless, most video containers only include those marks just before the file is closed.

The container chosen to fit our requirements is therefore MPEG Transport Stream *(.Ts)* [12]. This Transport stream, devoted to content broadcasting, specifies a container format encapsulating packetized elementary streams, with error correction and stream synchronization features for maintaining transmission integrity when the signal is degraded. This allows to read specific video segments while writing into the same file.

The open source ffmpeg[8] library has been used to compress, encode and encapsulate the video, as well as to retrieve video sections, and decode them. This package includes audio/video codec and audio/video container multiplexer and demultiplexer libraries.

3.3 Camera calibration

Camera calibration or resectioning is the process of finding the true parameters of the camera that produced a given photograph or video based on prior knowledge of the scene. The camera parameters are classified in extrinsic and intrinsic parameters.

Rotation and translation matrices (R, $\overrightarrow{t}$) contain the extrinsic parameters which denote the coordinate system transformations from 3D world coordinates to camera coordinates. On the other hand, the intrinsic parameter matrix (K) encompasses focal length, image format, and principal point (2).

$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Where,

f_x & f_y *Lens focal length*
c_x & c_y *Principal point (the image center)*

The camera calibration is carried out using the so-called pinhole camera model, on which Opencv[9] camera calibration routines are based. A scene view is formed

[8] www.ffmpeg.org

[9] http://opencv.willowgarage.com/wiki/

by projecting 3D points (x_p, y_p, z_p) into the image plane (x_i, y_i) using a perspective transformation.

$$P_i = \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}, \quad P_p = \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} \tag{3}$$

From the homography matrix (H), the matrices $(R, \overrightarrow{t})$ describing the rotation and translation parameters of the camera can be extracted.

In the mathematical development below the captured image points are identified by P_i (image coordinate, pixel) and playground plane points, where the ball will bounce, by P_p (real world plane coordinate, cm). P_p^* is an auxiliar point (real world plane coordinate, cm).

Since Opencv use homogenous coordinates:

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = K[R|t] \begin{pmatrix} x_p \\ y_p \\ z_p \\ 1 \end{pmatrix} \tag{4}$$

Where,

(x_p, y_p, z_p) *Real world 3D coordinates*
(x_i, y_i) *Projection point coordinates*

$$P_p^* = \begin{pmatrix} x_p^* \\ y_p^* \\ z_p^* \end{pmatrix} \tag{5}$$

$$H_{ip} = \begin{pmatrix} h_{(1,1)} & h_{(1,2)} & h_{(1,3)} \\ h_{(2,1)} & h_{(2,2)} & h_{(2,3)} \\ h_{(3,1)} & h_{(3,2)} & h_{(3,3)} \end{pmatrix} \tag{6}$$

$$P_p^* = H_{ip} \times P_i \tag{7}$$

$$H_{ip} = H_{pi}^{-1} \tag{8}$$

$$P_p = P_p^*/z_p^* = \begin{pmatrix} x_p^*/z_p^* \\ y_p^*/z_p^* \\ z_p^*/z_p^* \end{pmatrix} = \begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} \tag{9}$$

The homography matrix (H) needs to be calculated upon starting the system: it maps which pixels coordinates of captured image points P_i correspond with playground plane coordinate points P_p. Thus, once the bounce point in the captured image is identified, the position in the playground plane can be calculated.

The image points are selected using a calibration checkerboard as in Fig. 4. The playground plane points are predefined and they must correspond to the points of the checkerboard which are selected in the captured image. With this process, camera intrinsic parameters matrix (K) and plane homography matrix (H) are calculated. Consequently, the $(R, \overrightarrow{t})$ matrices are defined.

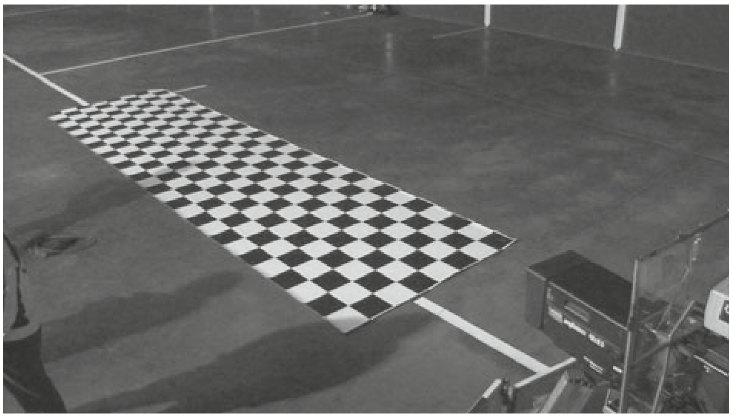


Fig. 4 Calibration checkerboard

The calibration information allows placing the camera (position and tilt) with respect to a reference point (x_p, y_p, z_p) of the pelota court world and determining its intrinsic distortion parameters. Undistort parameters and the geometric transformation, which establishes the relation between a captured image and the playground plane points parameters, are therefore established. Camera calibration makes the system robust in terms of different possible camera operating location. As a result, the system has no camera location/orientation limit, provided that line view is not occluded.

3.4 Image analysis and data processing

In this section, the image processing and accurate bounce point determination algorithms are explained. The development has been based on the open source Opencv and GSL—GNU[10] libraries.

Once the system has been started, the recording begins. The camera is acquiring the contentious area around the playground border line and storing the information in a laptop where data is also processed during the entire duration of the game. The data captured from the camera is stored as MPEG transport stream (.ts) and using H264 encoding. When a controversial play occurs, the operator triggers the system. To that end, it extracts the latest frames, which contain the controversial play.

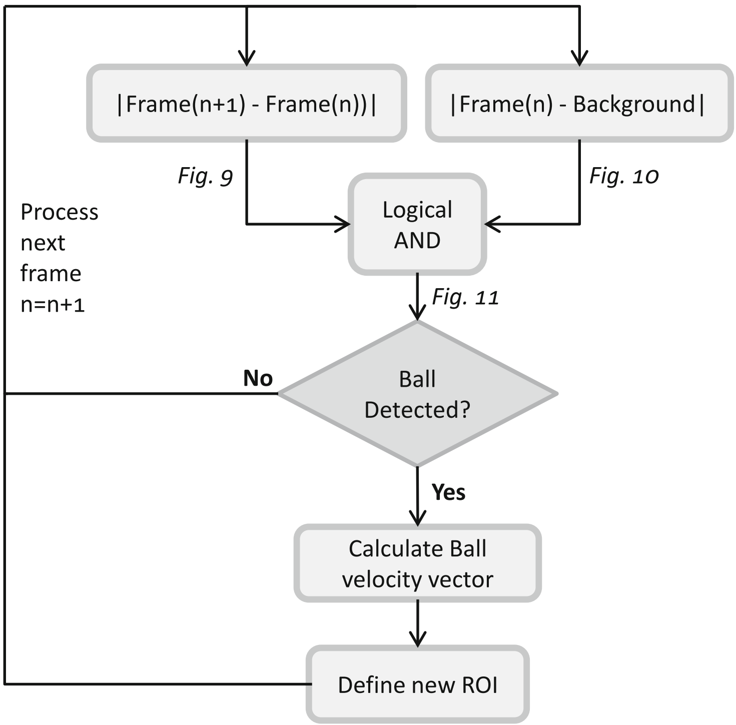
Once the set of images is extracted (Fig. 5), the first image of the sequence is set as the background image (Fig. 6). All the images of the sequence are converted from the color space, which is determined by the camera output, into a single component color space able to contrast the shape, movement and intensity descriptors to pinpoint the ball position in each frame.

After this, all the images are pre-processed using the camera intrinsic parameters matrix (K) to correct the distortion introduced by the optical lens.

After image preprocessing, the process for ball segmentation and tracking starts (Fig. 5) for each of the corrected images of the sequence (Fig. 7). Due to the knowledge of the probable initial ball position, this process is only applied for a

[10] www.gnu.org/software/gsl/

Fig. 5 Ball segmentation and tracking process

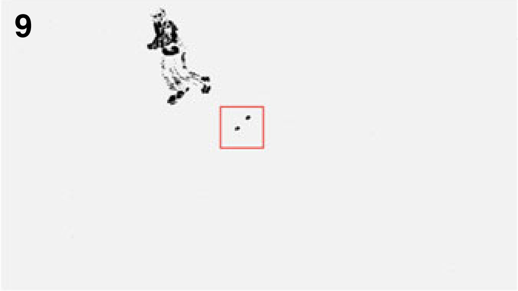


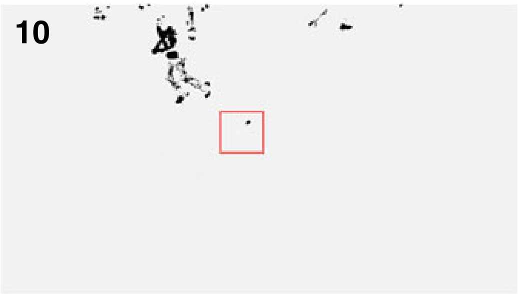
concrete image area. Image areas' difference is calculated pixel by pixel with respect to the previous image (temporally) (Fig. 8) and to the reference image (Fig. 6), such that two difference images (Figs. 9 and 10 respectively) are obtained. Broadcast camera position and orientation are statics from the beginning of the match. For this reason, the frame difference technique provides a background-free output.

These two substraction image areas are transformed into black and white image areas via thresholding. The logic operation AND is performed for each pair of image areas (Fig. 11), so that only regions which are present in both images are extracted. Regions identified as noise also have to be removed by the logical AND composition operation. In order to discard noisy regions, estimated shape and area are used. Furthermore, velocity of the ball, considerably greater that of rest of the objects present in the scene, is set as key characteristic for segmentation. This methodology is used for extracting the initial position of the ball. Once the initial position is determined, the tracking process of the ball is performed.

The tracking process is based on the calculation of a movement vector. This movement vector and the velocity vector of the ball are calculated taking into account its coordinates (x_i''', y_i''') in pixels with respect to the previous image and to the time that has elapsed between one image and the next. In order to calculate the movement vector, the difference between the coordinates (x_i''', y_i''') of the center of the ball is calculated for consecutive images.

The calculation of the movement vector allows predefining a ROI where the segmentation process occurs. Once the initial point of the ball has been extracted and the movement vector calculated, the system creates a ROI determining the prediction area for ball position. All the process steps of frames substraction and AND logical operation will be made in the extracted ROI. Therefore, the process of ball detection speeds up.





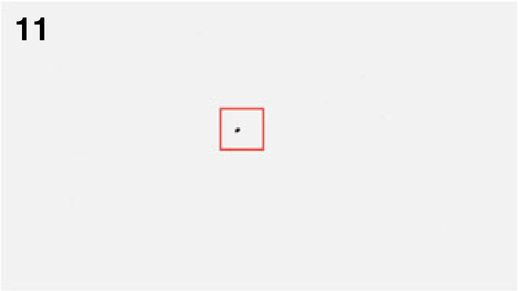


Fig. 6 Background reference image
Fig. 7 Frame(n+1)
Fig. 8 Extracted frame(n)
Fig. 9 |Frame(n+1) - Frame(n))|
Fig. 10 |Frame(n) - Reference background|
Fig. 11 Logical AND of (d) on (e) and resulting ROI

The combination of camera position/orientation and selected ROI size keep usually the players belonging regions (noise) out of ROI. However, if there are more than one ball candidate region after AND composition operation, eliptic shape, calculated area and predicted position are used to discard irrelevant ones.

Tracking prediction algorithms, like Kalman Filter or Particle Filters, have been implemented and tested but finally rejected because they do not offer any significant improvements comparing with less complex procedure assuming some approximation. This is due to the fact that ball trajectory can be considered quasi linear close to bounce point. In addition, the relation between capture frame rate and ball velocity makes vector module almost constant and smaller than ROI size. For this reason, the velocity vector information is enough to predict properly the future ROI position and ROI size to detect the ball even if changes its direction after bounce.

As shown in Fig. 12, once the velocity vector has been extracted for the entire sequence of frames, the sequence of positions of the ball (x_i''', y_i''') is divided into two segments. In order to define the limit of the segments, the difference in angle and modulus of the velocity vector is taken into account. The maximum value of the

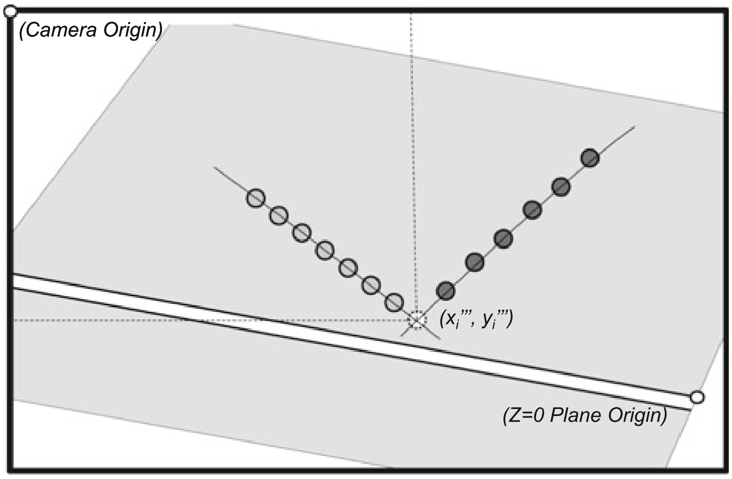


Fig. 12 Point sequence split & polynomial approximation

angle difference determines the limit which divides the two segments. If the angle values are similar, the modulus is used to break the deadlock.

Once the coordinates of the ball position of the ball have been determined for the two segments, a least-square fitting is performed for each of the two segments [1]. For the calculation of this fitting curve, the points which are above a minimum distance to the curve are iteratively discarded.

if,

$$|x_i'''(n) - lsf(x_i'''(n))| > \sum_{n=1}^{length} \frac{|x_i'''(n) - lsf(x_i'''(n))|}{length} \Rightarrow (x_i'''(n), y_i'''(n)) \text{ point discarded.} \tag{10}$$

Where,

lsf	*Least Square Fitted function*
length	*Each segment length*

The trajectory of the ball for each of the two segments is thus determined. The point of the intersection of the resulting curves is considered the bounce point of the ball (Fig. 13).

The position of the bounce point is now referenced with respect to the image coordinates (x_i''', y_i''') while the real distance of the bounce point to the playground border line is to be known. To that end, the geometric transformation obtained in the calibration procedure is applied to extract the coordinates (x_p, y_p) in the playground plane from the coordinates of captured image in pixels.

This geometric transformation, related to homography, can only be applied for the points of the playground plane $(z_p = 0)$. To that end, it is necessary to define the bounce point in the image. The point is defined by the coordinates (x'', y'') in pixels (Figs. 14 and 15).

$$x_i'' = x_i''' + R * \sin(\alpha) \tag{11}$$

$$y_i'' = y_i''' + R * \cos(\alpha) \tag{12}$$

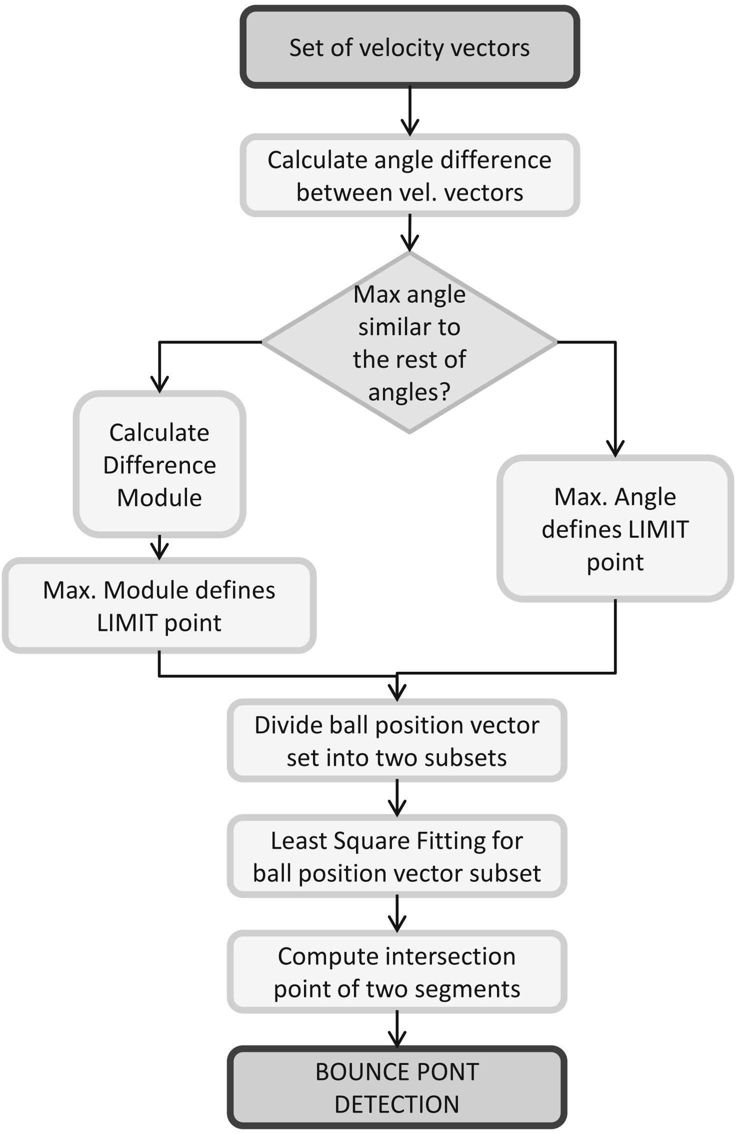

Set of velocity vectors
Calculate angle difference between vel. vectors
Max angle similar to the rest of angles?
Calculate Difference Module
Max. Angle defines LIMIT point
Max. Module defines LIMIT point
Divide ball position vector set into two subsets
Least Square Fitting for ball position vector subset
Compute intersection point of two segments
BOUNCE PONT DETECTION

Fig. 13 Bounce point detection using velocity vectors

A small error is produced at the exact point where the ball touches the ground It has not been reflected in the figures, since the ball is superimposed.

$$x_p = x'_p + v_x(R, \overrightarrow{t}) = x_{p_{bounce}} \tag{13}$$

$$y_p = y'_p + v_y(R, \overrightarrow{t}) = y_{p_{bounce}} \tag{14}$$

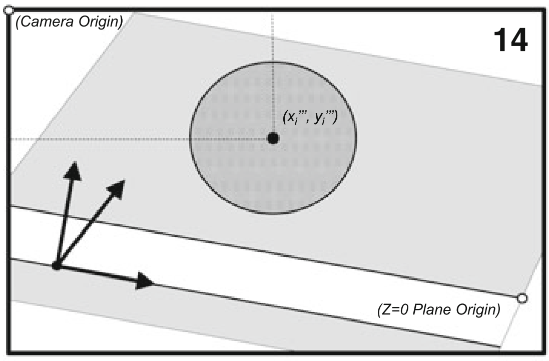
Where,

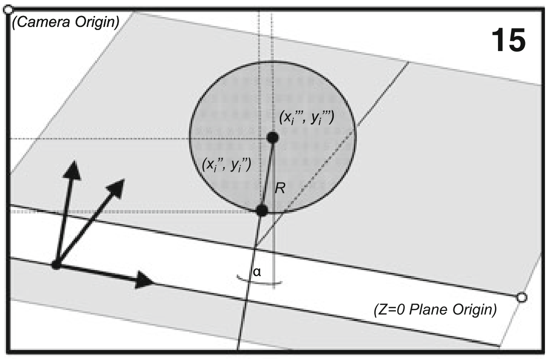
$(x_{p_{bounce}}, y_{p_{bounce}})$ *Bounce point at playground*

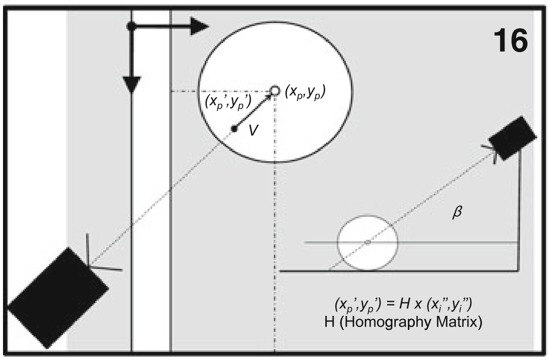
$$z_p = \begin{cases} \left(\dfrac{z_{p_0}}{x_{p_{bounce}}} + \dfrac{g}{v_x^2} * (x_p - (2 * (x_p - x_{p_{bounce}})))\right) * (x_p - x_{p_{bounce}}) & \text{if } x_p < x_{p_b} \\ \dfrac{g}{v_x^2} * \left(x^2_{p_{bounce}} - x^2_p\right) & \text{otherwise} \end{cases} \tag{15}$$

Where,

g *Gravity constant*







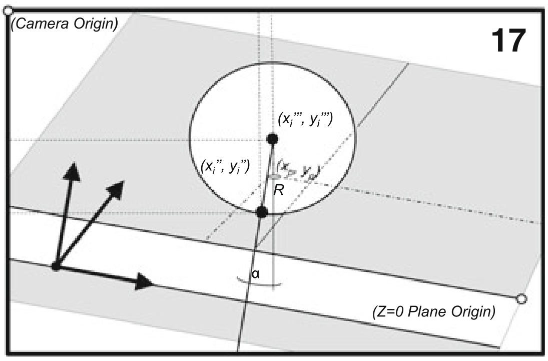


Fig. 14 The center of the ball (x_i''', y_i''') referenced to captured image origin
Fig. 15 The point (x_i'', y_i'') where the ball touch the ground referenced to image origin
Fig. 16 Transformed touch point (x_p', y_p') at ground $(z_p = 0)$, referenced to ground plane origin
Fig. 17 Bounce point $(x_{p_{bounce}}, y_{p_{bounce}})$ referenced to ground plane origin

Once the point (x_i'', y_i'') is calculated, the point on the real ground plane is obtained by multiplying it by the plane transformation matrix (H) (Fig. 16 and Eq. 7).

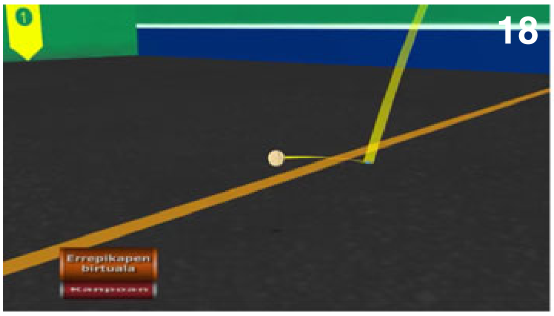
As seen in Fig. 16, the point (x_p', y_p') is not an exact projection of the center of the ball, so it is moved in the direction of the optical vector of the camera with a distance which depends on the position and tilt of the camera to the central point.

Once the real position of the bounce point, which is referenced to the field line, has been calculated it can be determined whether the ball has bounced outside, inside or on the line itself.

Since the correct geometric transformation provided by the two plane homography only can determine the relation between captured images and playground plane points, the only actual ball 3D positioning can be carried out when it touches the ground (at bounce point). From this data, the rest of replay ball trajectory is simulated. Its (x_p, y_p) components are computed from the bounce point $(x_{p_{bounce}}, y_{p_{bounce}}, z_{p_{bounce}})$ and ball direction vectors defined taking into account some ball positions parameters in the processed images close to the bounce moment. The (z_p) component (15) is based on parabolic model taking into account (x_p, y_p) points, approximate ball velocity vector and the z_{p_0} determined from the ball position at the first analyzed image frame (Fig. 17).

3.5 Virtual 3D replay

The visual result of the image analysis is the controversial play virtual 3D replay. Here one of the most performance demanding issues is the rendering engine,



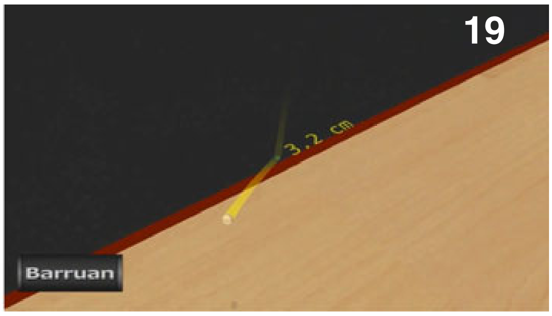


Fig. 18 3D virtual reconstruction of ball trajectory

Fig. 19 Virtual representation of bounce point

dedicated to the computational process of generating an image using 3D information. Firstly, this 3D shape information is converted into polygons and then into triangles. Secondly, these triangles are projected into a 2D image and, finally, each pixel inside the triangles is colored. The whole process takes too much time if no additional strategies or algorithms are used and live TV broadcast cannot be interrupted.

In order to address realtime 3D rendering, the approach is built on top of OpenSceneGraph[11] (OSG) library [11]. It is an open source, cross-platform graphics toolkit for the development of high performance graphic applications. It is based on the concept of a scene graph and uses OpenGL.[12]

OpenSceneGraph makes use of techniques that speed up the rendering computational process because the rendering motor deals with considerably reduced information: a Level of Detail (LOD) algorithm, culling techniques (frustum, occlusion and small feature culling) and a State Sorting strategy are employed to this end.

The basic LOD idea is to use simpler versions of an object as it makes less and less of a contribution to the rendered image. So, when an object is far away, less polygons will be used to define it, which reduces the number of triangles to be processed in the rendering. The criteria OSG uses to select a level of detail model depends on the distance of the object from eye point (range-based selection). And to stop the switching form one LOS to another being noticeable, a Continuous Level of Detail (CLOD) technique is used [8].

Culling techniques consist of removing portions of the scene that are not considered to contribute to the final image. The rest of the polygons are sent through the rendering pipeline. With the View frustum culling technique, all the polygon groups that are outside (the region of space in the modeled world visible form the eye point) are eliminated. When occlusion culling is used, all the objects hidden by groups of other objects are also eliminated from the sending-to-render process. And with Small Feature culling, small details that contribute little or nothing to the rendered images are not processed when the viewer is in motion [8].

State Sorting consists of sorting geometrical shapes with similar states into bins to minimize state changes in the rendering process [8].

[11] www.openscenegraph.org

[12] www.opengl.org

The 3D visualization module takes as input the ball 3D trajectory (Figs. 18 and 19), the exact bounce point, its distance from the line, its shape and if the bounce point is *in* or *out*. According to this incoming data, the scenario is loaded and the ball trajectory simulated creating an output video file with the controversial play reconstruction. Although the output video is rendered by a conventional camera view for standard TV broadcast, it can be rendered with stereoscopic cameras for future 3DTV broadcasts.

The module configuration defines the variable parameters which describe the scenario: playground border lines width and color, ground material texture, rendered measure number and arrow colors, etc. This makes it easier to configure the virtual scenario for the different real pelota courts where the game takes place.

3.6 Signal adaptation for PAL-quality digital TV signal playout

With regards to the output signal, the system is able to provide HD 1080p digital video throughput. However, it is required to also be compatible with nowadays Standard Definition (SD) broadcaster TV signal standards. Rendered images are adapted to these restrictions. The output video is rendered with a Matrox4 CG2000[13] video adaptor, since this hardware combines a 3D graphic accelerator with broadcast quality video I/O.

The system output signal can be adapted to lower quality formats (PAL 4:3, 16:9) if it is needed due to compatibility issues.

4 System evaluation

The initial assessment phase of the parameters described in Section 2.2 specifies in depth which parameters of the camera improve the further segmentation process.

According to this first assessment, the scene often is not illuminated enough for proper image acquisition. To improve the image quality, the balanced compromise between shutter speed, diaphragm aperture and electronic gain is set. Furthermore, the thresholds used in ball-player-background segmentation are set according to camera parameters and playground illumination. The evaluation of the system has been done before a professional TV broadcasted Basque Pelota match in three different stadiums. In sport broadcast, the courts must be well illuminated in order to maximize the contrast between foreground and background and make the ball visible over the playground for the audience and TV viewers. Accordingly, the illumination and its changes are under control during match time and our system take advantage of this stable environment.

The test set-up has been another considerable challenge. The ball physic behavior has been studied and tested, reaching that the ball bounce can be considered elastic, because its hardness, ruling out any deformation on its shape or track. In order to get to the conclusion that bounce area remains circular, tracing paper has been used. This tracing paper is set along the border line in order to determine real

[13] www.matrox.com

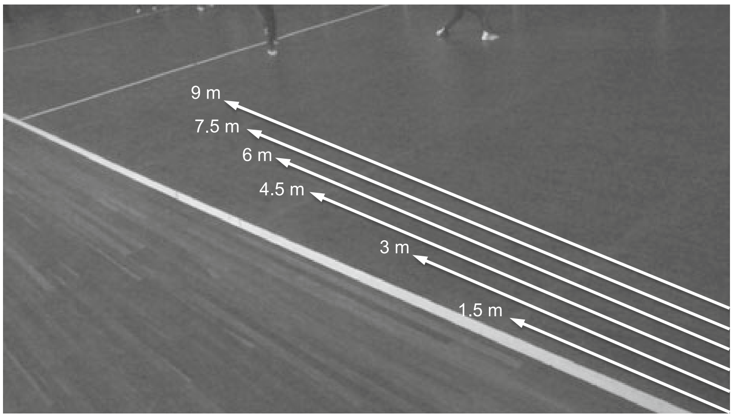


Fig. 20 Covered distances in the testing trials

distance measurements during testing period. Millimetric paper is set under the tracing paper to get the real distance from the bounce area center and the border line. Consequently, the real distance measurement error can not be above 0,5 mm to achieve market competence.

As mentioned before, the system has been tested in different playgrounds. For all tests, the selected parameters for the system are:

1. **Frame rate** = 50fps
2. **Shutter speed** = 1/500 sec
3. **Diaphragm aperture** = 1:1,7
4. **Image resolution** = 720p HD
5. **Camera lens focal length** = 35 mm
6. **ROI window size** = 100×100px
7. **Detection color space** = Gray scale
8. **File storage codec** = H.264

The camera was located 3 m from the ground and on one side of the line. The field of view allowed covering 9 m (as shown in Fig. 20), enough to cover the controversial action play zone. The camera pan and tilt were different for each test, determined to each playground. The system can be adapted for any kind of sport event, taking into account that the error can vary depending the referees requirements.

According to the results extracted from the tests made in one of the playgrounds (Table 2), the algorithm robustness is proved along different possible camera operating distances. Average error for both test measurements is 4,3 mm, which is below the target 5 mm deviation. The numerical error requirements listed in Section 2.1 are not fulfilled successfully, because the 80 % of measures should be below this 5 mm error threshold and only the 69,4 % of the errors satisfy this requirement. still, 80 % of the errors are below 7 mm.

Although Referee Committee considers this results as acceptable because the typical human eye incertitude in the appreciation from a distance, with millisecond duration, is considerably higher than that of the presented system, ongoing research is being developed in order to fulfill these requirements. One of the objectives of this ongoing research lays on the use of FullHD cameras instead of HD cameras. These cameras represent the same scene using a greater number of pixels and therefore the pixel-real distance ratio decreases, permitting a more precise calibration process and the minimization of error (in real cm measurement) when ball centre point is

Table 2 Accuracy test results for tests made in Ogueta(Vitoria) playground

Distance from camera [d] (m)	Test 1			Test 2		
	Real measure (cm)	System measure (cm)	Error (cm)	Real measure (cm)	System measure (cm)	Error (cm)
d < 1.5 m	8.40	8.65	0.25	−8.20	−8.34	0.14
d < 1.5 m	2.85	2.86	0.01	7.10	6.60	0.50
d < 1.5 m	3.75	4.12	0.37	0.80	0.57	0.23
1.5 m < d ≤ 3 m	0.75	−0.04	0.79	−0.50	−0.90	0.40
1.5 m < d ≤ 3 m	−3.75	−3.95	0.20	−5.20	−5.80	0.60
1.5 m < d ≤ 3 m	−0.10	−0.34	0.24	−7.80	−7.56	0.24
3 m < d ≤ 4.5 m	19.00	18.85	0.15	−3.00	−2.58	0.42
3 m < d ≤ 4.5 m	1.75	0.90	0.85	−0.70	−1.07	0.37
3 m < d ≤ 4.5 m	−3.20	−3.15	0.05	−3.55	−2.90	0.65
4.5 m < d ≤ 6 m	3.70	2.45	1.25	−4.75	−3.95	0.80
4.5 m < d ≤ 6 m	6.50	5.64	0.86	2.50	3.40	0.90
4.5 m < d ≤ 6 m	2.00	1.40	0.60	−6.20	−5.80	0.40
6 m < d ≤ 7.5 m	7.10	7.37	0.27	4.00	4.70	0.70
6 m < d ≤ 7.5 m	6.90	6.95	0.05	3.20	3.85	0.65
6 m < d ≤ 7.5 m	−1.70	−1.75	0.05	2.70	3.10	0.40
7.5 m < d ≤ 9 m	3.30	3.50	0.20	1.90	2.30	0.40
7.5 m < d ≤ 9 m	4.20	4.50	0.30	12.60	12.00	0.60
7.5 m < d ≤ 9 m	11.90	12.00	0.10	2.10	1.80	0.30

detected. The other major issue comes from the control of playground illumination, in order to improve the ball segmentation process.

The tests reveal that the precision in measurements is related to the accuracy in ball center pointing (in each captured frame) and to the accuracy in the homography matrix calculation, both of which are closely related to image resolution. Actually, the error in measurement is not constant across the field. Although the resolution of the image is constant, the real distance that a pixel represents (pixel/distance ratio) is different depending on camera location. The longer the distance between the line-point and the camera, the lower the (pixel/distance) ratio. Nevertheless, the experimental results show that this theoretical issue is not crucial for distances less than 9 m from the camera at HD 720p resolution.

In the springs of 2010, 2011 and 2012 the system was tested in the most important Basque Pelota competitions. Although the numerical measurements did not accomplish the goals of the Referee Committee, they considered the system ready to help them taking decisions during the match. The system worked as expected on professional platforms and the output signal was broadcasted live by the Basque public broadcaster (EiTB[14]) successfully. In 2010 it was watched by 219.000 spectators and the viewer share was 31,1 %.

The opportunity of broadcasting the virtual 3D repetition of the bounce permits to the spectator to get more information about the ongoing match. Due to the velocity of the ball and the limits of the broadcasting cameras, it's no viable to reproduce the last recorded frames and detect the bounce point of the ball. Only making a

[14] www.eitb.com

reconstruction of the followed track it's possible to determine this point. Therefore, the user experience is enhanced using the results of the described system.

5 Conclusions

In this work a low-cost automatic ball bounce detector and 3D virtual replay generator is proposed for sport event broadcast. The central engineering trade-off choice approach has been to reduce the bounce detection system set-up and hardware requirements to unique broadcast-type camera per border line as well as to reduce the system software to quasi real-time performance. The challenge of hardware complexity reduction keeping accuracy in results can be considered the main technical contribution of this work.

The algorithm introduces an additional advantage which makes it more flexible in terms of different possible camera operating location. Contrary to other approaches, the camera and the playground plane can form any angle, since the necessary transformations for calculating the point of impact in real coordinates are effective and accurate. For this reason, the system has no camera location constraint, provided that line view is not occluded.

The typical human eye incertitude in the appreciation of distant actions(a few meters from linesman to bounce events point), with millisecond duration (because of the ball speed) is considerably higher than the score of the presented system. Obtained results show that the measure errors are close to the demanded range in order to standardize the system for Basque Pelota events.

The use of 3D virtual reality for controversial action replays in sport event broadcasting enhances audiences and TV spectators' visual experience. Due to the reduction of the production costs, this contribution represents a new opportunity for less-known traditional/regional sport events to use this technology, as well as, for small producers, organizers and broadcasters to compete with well-known competition organizers and expensive broadcasting rights owners.

The experience of having developed a research effort applied to real world deployment for sports events has materialized a complete solution covering he whole production chain. Technical specifications and hardware requirements for a system that has to be included in a real world implementation are stronger than the ones required for a system with not so close relation with real world applications. Even more, several variables are no more under control of the researcher, which makes the work harder.

This work has been granted with the patent EP2455911 **Method for detecting the point of impact of a ball in sports events**.

Acknowledgements The authors would like to acknowledge the collaboration offered by G93 Telecomunicaciones[15] (Audio-Visual, Computer and Graphic Services for Television) and EiTB (Basque public broadcaster) for the help offered in the system development, test and broadcast processes.

The authors are also grateful for the collaboration offered by ASPE[16] (ASPE Jugadores de Pelota) in providing access to its professional pelota player training sessions and in advising in game

[15] www.g93.es

[16] www.aspepelota.com

rule issues as well as for the financial support offered by research project programs of the SPRI[17] (Society for Industrial Promotion and Restructuring of Basque Country).

Finally, the authors would like to thank the rest of *Begira* research team: Maider Laka, Julen García and Aritz Legarretaetxebarria. Also, Javier Barandiaran and Iñigo Barandiaran for their advice and the collegues of *Digital Television and Multimedia Services* department for the unconditional help offered.

References

1. Ahn S (2004) Least squares orthogonal distance fitting of curves and surfaces in space. In: Lecture notes in computer science. Springer (2004). http://books.google.es/books?id=we4cHJBFzLwC
2. Alper Yilmaz OJ, Shah M (206) Object tracking: a survey. ACM Comput Surv 38(4). doi:10.1145/1177352.1177355
3. Cavallaro R (1997) The foxtrax hockey puck tracking system. IEEE Comput Graph Appl 17:6–12
4. Erik Cueva DZ, Rojas R (2005) Kalman filter for vision tracking. Freie Univ., Fachbereich Mathematik und Informatik
5. Yan F, Christmas W, Kittler J (2005) A tennis ball tracking algorithm for automatic annotation of tennis match. In: British machine vision conference, vol 2, pp 619–628
6. Sullivan GJ, Topiwala PN, Luthra N (2004) The h.264/avc advanced video coding standard: overview and ntroduction to the fidelity range extensions. In: SPIE 49th annual meeting optical science and technology, international society for optics and photonics, pp 454–474
7. Gopal Pingali Agata Opalach YJ (2000) Ball tracking and virtual replays for innovative tennis broadcasts. In: Proceedings 15th international conference on pattern recognition, 2000, vol 4. IEEE, pp 152–156
8. Haines E, Akenine-Moller T (2002) Real-time rendering, 2nd edn. AKPeters
9. Innovations HE (2013) Hawk-eye accuracy and believability. http://www.hawkeyeinnovations.co.uk/
10. Isard M, Blake A (1998) Condensation—conditional density propagation for visual tracking. Int J Comput Vis 29(1):5–28
11. Inurrategi ML, Olaizola IG, Ugarte A, Macia I (2008) Tv sport broadcasts: real time virtual representation in 3d terrain models. In: 3DTV conference: the true vision—capture, transmission and display of 3D video, 2008. IEEE, pp 405–408
12. Miller FP, Vandome AF, McBrewster J (2009) MPEG-2: lossy compression, video compression, audio compression (data), ATSC (standards), MPEG transport stream, MPEG-1 audio layer II, H. 262/MPEG-2 Part 2, MPEG-4, advanced audio coding. Alpha Press. http://dl.acm.org/citation.cfm?id=1822909
13. Owens N, Harris C, Stennett C (203) Hawk-eye tennis system. In: International conference on visual information engineering, VIE 2003. IET, pp 182–185
14. Naidoo WC, Tapamo JR (2006) Soccer video analysis by ball, player and referee tracking. In: Proceedings of the 2006 annual research conference of the South African institute of computer scientists nd information technologists on IT research in developing countries. South African Institute for Computer Scientists and Information Technologists. pp 51–60
15. Richardson IE (2003) The H.264 advanced video compression standard, 2nd edn. Vcodex Limited, UK
16. Richardson IEG (2002) Video codec design: developing image and video compression systems. Wiley
17. Wu W (2010) Tennis touching point detection based on high speed camera and Kalman filter. Clemson University
18. Yu X, Leong HW, Xu C, Tian Qi (2006) Trajectory-based ball detection and tracking in broadcast soccer video. IEEE Trans Multimedia 8(6):1164–1178
19. Yu X, Xu C, Leong HW, Tian, Qi, Tang Q, Wan KW (2003) Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In: Proceedings of the eleventh ACM international conference on Multimedia. ACM, pp 11–20

[17] www.spri.es

4.2 Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology

- **Autores: Mikel Labayen** and Ricardo Vea and Julián Flórez and Naiara Aginako and Basilio Sierra
- **Revista:** IEEE Access
- **Volumen:** 9
- **Páginas:** 72398-72411
- **Año:** 2021
- **Editor:** IEEE

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 27, 2021, accepted May 5, 2021, date of publication May 11, 2021, date of current version May 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079375

Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology

MIKEL LABAYEN[1,3], RICARDO VEA[1], JULIÁN FLÓREZ[2], (Member, IEEE), NAIARA AGINAKO[3], AND BASILIO SIERRA[3]

[1] Smowltech, 20009 Donostia, Spain
[2] Vicomtech Research Center, 20009 Donostia, Spain
[3] Computer Sciences and Artificial Intelligence Department, University of the Basque Country, 20018 Donostia, Spain

Corresponding author: Mikel Labayen (mikel.labayen@ehu.eus)

This work was supported by the Spanish Ministry of Sciences, Research and Universities (Ministerio de Ciencia, Innovación y Universidades (MCIU)/Agencia Estatal de Investigación (AEI)/Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea (UE)) under Grant RTC-2016-5711-7.

ABSTRACT Identity verification and proctoring of online students are one of the key challenges to online learning today. Especially for online certification and accreditation, the training organizations need to verify that the online students who completed the learning process and received the academic credits are those who registered for the courses. Furthermore, they need to ensure that these students complete all the activities of online training without cheating or inappropriate behaviours. The COVID-19 pandemic has accelerated (abruptly in certain cases) the migration and implementation of online education strategies and consequently the need for safe mechanisms to authenticate and proctor online students. Nowadays, there are several technologies with different grades of automation. In this paper, we deeply describe a specific solution based on the authentication of different biometric technologies and an automatic proctoring system (system workflow as well as AI algorithms), which incorporates features to solve the main concerns in the market: highly scalable, automatic, affordable, with few hardware and software requirements for the user, reliable and passive for the student. Finally, the technological performance test of the large scale system, the usability-privacy perception survey of the user and their results are discussed in this work.

INDEX TERMS Biometric authentication, cloud computing, computer vision, data science applications in education, distance education and online learning, machine learning, security, computer vision.

I. INTRODUCTION

There is no doubt that online learning has been gaining popularity throughout the past years. This phenomenon is not surprising given that online learning allows education institutes to operate at a lower cost and with greater reach-out to more students. Educational institutions are offering courses online to leverage the benefits of online learning. This is especially so since the advent of Massive Open Online Courses (MOOC). On the other hand, COVID-19 has been a challenge for traditional institutes offering face-to-face teaching, and these institutions have had to migrate (in a very short period of time) to a fully online education model forced by the pandemic situation. However, online learning implementation presents challenges.

E-learning has a serious deficiency, which is the lack of efficient mechanisms that assure user authentication, in the system login as well as throughout the session. Especially for online certification and accreditation, the training organizations need to verify that the online learners who completed the learning process and received the academic credits are precisely those who registered for the courses. Inadequate methods of identity verification affect the reliability of credentials and certification earned online.

Without certainty of the authenticity of the online learner's identity, the aspiration towards fully online education is stymied and the evaluation of the knowledge and skills obtained by the online learner is unreliable. In order to prevent compromising the credibility of online accreditation,

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

validation must be carried out in a constant or continuous manner. At the same time, validation should be non-invasive and non-disruptive, and does not distract the learning process.

Online proctoring, generally refers to proctors (humans) monitoring an exam over the internet through a webcam. It includes as well the processes, occurring at a distance, for authenticating the examinee as the person who should be taking the exam. Online proctoring was first introduced by Kryterion [1], [2] in 2006, marketing it as a technological solution in 2008. Since then, several other organizations have followed Kryterion's lead creating more capable technology-based alternatives, which are gaining attention, such as online proctoring.

Nowadays, there are commercial solutions in the market as well as research publications that try to solve this problem. Some of them only authenticate the identity, others monitor, some in real time, others record the sessions. Some cover only exams or specific activities. Some are totally human based solutions (non-scalable) or fully automatic ones (non-reliable). There are also a few scientific approaches which develop the idea of combining some of the cited functionalities. However, there is no comprehensive and reliable solution which combines multi-biometric continuous authentication with continuous visual and audio monitoring, with device activity monitoring and lock-down options and human supervision (only when required) to guarantee 100% reliable results.

In this work we present a new system which gives commercial solutions to all that was needed. It is based on web applications which offer a continuous authentication identity service of online students through a constant biometric (face, voice, typing) recognition system (biometric traits cannot be lost, stolen, or recreated), as well as automatic continuous proctoring through automatic image and audio processing (device monitoring & lock-down and inappropriate behaviour detection) allowing online courses to gain value of what benefits both institutions and students. This solution is based on a high accuracy biometrics recognition and digital signal processing algorithms and it is complemented with human supervision for those situations in which the automatic algorithms are not able to determine reliable results. It can be used to continuously authenticate the learners, either throughout the entire learning process, or only at certain sensitive stages of e-learning. It is contactless and needs only a low level of user collaboration. In addition, the whole system is based on cloud computing technologies, which removes geographical and technological barriers for online learning providers.

The article is organized as follows. Section II gives an overview of some relevant related works and highlights the main differences with our approach. Section III describes the whole system overview and workflow. Section IV contains a scientific-technical description of core modules. Section V presents system tests to measure the algorithms' performance as well as a survey made for user experience evaluation. Section VI presents the results of the tests. Finally, section VII draws the conclusions and presents future works.

II. RELATED WORK

The ability to authenticate and monitor online users is becoming more important due to the increase of the internet world (e-learning, e-banking, e-gambling, e-government). Since first human based online proctoring systems, various fully or semi-automatic authentication and proctoring technologies based on biometric features have appeared in the last few years. Biometrics has proved itself to be one of the best methods for recognizing people based upon physiological or behavioural characteristics [3]. These technologies can be divided into two categories: those that are based on physical characteristics and those that are based on behaviour characteristics. The former includes face recognition, fingerprint scanners, iris scanners, vein matching, etc. The latter includes voice recognition, handwriting recognition, keystroke dynamics, etc. It is proved that no technology will provide the right answer on its own, but that the combination of different solutions will come up with the appropriate functionality depending on customer needs. In addition, most remote authentication proctoring technologies involve some level of human intervention for fully reliable service, thereby putting limitations on scale.

These biometric technologies have been widely used for various purposes, and they have become more and more common in our daily lives. However, very few of them have been successfully adopted for online learning validation.

A. COMMERCIAL SOLUTIONS

Some initial approaches have been brought to market as commercial solutions. The following is an overview of these services:

1) **Fully Live Online Proctoring:** Students are on video and watched remotely by a live proctor. Live proctoring is a live online service for students taking exams online. After making an appointment, the students are taken to the online proctoring room where they will connect with a live proctor from one of the two online proctoring centres via their web cameras. The students connect their screen to the proctor. This allows the proctor to see their computer screen. The proctor asks them to show a photo ID and to answer a few questions about themselves in order to verify they are in fact the right student. During the exam, the proctor looks at the student directly through a webcam. It is a secure and complete solution for exam proctoring, but since it is a non-automatic solution, it cannot deal with continuous identification during all learning process. Furthermore, it needs a high speed internet channel to transmit video data, probably unaffordable for different parts of the world and it is not passive for students. Some commercial solutions in the market are ProctorU [4], Examity [5] and Software Secure - PSI [6].
2) **Recorded and Reviewed Proctoring:** Sessions are recorded as the computer monitors students. A human can then review the video at any time afterward.

In these systems, students use their own computer and a webcam to record assessment sessions, the student and the surrounding environment are recorded during the entire exam. Instructors can quickly review details of the assessment, and even watch the recorded video. Recorded proctoring has the same limitations as live proctoring. In addition, it is a passive system. However, nobody analyzes the videos, so teachers must watch all of them in order to detect undesirable behaviours and maintain the live proctoring advantages. Some commercial solutions in the market are Kryterion [1], ProctorExam [7], Respondus [8], Remote Proctor [9], ProctorCam [10], B virtual [11] and Learner verified [12].

3) **Fully Automated Solutions:** The computer monitors students, it authenticates them and determines whether they are cheating. These are automatic and passive solutions. They just cover the beginnings of exams and work submission processes. However, users must be totally active in this kind of system (they must type a predefined paragraph and take an ID photo themselves). In addition, this kind of system does not cover all the learning process continuously. Some commercial solutions in the market are Proctorio [13], ProctorTrack [14], Comprobo [15], Sumadi [16], ProctorFree [17], HonorLock [18] and ExamSoft [19].
   a) **Authentication technologies:** Recognition technologies are used to authenticate a student based on a prior examination of some physical feature. They are typically built upon a before/during/after analysis to verify that the same student who initially registered for the course was actually the same student who took the exam. Commonly-known recognition technologies include facial, fingerprint, or voice recognition. In the last year, new biometric procedures such as keystroke dynamics (it recognizes typing patterns based on rhythm, pressure, and style) are gaining popularity. It is likely that recognition technologies will be most effective when used with some combination of other technologies available.
   b) **Monitoring technologies:**
      i) Webcams and microphones are one of the original technologies used to replace a live proctor and are present in most remote exam proctoring solutions on the market. They can record individual students when the camera is part of the computer, or groups when the camera is placed in a classroom. They can monitor the behaviour of the students, whether they are cheating, receiving help from other students, using mobile devices, books…Webcam/Microphone technologies often require significant storage capabilities so that video records can be reviewed if necessary.
      ii) Computer lockdowns are able to monitor the activity carried out by the student within their computer preventing them from ‘‘surfing the internet’’ while taking a test. This monitoring will be done only and exclusively when the student is doing an activity that can be evaluated.

None of the cited commercial solutions provides a multi-biometric authentication solution or continuous authentication/proctoring service (based on automatic analysis) through the whole learning course (not only exams). In addition, this work presents a completely new commercial approach to overcome barriers such as low-speed internet connection (using data samples, not continuous heavy video signals) or costly extra HW/SW requirements (using non-installable and fully integrated in LMS web applications).

B. SCIENTIFIC AND ACADEMIC APPROACHES

1) TECHNICAL WORKS

Nowadays, although there are still some non-biometric based authentication approaches [20], the latest attempts for online student authentication automation tends to use biometric technologies; facial [21]–[26], fingerprints [27] or typing [28], [29]. On the other hand, some approaches try some combination of them, such as face and voice [30] or face, voice and typing [31], [32]. All the approaches are focused mainly on student authentication without providing proctoring service.

It is through facial authentication complemented with other biometrics such as voice or typing recognition, that an opportunity appears in e-learning to verify the absence of frauds while the students do their activities on the platform.

The main novel contribution of the work we present in this article includes a completely new combination workflow of three main biometrics providing a continuous and non-intrusive authentication service. It also adds new automatic and continuous proctoring features based on image and audio signal processing to the system. Furthermore, it integrates computer activity monitoring and lock-down possibility and, finally, it even complements the service with automatic alarms which trigger minimal human supervision, guaranteeing the reliability of results.

Finally, the recent concern for safety and privacy has also provided recent research on this topic related to online proctoring [33].

2) USER EXPERIENCE RELATED WORKS

On the other hand, very few works completed the research about teachers and student user experience with this kind of authentication and proctoring approaches. One of them completed the research about the implementation of facial verification into education with a successful positive result [34]. The objective was to guarantee students authentication and to know exactly the amount of time that they spend in front of the computer reading or realizing their virtual activities.

TABLE 1. Commercial solutions vs SMOWL (solution described in this article). Service characteristics: 1-Authentication during whole exam or session; 2-Multi biometric authentication (at least 2 different); 3-Exam monitoring; 4-Continuous (full course) monitoring; 5-Dishonest behaviour detection; 6-Totally Passive and non-intrusive system; 7-Automatically analyzed results; 8-100% guaranteed and reliable results; 9-Personalised alarms; 10-Human real-time proctor; 11-Device monitoring. Technical features: 12-Scalable system; 13-Flexible access to students - no scheduled; 14-No extra SW/HW installation required for authentication and proctoring; 15-Works with low-speed connection; 16-Fully integrated in institution LMS; 17-Multi-Browser & device. Legal aspects: 18-EU-hosted solution; 19-GDPR compliance. ✓- Yes | X- No.

	Service characteristics											Technical features						Legal asp.	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Fully Live Online Proctoring																			
ProctorU	✓	X	✓	X	✓	X	✓	✓	X	✓	X	X	X	X	X	✓	✓	X	X
Examity	✓	X	✓	X	✓	X	X	✓	X	✓	X	X	X	X	X	X	X	X	X
PSI	✓	X	✓	X	✓	X	X	✓	X	✓	X	X	X	X	X	X	X	X	X
Recorded and Reviewed Proctoring																			
Proctoexam	✓	X	✓	X	✓	X	X	✓	X	✓	X	X	✓	X	X	X	X	✓	✓
Kryterion	✓	X	✓	X	✓	X	X	X	X	✓	✓	X	✓	X	X	X	X	X	X
Remote Proctor	✓	X	✓	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Proctorcam	✓	X	✓	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B Virtual	✓	X	✓	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fully Automated Solutions																			
Proctorio	✓	X	✓	X	✓	X	✓	✓	X	X	X	✓	✓	X	X	X	X	X	X
Proctortrack	✓	X	✓	X	X	X	✓	X	X	✓	✓	✓	✓	X	X	X	X	X	X
Respondus	✓	X	✓	X	✓	X	X	X	X	X	X	X	✓	X	X	✓	✓	X	X
Comprobo	✓	X	✓	X	✓	X	✓	X	X	X	X	✓	✓	X	X	✓	X	X	X
Sumadi	✓	X	✓	X	X	X	✓	X	X	X	✓	✓	✓	X	X	X	X	X	X
Proctorfree	✓	X	✓	X	X	X	✓	X	X	X	X	✓	✓	X	X	✓	✓	X	X
HonorLock	✓	X	✓	X	✓	X	✓	X	X	X	✓	✓	✓	X	X	X	X	X	X
ExamSoft	✓	X	✓	X✓	X	✓	X	X	X	✓	✓	✓	X	X	X	X	X	X	X
SMOWL	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓

In the same way, a facial authentication mechanism was also presented. This insured that the students are not impersonated to improve their marks in virtual tests [35].

III. SYSTEM OVERVIEW

The system we present in this work aims to provide a practical cyber-security solution for both a) continuous online user identification (using biometric technology) and b) monitoring using automatic signal processing and a computer monitoring system. The authentication process is based on automatic authentication of facial images (captured by webcams), audio clips (captured by the microphone) and keystroke dynamics (captured by the keyboard), checking that it is the person that it really should be during the entire online interaction. The monitoring process is supported by webcams and microphones too, checking continuously that the student is not making any inappropriate behaviour (using forbidden devices and applications, receiving help...). It also locks down the computers (with a previous installation in the learner computer and consent) during exams or training sessions preventing the user from visiting web pages or other documents while performing the course.

The system can be used for any online user authentication but it is specialized in the institutions that offer online courses providing training and degree certification, including verified MOOCs and corporate training for employees. This system can help e-learning providers in their objective to be awarded credit by Quality Educational Agencies for their courses by seeking traceability of evidence of student authenticity and their behaviour. It can be used to track the continuous authentication of the student in all or in sensitive stages of

TABLE 2. State-of-the-art solutions vs SMOWL (solution described in this article). Authentication method: 1-Face recognition; 2-Voice recognition; 3-Typing recognition; 4-Continuous authentication during whole session (not only at the beginning). Proctoring-Monitoring method: 5-Image processing; 6-Audio processing; 7-Screenshots capture; 8-Device information capture (active window, open processes, peripherals devices, copy/paste commands...). Proctoring-Device Lock-Down: 9-Device lock-down. Guarantee: 10-Human supervision to clarify doubts providing 100% guaranteed and reliable results. ✓-Yes | X- No.

	Auth.				Monit. & Proctor.					%
	1	2	3	4	5	6	7	8	9	10
[21] to [26]	✓	X	X	X	X	X	X	X	X	X
[28]	X	X	✓	X	X	X	X	X	X	X
[30]	✓	✓	X	X	X	X	X	X	X	X
[31] [32]	✓	✓	✓	X	X	X	X	X	X	X
SMOWL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

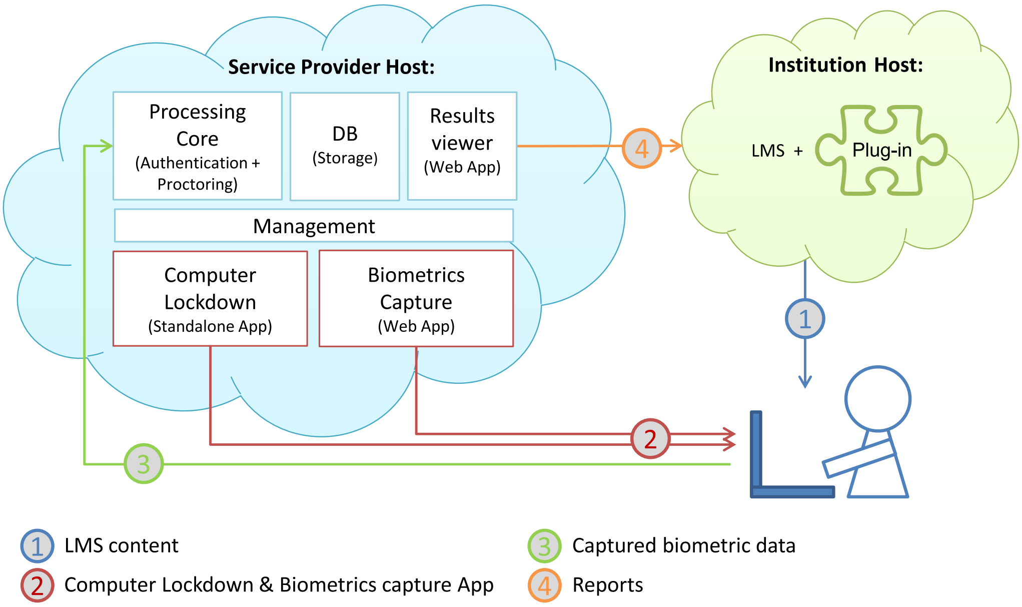

Service Provider Host:
Processing Core (Authentication + Proctoring)
DB (Storage)
Results viewer (Web App)
Management
Computer Lockdown (Standalone App)
Biometrics Capture (Web App)
Institution Host:
LMS + Plug-in
1
2
3
4
1 LMS content
2 Computer Lockdown & Biometrics capture App
3 Captured biometric data
4 Reports

FIGURE 1. **Authentication and proctoring system set-up.**

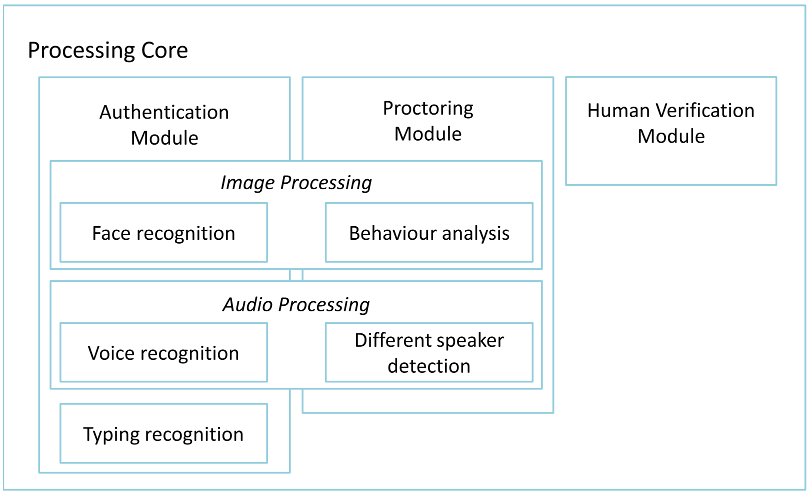


FIGURE 2. **Processing core description.**

e-learning. Figure 1 shows general set-up of the system and Figure 2 details the processing core description.

The complete system workflow is embedded in cloud computing applications, and can be used anywhere, removing geographical and technological barriers. The general scheme of operation is as follows and is given in more detail in Figure 3:

1) The system is integrated into the virtual campus of the training centre (available for different LMS platforms).
2) The training centre sends a code (unique student identifier) with an image of the student to register in the system. According to system data privacy policy, the system works with images, audio clips…not identities, so it lacks connection with the student personal data such as name, age or address [36].
3) The first time the student enters the virtual campus the system takes biometric samples (picture, short speech, predefined paragraph typing) which will help us create the tracking biometrical model.
4) Thereafter, whenever the student is connected to work, biometric samples will be taken randomly and continuously. This data is sent to servers in the cloud. The online management module stores and analyzes the data which is compared with the biometrical model that has been created previously for authentication purposes and analyzed to detect inappropriate behaviours. All storage, analysis and results report and alarm creation tasks are executed in online servers, making the integration, support and maintenance tasks for institutions easier and more transparent. During this period, the computer lockdown module can be activated for monitoring purposes.
5) The result leads to an individual user report that is updated constantly and to which the training centre has access.

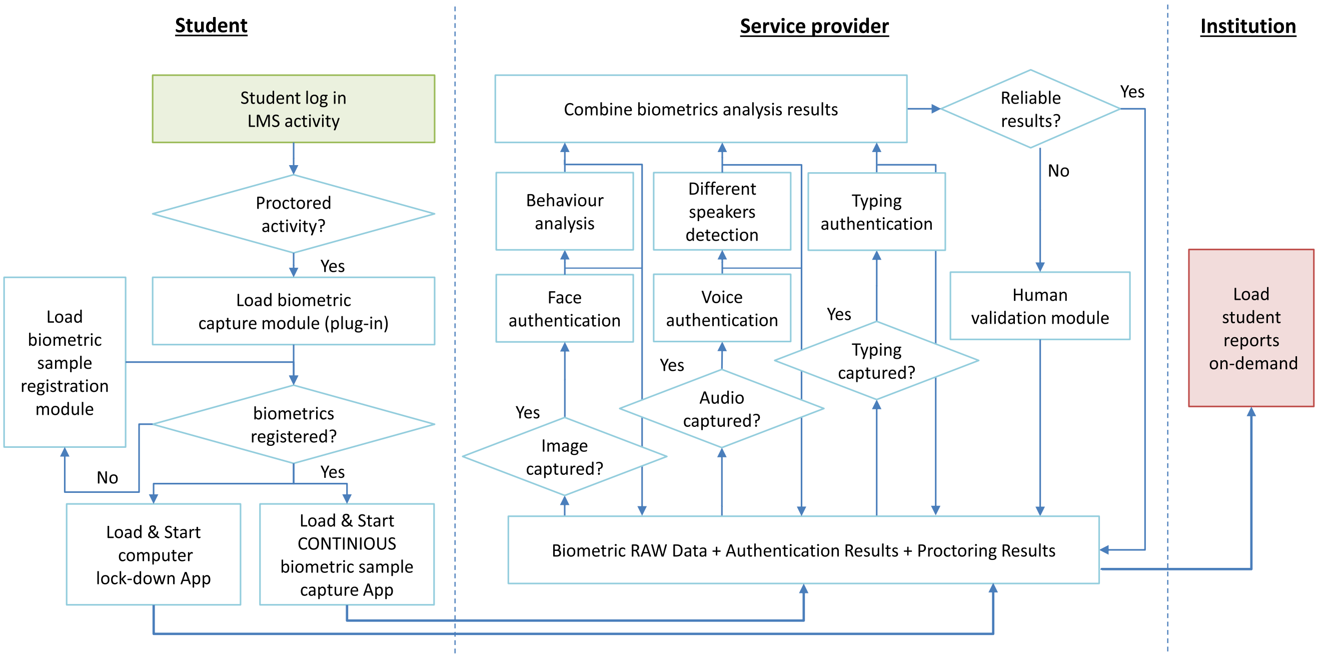

Student
Service provider
Institution
Student log in LMS activity
Proctored activity?
Yes
Load biometric capture module (plug-in)
Load biometric sample registration module
biometrics registered?
No
Yes
Load & Start computer lock-down App
Load & Start CONTINIOUS biometric sample capture App
Combine biometrics analysis results
Reliable results?
Yes
No
Behaviour analysis
Different speakers detection
Typing authentication
Face authentication
Voice authentication
Human validation module
Yes
Typing captured?
Yes
Audio captured?
Yes
Image captured?
Biometric RAW Data + Authentication Results + Proctoring Results
Load student reports on-demand

FIGURE 3. System workflow.

The key characteristics of the system are:

1) **Continuous and not scheduled system.** Proctoring and authentication processes are carried out throughout the entire session, not only when users log in. Furthermore, in the e-learning case, it can follow every session of the course, not only the assessments. It is very flexible. Service is given 24/7, anywhere. Previous schedule is not required.
2) **Passive & non-intrusive system.** The system offers a passive system for students when taking photos, audio clips or keystroke pattern. It does not need the collaboration of the student and it is contactless. For this reason, in the case of images, it properly works when the pose/appearance/complements/expressions of the students or the light conditions of the room are not controlled (in the wild), getting low-contrast images with partial occlusions due to wrong position or the appearance/compliments/expressions variations of the student. Regarding audio clips, the microphone only records when it detects some noise, nothing if the student is in silence. The clips are later analyzed and if voice is detected in the recording it is compared with the data gathered during registration of the student, to validate their identity, or to detect cheating when there are different voices in the recording.
3) **Automatic and scalable:** All capture, verification, data management and monitoring report modules are carried out with cloud computing technology as services in the cloud. Photos and patterns are taken automatically and randomly and compared with the biometric model made during registration. This scalable automatic setup makes it possible to bring this solution to overcrowded scenarios such as MOOCs.
4) **Few requirements for the end user.** Cloud-based (SaaS) automatic solution. Needed Hardware - Software (HW/SW): basic webcam, microphone, keyboard and any updated browser. Final users do not have to install anything. This system works over any device, platform, OS and browsers with no installation needed.
5) **Automatic analyzed results.** 100% guaranteed results with custom alarms. If automatic validation cannot be confirmed (if the pictures or audio clips do not compile with the quality needed to allow the system to automatically validate the student), a manual checking by staff will be set to certify the results 100%.
6) **Fully integrated in customer LMS.** It can be integrated in any Learning management system (LMS) using a general API but it has a specific plugin for Moodle, Moodlerooms, Blackboard, OpenedX, Canvas, etc. (most used LMS).
7) **Secure.** Data is transmitted under secure internet protocol and stored in safe cloud servers.
8) **Private.** The user's identity remains protected because we only handle data that are not linked to identities but to user codes provided by the online entity.

A. DATA CAPTURE AND STORAGE MODULE

This module captures data from the student webcam, microphone and keyboard. The core of this application has been developed using the latest HTML5 standard implementation in web browsers. The application is downloaded into the student's terminal and executed without any installation needed. Whenever the user is connected to the course, quiz or specific exercise into LMS, pictures, audio clips and keystroke dynamics samples will be taken randomly and continuously with predefined mean periodicity. This data is sent to servers in the cloud, through a SSL encrypted channel, with the user identification code. The system online management module stores and analyzes the images.

B. AUTHENTICATION MODULE

Once all data is stored in cloud servers, it is compared with the biometrical model, linked to student's identification code, which has been created at registration time and has been updated with recent positive data. The result is stored in the system database. The system recognition and training algorithms are developed using the latest algorithms in artificial intelligence (explained in Section IV) which are improving constantly their recognition precision and robustness facing light, position and student appearance (physical changes and complements such as hat, glasses...) change problems, noise in audio clips and variability in typing samples. The authentication result is a combination of each biometric authentication module result (face, voice and typing).

C. PROCTORING AND COMPUTER LOCK-DOWN MODULES

During monitoring sessions, the captured image and audio clips (which have been used for authentication purposes) are processed with different techniques in order to detect inappropriate behaviour of students during e-learning activities. For this reason, the system is able to detect if the student is receiving help (by phone, help from presential friend...) or is checking forbidden documentation (books, other devices connected to the internet...). All these actions can be strictly forbidden in some face-to-face learning activities according to the institution code of honour.

In addition, attempts to cheat are detected and reported if any student tries to trick the system, such as mounting a photograph in front of the camera or replacing the image of the ID card with someone else's. Attempts to insert another image or video signal into the camera are also detected.

On the other hand, the system contains a computer lockdown module. During all the online session, a computer lockdown module (Section IV) will monitor the computer of the student detecting connected peripherals, active windows, computer information (HW/SW), executing programs or processes, browsing history/webs and copy-paste commands. All the information captured in each session is stored in the database.

D. HUMAN VERIFICATION MODULE

As part of the quality warranty, a random data and results auditory must be set. This task will test try the quality assurance mechanism definition and implementation with a huge number of students connected at the same time. It will be based on a random data cross-verification (same images, voice and keystroke patterns validated by different persons) of images, voice and keystroke samples captured during the session with registered data. Besides, when the quality of the photos or audio does not reach the threshold needed, a human verification is made by trained staff delivering a 100% reliable verification of the student.

E. REPRESENTATION MODULE OF THE RESULTS

Final results are presented by the data representation module. It creates graphic charts and tables on demand, 24h/365d, as a dynamic web page. The final reports can be downloaded or printed in different formats. In addition, the data representation module also generates automated alarms when some predefined prohibited behaviour happens.

IV. AUTHENTICATION AND PROCTORING MODULES IMPLEMENTATION

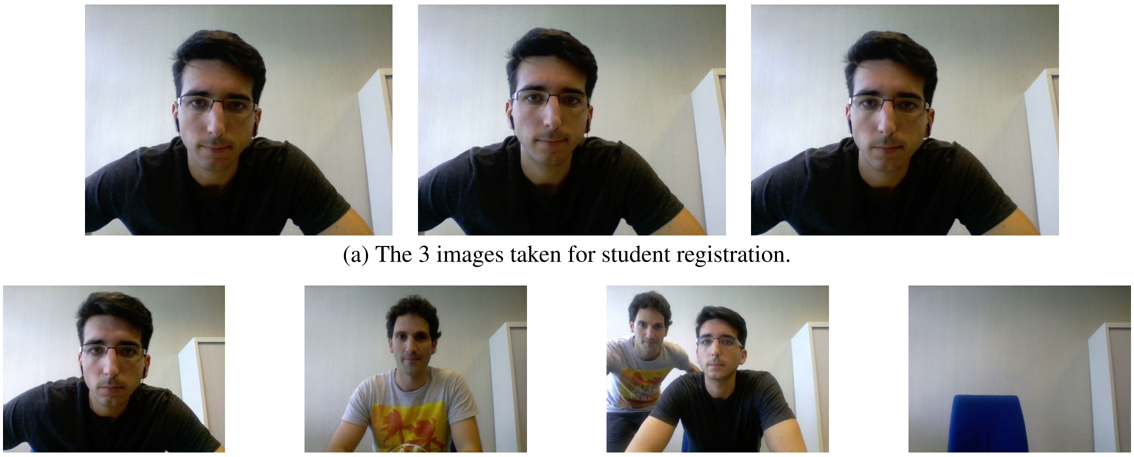
As explained in the previous sections, the system presented in this work contains artificial intelligence-based modules for user authentication as well as computer lockdown technologies for device monitoring. In this section the scientific algorithm behind authentication modules and technology and functionalities of the computer monitoring are explained and referenced in depth.

A. FACE DETECTION AND RECOGNITION

This system includes a facial detection and recognition module through a biometric model created using registration time face pictures. The module output results are clustered in five groups determining: a) If there is someone in front of the webcam or not, b) How many people (if any) are in front of the webcam, c) If one of these people is the person who should be in front of the screen, d) when only one person is in the image, whether this person is the person it should be, e) If the person who it should be is not involved in any inappropriate behaviour (book or electronic device use). Some examples are shown in Figure 4.

There are different approaches for face detection in the literature [37]. However, few of them are robust enough when dealing with variation in pose and lighting of captured images (remember that pictures are taken without student attention and randomly). The facial detection procedure presented in this work is based on the FaceBoxes methodology [38]. This methodology is known for being the most common "Deep Learning" based technique whose optimal deployment is based on use of GPUs. This methodology obtains better results in the Face Detection Data Set and Benchmark (FDDB) benchmark (Jain and Learned-Miller, 2010) than other methodologies tested in the development process of this module.

The image processing and authentication processes takes [39] as the base reference method for the extraction and normalization of facial texture. This algorithm contains the following subtasks: (1) face detection, (2) face characteristic points detection in the facial region and (3) deformable parametric 3D facial model adjustment based on the detected points. However, the requirement of system passiveness makes it necessary to have continuous improvements in the detection and authentication algorithm to deal with high



(a) The 3 images taken for student registration.

(b) Correct student authentication.

(c) Another person.

(d) More than one person.

(e) Nobody in front of the device.

FIGURE 4. Authentication and proctoring system captured and analyzed image examples.

variability of input images. Starting in this reference work, a series of improvements have been added:

1) **Pose and expressions correction:** A new method, called as M3L (Multi- level, Multi-modal, Multi-task Learning) [40], is used to improve efficiency in face points and other facial attributes detection (gestures of the face and eyes). M3L addresses the problem of extracting all these facial and ocular data through a hierarchy of neural networks using existing correlations between the data. Furthermore, a new multi-level deformable 3D facial model adjustment distributes the deformation error in an equitable way, distinguishing three stages with different levels of priority in estimation of (from greater to minor): (1) pose, (2) interpersonal deformations (user-specific facial shape) and (3) intra-personal deformations (deformations due to facial expressions).
2) **Aspect normalization, feature selection and classification:** The extraction of biometric features through a deep neural network [41] has been improved training a database with 10M of images of 100K individuals with great variability of appearances and facial shapes, lighting, facial expressions, accessories and poses (Guo *et al.*, 2016).
3) **Normalization of the lighting:** The procedure of normalization of the lighting has been carried out with a hierarchical method in which the facial region as a whole as well as specific and normalized regions of the face are analyzed. This normalization is performed using the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [42], which equalizes the image locally, highlighting the contrasts, applied to each RGB color channel.
4) **Robustness against partial occlusions:** Occlusions detection is based on the MobileNet-SSD neural network [2], [43]. Combining this person detector and the face detector, the system increases its robustness in detection when (at least) partial face occlusion is occurring. This people detector (body) is more robust than the face detector in these cases. Therefore, if a person is detected, but not a face, it is more likely that this face is at least partially occluded. In this case, the face detection alarm is considered. Additionally, the methodology proposed in [44] has been implemented and adapted to the framework of the needs of the project to handle the occluded normalized facial images. The facial detection returns more partially occluded facial cuts than desirable ones. Normally these occlusions are given either by the user's hands in front of the face or because the camera is only pointing to the top-half of the face. This occlusion negatively influences the later stages of facial point detection and biometric vector extraction. This system includes a facial image synthesis from Generative Adversarial Network [45], which fills the occluded part with close facial features obtained from the trained model. In this way, the negative impact of occlusion can be reduced.

B. VOICE DETECTION AND RECOGNITION

This module implements a continuous voice detection and authentication algorithm. The developments are based on the Kaldi tool [46] and the implementation of the method of [47]. Both include tools for the development of the biometric model, the vector representation of each speaker's characteristics. The algorithm works on four tasks:

1) **Analysis, interpretation and normalization of audio by VoIP:** Since the data used in VoIP (technology in which this system is based on) use the G.711 codec with a 64 kbps bit rate, which implies a loss of important information in order to compress the audio signal, all training data from the available acoustic databases are transformed into this encoding and format. In this way, the training and evaluation audio matches were obtained in the different frequency

ranges. Signal pre-processing is integrated to discard that acoustic segments that do not contain speech (silence, music or noise). The final version of the VAD vocal activity detection module has been developed using GMM Gaussian mixture models and processing functions proposed in the Kaldi code tool. A total of 3 model training level were performed. The difference between each of them is based on the transformation of training data for greater robustness versus the high acoustic variability of the application scenario.

2) **Background and speaker modelling:** The speaker modelling is based on d-vectors or speaker embeddings using deep neural networks. This solution offers better performance in terms of robustness and accuracy. The implementation follows the solution presented by Google in 2018 [47]. In this approach, a recurrent neural network based on LSTM cells is generated. It receives an acoustic characteristic of a specific audio (Mel filter bank) as input and returns its d-vector. Once the training is finished, the neural network can be used to generate d-vectors from the acoustic characteristics of the speaker. Then, a centroid is generated, which is considered as the speaker's biometric footprint.
3) **Patterns comparison:** For a verification or identification process, given a vector of acoustic characteristics and its associated d-vector, they are compared with the centroids of each of the speakers in a new similarity matrix.
4) **Speaker segmentation on streaming audio:** This diarization system employs d-vectors or speaker embeddings and an agglutination model based on recurrent neural networks [38]. This approach aims to overcome the traditional agglutination approach problems, which work with the sentences individually and independently, it being difficult to benefit from the information provided by large amounts of labelled data. This system is based on the work presented by [48]. An independent text announcer recognition network is used to extract d-vectors or speaker embeddings from 240 millisecond windows and a 50% overlap. A vocal activity detector based on Gaussian models is used to eliminate speechless parts and split the signal into segments less than 400 milliseconds. These segments are converted to d-vectors and included in the RNN network based diarization system.

C. TYPING RECOGNITION

Keystroke dynamics are an effective behavioural biometric, which captures the habitual patterns or rhythms an individual exhibits while typing on a keyboard. According to neurophysiological analysis [49], these typing styles are idiosyncratic, in the same way as handwriting or signatures, due to their similar governing neuronal mechanisms. For this reason, they can be used to authenticate an individual.

The system presented in this work applies keystroke dynamics in dynamic text, that is, the analysis occurs for any text that is typed by the user and continuously. Keystroke dynamics in static text requires less effort to be implemented and it also reached lower error rates in the literature [50]. However, a dynamic text analysis [51] is necessary to keep final student passiveness in the authentication process without bothering them by asking them to type a predefined paragraph (usually not related to the e-learning activities in progress). This approach considers the fact that the keystroke dynamics of one person may vary in different psycho-emotional states. For example, researches noticed [52] that tired people usually type more slowly and make more mistakes, for this reason, every typed sample is stored to make the recognition model more robust.

Two distinctive processes are involved in the keystroke dynamics analysis module:

1) **Feature extraction:** The extracted features (detailed timing information [53]) are time differences between the instants in which:
   a) DT: A key is pressed and released.
   b) PR: A key is pressed and the next key is released.
   c) FT: A key is released and the next is pressed.
   d) PP: A key is pressed and the next key is pressed.
   e) RR: A key is released and the next key is released.

   Based on different analysis carried out in develop and test cycles, DT (dwell time) and FT (flight time) features are considered the most relevant ones and they are weighted accordingly. In addition, a number of typing mistakes (number of presses of such keys such as "Delete" and "Backspace") are calculated separately as auxiliary parameter.
2) **Classification of the extracted features:** This module employs the CNN+RNN model [54] to learn a more complete personal keystroke input mode to carry out continuous authentication. The sequence length of 30 keystroke data (best performance) is vectorized and then divided into fixed-length keystroke feature sequences in order to enable keystroke sequences to be input into the RNN networks. The fact that the input data is pre-processed by CNN (extract a higher-level keystroke feature) improves the performance of the network model.

D. COMPUTER MONITORING

The needs of online proctoring have evolved. In recent times, the market not only seeks to identify students, but also to verify that they are not performing any type of cheating or behaviour that is not allowed with the device on which students perform the activity. In other words, one of the greatest changes is without any doubt the desire to monitor the activity within the device of the students who are doing evaluable activities.

The objective of this development is to obtain an application which is able to monitor the activity carried out by the student within their computer. This monitoring will be done only and exclusively when the student is doing an activity that can be evaluated and supervised by the proctoring system. Because clients can access exams from different operating systems, the objective is to develop a multi-platform application. The user interface is as small as possible so that it does not bother the student during the performance of the evaluable activity. However, it is large and visible enough so that the students know that they are being monitored. The data obtained through the application will be stored in the database or on the servers of the system, therefore, it is necessary that the application complies with all the standards and legislation related to confidentiality and data protection.

The software is developed using Electron JS, a framework that allows multi-platform application development in a simple way. In addition, it is based on web application technologies (as well as data capture modules) which means it does not need to be installed locally on the device in order to be executed. As far as requirements are concerned, the system monitoring tool complies with the following:

1) **Active window detection:** this functionality is one of the key aspects within the application. Not only does the system gets the name of the active window, but it also gets a screenshot of it.
2) **Detection of open/running processes:** This monitoring enables us to know what programs the students open and at what time they have opened them, as well as when they have closed them if the case arises.
3) **Peripheral devices:** A computer has different types of peripheral devices that can be connected. The system knows how many keyboards, microphones, screens and cameras the student has connected to the computer. In the case of cameras, the system also knows the name of them, in order to detect virtual cameras.
4) **Device Information:** Each computer has specific components that make it unique, such as the motherboard or processor. In order to identify if two users use the same computer, information about the computer and its connection is collected: the processor, the motherboard, the IP, the name of the manufacturer, operating system…
5) **Browsing history:** The tool is used especially during evaluable exams, where the students have to answer questions that are presented to them. The student can use any type of browser to look for these answers to these questions. For this reason, the user will be answering correctly without having the necessary knowledge. To combat this type of behaviour, or at least monitor it, the user's browsing history is collected during the activity. Not only the URL, but also the title of the website and the time of entry are registered in the system.
6) **Copy/Paste commands:** Closely linked to the previous point are copying and pasting events. To prevent the student from cheating and copying the answers or sending the test questions to other people, it is necessary to monitor these events. In particular, every event of copying and pasting of text that the user makes during the evaluable activity is recorded. In addition, the screenshots made by the student are collected, for example, if the student screenshots the quiz page to send the exam questions to another person.
7) **Screenshots:** In order to monitor behaviours that we have not yet contemplated, periodic screenshots are made. These screenshots allow the system to identify new methods of cheating.

TABLE 3. Number of captured samples for each type of biometrics.

Images	Audio clips	Keystroke dynamics
373.410	1.007	653

Taking into account that the online student usually uses the same device/browsers/connection to perform their online activities, the information related to computer HW/SW, as well as IP directions are analyzed and their variability in time for the same user is used to trigger more exhaustive automatic and manual authentication and proctoring analysis.

V. TEST

This system was tested through more than 57 activities in 5 different e-learning institutions (3 universities, 2 training centers) in 3 different countries (Latin America, Europe and Asia).

350 students did their assessment activities with the authentication and monitoring system, in three different generic categories: exams (22), short quizzes (10) and forum discussion (25) activities. These activities were chosen because they allow instructors to design activities that need students to spend more time on the platform and have a more complete experience of the biometric authentication and proctoring system.

The courses containing test activities had 3 types of pages: (1) pages of contents, which included texts, schemes and images about the main topic, (2) pages of short quizzes or more extensive exams where the students had to answer questions about what they had read or visualized before and (3) forum activities where instructors promoted discussion related course content through dynamic questions.

Furthermore, the activities were tested in 3 different LMS platforms: Moodle, Blackboard and OpenEdx in order to check the system's compatibility and integrability in the world's most used LMS platforms.

The average time students spent doing these activities was 1 hour and 42 minutes.

A. TECHNICAL TEST

The system captured images randomly every 5-8 second interval, and audio and typing samples every time one of the students spoke or typed text during the activity. The collected data is presented in Table 3. The image/audio/typing algorithms have been tested in depth in each of the captured samples.

All images contain at least 80% of face area (when a person is in the captured photo) and with enough illumination to distinguish facial features after applying brightness and contrast filters (if necessary). On the other hand, the audio samples signal to noise relation (SNR) is acceptable enough to identify the speaker by humans.

B. USER EXPERIENCE TEST

On the other hand, different surveys have been performed during these tests. The objective of this survey was to analyze the perception of students and teachers about the inclusion of these kinds of systems in order to be accepted in the future.

350 students and 50 teachers during the 2018-19 academic year were surveyed about the suitability of this technology. Once they had finished, the students replied to the questionnaire about their experience. In this work we present the most remarkable questions:

1 *Do you think it is appropriate to apply biometric authentication and proctoring to the learning activities?*
2 *Do you think this biometric authentication and proctoring should be implemented in e-learning?*
3 *Do you think this biometric authentication and proctoring should be implemented in all online universities?*
4 *If you could choose, would you prefer to carry out the activities with the incorporation of this software to demonstrate that you have done your activity and you will not be harmed in front of students who ask other people to do the activity?*
5 *Do you think it is fair to monitor distance education in order to avoid cheating?*
6 *Would you feel comfortable if authentication and the monitoring system was working while doing course activities?*

The most remarkable questions for teachers were the following ones:

7 *Would you like to introduce biometric authentication and proctoring tools in your activities?*
8 *Do you think the use of this kind of system will avoid fraud in e-learning activities?*
9 *In your opinion, would the use of the system increase the value and prestige of your courses?*
10 *Do you think authentication and proctoring systems, transparent applications which do not disturb the student, are needed in e-learning environment?*

The questions of the current research are answered with the seven-point Likert scale: Totally disagree (1), Disagree (2), Slightly Disagree (3), Neither agree nor disagree (4), Slightly Agree (5), Agree (6) and Strongly Agree (7).

VI. RESULTS

A. TECHNICAL RESULTS

In this section, the artificial intelligence modules processing results are presented. Since keystroke dynamics samples taken from students cannot be labelled manually (we cannot see or hear), the Table 4 only show an image and audio processing results. The precision and recall data are calculated based on a fully labelled database.

TABLE 4. Performance of authentication and automatic proctoring modules Vs artificial intelligence technologies: a) Authenticating student identity, b) Determining if student is alone or not c) Detecting inappropriate behaviour such as electronic device or book use during online exercises/exams).

	Image processing		Audio processing	
	Precision	Recall	Precision	Recall
Authentication	0.998	0.865	0.964	0.667
Student alone	0.996	0.985	0.963	0.865
Inappropriate behaviour	0.938	0.375	-	-

On the other hand, an analysis of the false positives and negatives of the automatic system has been carried out. Regarding facial authentication, 78% of the failures are a consequence of an excessive face occlusion due to an inappropriate pose and 12% due to poor lighting, mainly caused by the wrong placing of the student against the light. For voice authentication, 53% of failures are due to the low input amplitude of the signal and 33% due to background noise. When determining whether the student is alone or accompanied, motorization based on image processing has failed in 87% of cases due to occlusions (regarding the proximity between individuals or because part of the person protrudes from the image), and 5% because the non-student person is too far away in the image. Finally, sound monitoring has failed by 85% for confusing the second voice (usually with a lower signal amplitude) with background noise and 4% for those samples in which two or more voices have overlapped in the exact same instant. The rest of the errors (including most of the errors in detecting inappropriate behaviour) have been authentication and in monitoring errors made even when the conditions were acceptable for correct automatic operation.

As results table shows, the high precision and recall rates make human intervention almost unnecessary to guarantee 100% of accuracy in final result report. However, human verification is still required. During the tests, all false positive and false negatives (as well as a low rate of true positives and true negatives) were driven to human cross-verification. This action guarantees 100% accuracy in the given final results.

B. USER EXPERIENCE RESULTS

Among other questions, students they were asked whether this system was appropriate to verify the identity of students and proctoring their activities while learning online, which obtained an average of 6.01 in a seven-point Likert scale. However, the opinion of the teachers surveyed about the effectiveness and suitability of this kind of system in an e-learning environment is not as positive as that of the student.

In table 5, the results of the most remarkable questions are analyzed individually.

If we analyze the perceptions of the students based on the results of the most remarkable questions, most of the survey responses have been very positive and welcome. Firstly, students say that it is fair to have any type of biometric recognition software to monitor whether students cheat. Students give an average of 6,03 points in the seven-point Likert scale, in other words, this means that they think it is

TABLE 5. Technology suitability survey results.

	Totally disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly Agree	Agree	Strongly Agree
Q Nº	Student results (%)						
1	0	1.67	1.67	3.33	5	53.33	35
2	0	1.67	0	3.33	11.67	43.33	40
3	1.67	1.67	0	3.33	13.33	38.33	41.67
4	0	6.67	5	6.67	13.33	31.67	36.67
5	0	1.67	1.67	8.33	8.33	35	45
6	3.33	8.33	11.67	8.33	21.67	26.67	20
Q Nº	Teacher results (%)						
7	4	6	18	12	30	22	8
8	6	4	12	8	32	28	10
9	2	10	14	8	30	24	12
10	2	0	2	6	36	50	4

appropriate to rate this question with ‘‘I agree’’. Secondly, students were asked if face-to-face universities with a virtual learning platform should implement a software, and they had a good opinion of this question with an average of nearly 6 (specifically 5,79), which corresponds to ‘‘I agree’’ in the seven-point Likert scale.

The main reason why the implementation of the biometric recognition and proctoring software in education are so favourable for 87% of the student asked is that they are conscious about those students who cheat in their tasks and this is not fair for the rest of them.

In this experience, it is noticeable that there are quite positive average values. Thus, the students think that biometric authentication and proctoring is appropriate (in the range between agree and strongly agree on average) for Moodle lessons when these are used for evaluation, in the range between agree and strongly agree on average. In addition, they considered as a positive experience the one they had with the system presented in this work.

Finally, teachers have been asked (with a free answer type question) what are the main reasons that justify surveying the results of the teachers. It is remarkable that all of the reasons are related to privacy issues; they think student will feel a) observed (83%), b) not comfortable (58%), c) worried with the fact that a computer application is recording/managing their personal data (72%) (not real worries for students according to their survey results). Any given reason arguments lack of suitability, effectiveness or convenience of this kind of system use. Moreover, 78% of them explicitly recognise the need for this kind of application to authenticate and monitor online students in their e-learning activities in the near future.

VII. CONCLUSION AND FUTURE WORK

There is a need to know if the student who enrols in an e-learning course is the same student who completes the learning process and receives academic credit. In this work we present an application which offers a continuous authentication identity service of online student through constant biometrics (face, voice, typing) recognition system and a continuous online proctoring and monitoring system. Allowing online courses to take advantage of something that benefits both institutions and students.

The technical results shows that fully automated, continuous (not scheduled), passive (for students), scalable, fully integrated in LMS (with few HW requirements), secure and private biometric authentication and proctoring solutions are affordable and reliable. Furthermore, they exist in the current e-learning supplier market. As future work, more robust biometric models are needed to avoid undesirable deviations due to variance in face pose and light and noise conditions, and reduce human cross-verification needs only for quality warranty purposes (not to complement automatic system limitations).

The study, based on surveys of the uses of the system shows that the solution presented in this work is recognized as a system which is able to verify the identity of students while doing their activities with the purpose of preventing cheating, and as the system should be integrated in LMS as a needed and appropriate solution. Thus, this type of biometric system is positioned as a promising tool to be used in distance education, opening a variety of possibilities to improve the current LMSs. The results provided qualitative and quantitative data that support the use of this kind of software in distance education in order to prevent students from cheating when they are doing their virtual duties.

Institutions, teachers and students can take advantage of this system in their e-learning experience. Students are interested in better and more reliable academic credit for e-learning courses, despite the necessity of classroom exams, to take advantage with his/her competitor in the real-life professional market. The teachers can manage and take decisions during the subject period without having to wait for classroom exams. Finally, the respect of the institution is based on the quality of its study system and results, which are its students. It is crucial to make sure that the person who gets their academic credit in an e-learning environment is the person who completes all the study plan of the institution.

ACKNOWLEDGMENT

The system presented in this work is a commercial solution named Smiley Owl (SMOWL), which is patented under priority number and title: EP18382363-"Method and System for Verifying the Identity of a User of An Online Service Using Multi-Biometric". The authors would like to thank the Smowltech Company which develops the described system and markets it all over the world. They also thank the Vicomtech Research Center which supports the image and audio processing core algorithms explained in this article.

REFERENCES

[1] Kryterion. (2021). *Kryterion Global Testing Solutions*. [Online]. Available: https://www.kryteriononline.com/

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[3] A. A. Jain, A. K. Flynn, and P. J. Ross, *Handbook of Biometrics*. Springer, 2008. [Online]. Available: https://www.springer.com/gp/book/9780387710402#aboutBook

[4] ProctorU. (2021). *The Leading Proctoring Solution for Online Exams*. [Online]. Available: https://www.proctoru.com/

[5] Examity. (2021). *Better Test Integrity*. [Online]. Available: https://examity.com/

[6] PSIOnline. (2021). *Certification Testing Services and Programs*. [Online]. Available: https://www.psionline.com/en-gb/certification/

[7] ProctorExam. (2021). *Infrastructure for Online Proctoring & Invigilation*. [Online]. Available: https://proctorexam.com/

[8] (2021). *Assessment Tools for Learning Services*. [Online]. Available: https://web.respondus.com/

[9] RemoteProctor. (2021). *Remote Proctor*. [Online]. Available: https://remoteproctor.com/

[10] OnVUE. (2021). *OnVUE*. [Online]. Available: https://home.pearsonvue.com/Test-Owner/Deliver/Online-proctored.aspx

[11] BVirtual. (2021). *Online Proctoring Redefined*. [Online]. Available: https://bvirtualinc.com/

[12] L. Verified. (2021). *Make Your Online Learning Defensible*. [Online]. Available: https://learnerverified.com/

[13] Proctorio. (2021). *A Comprehensive Learning Integrity Platform*. [Online]. Available: https://proctorio.com/

[14] Proctortrack. (2021). *Trusted Exam Integrity | Remote Online Proctoring*. [Online]. Available: https://www.proctortrack.com/

[15] Comprobo. (2021). *OnlineValidation*. [Online]. Available: https://comprobo.co.uk/

[16] Sumadi. (2021). *AI-Powered Proctoring*. [Online]. Available: https://sumadi.net/

[17] ProctorFree. (2021). *Secure Online Proctoring*. [Online]. Available: http://proctorfree.com/

[18] HonorLock. (2021). *Honorlock On-Demand Online Proctoring Service*. [Online]. Available: https://honorlock.com/

[19] ExamSoft. (2021). *Learning Assessments Tools & Software*. [Online]. Available: https://https://examsoft.com/

[20] Y. Khlifi and H. El-Sabagh, "A novel authentication scheme for e-assessments based on student behavior over e-learning platform," *Int. J. Emerg. Technol. Learn.*, vol. 12, no. 4, pp. 62–89, 2017. [Online]. Available: https://online-journals.org/index.php/i-jet/article/view/6478

[21] Z. Zhang, M. Zhang, Y. Chang, S. Esche, and C. Chassapis, "A virtual laboratory system with biometric authentication and remote proctoring based on facial recognition," *Comput. Educ. J.*, vol. 7, no. 4, pp. 74–84, 2016.

[22] Z. Zhang, E.-S. Aziz, S. Esche, and C. Chassapis, "A virtual proctor with biometric authentication for facilitating distance education," in *Online Engineering & Internet of Things*, M. E. Auer and D. G. Zutin, Eds. Cham, Switzerland: Springer, 2018, pp. 110–124.

[23] H. S. G. Asep and Y. Bandung, "A design of continuous user verification for online exam proctoring on M-learning," in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Jul. 2019, pp. 284–289.

[24] L. K. Musambo and J. Phiri, "Student facial authentication model based on OpenCV's object detection method and QR code for Zambian higher institutions of learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, Jan. 2018.

[25] F. Guillen-Gamez, I. García-Magariño, and G. Palacios, *Comparative Analysis Between Different Facial Authentication Tools for Assessing Their Integration in m-Health Mobile Applications*. Cham, Switzerland: Springer, Mar. 2018, pp. 1153–1161. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-77712-2_110

[26] S. Sawhney, K. Kacker, S. Jain, S. N. Singh, and R. Garg, "Real-time smart attendance system using face recognition techniques," in *Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2019, pp. 522–525.

[27] A. Alshbtat, N. Zanoon, and M. Alfraheed, "A novel secure fingerprint-based authentication system for student's examination system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 515–519, 2019. [Online]. Available: https://thesai.org/Publications/ViewPaper?Volume=10&Issue=9&Code=IJACSA&SerialNo=68

[28] J. V. Monaco, J. C. Stewart, S.-H. Cha, and C. C. Tappert, "Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works," in *Proc. IEEE 6th Int. Conf. Biometrics, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.

[29] E. Flior and K. Kowalski, "Continuous biometric user authentication in online examinations," in *Proc. Int. Conf. Inf. Technol.*, Jan. 2010, pp. 488–492.

[30] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu, "Automated online exam proctoring," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1609–1624, Jul. 2017.

[31] A. Okada, I. Noguera, L. Alexieva, A. Rozeva, S. Kocdar, F. Brouns, T. Ladonlahti, D. Whitelock, and A. Guerrero-Roldán, "Pedagogical approaches for e-assessment with authentication and authorship verification in higher education," *Brit. J. Educ. Technol.*, vol. 50, no. 6, pp. 3264–3282, Nov. 2019.

[32] G. Fenu, M. Marras, and L. Boratto, "A multi-biometric system for continuous student authentication in e-learning platforms," *Pattern Recognit. Lett.*, vol. 113, pp. 83–92, Oct. 2018.

[33] L. Slusky, "Cybersecurity of online proctoring systems," *J. Int. Technol. Inf. Manage.*, vol. 29, no. 3, pp. 56–83, 2020.

[34] F. Guillen-Gamez, J. Bravo, and I. García-Magariño, "Students' perception of the importance of facial authentication software in moodle tools," *Int. J. Eng. Educ.*, vol. 33, pp. 84–90, Jan. 2017.

[35] A. Ullah, H. Xiao, and T. Barker, "A dynamic profile questions approach to mitigate impersonation in online examinations," *J. Grid Comput.*, vol. 17, no. 2, pp. 209–223, Jun. 2019, doi: 10.1007/s10723-018-9442-6.

[36] S. A. Razak, N. H. M. Nazari, and A. Al-Dhaqm, "Data anonymization using pseudonym system to preserve data privacy," *IEEE Access*, vol. 8, pp. 43256–43264, 2020.

[37] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020.

[38] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "FaceBoxes: A CPU real-time face detector with high accuracy," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 297–309. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231219310719

[39] L. Unzueta, W. Pimenta, J. Goenetxea, L. P. Santos, and F. Dornaika, "Efficient generic face model fitting to images and videos," *Image Vis. Comput.*, vol. 32, no. 5, pp. 321–334, May 2014.

[40] X. Liu, X. Ma, J. Wang, and H. Wang, "M3L: Multi-modality mining for metric learning in person re-identification," *Pattern Recognit.*, vol. 76, pp. 650–661, Apr. 2018.

[41] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823. [Online]. Available: https://ieeexplore.ieee.org/document/7298682

[42] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. 1994. [Online]. Available: https://dl.acm.org/doi/10.5555/180895.180940

[43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, *SSD: Single Shot Multibox Detector* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, 2016. [Online]. Available: https://www.springer.com/gp/book/9783319464770

[44] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6882–6890. [Online]. Available: https://ieeexplore.ieee.org/document/8100211

[45] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680. [Online]. Available: http://dl.acm.org/citation.cfm?id=2969033.2969125

[46] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Waikoloa, HI, USA. Piscataway, NJ, USA: IEEE Signal Processing Society, Dec. 2011, p. 30. [Online]. Available: https://dblp.org/db/conf/asru/asru2011.html

[47] L. Wan, Q. Wang, A. Papir, and I. L. Moreno, "Generalized end-to-end loss for speaker verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4879–4883. [Online]. Available: https://ieeexplore.ieee.org/document/8462665

[48] Q. Wang, C. Downey, L. Wan, P. A. Mansfield, and I. L. Moreno, "Speaker diarization with LSTM," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5239–5243. [Online]. Available: https://ieeexplore.ieee.org/document/8462628

[49] Y. Zhong, Y. Deng, and A. K. Jain, "Keystroke dynamics for user authentication," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 117–123. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6239225

[50] H. Crawford, "Keystroke dynamics: Characteristics and opportunities," in *Proc. 8th Int. Conf. Privacy, Secur. Trust*, Aug. 2010, p. 20108.

[51] A. T. Kiyani, A. Lasebae, K. Ali, M. U. Rehman, and B. Haq, "Continuous user authentication featuring keystroke dynamics based on robust recurrent confidence model and ensemble learning approach," *IEEE Access*, vol. 8, pp. 156177–156189, 2020.

[52] A. F. M. N. H. Nahin, J. M. Alam, H. Mahmud, and K. Hasan, "Identifying emotion by keystroke dynamics and text pattern analysis," *Behav. Inf. Technol.*, vol. 33, no. 9, pp. 987–996, Sep. 2014.

[53] R. Moskovitch, C. Feher, A. Messerman, N. Kirschnick, T. Mustafić, A. Camtepe, B. Löhlein, U. Heister, S. Möller, L. Rokach, and Y. Elovici, "Identity theft, computers and behavioral biometrics," in *Proc. IEEE Int. Conf. Intell. Secur. Informat.*, Jun. 2009, pp. 155–160. [Online]. Available: https://ieeexplore.ieee.org/document/5137288

[54] L. Xiaofeng, Z. Shengfei, and Y. Shengwei, "Continuous authentication by free-text keystroke based on CNN plus RNN," *Procedia Comput. Sci.*, vol. 147, pp. 314–318, Jan. 2019.

4.3 Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case

- **Autores:** Mikel Etxeberria-Garcia and Maider Zamalloa and Nestor Arana-Arexolaleiba and **Mikel Labayen**
- **Revista:** IEEE Access
- **Volumen:** 10
- **Páginas:** 69200-69215
- **Año:** 2022
- **Editor:** IEEE

Received 31 May 2022, accepted 21 June 2022, date of publication 29 June 2022, date of current version 5 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187209

APPLIED RESEARCH

Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case

MIKEL ETXEBERRIA-GARCIA[1], **MAIDER ZAMALLOA**[1], **NESTOR ARANA-AREXOLALEIBA**[2,3], **AND MIKEL LABAYEN**[4,5]

[1]Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), 20500 Arrasate/Mondragón, Spain
[2]MGEP, Mondragon Unibertsitatea, Loramendi Kalea, Arrasate/Mondragón, 20500 Gipuzkoa, Spain
[3]Department of Materials and Production, Aalborg University, 9220 Aalborg, Denmark
[4]CAF Signaling, Donostia, 20018 Gipuzkoa, Spain
[5]Faculty of Informatics, UPV/EHU, Manuel Lardizabal Ibilbidea, Donostia, 20018 Gipuzkoa, Spain

Corresponding author: Mikel Etxeberria-Garcia (mikel.etxeberria@ikerlan.es)

This work was supported in part by the Basque Government through BIKAINTEK2018 Program and CAF Signaling in Collaboration Between Ikerlan, CAF Signaling, and Mondragon Unibertsitatea.

ABSTRACT Localization is one of the most critical tasks for an autonomous vehicle, as position information is required to understand its surroundings and move accordingly. Visual Odometry (VO) has shown promising results in the last years. However, VO algorithms are usually evaluated in outdoor street scenarios and do not consider underground railway scenarios, with low lighting conditions in tunnels and significant lighting changes between tunnels and railway platforms. Besides, there is a lack of GPS, and it is not easy to access such infrastructures. This research proposes a method to create a ground truth of images and poses in underground railway scenarios. Second, the EnlightenGAN algorithm is proposed to face challenging lighting conditions, which can be coupled with any state-of-the-art VO techniques. Finally, the obtained ground truth and the EnlightenGAN have been tested in a real scenario. Two different VO approaches have been used: ORB-SLAM2 and DF-VO. The results show that the EnlightenGAN enhancement improves the performance of both approaches.

INDEX TERMS Visual Odometry, autonomous vehicles, computer vision, data enhancement, simultaneous localization and mapping, image processing, railway domain.

I. INTRODUCTION

Visual Odometry (VO) is a particular case of odometry based on Computer Vision (CV), where the position and motion information are acquired through camera images [1]. VO algorithms aiming to derive localization data through visual sensors are usually evaluated and compared by reference standard datasets such as KITTI [2], [3] and EuRoC-MAV [4]. This situation leads solutions adapted to the visual characteristics contained on those scenarios with adequate lighting conditions (good illumination and similar lighting conditions in subsequent frames), relatively sufficient textures and Lambertian surfaces. However, few algorithms, datasets, and benchmarks can be found in challenging scenarios with varying light conditions, low illumination, low textures, or non-Lambertian surfaces.

For instance, one of the latest benchmark challenges in visually challenging odometry is the Subterranean Challenge (SubT), organized by the Defense Advanced Research Projects Agency (DARPA). Perceptually challenging scenarios and tasks were stated in this challenge, such as navigation through tunnel systems, cave networks, or urban underground environments. The participating teams presented several approaches [5]–[8] to study the robotics autonomy in underground scenarios exploration and navigation. These works emphasize the complexity of localization and navigation in underground environments due to their perceptually-degraded conditions. They also emphasize on the importance of field testing.

The associate editor coordinating the review of this manuscript and approving it for publication was Kegen Yu.

The railway domain is also moving towards the *Intelligent Transportation Systems* (ITS) and the *Advanced Driving Assistance Systems* (ADAS) industry. A train that implements autonomous operations requires accurate localization estimation to carry out operations as precise stop operation or coupling successfully. Algorithms applied in urban underground railway scenarios must deal with significant light changes from tunnel areas to platforms, with insufficient illumination and low textures in tunnels.

In this context, the application of state of the VO algorithms and data enhancement techniques was analyzed in a perceptually challenging driving car scenario [9]. The results showed that the Generative Adversarial Network (GAN)-based image enhancement methods can improve the performance achieved by state-of-the-art VO solutions.

In this paper, an analysis of state-of-art VO algorithms is performed and the use of a data enhancement method in underground railway VO solutions is evaluated. Algorithms applied in these scenarios must deal with significant light changes from tunnel areas to platforms, with insufficient illumination and low textures in tunnels. Therefore, an image enlightening technique is integrated to improve the results of state-of-the-art VO algorithms.

A dataset with challenging characteristics is really needed in order to evaluate VO performance in such scenarios. From an analysis of datasets used in CV for localization (datasets labeled with 6-DoF pose), no standard dataset of the railway domain was found; hence, an ad-hoc underground railway dataset generation was pursued.

The following section (II) includes a literature review of the main VO algorithms, a description of the applied enlightening data enhancement technique, and a list of reference VO datasets. Section III depicts the urban underground railway dataset generation process. Then, the results of state-of-art VO algorithms in the underground railway dataset and the influence of an enlightening technique are shown in sections IV and V, respectively. Finally, some conclusions are drawn in section VI.

II. LITERATURE REVIEW

A. VISUAL ODOMETRY

The term Visual Odometry was first introduced by Niester *et al.* [10] proposing a technique to estimate camera motion using RANSAC [11] outlier refinement method and tracking extracted features across the frames. Previously, feature matching was done just in consecutive frames. Later works have shown that VO methods might perform as well as wheel odometry while the cost of cameras is much lower compared to wheel sensors [1].

The VO research community started from the robotics domain to, later, focus on the localization in other subdomains. In this context, different types of vehicles from distinct sub-domains and diverse characteristics have been studied, such as, cars [12], [13], trains [14], or lately UAVs [15].

Depending on the algorithm used to estimate odometry data, VO techniques can be classified as learning-based and geometry-based [16], [17]. *Geometry-based VO* is usually divided into appearance-based VO (also referred to as direct), feature-based VO, and a hybrid approach that mixes the two of them.

Direct VO techniques operate directly on intensity values. In feature-based VO methods features are extracted from the image and a tracking-matching process is done. Feature-based methods have good accuracy, are robust in dynamic scenes, and can deal with variances in viewpoint [18]; however, in contrast to direct methods, feature-based techniques are inadequate in low texture areas. However, the performance of direct VO algorithms degrades if the dataset is not photometrically calibrated and is sensitive to geometric distortions as those induced by the camera speed [19]. Furthermore, as mentioned in [20], direct methods require a constant irradiation appearance between matched pixels, which hinders its application in some scenarios.

Geometry-based VO approaches rely on image geometric characteristics and camera model to reconstruct the ego-motion between consecutive frames. One of the most standard geometric VO approach is ORB-SLAM2 [21]. It is based on the ORB [22] feature matching and a bundle adjustment algorithm. It is the reference geometric solution in the VO community [19], [23]–[28].

Geometry-based VO is reliable and accurate under favorable conditions, when there are enough illumination and textures to make the feature matching among consecutive frames. As stated in [29], monocular VO experiences a scale drift issue and global bundle adjustment algorithms needs to be applied. Furthermore, monocular VO algorithms have a depth-translation scale ambiguity issue [30].

Stereo geometry-based VO works have been also targeted lately. Semi-direct visual odometry (SVO) [31] is one of the most predominant approaches among direct monocular and stereo VO algorithms. It uses a probabilistic mapping method to estimate ego-motion and explicitly models outlier measurements. In 2017, Wang *et al.* presented Stereo Direct Sparse Odometry (Stereo DSO) [19], a method for VO estimation from stereo cameras based on the previously proposed monocular DSO algorithm [32]. Lately, Koestler *et al.* presented TANDEM [33], a SLAM system that estimates ego-motion based on a direct VO pipeline and deep multiview stereo.

The expansion of Deep Learning-based Computer Vision techniques carried the emergence of *Deep Learning-based VO* solutions. Learning-based VO/vSLAM algorithms usually rely on learning parts of a standard VO/vSLAM pipeline or designing end-to-end trainable algorithms for ego-motion estimation.

One of the first and most relevant learning-based VO algorithms was PoseNet proposed by [34] Kendall *et al.*, a robust and real-time monocular re-localization system based on an end-to-end trained CNN. This approach was later improved by introducing loss functions based on geometry and scene reprojection error [35]. Following this end-to-end pose estimation networks, DeepVO [36] was published, a solution that infers camera poses directly in an end-to-end manner from a sequence of RGB frames through a supervised Deep Recurrent Convolutional Neural Network (RCNN).

Some research works have tried to adapt traditional non-learning approaches into Deep Learning pipelines. Brachmann *et al.* introduced DSAC (Differentiable Sample Consensus) [37] algorithm based on previously proposed RANSAC [11]. They applied DSAC in a camera localization solution, learning an end-to-end camera localization pipeline.

However, most of the research works from the literature emphasize the importance of an accurate depth and flow estimation for VO/vSLAM. Depth information is crucial for the localization as it enables the inference of the scene geometry from 2D images. Moreover, it allows scale recovery [38] and the distinction of foreground and background points, allowing a better environment understanding. Together with depth estimation, the optical flow estimation is also a critical component of some VO/vSLAM algorithms as it models the motion between consecutive images. Therefore, most of learning-based VO/vSLAM algorithms have focused on learning depth and flow estimation for the pose inference process.

Following this research line, several works have focused the depth estimation [39], [40], [41]. In 2018, Zhan *et al.* presented Depth-VO-Feat [42], where stereo training was introduced to reduce the spatial and temporal photometric error. At the same time, DVSO was presented by Yang *et al.* [29], introducing deep depth predictions in Direct Sparse Odometry (DSO). D3VO [43] algorithm was also proposed in this direction, including the uncertainty estimation with camera pose and depth.

Zhan *et al.* proposed the unsupervised VO algorithm DF-VO [17]. This algorithm applies a deep learning-based depth and flow estimation, and, geometric image information to estimate the camera pose. As shown in [17], DF-VO outperforms most learning-based state-of-the-art algorithms in standard datasets.

Some works have proposed loss functions to handle challenging scenario characteristics. Yin *et al.* proposed GeoNet [44], to increase robustness towards outliers and non-Lambertian surfaces. After GeoNet, more works were proposed in this direction [45], [46].

However, as mentioned in [47], literature VO solutions have limitations in challenging scenarios that contain insufficient illumination and textures, or, variable lighting conditions. Literature VO solutions, as they are adapted to the characteristics of standard datasets, require sufficient illumination and enough textured surfaces for a correct feature matching. A good illumination allows motion extraction from images, as pixel displacement can not be accurately estimated otherwise. Therefore, the lighting issue needs to be handled in scenarios that contain low illumination or varying illumination conditions. These are the conditions that face the urban underground railway scenario.

DF-VO and *ORB-SLAM2* have been selected from the literature review as reference VO algorithms. As stated before, the DF-VO algorithm outperforms most learning-based state-of-the-art algorithms, while ORB-SLAM2 is the most referenced geometric algorithm. Moreover, these algorithms represent two distinct types of VO algorithms (learning-based and geometric). Both solutions can use mono-vision or stereo-vision camera frames as input. The stereo-vision input was chosen for the analysis, as stereo-vision solutions keep the real-world scale, i.e. the predictions are directly aligned to a real-world scale.

B. DATA ENHANCEMENT FOR VISUAL ODOMETRY IN CHALLENGING ENVIRONMENTS

In order to afford the scenario limitations of VO in challenging environments, the application of a data enhancement technique was considered. In this work, the data enhancement process is dedicated to the lighting limitations of the target domain. It aims to reduce the impact of the drastic lighting conditions found in the underground railway scenario.

In this paper, the work published in [9] is extended. In the previous work the application of *EnlightenGAN* [48] data enhancement approach in an outdoor driving car scenario with varying lighting conditions was evaluated. This previous research was focused on a driving car scenario where the lighting conditions of the underground railway domain where replicated driving by night. The results showed that the performance of DF-VO algorithm is improved when EnlightenGAN is applied in the recorded frames.

EnlightenGAN is based on machine learning models proposed by Ian Goodfellow *et al.* [49]. The algorithm uses an unsupervised Generative Adversarial Network (GAN) pretrained on the ImageNet dataset [50] and then trained on several datasets [51]–[54] to improve input image lighting.

EnlightenGAN was previously used for several tasks such as image reconstruction [55], photo exposure correction [56], image quality assessment [57] or illumination enhancement [58]. However, to our knowledge, the use of data enhancement methods to handle specific problems of VO methods in such challenging scenarios has not been researched yet.

In this paper, the application of EnlightenGAN in the underground railway domain when using geometric and hybrid VO solutions is evaluated. The study aims to explore if EnlightenGAN technique can afford the lighting limitations of reference VO approaches (DF-VO and ORB-SLAM2). The evaluation procedure and results are detailed in section V.

C. DATASETS FOR UNDERGROUND RAILWAY VISUAL ODOMETRY

In this work, a propietary dataset is generated as no standard or reference railway dataset fitted to the underground railway scenario was identified. Table 1 resumes the reference datasets used by starte-of-the art VO approaches.

Most state-of-the-art VO approaches are evaluated in the standard KITTI [2], [3] vision benchmark [17], [29], [36], [42], [43], [81]. This benchmark includes several datasets for

TABLE 1. Referenced datasets for Computer Vision-based VO approaches application and evaluation ordered by domain or motion type.

Dataset	Domain	Sensor configuration	Pose ground truth	Environment
Cambridge Landmarks [34]	Handheld sensor	Monocular	SfM	outdoors
7-scenes [59]	Handheld sensor	RGB-D	MoCap	indoors
BigSFM [60]	Handheld sensor	Monocular	GPS	outdoors
ICL-NUIM [61]	Handheld sensor	RGB-D	SLAM	indoors
ADVIO [62]	Handheld sensor	Stereo/IMU	IMU	in/outdoors
OIVIO [63]	Handheld sensor	Stereo/IMU	Total station	in/outdoors
Rawseeds [64]	Robot	Stereo/IMU	GPS	in/outdoors
SUN3D [65]	Robot	RGB-D	SfM	indoors
TUM-VI [66]	Robot	Stereo/IMU	MoCap	in/outdoors
TUM-RGB-D SLAM [67]	Robot	RGB-D	MoCap	indoors
TUM-Monocular VO [68]	Robot	Monocular	LSD-SLAM/MoCap	in/outdoors
NavVis [69]	Robot	Monocular	GPS	indoors
MIT Stata [70]	Robot	Stereo/RGB-D/Laser	Laser	indoors
The Wean Hall [71]	Robot	Stereo/IMU/Laser/Wheel odometry	GPS	in/outdoors
RGB-D SLAM [67]	Robot	RGB-D	MoCap	indoors
ETH3D [72]	Robot	Stereo/RGB-D/Laser/IMU	MoCap/SfM/LIDAR	in/outdoors
NCLT [73]	Segway	Stereo/IMU/Laser	GPS/IMU/Laser	in/outdoors
KITTI [2, 3]	Car	Stereo/IMU/Laser	GPS/IMU	outdoors
Málaga Urban [74]	Car	Stereo/IMU/Laser	GPS	outdoors
Oxford RobotCar [75]	Car	Stereo/Laser	GPS	outdoors
Ford Campus [76]	Car	Stereo/Laser/IMU	GPS	outdoors
KAIST Urban [77]	Car	Stereo/IMU	GPS/Laser	outdoors
Nordland [78]	**Railway**	**Monocular**	**GPS**	**outdoors**
Zurich Urban [79]	MAV	Monocular/IMU	GPS	outdoors
EuroC/MAV [4]	MAV	Stereo/IMU	MoCap/Laser	indoors
MVSEC [80]	Multi Vehicle	Stereo/IMU/Laser	GPS/MoCap/Laser	in/outdoors

* MoCap=Motion Capture System. SfM=Structure From Motion

tasks like VO, optical flow estimation, 3D object detection, or 3D tracking. The data is captured from a moving car in outdoor urban scenarios, and they provide datasets and evaluation metrics for each task. However, as the KITTI odometry dataset contains images from an outdoor environment with good lighting conditions, it is not adequate to evaluate the VO algorithms in the pursued scenario. Among the other analyzed datasets, it should be noted that only one database (Norland [78]) covers the railway domain; however it only covers outdoor scenarios, which is also out of the scope of this research work. Searching for a publicly available VO dataset from an indoor urban railway domain, no dataset was found. Following the idea that the evaluation of the VO approaches that have previously been evaluated in standard datasets is essential to adapt the algorithms to other industrial scenarios. Therefore, the generation of a proprietary database was considered.

The data for a proprietary dataset can be collected from different sources: from real scenarios or simulated environments. Real environment datasets are based on real-world scenarios, and therefore, the performance of algorithms can be effectively evaluated in the target scenario. However, the database generation in real-world scenarios increases recording and processing time, effort, and cost. In addition, it also depends on the access and permission to make the recordings in the target scenario.

Simulated environments can overcome these problems. The drawback of simulated environments is that it can not be assured that an algorithm trained and validated in a simulated environment will perform the same way in a real-world scenario. As stated in [82], all the challenging conditions inherent to underground environments can not be recreated in virtual scenarios.

Consequently, and as a real-world underground railway scenario was accessible, a proprietary dataset was generated from a real underground railway scenario. The definition, generation and validation processes of the proprietary *CAF* dataset is explained in the next section III.

III. URBAN UNDERGROUND RAILWAY DATASET GENERATION

The proprietary (*CAF*) was generated for the evaluation of VO algorithms in underground railway scenarios. The sensor set validation and camera calibration procedure was done by generating a complementary dataset (*CarDriving*) in an urban driving car domain. *CarDriving* dataset generation is described in [9].

The *CAF* dataset was recorded in an underground scenario in the railway *Line 3* of Euskotren-Bilbao. The line is composed by seven stations from Matiko to Kukullaga and it has a whole track length of 5.8km. It contains poor lighting conditions in tunnel areas and significant light changes in platform areas. Furthermore, the images captured in the tunnels contain repetitive and light dependent textures, and therefore, they are challenging for feature extraction algorithms. Figure 1 shows two frames of this scenario: (a) tunnel frame and (b) platform frame.

The camera was placed in the front of the train, inside the driving cabin according to the safety requirements of the railway domain. Figure 2 shows the camera placement in the active cabin.

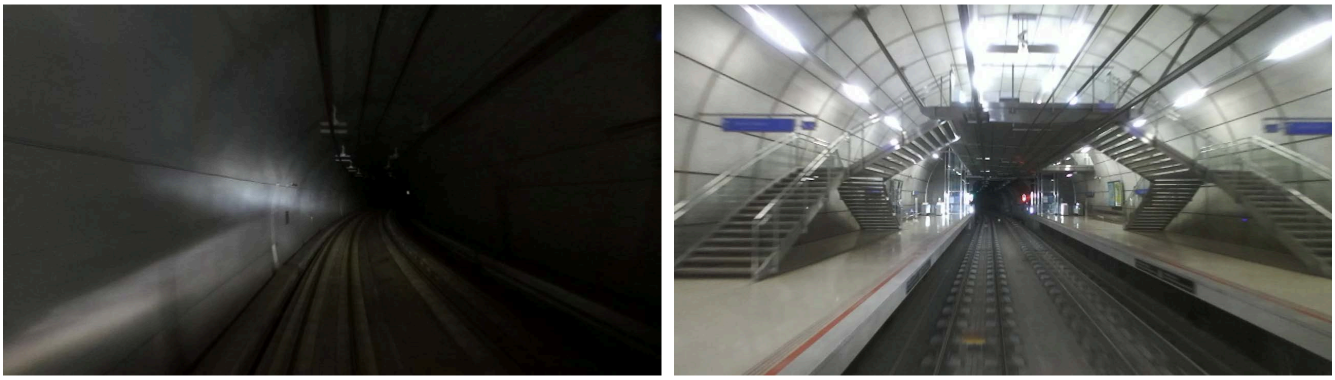


FIGURE 1. The *CAF* dataset's tunnel and platform areas where the poor light conditions and textureless areas can be appreciated.

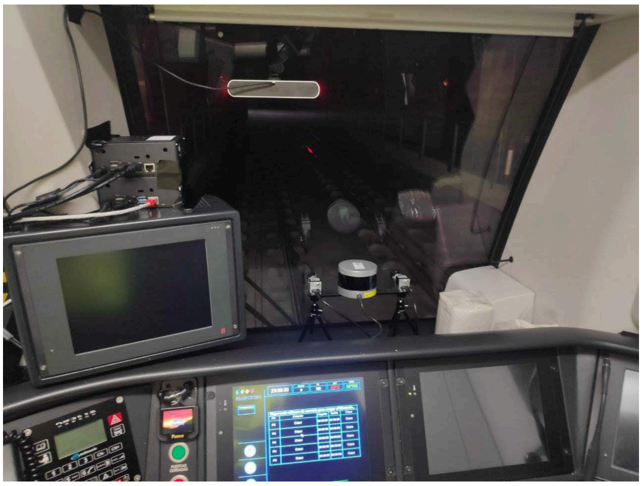


FIGURE 2. Camera setup for *CAF* dataset, placed in the cabin of a train moving through an underground urban railway scenario.

The recording camera is a ZED Stereo Camera. The image's resolution is 1280 × 720 pixels at 30 Hz with an electronic synchronized rolling shutter, automatic gain and a lens aperture of F2.0.

A. CAF DATASET

The dataset is composed by 19 sequences captured in the two directions of the rail Matiko-Kukullaga. A sequence is a record that begins at one station and ends in the stations the train stops. A 6-DoF pose is estimated for each captured frame. The dataset format follows the standard KITTI odometry dataset format and naming convention. The frames are rectified RGB color images stored with lossless compression using 8-bit PNG files.

The camera calibration parameters and the poses are stored in files specified by the KITTI format [3]. Each row of the pose file contains the first three rows of a 4 × 4 homogeneous pose matrix flattened into one line. The homogeneous pose matrix p_n can be represented as:

$$p_n = [r_n|\mathrm{tr}_n] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & x_n \\ r_{21} & r_{22} & r_{23} & y_n \\ r_{31} & r_{32} & r_{33} & z_n \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

TABLE 2. *CAF* dataset resume with recorded sequences, the direction of the sequences, arriving station for each sequence, frame quantity, and sequence length.

Direction	Arrival station	Sequence	Frames	Length (m)
Matiko	Otxakoaga	01_50	3048	1420
	Txurdinaga	01_53	1977	699
	Zurbaranbarri	01_54	2663	1029
		03_49	6700	3148
	Zazpikaleak	02_22	3260	1011
	Uribarri	01_15	2724	903
		02_25	2639	
		03_54	5904	1913
	Matiko	01_17	2532	505
		02_27	2505	
Kukullaga	Uribarri	01_31	2140	524
	Zazpikaleak	01_33	2830	979
	Zurbaranbarri	01_35	2494	1007
		03_36	6560	2449
	Txurdinaga	01_37	2550	1032
	Otxarkoaga	01_39	2126	695
		03_36	4493	1729
	Kukullaga	01_40	4095	1405
		03_44	4144	
TOTAL			65384	23261

where r_n and tr_n are the rotation matrix and the translation matrix of the n-th frame, respectively. The translation component of the pose matrix follows the right-hand rule when defining axes in a 3D space (x-axis forward, y-axis right and z-axis up).

The dataset generated in this domain is represented in table 2 where the recorded sequences, recording direction, the arrival station for each sequence, the number of frames, and the track length of each sequence are depicted. The entire set of sequences yields 65.384 frames, with varying speed and length.

B. GROUND TRUTH GENERATION ALGORITHM DATA SOURCES

In general, the ground truth of VO datasets is generated using a GPS sensor [3], [74]–[77] (refer to Table 1). But, the GPS signal is unavailable in underground zones like the urban underground railway domain. Thus, a method that computes the 6-DoF pose of each frame from the train ERTMS/ETCS ATP data, geodetic map coordinates, and railway infrastructure gradient profile data was defined and implemented (see figure 3).

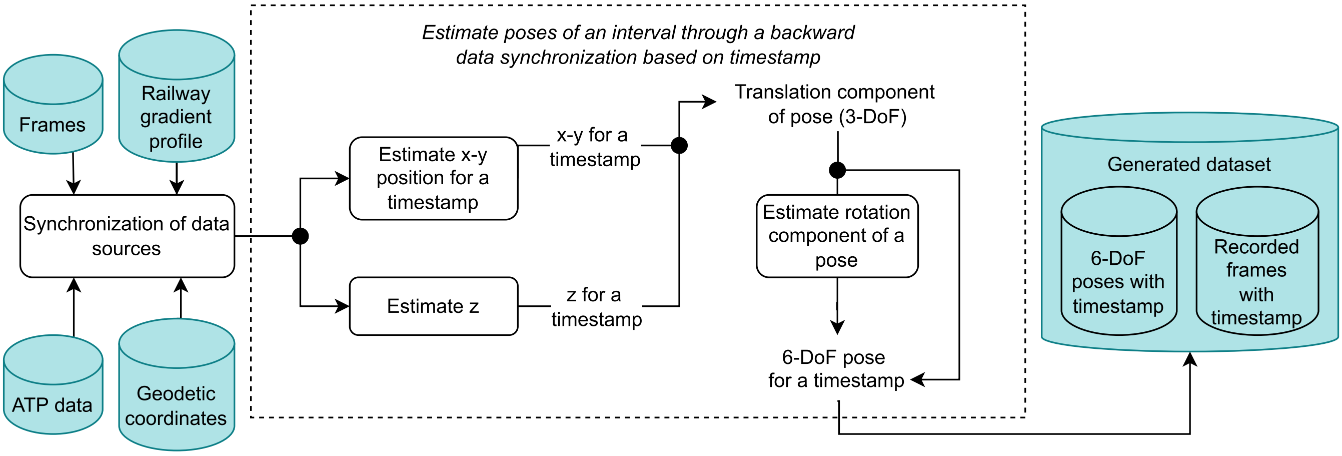

Frames
Railway gradient profile
ATP data
Geodetic coordinates
Synchronization of data sources
Estimate poses of an interval through a backward data synchronization based on timestamp
Estimate x-y position for a timestamp
x-y for a timestamp
Estimate z
z for a timestamp
Translation component of pose (3-DoF)
Estimate rotation component of a pose
6-DoF pose for a timestamp
Generated dataset
6-DoF poses with timestamp
Recorded frames with timestamp

FIGURE 3. **Diagram of the algorithm processes, with the data sources and the outputs.**

The algorithm first estimates (x,y) positions based on geodetic coordinates, then z is added through the gradient profile. Afterwards the (x,y,z) translation data is estimated for each frame by using ERTMS ATP data, and, finally, the rotation data of each pose is calculated.

1) GEODETIC COORDINATES

The geodetic coordinates are represented by a pair (ϕ, λ) expressing *Latitude (Lat.)* and *Longitude (Lon.)* in decimal degrees. These coordinates use an ellipsoid to approximate the the earth's surface locations [83].

In this research, the geodetic coordinates define the coordinates followed by the trains in the target railway and have been extracted from a Geomap called ÖPNVKarte [84]. This Geomap contains public data that includes worldwide public transport facilities on a uniform map with information concerning several transport methods such as train, railway, ferry or bus. It is derived from OpenStreetMap [85], an initiative to create and provide accessible geographic data (i.e. street maps, etc.). It also contains railway-related information, such as platforms, stop positions, and routes.

The entire trajectory of an underground train in L3 extracted from ÖPNVKarte is shown in figure 4. As stated before, the trajectory of L3 is made up of seven stations in the route Kukullaga - Matiko, where some route positions, the station entrances, and train stop positions of each station are known in geodetic coordinates. However, the frequency of the camera is higher than the geodetic coordinates defined in the Geomap, and, therefore, a method based on ERTMS ATP data has been designed and implemented in order to generate the poses of the frames that were recorded between the geodetic coordinates.

The geodetic coordinates must be transformed from 3D plane to a 2D plane to assign an equal-area (x,y) position to each geodetic coordinate. Figure 5 shows a trajectory sample in geodetic coordinates and the generated equal-area (x,y) coordinates. In the ground truth generation algorithm, an equal-area (x,y) coordinate refers to tr_x and tr_y components of a 6-DoF pose.

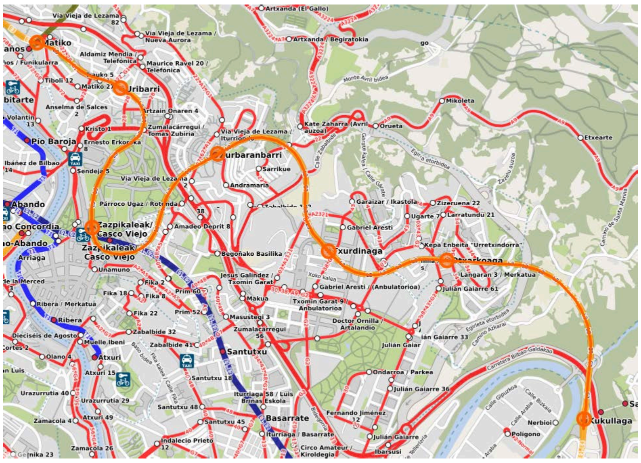


FIGURE 4. **Line 3 railway extracted from ÖPNVKarte map [84]. Each circle represents one station from Line 3.**

2) RAILWAY GRADIENT PROFILE

The railway gradient profile provided by the railway infrastructure managers, defines how the slope of the railway varies in predefined sections and allows the estimation of the height (z) for each 6-DoF pose. For that, a height profile can be constructed with this gradient profile. The initial height is initialized as 0, and then the height for each 1m section is calculated using the Equation 1.

$$h(d_n) = h(d_{n-1}) + (0.01 \cdot \text{grad}_n)\,, \tag{1}$$

where h refers to height, d_n refers to 1m railway sections and grad$_n$ is the gradient value corresponding to that section from the gradient profile. Figure 6 shows the obtained railway gradient profile of the whole L3 railway.

3) ATP DATA: TRAIN's DYNAMICS AND SPEED DATA

The ERTMS/ETCS ATP train speed estimation process is based on redundant wheel encoder and radar sensor in order to get a safe and accurate estimation. By using these sensors, the ATP subsystem embedded in the train estimates the train position in the track, i.e. the distance traveled from an station or a beacon of the track. Track beacon position or

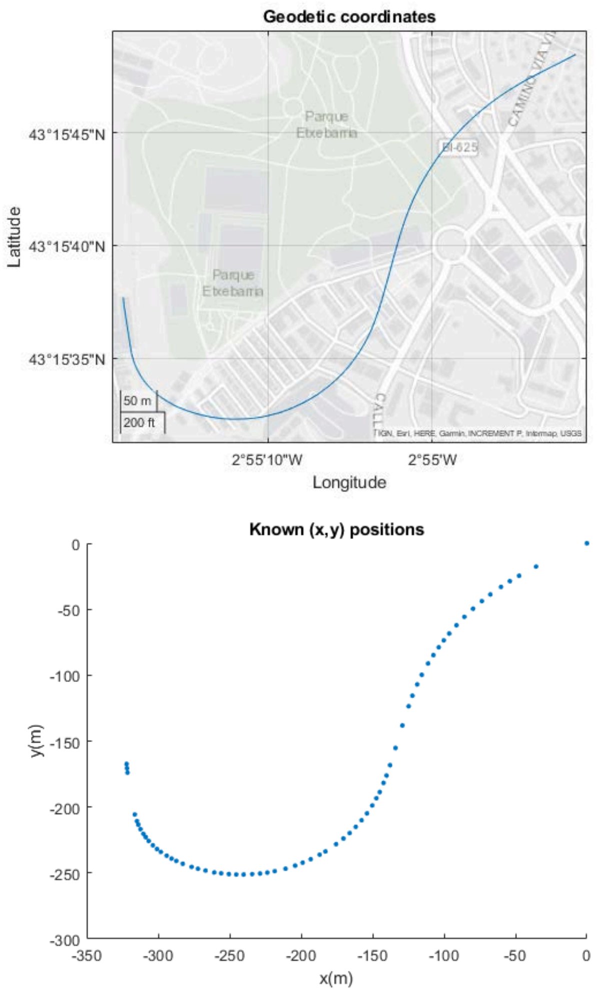


FIGURE 5. Transformation of a given sequence from L3 railway defined by geodetic coordinates into equal-area (x,y) positions.

inter-beacon distance is predefined and known by railway infrastructure managers, even by the ATP subsystem, and therefore the ATP train position is re-adjusted when a beacon signal is received obtaining a precise estimation. The 6-DoF pose estimation of each frame is made by synchronizing the ATP system monitoring process with the image recording process as both are installed in the train. The objective of this process is to obtain a synchronized train position information for each frame. The data monitored from the ATP system is the following one:

- *timestamp* (s): time measured in the Coordinated Universal Time (UTC) standard read from the train's internal clock.
- *linear position estimation* (cm): distance traveled by the train from a previous station.
- *train speed* (m/s): train speed calculated by ATP.
- *train acceleration* (cm/s^2): train acceleration calculated by ATP.
- *train stopped*: boolean reflecting whether the train has reached stopping point or not.

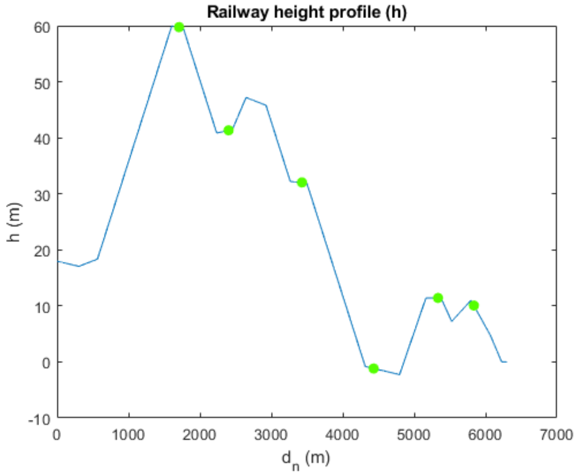


FIGURE 6. Results of height generation process. Height profile (*h*) is generated from gradient profile provided by railway constructor. The green circles represent the stations.

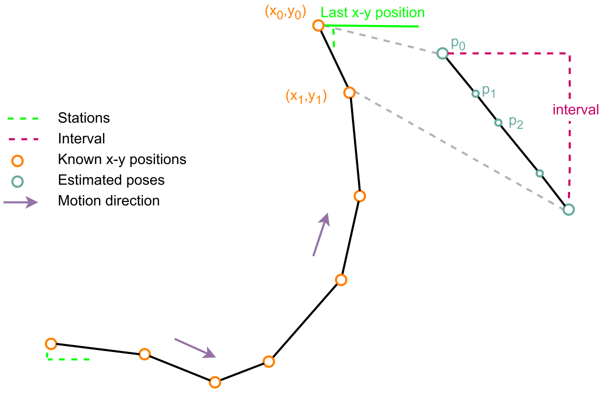


FIGURE 7. Railway with the known (x,y) positions, the intervals and estimated poses.

All those variables are extracted from a ATP monitoring proprietary application that monitors ATP data with a frequency of 128000 Hz. The data acquisition frequency higher than the camera frequency (30Hz), and, consequently, they have been synchronized and a pose estimated for each frame.

C. ESTIMATE POSES OF AN INTERVAL THROUGH A BACKWARD DATA SYNCHRONIZATION BASED ON TIMESTAMP

The main idea of the synchronization algorithm is the estimation of poses in the trajectory sections between the known (x,y) positions obtained by transforming the known geodetic coordinates. These known (x,y) positions define the trajectory, but they are not enough for camera frequency and, therefore, more poses must be estimated between them. The *interval* has been defined to represent the idea of the trajectory sections, and it is a straight line between two consecutive known (x,y) positions. The estimated poses are located in the intervals. Figure 7 represents the intervals, known (x,y) positions and estimated poses in the railway. The main concepts of the ground truth generation algorithm are described in 1.

As the data sources are synchronized at the sequence ending, from now on, the ground truth generation is done in a

Algorithm 1 Ground Truth Data Generation Algorithm

```
Input: Given an interval (i) defined as a straight line between two known (x,y) positions
Phase 1 - Synchronize last (x,y) position, last image and ATP data of an interval
1: if $i = 0$ then                                              ▷ First interval
2:     Last image ← $SSIM > threshold$                          ▷ SSIM [86]
3:     Last $(x_i, y_i)$ position ← given in the interval definition
4:     ATP data ← $train\_stopped = 1$
5: else                                                         ▷ Following intervals
6:     Last image, last $(x_i, y_i)$ position and ATP data ← taken from $i - 1$
7: end if
Phase 2 - Estimate poses on an interval through a backward data synchronization based on timestamps
Input: $V_n$: train speed, $a_n$: train acceleration, $t$: timestamp, $h$: height profile, $d_n$: linear position estimation
8: Estimate translation component of poses $(tr_n)$
    a: $(x_n, y_n) \leftarrow f(v_n, a_n, t)$                   ▷ Eqn. 2
    b: $z_n \leftarrow h(d_n)$                                  ▷ Eqn. 1
9: Estimate rotation component of poses $(r_n)$
    a: $r_n \leftarrow g(tr_{n-1}, tr_n)$                       ▷ Eqn. 3, 4, 5
```

backward data synchronization process of an *interval* based on the images timestamps. The last (x,y) position, last image and ATP data are taken for a given interval and the poses for all timestamps in that interval are estimated. Then, the poses of the following interval are estimated by taking the last (x,y) position and the last image of the previous interval as the initial position.

However, the train speed is variable and, therefore, the distribution of these poses can not be linear in different intervals. The total number of poses within the whole sequence should match the record frame amount.

1) SYNCHRONIZE LAST (x,y) POSITION, LAST IMAGE AND ATP DATA

The first step is to synchronize the different data sources using the last (x,y) position, last image and ATP data. The algorithm generates ground-truth poses for each recorded sequence using the position where the train has stopped as origin. For that, first the image where train stops (last image of the sequence) must be estimated. When there is motion, the similarity between consecutive frames is very low, however the similarity increases when the train has stopped. Due to the similarity of the frames corresponding to the train stopping point, the last frame is selected using the Structural Similarity Index (SSIM) [86]. SSIM is one of the most standard algorithms for image quality assessment [57], and therefore, for image similarity measure. It has shown that can outperform other common image similarity measurements as MSE [87] and has been previously referenced [88]. The SSIM measures the luminance, contrast, and structure of two given images and returns a similarity value between them.

Also, it only requires a starting optimization phase where the threshold is selected. Furthermore, the index was used to find just the first image within the threshold in each sequence, which gives a little number of results totally. Although SSIM is sensitive to image distortions, the environment being static, and the view fixed enables the SSIM application in underground railway scenarios.

The threshold was selected by exploratory testing. A predefined threshold was stated and iterated it until a SSIM threshold that best fitted to the lighting conditions of the scenario was identified. In this case a $SSIM > 0.965$ has been used as similarity threshold at the train stopping point.

The last (x,y) coordinates refer to the train stopping position; therefore, this coordinate pair and the last image are already synchronized. Finally, ATP monitored data is synchronized using the *train stopped* variable.

2) ESTIMATE POSES OF AN INTERVAL THROUGH A BACKWARD DATA SYNCHRONIZATION BASED ON TIMESTAMPS

A ground truth pose is generated for each recorded image in an interval using a backward synchronization process based on the timestamp. This process has two steps; first, the translation component is estimated, and then, the rotation is calculated from that translation.

a: ESTIMATION OF TRANSLATION COMPONENT

Translation component $T = \{tr_0, tr_1, \ldots, tr_m\}$ is defined as a set containing all the 3-DoF poses ($tr_n = [x_n, y_n, z_n]$) of an interval where n is the pose number ($0 \leq n \leq m$) and m is the total number of poses for that interval. For the translation component of a pose, first, the (x,y) position is estimated, and then the height (z) is added. The translation is estimated by taking an initial (x,y) position and calculating the motion to the next one using the ATP data *train speed* and *train acceleration*. The translation between two consecutive (x,y) positions in a straight line that forms the interval can be calculated using *Uniformly Accelerated Motion (UAM)* equations. This estimation is possible because it is considered that the poses follow a motion in a straight line and with a constant acceleration between them. Equation 2 shows the application of UAM equations in this case.

$$d_n = v_{n-1}t + \frac{1}{2}a_{n-1}t^2\,, \tag{2}$$

where t refers to the timestamp, v_n and a_n refer to ATP data train speed and acceleration respectively. The initial (x,y) translation component is set as [0, 0].

After calculating the (x,y) positions, the z or height is estimated using the height profile estimated from the gradient profile and ATP data. The railway height profile can be synchronized with the train stopping point, and therefore, with the first (x,y) position.

Then, previously calculated (x,y) positions can be used to extract the Euclidean distance traveled from position to position. Each pose's height (z) is calculated using traveled distances and the height profile. Therefore, after height estimation, the translation component of a pose has been estimated with respect to a timestamp.

b: ESTIMATION OF ROTATION COMPONENT

Rotation component $R = \{r_0, r_1, \ldots, r_m\}$ is defined as a set containing all the rotation matrices (r_n) within an interval where n is the pose number ($0 \leq n \leq m$) and m is the total number of poses for that interval calculated in the previous steps.

To calculate the rotation component r_n for each translation tr_n the transformation between two consecutive orientation vectors or_{n-1} and or_n is estimated. or_n defines the orientation of the train in tr_n and represents the vector between consecutive translations tr_{n-1} and tr_n. It is calculated as shown in 3:

$$\mathrm{or_n}(\mathrm{tr}_{n-1}, \mathrm{tr}_n) = (x_n - x_{n-1}, y_n - y_{n-1}, z_n - z_{n-1}), \quad (3)$$

where x, y and z represent the translation components of tr_{n-1} and tr_n. Then, using the axis-angle representation, the transformation between consecutive orientation vectors or_{n-1} and or_n can be calculated. For that, first the orientation vectors are normalized by dividing their value with the Euclidean norm (vector magnitude) $\|or_n\|$ of each vector (Eqn. 4) to align them at the same origin. The Euclidean norm can also be defined as the Euclidean distance of a vector from the origin to a point.

$$\mathrm{normalize}(\mathrm{or}_n) = \frac{\mathrm{or}_n}{\|\mathrm{or}_n\|}, \quad (4)$$

Then, the Euclidean norm of the cross product between the normalized consecutive orientations is estimated to get the axis. Finally, the rotation component is estimated using the inverse tangent function as shown in equation 5, where the angle between the orientations vectors is calculated trough the dot product:

$$r_n = \arccos\left(\frac{\|\mathrm{or}_n \times \mathrm{or}_{n-1}\|}{\mathrm{or}_n \cdot \mathrm{or}_{n-1}}\right), \quad (5)$$

where arccos refers to the inverse cosine function and or_{n-1} and or_n to two consecutive orientation vectors. This rotation estimation method accumulates an error relative to the previous estimations. However, as the train is tied to the rails, the trains' orientation is always fixed, and the orientation estimation is not critical.

The previously calculated translation component is added to the newly calculated rotation component to obtain the target 6-DoF ground truth pose. This is done by following the representation in equation 1.

Once all the poses from a given interval have been estimated, the next interval is taken and the process is repeated until all the intervals of a sequence have been covered.

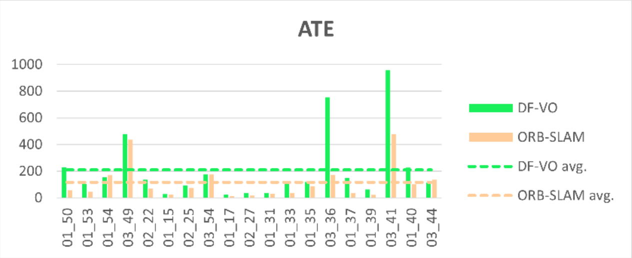


FIGURE 8. **ATE of DF-VO and ORB-SLAM2 application on the generated *CAF* dataset.**

IV. VO APPLICATION IN URBAN UNDERGROUND RAILWAY ENVIRONMENT

In this section the application of DF-VO and ORB-SLAM2 in the CAF dataset is evaluated.

In the following subsection, the standard VO evaluation metrics are explained. Then, the experimentation setup is described. Finally, the experimental results are discussed.

A. VO EVALUATION METRICS

The metrics used to evaluate the performance of the experiments are the following: Absolute Trajectory Error – *ATE* [67], Relative Pose Error – *RPE* [67], Average Translational Error – t_{err} and Average Rotational Error – r_{err}.

All the sequences were transformed with a 6-DoF Umeyama alignment [89], a standard alignment method used in most VO and SLAM evaluation benchmarks. [2]. A 6-DoF alignment is recommended to evaluate shape similarities of trajectories [90].

Given this transformation, ATE evaluates the global consistency of an estimated trajectory compared to the ground-truth trajectory. The RPE measures the drift error for each pose of the trajectory and the rotation and the translation components are calculated separately.

Finally, following KITTI evaluation benchmark criteria, the Average Translational Error (t_{err}) and the Average Rotational Error (r_{err}) are calculated on sub-sequences of different lengths. These errors measure the average relative pose error at a fixed distance. The sub-sequences length in meters is (100,200,…,800) because the error for smaller sub-sequences was large and hence biased the evaluation results.

B. EXPERIMENTATION SETUP

These experiments extend the evaluation done at [9], where ORB-SLAM2 and DF-VO were evaluated in an outdoor urban car driving scenario. In those experiments, the bad lighting conditions were replicated by car driving recordings in the night.

DF-VO implementation [91] flow-weights and depth estimation deep models were selected from the authors' trained models. The flow model is trained by the authors in the synthetic dataset Scene Flow [92].

To handle the non-deterministic nature of the ORB-SLAM2 algorithm, each sequence is run five times, and the

TABLE 3. **DF-VO and ORB-SLAM2 application evaluation using standard VO evaluation metrics: Average Translational Error (t_{err}), Average Rotational Error (r_{err}), ATE and RPE. The sequences are organized by the direction they are recorded. The average errors for all 19 sequences are calculated, and the best result is in bold.**

Algorithm	Record	14_11_2021 (->Matiko)									
	Seq	01_50	01_53	01_54	03_49	02_22	01_15	02_25	03_54	01_17	02_27
DF-VO	t_{err} (%)	80	52.97	85.74	86.27	77.94	64.09	94.52	111.55	56.61	55.69
	r_{err} (°/100m)	13.21	21.25	29.92	18.48	35.88	24.43	33.5	41.71	21.34	21.33
	ATE	230.66	106.38	157.46	478.76	135.36	29.11	94.27	175.64	26.1	36.39
	RPE (m)	0.402	0.232	0.354	0.423	0.269	0.236	0.314	0.34	0.132	0.133
	RPE (°)	0.156	0.135	0.156	0.176	0.124	0.13	0.157	0.143	0.057	0.062
ORB-SLAM2	t_{err} (%)	68.79	54.93	177.16	125.85	80.28	55.03	94.26	136.06	51.89	53.88
	r_{err} (°/100m)	5.95	19.19	50.42	17.2	25.6	20.17	31.97	39.34	15.21	14.24
	ATE	56.58	44.98	169.39	435.36	72.35	26.71	74.61	177.23	15.61	17.1
	RPE (m)	0.34	0.192	0.641	0.646	0.277	0.204	0.302	0.371	0.116	0.122
	RPE (°)	0.081	0.092	0.429	0.264	0.125	0.104	0.11	0.121	0.065	0.064

Algorithm	Record	14_11_2021 (->Kukullga)									Avg. Err.
	Seq	01_31	01_33	01_35	03_36	01_37	01_39	03_41	01_40	03_44	
DF-VO	t_{err} (%)	63.64	135.76	175.28	148.11	136.82	58.1	110	93.32	99.36	**93.9879**
	r_{err} (°/100m)	17.81	28.84	29.64	23.51	27.19	21.01	22.21	31.19	41.2	26.50789
	ATE	38.38	104.98	119.08	754.45	150.64	62.88	957.69	226.86	114.75	210.5179
	RPE (m)	0.183	0.467	0.583	0.541	0.594	0.192	0.365	0.332	0.35	0.35389
	RPE (°)	0.088	0.13	0.132	0.116	0.166	0.103	0.111	0.103	0.147	0.125895
ORB-SLAM2	t_{err} (%)	58.49	145.36	185.8	101.98	155.7	51.05	125.01	94.92	96.31	100.6711
	r_{err} (°/100m)	16.71	22.53	24.31	20.11	17.51	17.15	18.82	10.41	10.41	**20.9079**
	ATE	30.67	38.47	88.96	172.66	36.31	22.28	478.87	103.74	137.45	**115.754**
	RPE (m)	0.167	0.498	0.649	0.388	0.672	0.154	0.362	0.311	0.312	**339053**
	RPE (°)	0.09	0.099	0.113	0.11	0.113	0.079	0.097	0.07	0.07	**0.12084**

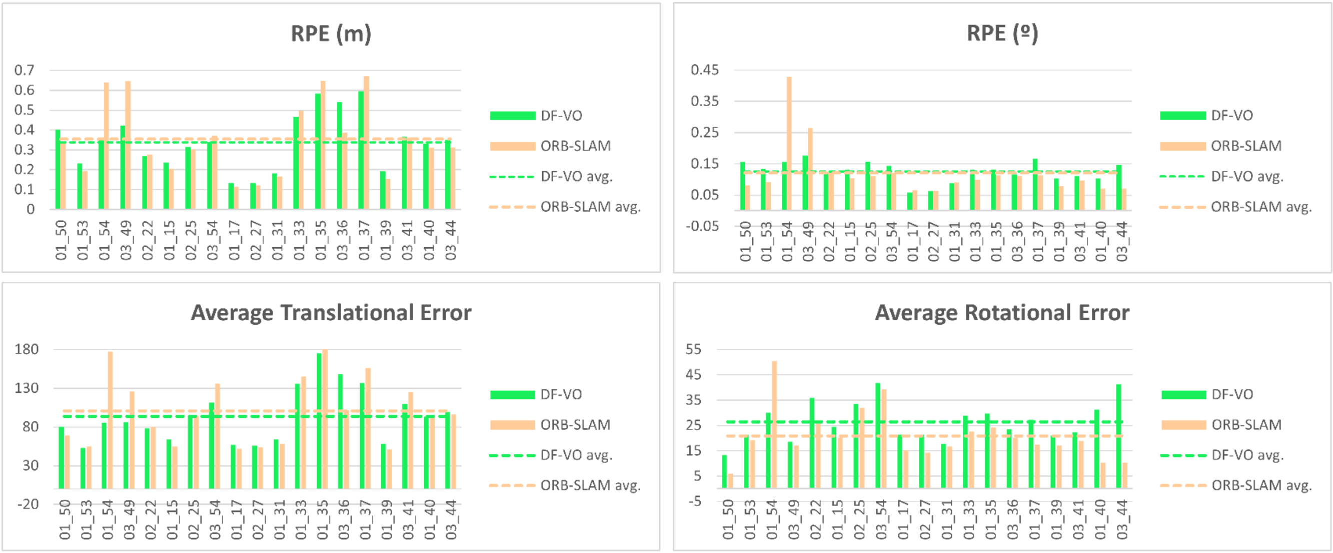


FIGURE 9. **Comparison of relative VO evaluation metrics when applying DF-VO and ORB-SLAM2 algorithms in CAF datasets. Translational and rotational components of relative errors are shown separately.**

median accuracy is evaluated as proposed by authors in [21]. The VO evaluation is done using the *KITTI Odometry Evaluation Toolbox* [17].

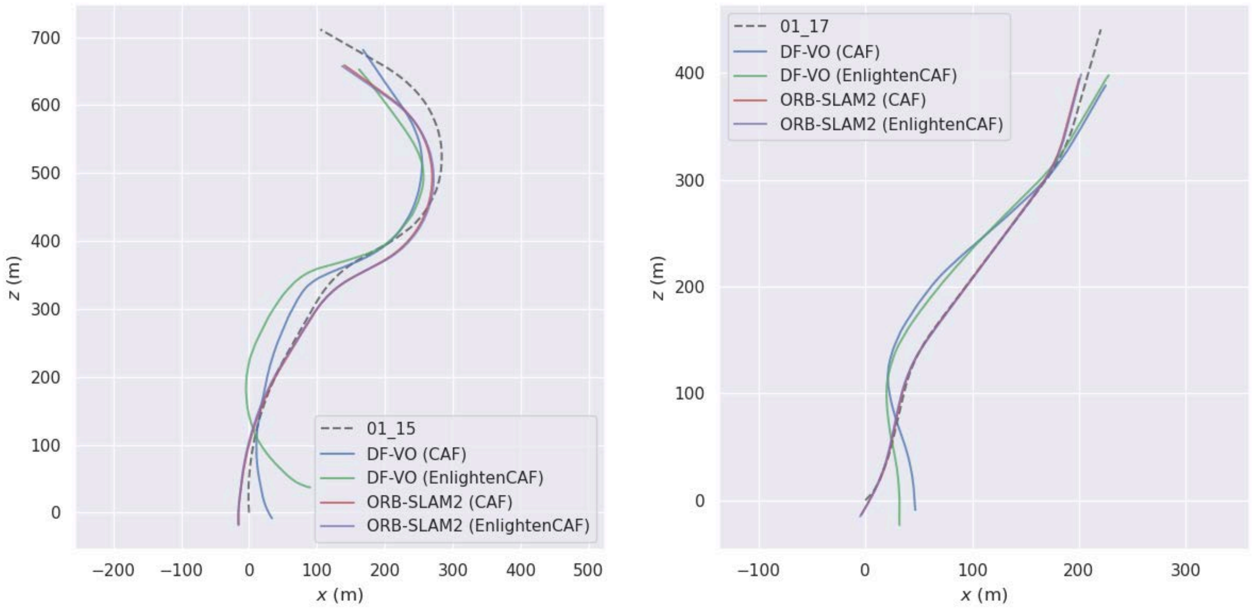
C. VO RESULTS IN CAF DATASET

Table 3 shows the results of DF-VO and ORB-SLAM2 in the CAF dataset. Figures 8 and 9 represent the results depicted in table 3. The visual representation can be found in Figure 9.

Previously, DF-VO and ORB-SLAM2 were evaluated in the KITTI Odometry dataset; however, KITTI does not contain those perception challenges as it contains considerably different properties related to the sequence length and visual characteristics. Results in *CAF* dataset show that the errors of both algorithms are higher than those found in the KITTI dataset. The RPE for DF-VO is 0.038 and 0.339 in KITTI dataset and *CAF* dataset, respectively. While for ORB-SLAM2, RPE measures are 0.130 and 0.353.

In the case of the ATE, the error of DF-VO in KITTI dataset is 6.344 while in the *CAF* dataset is 210.517. For ORB-SLAM2, the ATE is 26.48 and 115.754 in KITTI dataset and *CAF* dataset, respectively.

It can be seen that ORB-SLAM2 outperforms DF-VO in this challenging scenario, where the sequences are longer



(a) Sequence 01_15

(b) Sequence 01_17

FIGURE 10. Comparison of ORB-SLAM2 and DF-VO application on two sample sequences in both CAF and EnlightenCAF datasets and the ground truth for each trajectory.

TABLE 4. Average standard VO errors in *CAF* dataset when reducing the sequences to platform areas without lighting constraints.

Algorithm	Metric	Avg. Err
DF-VO	t_{err} (%)	18.520
	r_{err} (°/100m)	6.975
	ATE	2.298
	RPE (m)	0.049
	RPE (°)	0.037
ORB-SLAM2	t_{err} (%)	19.484
	r_{err} (°/100m)	14.681
	ATE	4.113
	RPE (m)	0.0798
	RPE (°)	0.126

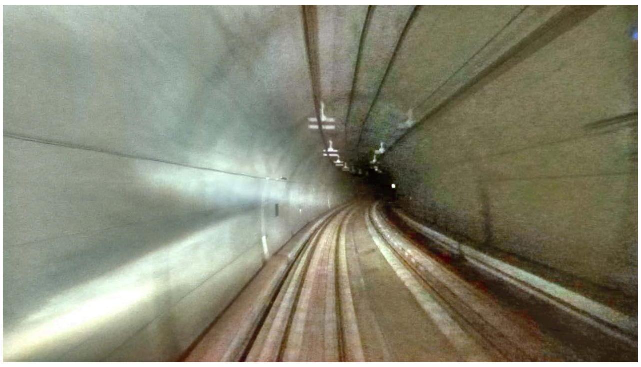


FIGURE 11. A frame from the *CAF* dataset enhanced by EnlightenGAN.

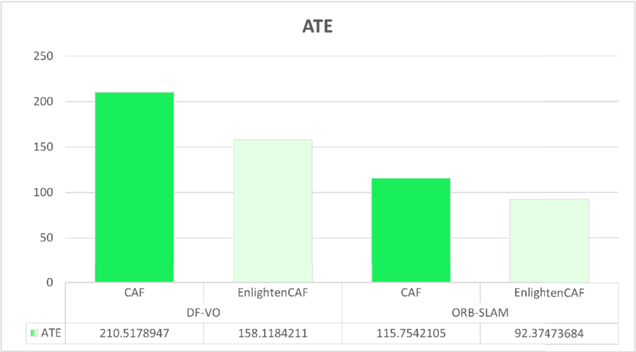


FIGURE 12. Comparative of ATE when applying DF-VO and ORB-SLAM2 algorithms in CAF and EnlightenCAF datasets.

than the standard KITTI dataset. If the *CAF* sequences are shortened to just platform areas where the lighting challenges are more limited, and more similar to the lighting conditions of the KITTI dataset, the errors are reduced to similar values (see Table 4) of executing DF-VO, and ORB-SLAM2 in KITTI dataset [17], [21]. For instance, DF-VO achieves an RPE (m) of 0.027 in KITTI dataset and 0.049 in shortened *CAF* dataset. ORB-SLAM2 achieves an ATE of 9.464 in KITTI dataset while 4.113 is achieved in shortened *CAF* dataset. Furthermore, the same behavior as in KITTI dataset is observed: DF-VO performance is higher than ORB-SLAM2. These results seem to support that the challenging scene conditions hinder the application of VO algorithms in such scenarios.

Results are visually shown in figure 10. In the case of DF-VO, a scale misalignment can be appreciated as the shape of most estimated trajectories is similar to the ground truth shape, but a dimensionality error appears.

As mentioned in [17], geometry-based VO algorithms as ORB-SLAM2 suffer from a scale drift when ideal visual conditions are not met. In the case of DF-VO, being a

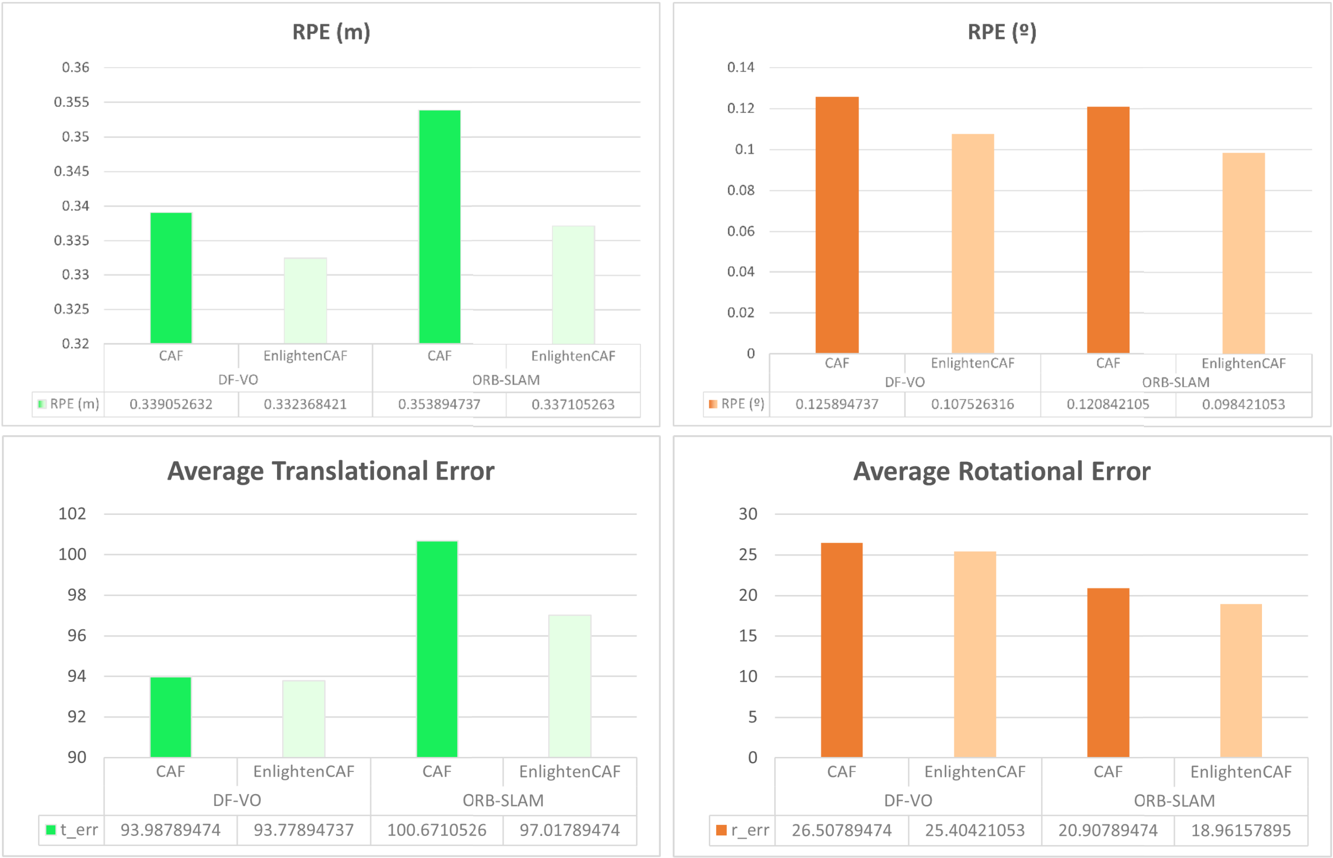


FIGURE 13. Comparison of relative VO evaluation metrics when applying DF-VO and ORB-SLAM2 algorithms in CAF and EnlightenCAF datasets. Translational and rotational components of relative errors are shown separately.

hybrid algorithm, the scale may be wrongly estimated due to issues related to the geometric characteristics of the underground visual domain or deep-learning training process. The estimation error of the learning part of the algorithm could be reduced by training the deep models in the target scenario.

Nevertheless, these results require an adaptation of reference VO solutions to increase the performance in the underground railway domain. Image enhancement techniques or solutions based on the fusion of different odometry sensors could provide the precision required by autonomous train operations.

V. ENLIGHTENGAN IN VO APPLICATION

This section explores the application of the image enhancement technique EnlightenGAN in ORB-SLAM2 and DF-VO algorithms.

VO algorithms are based on minimizing the reprojection error of consecutive frames captured by the camera. The error is estimated by solving the essential matrix, which depends on the intrinsic camera parameters, and assuming the camera satisfies the pinhole camera model. In a previous work, the enhanced images calibration procedure was pursued to assess the EnglithenGAN architecture's effect on the camera's calibration. The experimental results showed that the GAN architecture did not significantly disturb the camera calibration parameters. Therefore, it was concluded that VO algorithms could be applied directly to the dataset enhanced by EnlightenGAN.

In the following section the enhanced dataset generation, the experimental configuration, and, finally, the results are explained.

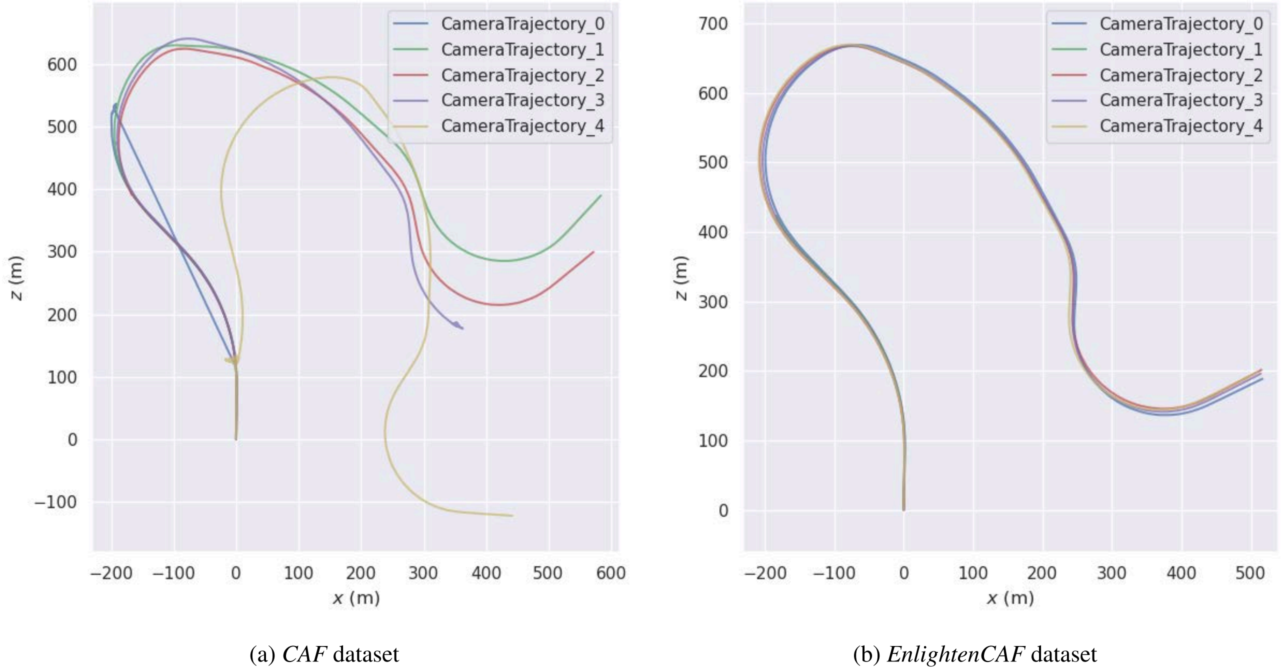
A. ENHANCED DATASET GENERATION: EnlightenCAF

The CAF dataset enhanced by EnlightenGAN is named *EnlightenCAF*. Figure 11 shows the result of the enhancement in the same tunnel zone frame as in figure 1.

The same algorithm configuration from *CAF* dataset experimentation has been used. The enhancing inference model is composed of pretrained weights from original authors.

B. RESULTS IN EnlightenCAF

In the previous work [9], the experimental results showed that *EnlightenGAN* improves the DF-VO performance in low-light car scenarios. In this case, the same behavior was confirmed: quantitative results show that EnlightenGAN reduces the VO errors for both algorithms. Figure 12 shows the reduction in the mean ATE and mean RPE of both algorithms for all the sequences in *EnlightenCAF*.



(a) *CAF* dataset

(b) *EnlightenCAF* dataset

FIGURE 14. Pose dispersion analysis on sample trajectory *01_54*. ORB-SLAM2 algorithm is executed five times on each dataset.

A relative ATE reduction of 24.89% and 20.20% is observed, respectively, when DF-VO and ORB-SLAM2 are applied in the enhanced sequences. Figure 13 shows RPE, t_{err} and r_{err} evaluation metrics in EnlightenCAF dataset.

In the case of RPE, DF-VO algorithm obtains a relative improvement of 1.97% and 4.74% for translation and rotation components, respectively. ORB-SLAM2 gets a relative improvement of 14.59% for the RPE translation component and a relative improvement of 18.55% for the rotation component. t_{err} and r_{err} present a relative reduction of 0.22% and 4.16% when applying DF-VO, and a relative reduction of 3.63% and 9.31% when applying ORB-SLAM2.

Figure 10 shows a result comparison of DF-VO and ORB-SLAM2 in the sequences of CAF and EnlightenGAF. As in *CAF* dataset, it can be seen that the algorithms can estimate the shape of the *EnlightenCAF* trajectories. However, a scale underestimation problem appears again. Furthermore, DF-VO results show that the rotation estimation is affected in the *EnlightenCAF* dataset.

The results demonstrate that EnlightenGAN improves VO algorithms performance in the underground railway domain. Furthermore, the relative error is reduced more for the geometric-based VO algorithm, while absolute error is reduced more in the learning-based algorithm.

However, as in the *CAF* dataset, the errors continue being higher than the results obtained by the algorithms in the KITTI dataset. Therefore, an affection of lighting conditions of the scenario can still be appreciated. This affection could be related to scale underestimation problems found in both algorithms, especially in the hybrid DF-VO.

Additionally, when evaluating the VO algorithms, it has been seen that the dispersion of the poses estimated by ORB-SLAM2 in different runs is reduced when enhancing the frames with EnlightenGAN.

The dispersion of poses among different executions of ORB-SLAM2 has been evaluated using standard metrics [93]. These metrics include the *variance* (σ^2) and the *Coefficient of Variation* (*cv*).

The evaluation procedure has been to run ORB-SLAM2 five times in each dataset, the original *CAF* and the enhanced *EnlightenCAF*. Figure 14 shows the results of applying ORB-SLAM2 five times for a given sequence (*01_54*) in the *CAF* and the enhanced *EnlightenCAF* datasets. It can be seen that the distribution of the poses through the trajectory is more constant in the enlightened dataset.

From the results, it can be seen that enlightening the datasets with *EnlightenGAN* increases the VO performance and tends to reduce ORB-SLAM2 dispersion. An analysis of the trouble spots in the dispersion results could better understand the high dispersion in such frames and detect further possible improvements for VO algorithms in such scenarios.

VI. CONCLUSION

This paper has presented a method to create a ground truth database for underground railway scenarios, where the GPS is unavailable, or the access to the infrastructure is not easily granted. The ground truth data generation is based on camera frames, ERTMS/ETCS ATP data, the railway gradient profile map, and geodetic coordinates of the target railway. Second, it has proposed to enhance image lighting conditions with EnlightenGAN, which can be used with any state-of-the-art VO. Finally, it has presented the result of the experiment performed within a real urban underground railway scenario. The scenario was characterized by varying lighting conditions (tunnel vs. platform), low illumination (in tunnels), or textureless areas that challenged the state-of-the-art VO algorithms. The experiments were performed using two VO approaches: geometric (ORB-SLAM) and hybrid (DF-VO). The results show that the data enhancement increases the performance of both VO algorithms, reducing the translational error by at least 18%.

Future research proposes to apply the proposed dataset generation method and image enhancement algorithm in more underground railway scenarios. Sensor fusion is also a promising research direction. It is expected that the inclusion of new sensors will reduce uncertainty and increase accuracy, which will be welcome for autonomous train operations requiring higher localization accuracy (e.g., precise train stop operation).

ACKNOWLEDGMENT

The authors would like to thank Euskotren and Eusko Trenbide Sarea.

REFERENCES

[1] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An overview to visual odometry and visual SLAM: Applications to mobile robotics," *Intell. Ind. Syst.*, vol. 1, no. 4, pp. 289–311, Dec. 2015.

[2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[4] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016.

[5] C. G. Atkeson, P. W. Benzun, N. Banerjee, D. Berenson, C. P. Bove, X. Cui, M. DeDonato, R. Du, S. Feng, P. Franklin, and M. A. Gennert, "Achieving reliable humanoid robot operations in the DARPA robotics challenge: Team WPI-CMU's approach," in *Proc. Int. Conf. Modeling Simul. Auton. Syst.* Cham, Switzerland: Springer, 2018, pp. 271–307.

[6] T. Rouček, M. Pecka, P. Čížek, T. Petrícek, J. Bayer, V. Šalanský, D. Hert, M. Petrlík, T. Báca, V. Spurný, and F. Pomerleau, "Darpa subterranean challenge: Multi-robotic exploration of underground environments," in *Proc. Int. Conf. Modeling Simul. Auto. Syst.* Cham, Switzerland: Springer, 2019, pp. 274–290.

[7] K. Ebadi, Y. Chang, M. Palieri, A. Stephens, A. Hatteland, E. Heiden, A. Thakur, N. Funabiki, B. Morrell, S. Wood, L. Carlone, and A.-A. Agha-mohammadi, "LAMP: Large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2020, pp. 80–86.

[8] A. Agha, K. Otsu, B. Morrell, D. D. Fan, R. Thakker, A. Santamaria-Navarro, S.-K. Kim, A. Bouman, X. Lei, J. Edlund, and M. F. Ginting, "NeBula: Quest for robotic autonomy in challenging environments; TEAM CoSTAR at the DARPA subterranean challenge," 2021, *arXiv:2103.11470*.

[9] J. Z. Ansorregi, M. E. Garcia, M. Z. Akizu, and N. A. Arexolaleiba, "Image enhancement using GANs for monocular visual odometry," in *Proc. IEEE Int. Workshop Electron., Control, Meas., Signals Appl. Mechatronics (ECMSM)*, Jun. 2021, pp. 1–6.

[10] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, p. 1.

[11] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] G. P. Stein, O. Mano, and A. Shashua, "A robust method for computing vehicle ego-motion," in *Proc. IEEE Intell. Vehicles Symp.*, Oct. 2000, pp. 362–368.

[13] K. Yamaguchi, T. Kato, and Y. Ninomiya, "Vehicle ego-motion estimation and moving object detection using a monocular camera," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006, pp. 610–613.

[14] F. Tschopp, T. Schneider, A. W. Palmer, N. Nourani-Vatani, C. Cadena, R. Siegwart, and J. Nieto, "Experimental comparison of visual-aided odometry methods for rail vehicles," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1815–1822, Apr. 2019, doi: 10.1109/LRA.2019.2897169.

[15] V. Grabe, H. H. Bŭlthoff, D. Scaramuzza, and P. R. Giordano, "Nonlinear ego-motion estimation from optical flow for online control of a quadrotor UAV," *Int. J. Robot. Res.*, vol. 34, no. 8, pp. 1114–1135, 2015.

[16] Y. Zou, P. Ji, Q.-H. Tran, J.-B. Huang, and M. Chandraker, "Learning monocular visual odometry via self-supervised long-term modeling," in *Computer Vision*. Glasgow, U.K.: Springer, Aug. 2020, pp. 710–727.

[17] H. Zhan, C. S. Weerasekera, J.-W. Bian, R. Garg, and I. Reid, "DF-VO: What should be learnt for visual odometry?" 2021, *arXiv:2103.00933*.

[18] H. Gaoussou and P. Dewei, "Evaluation of the visual odometry methods for semi-dense real-time," *Adv. Comput., Int. J.*, vol. 9, no. 2, pp. 01–14, Mar. 2018, doi: 10.5121/acij.2018.9201.

[19] R. Wang, M. Schworer, and D. Cremers, "Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3903–3911.

[20] H. Alismail, M. Kaess, B. Browning, and S. Lucey, "Direct visual odometry in low light using binary descriptors," *IEEE Robot. Automat. Lett.*, vol. 2, no. 2, pp. 444–451, Apr. 2016.

[21] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[22] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.

[23] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular SLAM with map reuse," *IEEE Robot. Automat. Lett.*, vol. 2, no. 2, pp. 796–803, Apr. 2017.

[24] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual SLAM algorithms: A survey from 2010 to 2016," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, pp. 1–11, Dec. 2017.

[25] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2502–2509.

[26] B. Bescos, J. M. Fácil, J. Civera, and J. L. Neira, "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4076–4083, Oct. 2018.

[27] C. Yu, Z. Liu, X.-J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, "DS-SLAM: A semantic visual SLAM towards dynamic environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1168–1174.

[28] F.-A. Moreno, D. Zuñiga-Noël, D. Scaramuzza, and J. Gonzalez-Jimenez, "PL-SLAM: A stereo SLAM system through the combination of points and line segments," *IEEE Trans. Robot.*, vol. 35, no. 3, pp. 734–746, Jun. 2019.

[29] N. Yang, R. Wang, J. Stuckler, and D. Cremers, "Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 817–833.

[30] H. Zhan, C. S. Weerasekera, J.-W. Bian, and I. Reid, "Visual odometry revisited: What should be learnt?" in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2020, pp. 4203–4210.

[31] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2014, pp. 15–22.

[32] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, Mar. 2017.

[33] L. Koestler, N. Yang, N. Zeller, and D. Cremers, "Tandem: Tracking and dense mapping in real-time using deep multi-view stereo," in *Proc. Conf. Robot Learn.*, 2022, pp. 34–45.

[34] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DOF camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2938–2946, doi: 10.1001/jama.284.15.1980.

[35] A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6555–6564, doi: 10.1109/CVPR.2017.694.

[36] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 2043–2050.

[37] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother, "DSAC-differentiable ransac for camera localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6684–6692.

[38] S. J. Lee, H. Choi, and S. S. Hwang, "Real-time depth estimation using recurrent CNN with sparse depth cues for SLAM system," *Int. J. Control, Autom. Syst.*, vol. 18, no. 1, pp. 206–216, Jan. 2020.

[39] G. Costante and T. A. Ciarfuglia, "LS-VO: Learning dense optical subspace for robust visual odometry estimation," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1735–1742, Jul. 2018.

[40] B. Ummenhofer, H. Zhou, J. Uhrig, N. Mayer, E. Ilg, A. Dosovitskiy, and T. Brox, "DeMoN: Depth and motion network for learning monocular stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, p. 50.8–5047.

[41] K. Tateno, F. Tombari, I. Laina, and N. Navab, "CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6243–6252.

[42] H. Zhan, R. Garg, C. S. Weerasekera, K. Li, H. Agarwal, and I. M. Reid, "Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 340–349.

[43] N. Yang, L. von Stumberg, R. Wang, and D. Cremers, "D3 VO: Deep depth, deep pose and deep uncertainty for monocular visual odometry," 2020, *arXiv:2003.01060*.

[44] Z. Yin and J. Shi, "GeoNet: Unsupervised learning of dense depth, optical flow and camera pose," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1983–1992.

[45] R. Mahjourian, M. Wicke, and A. Angelova, "Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5667–5675.

[46] T. Shen, Z. Luo, L. Zhou, H. Deng, R. Zhang, T. Fang, and L. Quan, "Beyond photometric loss for self-supervised ego-motion estimation," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 6359–6365.

[47] M. O. A. Aqel, M. H. Marhaban, M. I. Saripan, and N. B. Ismail, "Review of visual odometry: Types, approaches, challenges, and applications," *SpringerPlus*, vol. 5, no. 1, pp. 1–26, Dec. 2016.

[48] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[51] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE: A raw images dataset for digital image forensics," in *Proc. 6th ACM Multimedia Syst. Conf.*, Mar. 2015, pp. 219–224.

[52] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–144, 2017.

[53] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.

[54] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," 2018, *arXiv:1808.04560*.

[55] Y. Gong, P. Liao, X. Zhang, L. Zhang, G. Chen, K. Zhu, X. Tan, and Z. Lv, "Enlighten-GAN for super resolution reconstruction in mid-resolution remote sensing images," *Remote Sens.*, vol. 13, no. 6, p. 1104, Mar. 2021.

[56] M. Afifi, K. G. Derpanis, B. Ommer, and M. S. Brown, "Learning multi-scale photo exposure correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9157–9167.

[57] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Trans. Image Process.*, vol. 29, pp. 9140–9151, 2020.

[58] W. Xiong, D. Liu, X. Shen, C. Fang, and J. Luo, "Unsupervised low-light image enhancement with decoupled networks," 2020, *arXiv:2005.02818*.

[59] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi, "Real-time RGB-D camera relocalization," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2013, pp. 173–179.

[60] Y. Li, N. Snavely, and D. P. Huttenlocher, "Location recognition using prioritized feature matching," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 791–804.

[61] A. Handa, T. Whelan, J. McDonald, and A. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM," in *Proc. IEEE Intl. Conf. Robot. Automat. (ICRA)*, Hong Kong, May 2014, pp. 1524–1531.

[62] S. Cortés, A. Solin, E. Rahtu, and J. Kannala, "ADVIO: An authentic dataset for visual-inertial odometry," in *Computer Vision*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 425–440.

[63] D. Zuñiga-Noël, A. Jaenal, R. Gomez-Ojeda, and J. Gonzalez-Jimenez, "The UMA-VI dataset: Visual–inertial odometry in low-textured and dynamic illumination environments," *Int. J. Robot. Res.*, vol. 39, no. 9, pp. 1052–1060, Aug. 2020, doi: 10.1177/0278364920938439.

[64] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti, and P. Taddei, "Rawseeds ground truth collection systems for indoor self-localization and mapping," *Auton. Robots*, vol. 27, no. 4, pp. 353–371, 2009. [Online]. Available: http://dblp.uni-trier.de/db/journals/arobots/arobots27.html#CerianiFGMM%MRST09

[65] J. Xiao, A. Owens, and A. Torralba, "SUN3D: A database of big spaces reconstructed using sfm and object labels," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1625–1632.

[66] S. Klenk, J. Chui, N. Demmel, and D. Cremers, "TUM-VIE: The TUM stereo visual-inertial event dataset," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 8601–8608.

[67] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 573–580.

[68] J. Engel, V. Usenko, and D. Cremers, "A photometrically calibrated benchmark for monocular visual odometry," 2016, *arXiv:1607.02555*.

[69] F. Walch, C. Hazirbas, L. Leal-Taixe, T. Sattler, S. Hilsenbeck, and D. Cremers, "Image-based localization using LSTMs for structured feature correlation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 627–637.

[70] M. Fallon, H. Johannsson, M. Kaess, and J. J. Leonard, "The MIT stata center dataset," *Int. J. Robot. Res.*, vol. 32, no. 14, pp. 1695–1699, 2013.

[71] H. Alismail, B. Browning, and M. B. Dias, "Evaluating pose estimation methods for stereo visual odometry on robots," in *Proc. 11th Int. Conf. Intell. Auton. Syst. (IAS)*, vol. 3, 2010, p. 2.

[72] T. Schops, J. L. Schonberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger, "A multi-view stereo benchmark with high-resolution images and multi-camera videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3260–3269.

[73] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of Michigan north campus long-term vision and lidar dataset," *Int. J. Robot. Res.*, vol. 35, no. 9, pp. 1023–1035, Aug. 2015.

[74] J.-L. Blanco-Claraco, F.-Á. Moreno-Dueñas, and J. González-Jiménez, "The Málaga urban dataset: high-rate stereo and LiDAR in a realistic urban scenario," *Int. J. Robot. Res.*, vol. 33, no. 2, pp. 207–214, Feb. 2014, doi: 10.1177/0278364913507326.

[75] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford robotcar dataset," *Int. J. Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2017.

[76] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and LiDAR data set," *Int. J. Robot. Res.*, vol. 30, no. 13, pp. 1543–1552, 2011.

[77] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *Int. J. Robot. Res.*, vol. 38, no. 6, pp. 642–657, May 2019.

[78] D. Olid, J. M. Fácil, and J. Civera, "Single-view place recognition under seasonal changes," in *Proc. PPNIV Workshop IROS*, 2018, pp. 1–6.

[79] A. L. Majdik, C. Till, and D. Scaramuzza, "The Zurich urban micro aerial vehicle dataset," *Int. J. Robot. Res.*, vol. 36, no. 3, pp. 269–273, 2017, doi: 10.1177/0278364917702237.

[80] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2032–2039, Jul. 2018, doi: 10.1109/LRA.2018.2800793.

[81] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1851–1858.

[82] M. T. Ohradzansky, E. R. Rush, D. G. Riley, A. B. Mills, S. Ahmad, S. McGuire, H. Biggie, K. Harlow, M. J. Miles, E. W. Frew, C. Heckman, and J. S. Humbert, "Multi-agent autonomy: Advancements and challenges in subterranean exploration," 2021, *arXiv:2110.04390*.

[83] H. Zhao, B. Zhang, C. Wu, Z. Zuo, and Z. Chen, "Development of a Coordinate Transformation method for direct georeferencing in map projection frames," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 94–103, Mar. 2013.

[84] (2017). ÖPNVKarte Map. *Planet Dump*. [Online]. Available: https://planet.osm.org

[85] (2017). OpenStreetMap Contributors. *Planet Dump*. [Online]. Available: https://planet.osm.org

[86] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[87] Y. Gao, A. Rehman, and Z. Wang, "CW-SSIM based image classification," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1249–1252.

[88] J. Søgaard, L. Krasula, M. Shahid, D. Temel, K. Brunnström, and M. Razaak, "Applicability of existing objective metrics of perceptual quality for adaptive video streaming," *Electron. Imag.*, vol. 28, no. 13, pp. 1–7, Feb. 2016.

[89] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 376–380, Apr. 1991.

[90] (2017). Michael Grupp. *EVO: Python Package for the Evaluation of Odometry and Slam*. [Online]. Available: https://github.com/MichaelGrupp/evo

[91] H. Zhan, C. S. Weerasekera, J. Bian, and I. Reid. (2021). *DF-VO*. [Online]. Available: https://github.com/Huangying-Zhan/DF-VO

[92] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.

[93] C. S. Rayat, "Measures of dispersion," in *Statistical Methods in Medical Research*. Singapore: Springer, 2018, pp. 47–60.

4.4 Semi-Automatic Validation and Verification Framework for CV & AI enhanced Railway Signalling and Landmark Detector

- **Autores: Mikel Labayen** and Xabier Mendialdua and Naiara Aginako and Basilio Sierra
- **Revista:** IEEE Transactions on Instrumentation and Measurement
- **Volumen:** 72
- **Páginas:** 1-13
- **Año:** 2023
- **Editor:** IEEE ↗

Semi-Automatic Validation and Verification Framework for CV&AI-Enhanced Railway Signaling and Landmark Detector

Mikel Labayen, Xabier Mendialdua, Naiara Aginako, and Basilio Sierra

***Abstract*— The automation of railway operations is an activity in constant growth. Different railway stakeholders are already developing their research activities for the future driverless autonomous driving based on computer vision (CV) and artificial intelligence (AI)-enhanced perception technologies (e.g., obstacle detection). Unfortunately, the AI models are opaque in nature, and there are no certification accepted rules for CV&AI-enhanced functionality certification. Capturing and labeling camera image in real environment is expensive in terms of time and resources and it does not guarantee enough variation in edge visibility conditions, which makes the resulting database less valuable for the validation and verification (V&V) processes. To meet the increasing needs of trusted CV&AI-based solutions, numerous V&V approaches have been proposed in other sectors such as automotive, most of them based on virtual simulators. Unfortunately, there is currently no virtual perception simulator for railway scenario. This work aims to create a semi-automatic system based on virtual scenarios measuring the CV&AI-enhanced system performance facing different visibility conditions. It will be based on the global accuracy metrics and detected potential safety and operation rules' violations. This work also demonstrates the quantitative and qualitative improvements while reducing current V&V cost.**

***Index Terms*— Artificial intelligence (AI), autonomous train, certification, perception system, validation, verification.**

I. Introduction

THE European standardization group of Europe Rail is currently working on a system definition of a future grade of automation level 4 (GoA4) driverless automatic train operation (ATO). This solution is in high demand due to the benefits of a reduction in operation cost, prolonging the life cycle of railway products, and an increase in safety. For this reason, and as part of the perception (PER) system of future autonomous trains, railway technology suppliers have already started exploring computer vision and artificial intelligence (CV&AI)-enhanced technologies for artificial sensing of the environment. This artificial perception is based on automatic object detection technologies [1], [2], and it is required for different autonomous driving functionalities, such as obstacle detection [3], [4], [5], railway signal/side-signs' detection [6], [7], railroad and switch detection [8], [9], or visual odometry-localization [10], [11]. On the other hand, this technology is also powering the emerging automatic infrastructure monitoring functionalities, allowing failure detections [12], [13], [14] and avoiding future accidents.

However, the CV&AI-enhanced object detector algorithms for driverless autonomous train operation will need a further substantial effort to increase the technology readiness level (TRL) before bringing it to the market. Railway domain technology will have to satisfy strict standards and safety regulations to be certified. Presently, there are no certification accepted rules for CV&AI-enhanced functionality certification according to the EN-5012x standards which work for deterministic algorithms. This is an ongoing work [15], so, by now, the adoption of solutions based on probabilistic models is still a challenge for railway suppliers.

A common understanding in the CV&AI technology community is that the AI models are as robust as the dataset used to train and test them. However, the generation of specific databases for certification entails the necessity to capture and label very large datasets. In addition, these datasets must contain balanced information about the elements to be detected or identified (e.g., obstacles, railway signs), but also in a range of conditions that may affect their appearance under operative circumstances (e.g., weather conditions, illumination, conservation state, position and size, angle of view, partial occlusions). For this reason, it is then mandatory that the well-validated and verified system has been tested using databases containing different videos/footage representing all kinds of: 1) visibility conditions and 2) situations and behaviors of the static/dynamic objects that are present in the railway environment (e.g., pedestrians, vehicles).

Furthermore, depending on the expected operational design domain (ODD), which determines the situations under which the model is assumed to operate correctly, the preparation of the datasets might be unaffordable, extremely expensive, or very difficult to manage and created due to a hazardous situation which only happens once in a long time or never. In addition, labeling them is usually the main bottleneck. Huge efforts must be made to orchestrate large teams of human annotators to create the labels attached to the images, validate

Manuscript received 30 April 2023; accepted 27 May 2023. Date of publication 13 June 2023; date of current version 27 June 2023. This work was supported in part by the Electronic Components and Systems for European Leadership (ECSEL) and in part by the Industrial Leadership [Information and Communication Technology (ICT)] Programs, under Grant 876852. The Associate Editor coordinating the review process was Dr. Liuyang Zhang. *(Corresponding author: Mikel Labayen.)*

Mikel Labayen is with CAF Signalling, Autonomous Vehicle Department, 20018 Donostia, Spain, and also with the Computer Sciences and Artificial Intelligence Department, University of Basque Country, 20018 Donostia, Spain (e-mail: mlabayen@cafsignalling.com).

Xabier Mendialdua is with the Dependable Embedded Systems, Ikerlan Technology Research Centre, Basque Research and Technology Alliance, 20500 Arrasate, Spain (e-mail: xmendialdua@ikerlan.es).

Naiara Aginako is with the Computer Sciences and Artificial Intelligence Department, University of Basque Country, 20018 Donostia, Spain (e-mail: naiara.aginako@ehu.eus).

Basilio Sierra is with the Computer Sciences and Artificial Intelligence Department, University of Basque Country, 20018 Donostia, Spain (e-mail: b.sierra@ehu.eus).

Digital Object Identifier 10.1109/TIM.2023.3284928

whether they are correct, and that they meet the expected quality requirements.

In summary, these limitations are focalizing the attention into simulated environments that can create synthetic or virtual scenarios. They have sufficient fidelity or realism compared with real material and automatically come with labels which have been produced by a computer program that renders them. However, there is not any available simulator solution for railway environment.

In this context, CAF signaling, as a railway signaling supplier, has been working on CV&AI-based railway signal detector/identifier techniques. After recording data in the field, it trained different object detectors/identifiers. Light signals, static speed restriction panels, platform stopping point signals, platform proximity signals, and so on have been labeled in different video databases to train these custom models. Although the resulting system shows accurate performances in nominal scenarios, it must be tested in a higher variety of situations, extreme conditions, and hazardous situations to consider it validated and verified.

Its current V&V process is based on videos recorded in the field, where lighting and weather conditions cannot be controlled, and which are not properly protected sometimes. In addition to this difficulty for setting a controlled scenario, the video recording process is cumbersome. It is carried out in operator facilities, which demands train, personal, and track availability. This generally requires advance planning to carry out these activities without interfering with the normal operation of the trains, as well as some permissions that can delay the process.

The work presented in this article aims to set the first semi-automatic V&V method in railway environment, based on virtually generated scenarios to test the robustness of the CV&AI-based visual object detector when faced with reduced visibility conditions. Although the solution is suitable for any functionality based on a visual object detector, this method and framework have been validated with railway signaling detector use case. In addition, this work presents the results of different tests which have been carried out to quantify the improvement introduced by the method both qualitatively and quantitatively.

This article is organized as follows. Section II gives an overview of related works and highlights the main differences with our approach. Section III introduced the CV&AI-enhanced system under validation and which is used to evaluate the new proposed V&V method. Section IV describes the whole system overview, the new generated tools, and workflow. Section V contains a description of evaluation criteria established for method validation. Section VI presents system tests to measure the new method performance. Section VII presents the results of the tests and a comparison with the current standard manual V&V method. Finally, Section VIII draws the conclusions and presents future works.

II. Related Work

The CV&AI-enhanced external perception technologies, in particular deep learning (DL), are a critical enabling technology for many of the highly automated applications today. Typical examples include intelligent transport systems (ITSs), where these solutions are used to extract a digital representation of the environment context from the highly dimensional sensor inputs. Unfortunately, the AI models are opaque in nature, with limited output interpretability, while functional safety requirements are strict and require a corresponding safety case [16]. Furthermore, the development of systems that rely on DL introduces new types of faults [17]. To meet the increasing needs of trusted CV&AI-based solutions [18], numerous V&V approaches have been proposed. Most of them are based on synthetic data generation, either by creating limited discrete samples or sequences from real samples or by creating a completely new scenario from a virtual simulator. These techniques can be tackled considering three main approaches as follows.

Fig. 1. Railway virtual environment example: Train Simulator videogame.

1) *Generation From Canonical Images:* This approach implies the utilization of seed examples of the elements of interest from validated sources, where the “canonical” or gold reference examples exist (e.g., real railway signs). These images can then be processed using data augmentation techniques by creating modifications on the canonical examples to cover all the expected appearance variability the elements may have in real life (e.g., adding noise, size, position and perspective transforms, partial occlusion, blurring). A possible drawback of this approach is that the used transforms might represent variants which will never occur in real life or they do but cannot be easily recreated, with the risk of biasing the generated model.
2) *Generation From Virtual Environments (Simulator Engines):* Another method is to simulate an entire recording campaign, running a simulator engine which creates a virtual world where the elements of interest are represented naturally and thus showing all the required variability. In the case of traffic-related domains, there exist several driving simulators and videogames [19] (see Fig. 1) that can be useful for the synthetic generation of data. However, depending on the domain, some realism is necessary, not only in appearance but also in behavior. The rendering effort must be taken into account.
3) *Generation From a Generative Deep Neural Network:* Novel techniques use the power of DL to create the so-called generative adversarial networks (GANs) to create synthetic samples. They are inspired by real examples but randomized to some extent to increase the

variability of the resulting dataset. This approach can combine the advantages of the two previous methods creating large amounts of samples with real-life examples fidelity, while minimizing the disadvantages. GAN is a relatively new technique and paradigm in the AI ecosystem. There is not an established framework or toolset which focuses on GAN generation for synthetic dataset preparation. Nevertheless, many researchers such as Saito et al. [20], Isola et al. [21], Zhu et al. [22], and Ramesh et al. [23] have proposed different GAN models which fulfill such purpose.

Nowadays, the major bulk of system-level testing of autonomous features, for example, in the automotive industry, is carried out through on-road testing [24], [25], [26]. These activities are expensive, dangerous, and ineffective [27]. However, conducting system-level testing through computer simulations is gaining popularity and it seems a feasible and efficient alternative. In fact, automotive V&V as simulation is recognized as one of the main techniques in ISO/PAS 21448, but also in the railway environment. There are a growing number of public domain and commercial simulators that have been developed over the past few years.

As the possible input space when testing ITS external PER systems is practically infinite, attempts to design test cases for comprehensive testing over the space of all possible simulation scenarios are futile. Hence, search-based software testing has been advocated as an effective and efficient strategy to generate test scenarios in simulators [28], [29]. Another line of research proposes techniques to generate test oracles, i.e., mechanisms for determining whether a test case has passed or failed [30]. Related to the oracle problem, several authors proposed using metamorphic testing of AI-based PER systems [31], [32], i.e., executing transformed test cases while expecting the same output. Such transformations are suitable to test in simulated environments, e.g., applying filters on camera input or modifying images using GAN.

A. Virtual Environments and Simulators

Virtual or synthetic data are used to simulate environments, vehicle dynamics, and scenarios which are difficult, expensive, or risky to execute in the real world. Simulation engines are used to generate these simulated conditions, including virtual sensors, which are attached to the virtual vehicles participating in the simulation.

Simulators can also be used to support system testing as part of V&V of functional and safety requirements. An ideal simulator to test perception, planning, and decision-making components of an autonomous system must realistically simulate the environment, sensors, and their interaction with the environment through actuators. Simulated environments bring several benefits to V&V of the AI-based systems.

1) *Cost Efficient:* Using simulation for V&V of autonomous systems reduces the cost of using a real track, vehicles, and instruments that could risk damage during the testing process.
2) *Time:* Having an immediate response from a simulator shortens the software development cycle.
3) *Safety:* Currently, testing many vehicle collisions and accident scenarios is done using safe dedicated test and assessment protocols. Using simulators, the risks of test driving of an autonomous vehicle in urban areas will be substantially reduced.
4) *Edge Cases:* Many low-probability safety critical situations and hazards that would not be encountered on a test track can be generated in simulated environments.

These benefits and strengths make simulation-based V&V methods particularly useful when:

1) data collection or data annotation is difficult, costly, or time-consuming;
2) real-world testing is endangering human safety;
3) coverage of collected data is limited;
4) reproducible and scalability are important.

However, the difference between synthetic images from simulators and images in the real world can lead to different results in the validation of the system in the laboratory and in the field, and for this reason, the gap between simulation and reality is always a key point to be taken into account. Fortunately, the need for the recreation of real-life scenarios is not only necessary for AI contexts but also most of the demands from the entertainment industry, videogames, and animated movie designers are constantly developing virtual characters and scenarios. For that reason, 3-D modeling software for videogames such as unreal engine or blender focus on making the creation of complete virtual environments easier, quicker, more customizable, and realistic.

These are the most important application-specific toolsets and environments for particular purposes given as follows.

1) *Driving and Autonomous Driving Simulations:* The automotive sector has evolved rapidly to support the simulation of complex multisensor setups in vehicles, and a wide range of simulation tools exist: CARLA [33], LGSVL [34], or professional-level development suites such as Prescan [35], CarMaker [36], 4-D Virtualiz [37], and ESI Pro-SiVIC [38]. From these simulators, data can be gathered, manipulated, and processed as desired.
2) *Robotic Simulations:* ROS [39] and Gazebo [40].
3) *Drone-Like View:* AirSim [41].

Unfortunately, there is currently no virtual simulator for railway scenario generation in V&V CV&AI-based outdoor PER systems in the railway domain. The only tools that generate realistic railway virtual environments are videogames such as Train Simulator [19] that do not support full control over scenarios, vehicle dynamics, and/or sensor simulation. This work aims to create an in-between system that takes advantage of the scenario design parameterization possibilities of train driving videogames by exploiting their virtual cameras as sensors.

III. System Under Validation

The next step in ATO specification and standardization is to define functionalities up to GoA level 4 (driverless and unattended autonomous train). An essential feature of future autonomous train is to monitor the train environment, drive according to signaling rules, and react in the case of any detectable and severe anomaly, like a driver does today.

For this reason, a new set of on-board modules sensing the physical railway environment in place of the driver are required. This subsystem is called the PER module and it relies on sensors that must emulate or improve the driver's perception.

The system under evaluation in this work is a CV&AI-enhanced railway lateral signaling detector (see Fig. 2), specially needed in case there is not an automatic train protection (ATP) system underlying the train operation as follows.

1) *Signal Passed at Danger (SPAD) Control:* This is an event where a train passes a red stop signal without authority. Depending on the underlying ATP system, this control can rely on the on-board ATP module, but in other legacy or class B ATPs this responsibility can still rely on the driver. The PER system should be recognized and identifies in real-time safety critical signal aspects.
2) *Overspeed Control:* This is an event when the driver or automatic driving system exceeds the speed limit in train operation. Once again, depending on the underlying ATP system, this control can still rely on the driver. In some signaling systems, the speed limitation can be indicated with dynamic-light number panels, with static side-sign fixed panels or signals in different colors and with flashing/blinking events. The PER system should be recognized and identifies in real-time the speed restrictions that may change dynamically during the train's journey.
3) *Accurate Localization:* Nowadays, trains' automatic operations over legacy or class B ATPs determine the train position using global navigation satellite system (GNSS) and inertial sensors. The error, when using just satellite data, can remain around 5–15 m, something common with this technology. For this reason, alternative solutions are required to increase position precision. The PER system can reduce this localization error at specific low-speed scenarios such as stopping point or red signal approaching phases. Landmarks can be used to estimate visual distance measurement, which calculates the remaining distance to detected objects.

A. Evaluation Criteria of the System

Regarding the validation metrics to be used in the V&V system, the PER system will be evaluated taking into account the following cases.

1) *Safety Violations on Signal (Light) Detection:* A railway light signal can have a different meaning depending on the color of the light. In addition, its state can change between different journeys. Not detecting a more restrictive light signal can result in a safety violation.
2) *Safety Violations in Speed Restriction Sign Detection:* Speed restriction signs can be fixed signs or dynamic light signs. Dynamic sign aspect can change between different journeys. Not detecting a more restrictive speed signal will lead to a risk situation.
3) *General Operational Limitations:* Comparison of the validation results obtained using test datasets having different visibility conditions for the same journey will provide useful information that will help determine the conditions under which the system does not meet the safety requirements.

Distance estimator system is beyond the scope of this V&V system due to the impossibility of collecting/capturing the distance measures in real scenarios or creating synthetic measurable data as ground truth against which to validate the results.

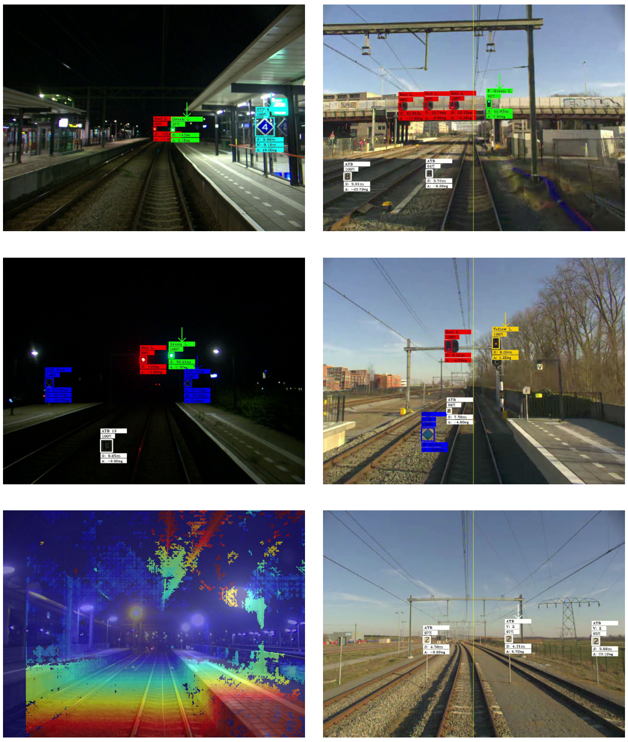


Fig. 2. Examples of railway signal, side-signs, and landmark detection over real images and distance estimation based on stereo depth map.

IV. V&V Solution Overview

The current V&V system is based on videos recorded on track. Each video is labeled manually, one by one without being able to take advantage of any template, as it is impossible to reproduce the same journey. In addition, the capacity to control the atmospheric weather or create real signal occlusions is not going to be guaranteed. The aim of this semi-automatic V&V framework is to improve the current manual process providing a new method and new tools that automate some of the activities that are involved. There are two main activities that the proposed method will automate to get this goal as follows.

1) *Data Generation for Validation Scenario:* To speed up this process, image datasets for system validation are obtained using a simulator, where scenarios are designed and run under different light, weather, and occlusion conditions. This enables to test the behavior of the system under different visual conditions, increasing the test coverage for the system under validation.
2) *Semi-Automatic Ground-Truth Labeling:* The possibility of generating synthetic data using a simulator allows the generation of videos and scenes fully synchronized in time and space. Even if the visibility conditions of the

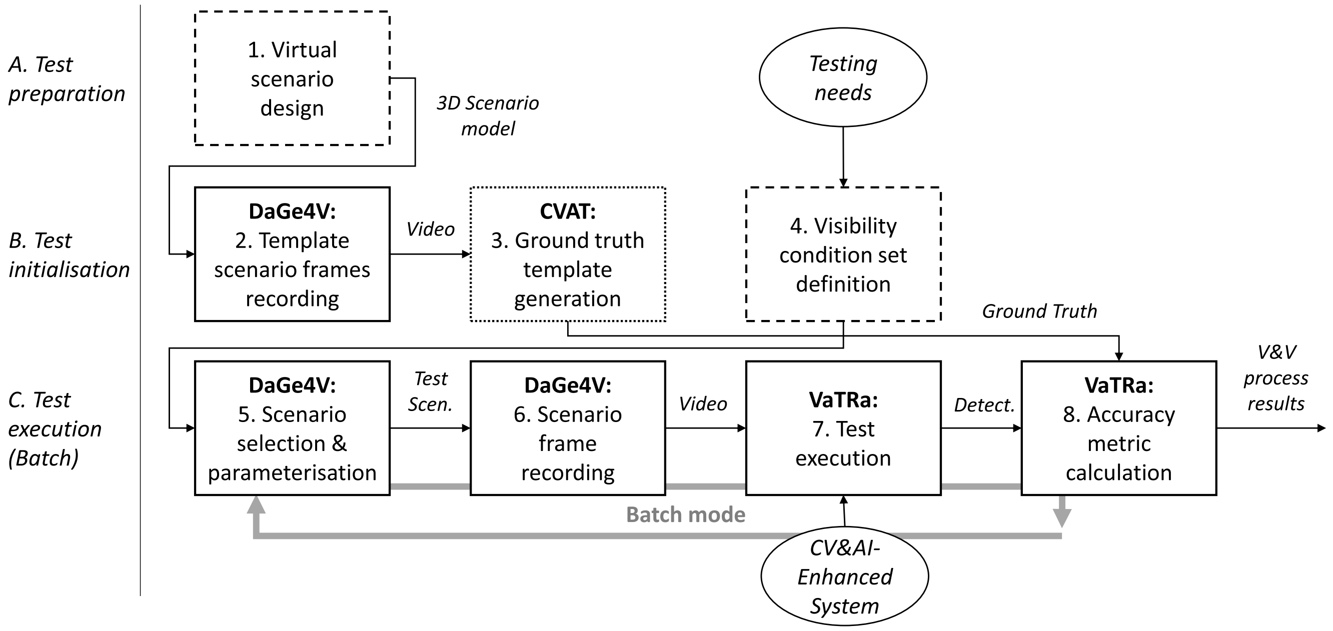

A. Test preparation
B. Test initialisation
C. Test execution (Batch)
1. Virtual scenario design
3D Scenario model
Testing needs
DaGe4V: 2. Template scenario frames recording
Video
CVAT: 3. Ground truth template generation
4. Visibility condition set definition
Ground Truth
DaGe4V: 5. Scenario selection & parameterisation
Test Scen.
DaGe4V: 6. Scenario frame recording
Video
VaTRa: 7. Test execution
Detect.
VaTRa: 8. Accuracy metric calculation
V&V process results
Batch mode
CV&AI-Enhanced System

Fig. 3. Workflow of the V&V method and framework. *Continuous line, automatic; round dots, semi-automatic; dashes, manual.*

objects to be detected vary, their position and presence in time and position into image can be synchronized. This makes it possible to reuse the same labeling template for tens and hundreds of videos and scenes, reducing the manual workload when creating the necessary ground truth.

The approach presented in this work (see Fig. 3) is a framework which will test the system over same railway journey but under different visibility conditions. This method not only comprises the generation of image datasets for test cases but also the analysis of the results obtained in the performed tests. The results obtained from the execution of the tests are analyzed to obtain accuracy metrics, to identify potential safety violations, and to evaluate the behavior of the system in the different operation conditions.

A. Solution Specification

The design of the solution proposed in this work has been based on the following basic specifications.

1) Tests shall be executed under different environment conditions that provoke diverse and enough variety of visibility conditions for object detection V&V.
2) An analysis of the results obtained during the tests carried out for a set of validation tests shall provide information to determine the conditions for a safe operation of the system.
3) The proposed method shall reduce CV&AI-enhanced object detector V&V costs and required time.

To be able to handle these generic specifications, the designed framework shall:

1) provide tools to design test scenarios to create datasets for the validation tests in a semi-automatic way;
2) provide a simulation environment with which to create a set of time synchronized videos with the same train journey under different visibility conditions:
    a) several different daytime journeys;
    b) different meteorological conditions (clouds, rain, snow, fog, and so on);
    c) partial occlusions over detectable objects (tree branches, and so on).
3) provide tools to label the ground-truth template in a semi-automatic way;
4) offer inference capabilities for the CV&AI-enhanced algorithm over the created synthetic dataset;
5) record information about all the detected objects by the system during the execution of a validation test for further analysis;
6) record the evidence of the execution of validation tests for light signals, side-signs, and landmark detection;
7) provide tools to calculate accuracy metrics for each test execution, comparing the test execution results and the expected results, defined by the ground truth for the journey;
8) provide tools for batch operation and data storage.

B. Scenario Definition

Before starting the implementation of the required tools, the scenarios and visibility conditions under which the object detector should be tested has to be defined. This definition will configure the virtual simulator for a specific train journey scenario generation.

1) *Daytime Periods:* During different daytime periods, object coloring and the amount of light reflected in them will vary. In extreme or edge situation, cameras may also blink as they are facing the sun directly (see Figs. 4 and 5). The scenarios should incorporate at least train journeys as follows.
    a) *Day:* This is a nominal scenario with ideal lighting conditions for object detection and identification.

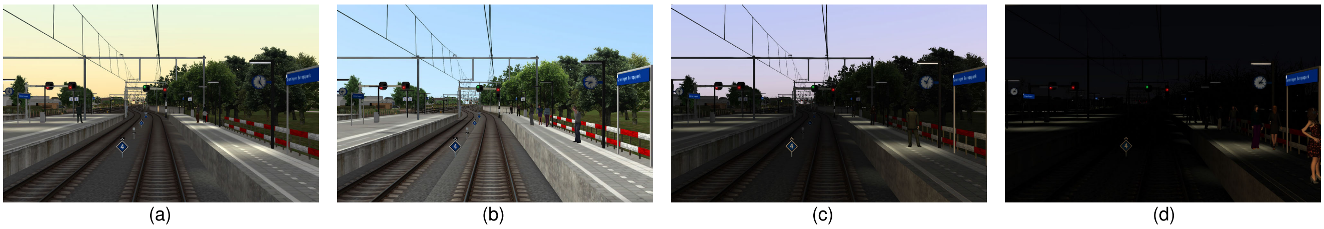


Fig. 4. DaGe4V tool: virtual scenario at summer with different day-light conditions. (a) Dawn (05:03). (b) Day (08:45). (c) Sunset (22:05). (d) Night (03:10).

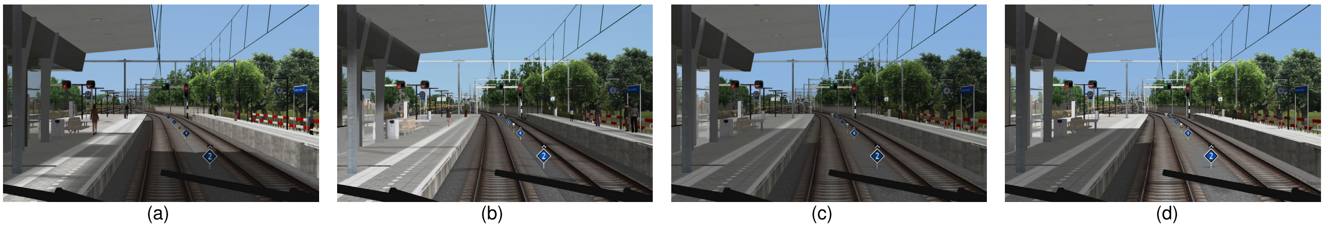


Fig. 5. DaGe4V tool: virtual scenario at summer day. Different shadow and sun intensity combinations. (a) Shadow at 10:00. (b) Shadow at 18:00. (c) Normal sun intensity. (d) High sun intensity.

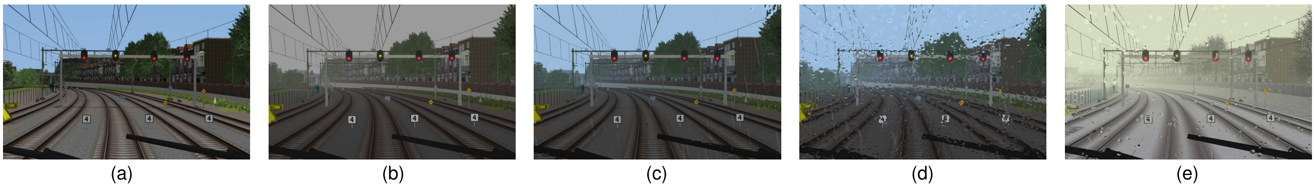


Fig. 6. DaGe4V tool: virtual scenario at summer (snowy at winter) 11:00 A.M. with different meteorological conditions. (a) Sunny. (b) Foggy. (c) Rainy with wiper. (d) Rainy without wiper. (e) Snowy.

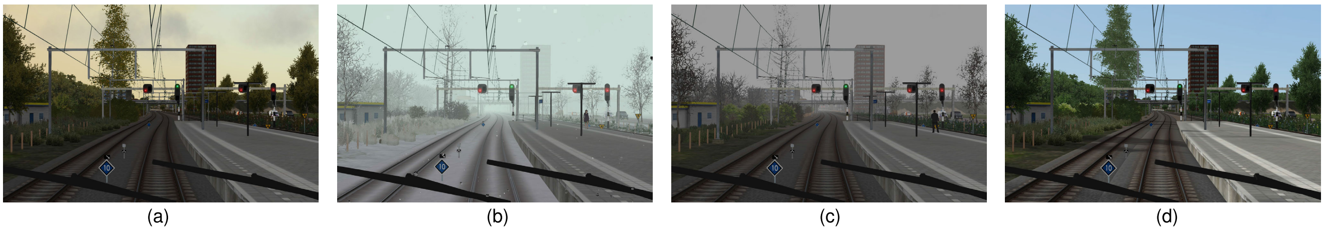


Fig. 7. DaGe4V tool: virtual scenario with different season–meteo–daylight combination conditions. (a) Autumn–cloudy. (b) Winter–cloudy/snowy. (c) Spring–foggy. (d) Summer–sunny.

   b) *Dawn and Sunset:* The sun's low altitude presents strong coloring, large shadows, and blinking possibilities that make detection task more difficult.
   c) *Night:* The lighting conditions significantly reduce the operating distance. This scenario must be present in the V&V dataset.

2) *Meteorological Conditions:* Adverse weather conditions reduce visibility and have a negative effect on both the detection and identification accuracy and the operational performance of the CV&AI-based object detectors (see Fig. 6). The scenarios should incorporate at least train journeys in the following.
   a) *Sunny Day:* Although this is a nominal scenario with ideal conditions for object detection, sun position situates some object in shadows at areas in some point during the day reducing the possibilities for correct identification.
   b) *Covered (or Partially Covered) Sky:* This situation will change the coloring and lighting conditions of the objects making the variance possibilities greater over detectable objects. Partial cloud cover can also simulate partially shadowed objects which makes their identification more difficult.
   c) *Snow and Rain:* Apart form lighting reduction and coloring change situation, the wiper working on the front window to remove snowflakes and rain drops can create partial occlusions in detectable objects.
   d) *Fog:* Dense fog can significantly reduce visibility, and hence the capacity to detect the objects. In addition, it also will reduce the operating distance.

3) *Seasons:* The appearance of the environment is also dependent on the season (see Fig. 7).

4) *Partial Occlusions:* This situation can partially cover detectable objects making their physical appearance different from the AI models that have been trained (see Fig. 8). They may happen due to the following.
   a) *Vegetation:* Tree branches or bushes may be partially occluding the side-signs, signals, and

landmarks to be detected, reducing the detection capability of the system under validation.

b) *Physical Deterioration:* Detectable objects may be damaged changing their physical appearance by the passage of time or by inclement weather.

c) *Vandalism:* They can also be partially broken, or painted due to vandalism actions.

C. Evaluation Metric Definition

While data preparation and training an AI model is a key step in the CV&AI-enhanced system pipeline, it is equally important to measure the performance and accuracy of this trained model. If AI models do not achieve the expected performance in certain situations, the overall system that uses the outcomes can face safety problems. To make sure the model learns, it is significant to use multiple evaluation metrics to evaluate the model. In this specific case, looking only at a generic hit rate may mask poor performance in critical situations. Mistaking a red signal for a green one is not the same as mistaking a green one instead of a red one. The first situation has operational consequences and the second can lead to an accident with a train crash. The use of evaluation metrics is critical in ensuring that the system is operating correctly, optimally and safely.

These metric definitions will configure the VaTRA of the proposed V&V framework as follows.

1) *Safety Requirement Violations Ratio (SRVR):* Incorrect object identification can cause a safety risk, e.g., not identifying a red signal can cause an accident. This measure takes into account those missed detections that cause safety-related risk situations

$$\text{SRVR}(\%) = \frac{N^{\circ}\ \text{safety\_req violations}}{N^{\circ}\ \text{safety\_req}} * 100. \quad (1)$$

To determine the safety requirement violation, the classification matrix is defined (see Table I). This is a standard tool to evaluate object detection and classification algorithms. Any given cell in row (i) and column (j) represents the number of detections of actual class (i) detected and classified as detection of class (j). Therefore, the goal is to maximize the percentages in the main diagonal $(C(i, j),$ where $i = j)$ (correct behavior). The last cell of each row shows the false negatives (FNs) (i.e., nothing was detected where an actual signal stands), while in the last row the false positives (FPs) can be seen (i.e., a signal was detected where there was not any).

Then, classification of confusions are clustered into four different risk/operational levels as shown in Table II: 1) correct detections and errors that do not have any effect (green); 2) incorrect detections that might affect the operation behavior (orange); 3) incorrect detections that imply safety risks (red); and 4) impossible or unacceptable miss-detection due to nonprobable match (gray), such as confusion between signal and lateral panel.

To determine whether the detected object corresponds with enough accuracy to the object presented in the ground-truth template, the intersection over union (IoU) criteria are used. Using this metric, the overlapping area between the predicted bounding box and the ground-truth bounding box divided by the area of union between them can be measured. According to the prefixed threshold, this overlapping can be considered enough to consider the detection correlated with the same in the ground-truth template.

TABLE I

AI Model Classification Matrix. *Nothing Detected (ND) and Nothing to Detect (NTD)*

	Class 1	Class 2	...	Class n	NDT
Class 1	C(1,1)	C(1,2)	...	C(1,n)	C(1,N)
Class 2	C(2,1)	C(2,2)	...	C(2,n)	C(2,N)
...	...	...	...	...	...
Class n	C(n,1)	C(n,2)	...	C(n,n)	C(n,N)
ND	C(N,1)	C(N,2)	...	C(N,n)	C(N,N)

TABLE II

AI Model Classification Matrix by Safety and Operational Violations. *More Restrictive Signal (MRS), Less Restrictive Signal (LRS), No Signal (NS), More Restrictive Speed Sign (MRS), Less Restrictive Speed Sign (LRS), and No Speed Sign (NS)*

	MRS	LRS	NS	MRSs	LRSs	NSs
MRS	🟩	🟧	🟧	⬜	⬜	⬜
LRS	🟥	🟩	🟥	⬜	⬜	⬜
NS	🟥	🟧	🟩	⬜	⬜	⬜
MRSs	⬜	⬜	⬜	🟩	🟧	🟧
LRSs	⬜	⬜	⬜	🟥	🟩	🟥
NSs	⬜	⬜	⬜	🟥	🟧	🟩

🟩 No risk

🟥 Safety impact risk

🟧 Operation impact risk

⬜ Unlikely confusion

2) *Operational Requirement Violations Ratio (ORVR):* Incorrect object identification can also lead to behavior that increases energy consumption or reduces track capacity, due to unnecessary braking/stop identifying a red signal when it is green or due to inefficient time management coming from incorrect speed reductions. This measure takes into account those miss-detections that cause operational-related undesirable situations

$$\text{ORVR}(\%) = \frac{N^{\circ}\ \text{op\_req violations}}{N^{\circ}\ \text{op\_req}} * 100. \quad (2)$$

To determine safety requirement violation, the same classification matrix for SRVR metric (Table I) is used.

3) *Mean Average Precision (mAP) for Object Detection:* Average precision (AP) [42] is a popular metric in measuring the accuracy of object detectors. The mAP is calculated by finding the AP for each class and then averaging over a number of classes. It incorporates the tradeoff between precision and recall and considers both FP and FN. This property makes mAP a suitable metric

for most detection applications

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i. \tag{3}$$

This metric will be used to measure the average detection accuracy of all the classes present in the object detector.

D. Framework Tools

The framework that has been created to develop the V&V method integrates different tools. Some of them include third-party applications that have been included in the system, but others have been developed from scratch.

For example, the tool for video frame annotation is based on the well-known CV annotation tool (CVAT) [43] applications which allow a simple and intuitive labeling. This tool is semi-automatic as it uses the same AI model as the system under validation for automatic prelabeling. Afterward, two independent users correct the detections by hand to create a cross-validated ground truth template for a set of videos containing the same train journey.

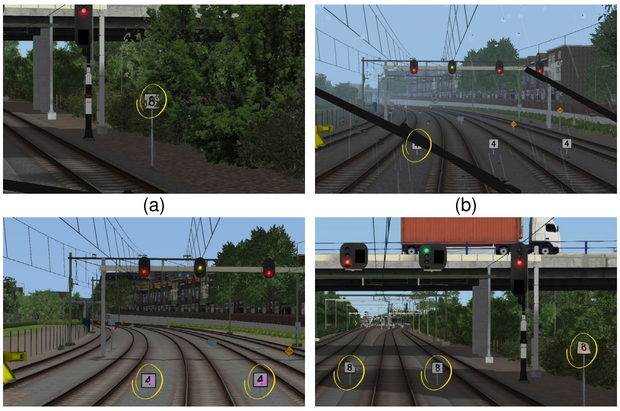
However, the main two tools, generated specifically for this framework, are: 1) data generation for validation tool (DaGe4V) and 2) validation test result analyzer (VaTRA).

1) Dataset Generator for Validation: DaGe4V incorporates the Train Simulator videogame as a virtual environment for synthetic data generation, and it aims to automate the process of creating a custom virtual scenario set according to test definition and recording frameset/videos from it. As an output, it generates the datasets with frameset/videos of the same train journey but recorded under different visibility conditions.

The only input it needs is the route of the journey. Once this route is loaded into the train simulator scenario, DaGe4V tool enables the selection of the visibility conditions under which the video dataset must be generated. After this, it activates the frame capturing subsystem and it executes the journey in the virtual environment with different visibility conditions selected. After journey execution, frameset/videos for validation are generated.

Train Simulator: Train Simulator is the simulation environment used by DaGe4V for designing routes. It is a Train Simulation game developed by Dovetail Games [19].

A complete suite of tools is available to customize content. This allows terrain modeling either by hand using the tools provided (see Fig. 9) or by importing digital elevation model (DEM) data from NASA. Track construction based on a system of straights and arcs allows an infinite amount of possible junction configurations and scenery placement. A scenario editor allows the creation of tasks such as picking up passengers, hauling cargo, and shunting wagons around yards. These tools also allow players to build unlimited sizes of layouts, create their own scenery and rolling stock, and modify the provided content by adding features or re-skins.



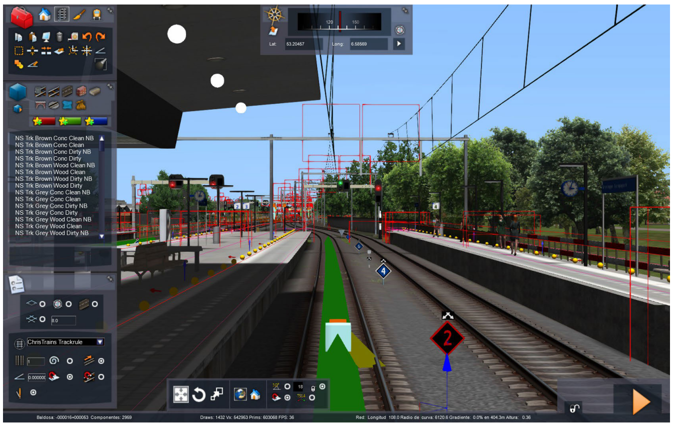
(a)

(b)

(c)

(d)

Fig. 8. DaGe4V tool: virtual scenario with different partial occlusion conditions. (a) By tree fork. (b) By windscreen wiper. (c) Due to vandalism act. (d) Due to deterioration.



NS Trk Brown Conc Clean NB
NS Trk Brown Conc Clean
NS Trk Brown Conc Dirty NB
NS Trk Brown Conc Dirty
NS Trk Brown Wood Clean NB
NS Trk Brown Wood Clean
NS Trk Brown Wood Dirty NB
NS Trk Brown Wood Dirty
NS Trk Grey Conc Clean NB
NS Trk Grey Conc Clean
NS Trk Grey Conc Dirty NB
NS Trk Grey Conc Dirty
NS Trk Grey Wood Clean NB
NS Trk Grey Wood Clean
NS Trk Grey Wood Dirty NB

ChrisTrains Trackrule

Baldosa: -000016+000053 Componentes: 2959 Draws: 1432 Vx: 542953 Prims: 603068 FPS: 36 Red: Longitud 108.0 Radio de curva: 6120.6 Gradiente: 0.0% en 404.3m Altura: 0.36

Fig. 9. Inserting landmark panel with the Train Simulator editor.

However, the most interesting point to integrate this solution as a toolchain in DaGe4V application is its application programming interface (API) with documentation that allows a third party to program custom scenarios by changing the available parameters. This detailed reference [44] follows on from the creator manual supplied with the Train Simulator and gives detailed information on how to create your own routes, scenarios, rolling stock, in-game assets, and signaling systems—in short, everything you need to create your own virtual railway scenario.

2) Validation Test Result Analyzer: VaTRA has two main functionalities: it executes the batch scenario test (see Figs. 10–13) and it compares the results of V&V with the expected results according to the ground-truth template. As an output, it provides evaluation metrics of the CV&AI-enhanced system behavior, automatically detects potential safety violations, and compares the results of the tested system for the same route under different visibility conditions.

Image datasets recorded by DaGe4V and the ground-truth template generated using CVAT are used as inputs for testing the CV&AI-enhanced system for railway signaling detection. VaTRA analyses the results based on predefined evaluation criteria defined in Section V.

E. Solution Workflow

The semi-automatic V&V method system is executed in a single TestBench (see Fig. 3). A task loads the specific scenarios definitions with their parameters, and it creates

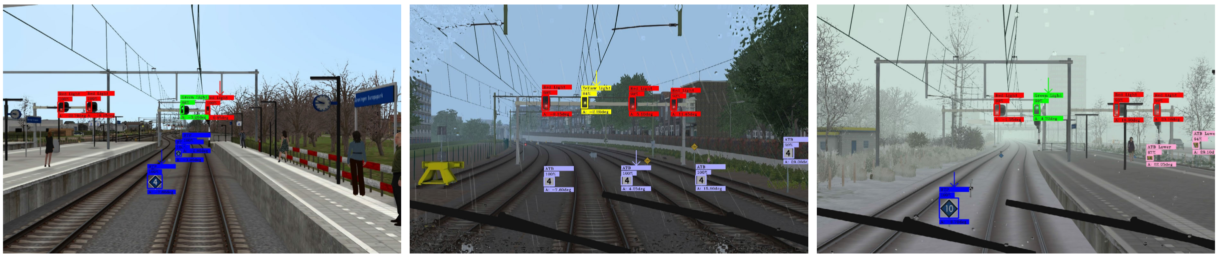


Fig. 10. VaTRA tool: CV&AI-enhanced system correct detections.

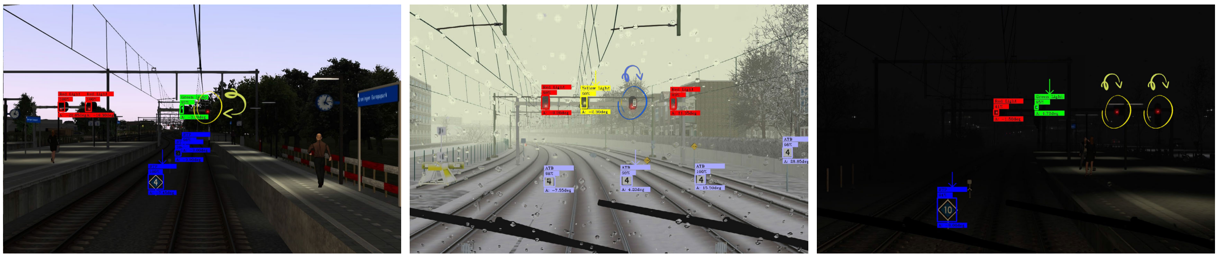
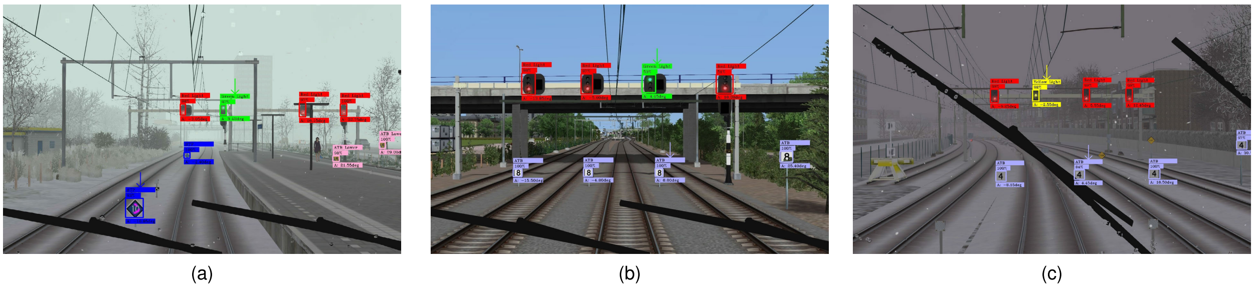


Fig. 11. VaTRA tool: CV&AI-enhanced system misdetections.

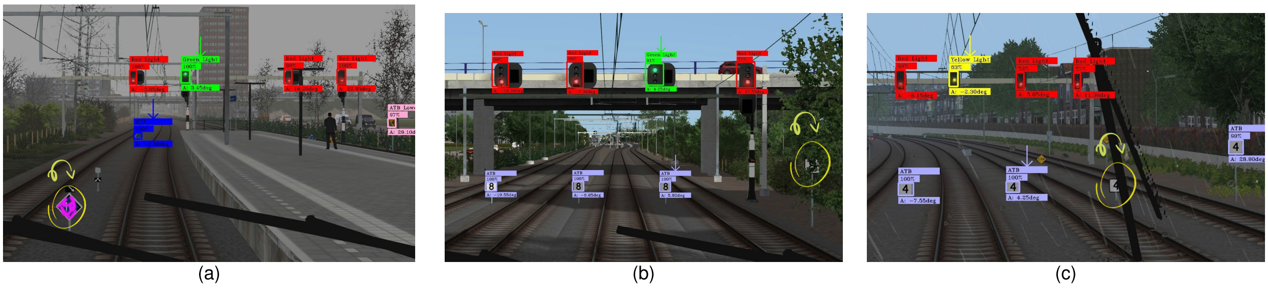


(a)

(b)

(c)

Fig. 12. VaTRA tool: CV&AI-enhanced system correct detections even partial occlusion. (a) Due to vandalism act. (b) Due to tree fork. (c) Due to wiper.



(a)

(b)

(c)

Fig. 13. VaTRA tool: CV&AI-enhanced system misdetections due to occlusions. (a) Due to vandalism act. (b) Due to tree fork. (c) Due to wiper.

a corresponding dataset for system testing. After this, the CV&AI-enhanced system is passed over these videos, obtaining its results. Finally, the ground-truth information is loaded and compared with the CV&AI-enhanced system results. This generates two analysis results, one related to classification and precision, and the other related to safety violation analysis. The combination of those results is the V&V process results.

The activities clustered in different artifacts that compose the workflow of the V&V method are described as follows.

1) *Test Session Preparation (Once):*
   a) *Virtual Scenario Design (Manual):* The first activity comprises the design of the set of scenarios that will be created by the simulator. At this point, by means of a simulation environment, the objects that the CV&AI-enhanced system should detect are placed in different real-world scenarios and the different visibility conditions are defined. The editor of the train simulator videogame is used for this task.
2) *Test Initialization (Once):*
   a) *Template Scenario Frames Recording (Automatic):* During a simulation execution, the DaGe4V tool records the frames and related metadata from the simulator. As a result, the template scenario video is generated for labeling.
   b) *Ground-Truth Template Generation (Semi-Automatic):* This activity comprises the creation of semi-automatically labeled ground-truth template

for dataset validation. This tool uses external labeling tools powered by the CV&AI-enhanced object detector as a helping tool for first candidate detection in an automatic way.

c) *Visibility Condition Set Definition (Manual):* A set of different scenarios are parameterized according a chosen specific visibility conditions for test.

3) *Test Execution (Batch Operation):*

a) *Visibility Condition-Based Scenario Selection and Parameterization (Automatic):* This activity will comprise the selection of a previously designed set of visibility conditions based on a set of scenarios that will be executed in the simulator. Setting the configuration parameters will enable simulation in different conditions to be carried out. It is set by DaGe4V tool.

b) *Scenario Frame Recording (Automatic):* During simulation execution, DaGe4V tool records the frames and related metadata from the simulator for all the tests. As a result, the validation datasets for testing the system under test are generated.

c) *CV&AI-Enhanced System Test Execution (Automatic):* After the simulated videos are available, VaTRA executes test to extract CV&AI-enhanced system's detections on railway signals and landmarks. All the outputs are stored and exported as detection files.

d) *Accuracy Metric Calculation (Automatic):* In this activity, the VaTRA tool will analyze the results obtained in several tests carried out for the CV&AI-enhanced system. It will compare the results obtained on each test with the ground-truth template. Finally, it provides several metrics and identifying potential safety violations due to incorrect object detection during the tests, generating the V&V results' artifact.

V. Evaluation Criteria for V&V Framework

This section defines the parameters and criteria that will be used for the evaluation of the V&V framework. In this case, the system under evaluation will be the semi-automatic V&V system itself and not the CV&AI-enhanced system that is validated and verified with the framework.

The objectives in designing the V&V framework were to extend the diversity and number of test types while reducing costs, making them affordable for the process. Therefore, the following V&V quality criteria have been selected to evaluate the framework.

1) *Number of Test Cases (NTCs):* This criterion aims to evaluate the capacity of detection of misidentified objects and the potential safety and operational violations. It also aims to evaluate the coverage obtained by the validation tests. NTC is calculated by counting the number of tests created and performed using the V&V method.

2) *Variety of Test Cases (VTCs):* This metric measures the quantity of visibility conditions' variety that can be used to V&V the system under test. Visibility conditions are not discrete states, but analog since from a sunny day to fully cloudy, there is a continuous gradual transition. However, to facilitate the counting of different types of tests, predefined and limited scenarios will be set up during dawn, day, sunset, and night, and each of these scenarios with sunny, partly cloudy, rainy, snowy, and foggy situations. In total, there are 24 different combinations. VTC is measured counting a number of different visibility conditions under which the system is tested using the V&V framework.

On the other hand, the following parameters have also been defined to evaluate the framework in terms of effort and costs.

1) *Effort Needed for Test (ENT):* These evaluation criteria are used to measure the effort in terms of person-hours required to perform a test on a system. This measure is especially useful to compare the effort spent doing manual work versus semi-automated work. This encompasses the entire process for doing a test, from specification and data preparation to execution of test cases and the extraction of the results. The following parameters should be taken into account.

a) *Staff Hours During Data Creation/Gathering (Technicians and Drivers)*: Total working hours, used to create synthetic data coming from scenario simulator and used to gather images during real recording runs.

b) *Time Dedicated to Manual Labeling of Recorded Data*: For manual nonautomated method, each run will be different, and therefore, all runs have to be labeled one by one independently. For the semi-automatic process, only one run template is needed to label since all the remaining tests will have synchronized runs.

It is assumed that the effort and cost of setting up the required infrastructure for the real data gathering system and the virtual environment can be matched. The inference times for testing execution and results' comparison to validate the CV&AI-enhanced system can also be matched. For this reason, the effort needed for the whole V&V process can be measured by considering the amount of people involved for the work and by the number of hours (normalized to full-time employees) needed to complete it.

2) *Operational Costs (OCs):* This indicator adds up all the costs of operating the V&V process, excluding other costs related to framework development and maintenance, required equipment, and so on It is not easy to quantify and standardize the cost of a V&V process based on real-world data captured in the operational scenario, but as a minimum the following parameters should be taken into account.

a) *Travel Cost:* Technicians from the supplier and drivers from the operator must be translated to the operation scenario. It usually involves international travel for supplier technicians and domestic travel for drivers.

b) *Track Availability:* The use of tracks by reserving a time slot is an important cost of train operation and also of testing.

c) *Personnel Costs (Technicians and Drivers):* During the data gathering process and manually labeling the captured data.

OCs of the manual nonautomated V&V process are calculated as the sum of operational and personnel costs for one track test session. For the semi-automatic V&V process, the OCs are practically zero as they are carried out on an online server.

To evaluate the improvement introduced by the semi-automatic V&V process, it will be compared against a manual nonautomated V&V process.

VI. Test Description

Currently, there is no other virtual validation tool for PER systems in the railway sector against which the results obtained can be compared. Nor is there any work that has quantified the cost of on-track validation of any type of autonomous railway systems or subsystems.

On the other hand, although the automotive sector is a few steps ahead, its most recent works are using field tests to validate the whole autonomous driving system [25], [26]. These tests do not focus on comprehensive validation with future certification purposes for the PER module in particular. Therefore, the test presented in this work will compare the semi-automatic system presented in this work with the current manual system used to validate the PER module in the railway company from which this idea was born.

However, it is not an easy task to make an equitable direct comparison between the current manual nonautomated V&V method and the new proposed semi-automatic method. The difficulty lies in quantifying the metrics such as visibility conditions' variability and the total cost of a validation session in a real environment given its variability, especially with regard to the data capture process.

To carry out these tests and obtain truly comparable data, a current manual nonautomatic validation test session followed until now in the company has been established as a reference. It consists of a group of two workers traveling to the real operating environment at the European Union for three working days with 8-h working hours each. They record a total of 15 h of valid video per test session in the conditions they are exposed to. Normally, daylight and nighttime conditions are covered, but weather variability is not guaranteed. Two drivers, working 4 h per day, are also needed.

Taking into account this scenario, the reference values for current nonautomated validation session in real tracks have been calculated as follows.

1) *For NTE*: 30 test cases (5 h for valid journeys; 30 min per journey; ten test runs per validation day; three validation days).

2) *For VTE*: Eight combinations of different visibility conditions (dawn, day, sunset, and night combined with annual probability of snow, rain, cloud, or sun in the place).

TABLE III

V&V Method Evaluation Results. *Manual V&V Method (M), Semi-Automatic V&V Method (SA), and Obtained Improvement (%)*

	NTC (Nº)			VTC (Nº)			ENT (h)		
	M	SA	(xTime)	M	SA	(xTime)	M	SA	(xTime)
NTC (=30)				8	30	3.8	252	30	8,4
VTC (=8)	45	8	5,6				450	12	37,5
ENT (=252)	30	252	8.4	8	252	31,5			

3) *For ENT*: 252 h in total. 102 h of staff working on tracks (two staff member 8-h working day; two drivers 4-h working day; three validation days; 12 h travel time (return trip) per staff member; 3 h travel time for drivers) and 150 h labeling the 30 test cases one by one (5 h per run).

4) *For OC*: 5000 € operation cost (no overtime assumed, travel cost per person including accommodation and subsistence allowance, and cost of the operator related to track utilization).

Once the current manual and nonautomated process has been defined, three comparable scenarios have been established. In each of these scenarios, one evaluation metric/criterion has been matched for both the processes, and the benefit of the new proposed method is evaluated against the other two metrics. For example, in the first comparison, the NTC obtained by both the systems has been equated and the other metrics (VTC and ENT) are set following this assumption. As a result, the comparison of both the systems is shown when the objective is to get a specific NTC.

Since the OCs are considerably negligible for the semi-automated system compared with the nonautomated manual system, the improvement with respect to this indicator remains constant during different experiments.

VII. Results

Table III shows the numerical results of the tests performed.

The first row of Table III shows the results taking into account that the same NTC has been assumed for both the validation methods. The value is a number of use cases that can be generated in a standard validation test with manual nonautomated method. As can be seen, the ENT is higher (×8, four times) for the manual method and, although much more time is invested, the variety in the recorded tests is not guaranteed (×3, eight times lower). This is due to the frameworks' batch operation capability.

In the second test set, the same VTC has been considered, trying to make both the methods have the same variety of test cases. To ensure a representative comparison, a minimum of 12 scenario variants are estimated, which requires more than one field test session at different seasons of the year. The ENT metric skyrockets (×37, five times higher) for the manual method. In addition, far fewer tests are required to cover the required variability (×5, six times fewer) for the semi-automatic method. This is due to the ability to parameterize the elements that create the virtual scene of the simulator.

Finally, in the third case, the same effort cost is assumed. It is equal to the effort cost of a track validation session. The NTC (×8, four times) and VTC (×31, five times) decrease drastically in the manual system. This is due to the high time cost of replicating a scenario with desired lighting conditions and running it in a real environment where weather factors are not controllable.

For all the tests, a significant reduction in OC is estimated (×20 times for tests 1 and 3 and ×30 times for test 2, which requires more than one session in tracks). Moreover, although it cannot be quantified, the reduction of personal and material risk guaranteeing safety in virtual environment testing is evident.

These results reflect all the advantages already attributed to V&V systems in virtual environments. The possibilities of virtualization and parameterization of scenarios, without the need for real operational scenarios, makes possible to reduce the effort, costs, and time needed to operate an equal number of tests. In addition, for a much lower cost and time, the variability in visibility conditions, which is essential for PER module accurate validation, is guaranteed and is enormous, as there is absolute control over the content of the rendered scenario.

The results also show that such systems are valid and indispensable tools to accelerate the inclusion and commercialization of AI technologies for perception in the railway sector in the short term.

VIII. Conclusion

This work presented a semi-automatic V&V framework for CV&AI-enhanced railway signaling and landmark detector. The design and implementation of the framework have been detailed, as well as the CV&AI-based system that has been evaluated with it. Moreover, the metrics have been identified and defined to compare the quantity and variety of tests performed, and their associated costs of both the current validation process and the one proposed in this work.

In addition, several tests have been carried out to evaluate a V&V framework. To compare them, different tests have been developed where the same time/effort invested or the number/variety of different scenarios that have been used in both the methods have been equalized.

The results obtained show a significant increase in the amount of tests that can be performed in the same time for a more robust validation of the system. In addition, even if the quantitative improvement is enormous, the most important change comes in qualitative matters. It is clear that the variety in test scenarios increases considerably as the simulated data generation environment allows to simulate a wide variety of visibility conditions that could only be achieved in real recordings after much time and cost. Finally, resource savings are another key point to take into account with a significant effort reduction for the required same workload. The clear advantage of the current track validation system is the realism of the test. The gap between the simulated scenario and the real journey can still be very large, so this is something to be taken into account.

As described in the article, this is an intermediate work that provides a solution to a current problem but it is still a semi-automatic and therefore temporary solution. As a future work, it is intended to transfer the railway environment to open-source tools such as CARLA. This will allow the whole process of data generation to be fully automated, manipulation of sensor data sources, and the total automation of the labeling process without having to follow the same train journey for each batch of tests. Furthermore, the option of providing the framework with the capability to perform ablation tests will be evaluated. It can be interesting to test AI-based functionalities performance by removing certain components to understand the contribution of the component to the overall system.

These V&V solutions will permit the simulation of all possible scenarios in virtual environments, especially those scenarios that are improvable but critical from a safety point of view such as people crossing railways and reduced visibility due to meteorological conditions. It can be predicted that future automatic V&V based on virtual environments will reduce costs of CV&AI-enhanced algorithms and it will support an easier marketing process of them for railway functionalities avoiding important initial barriers.

Acknowledgment

The novel approach presented in this work is being developed as a specific railway use-case for autonomous train operation into "VALU3S—Verification and Validation of Automated Systems' Safety and Security" European research project. The authors would also like to thank the whole PER Development Team of CAF Signaling for their help and support.

References

[1] Y. Cai et al., "YOLOv4-5D: An effective and efficient object detector for autonomous driving," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[2] L. Wang, H. Qin, X. Zhou, X. Lu, and F. Zhang, "R-YOLO: A robust object detector in adverse weather," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[3] D. He, Z. Zou, Y. Chen, B. Liu, and J. Miao, "Rail transit obstacle detection based on improved CNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.

[4] L. Guan, L. Jia, Z. Xie, and C. Yin, "A lightweight framework for obstacle detection in the railway image based on fast region proposal and improved YOLO-tiny network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.

[5] Y. Dai, W. Liu, H. Wang, W. Xie, and K. Long, "YOLO-Former: Marrying YOLO and transformer for foreign object detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[6] A. Staino, A. Suwalka, P. Mitra, and B. Basu, "Real-time detection and recognition of railway traffic signals using deep learning," *J. Big Data Anal. Transp.*, vol. 4, no. 1, pp. 57–71, Apr. 2022, doi: 10.1007/s42421-022-00054-7.

[7] T. Ye, X. Zhang, Y. Zhang, and J. Liu, "Railway traffic object detection using differential feature fusion convolution neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1375–1387, Mar. 2021.

[8] M. Hadded, A. Mahtani, S. Ambellouis, J. Boonaert, and H. Wannous, "Application of rail segmentation in the monitoring of autonomous train's frontal environment," in *Proc. Int. Conf. Pattern Recognit. Artif. Intell.*, 2022, pp. 185–197.

[9] Y. Zhang, K. Li, G. Zhang, Z. Zhu, and P. Wang, "DFA-UNet: Efficient railroad image segmentation," *Appl. Sci.*, vol. 13, no. 1, p. 662, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/1/662

[10] H. Yin, P. X. Liu, and M. Zheng, "Stereo visual odometry with automatic brightness adjustment and feature tracking prediction," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[11] M. Etxeberria-Garcia, M. Zamalloa, N. Arana-Arexolaleiba, and M. Labayen, "Visual odometry in challenging environments: An urban underground railway scenario case," *IEEE Access*, vol. 10, pp. 69200–69215, 2022.

[12] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

[13] H. Yang, Y. Wang, J. Hu, J. He, Z. Yao, and Q. Bi, "Deep learning and machine vision-based inspection of rail surface defects," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[14] Y. Zhang, M. Liu, Y. Chen, H. Zhang, and Y. Guo, "Real-time vision-based system of fault detection for freight trains," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 5274–5284, Jul. 2020.

[15] EU-Commission. (2023). *TAURO: Technologies for Autonomous Rail Operation*. [Online]. Available: https://cordis.europa.eu/project/id/101014984

[16] M. Borg et al., "Safely entering the deep: A review of verification and validation for machine learning and a challenge elicitation in the automotive industry," 2018, *arXiv:1812.05389*.

[17] N. Humbatova, G. Jahangirova, G. Bavota, V. Riccio, A. Stocco, and P. Tonella, "Taxonomy of real faults in deep learning systems," 2019, *arXiv:1910.11015*.

[18] High-level expert group on artificial intelligence. (2023). *Assessment List for Trustworthy AI, High-Level Expert Group on AI (AI HLEG)*. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[19] DLive. (2023). *Train Simulator*. [Online]. Available: https://live.dovetailgames.com/live/train-simulator

[20] K. Saito, K. Saenko, and M.-Y. Liu, "COCO-FUNIT: Few-shot unsupervised image translation with a content conditioned style encoder," 2020, *arXiv:2007.07431*.

[21] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*.

[22] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*.

[23] A. Ramesh et al., "Zero-shot text-to-image generation," 2021, *arXiv:2102.12092*.

[24] H. Gao et al., "Situational assessment for intelligent vehicles based on stochastic model and Gaussian distributions in typical traffic scenarios," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 3, pp. 1426–1436, Mar. 2022.

[25] D. Li and H. Gao, "A hardware platform framework for an intelligent vehicle based on a driving brain," *Engineering*, vol. 4, no. 4, pp. 464–470, Aug. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2095809917303648

[26] H. Gao, H. Yu, G. Xie, H. Ma, Y. Xu, and D. Li, "Hardware and software architecture of intelligent vehicles and road verification in typical traffic scenarios," *IET Intell. Transp. Syst.*, vol. 13, no. 6, pp. 960–966, Jun. 2019.

[27] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE Int. J. Transp. Saf.*, vol. 4, no. 1, pp. 15–24, Apr. 2016.

[28] R. Ben Abdessalem, S. Nejati, L. C. Briand, and T. Stifter, "Testing vision-based control systems using learnable evolutionary algorithms," in *Proc. IEEE/ACM 40th Int. Conf. Softw. Eng. (ICSE)*, May 2018, pp. 1016–1026.

[29] A. Gambi, M. Mueller, and G. Fraser, "Automatically testing self-driving cars with search-based procedural content generation," in *Proc. 28th ACM SIGSOFT Int. Symp. Softw. Test. Anal.*, New York, NY, USA, Jul. 2019, pp. 318–328, doi: 10.1145/3293882.3330566.

[30] A. Stocco, M. Weiss, M. Calzana, and P. Tonella, "Misbehaviour prediction for autonomous driving systems," in *Proc. IEEE/ACM 42nd Int. Conf. Softw. Eng. (ICSE)*, Oct. 2020, pp. 359–371.

[31] Y. Tian, K. Pei, S. Jana, and B. Ray, "DeepTest: Automated testing of deep-neural-network-driven autonomous cars," 2017, *arXiv:1708.08559*.

[32] M. Zhang, Y. Zhang, L. Zhang, C. Liu, and S. Khurshid, "DeepRoad: GAN-based metamorphic autonomous driving system testing," 2018, *arXiv:1802.02295*.

[33] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," 2017, *arXiv:1711.03938*.

[34] LEAR Lab. (2023). *SVL Simulator by LG—Autonomous and Robotics Real-Time Sensor Simulation, LiDAR, Camera Simulation for ROS1, ROS2, Autoware, Baidu Apollo. Perception, Planning, Localization, SIL and HIL Simulation, Open Source and Free*. [Online]. Available: https://www.svlsimulator.com/

[35] S Software. (2023). *Simcenter Prescan*. [Online]. Available: https://www.lm.automation.siemens.com/global/en/products/simcenter/prescan.html

[36] I Automotive. (2023). *CarMaker*. [Online]. Available: https://ipg-automotive.com/en/products-solutions/software/carmaker/

[37] D. Virtualiz. (2023). *4D Virtualiz*. [Online]. Available: https://www.4d-virtualiz.com/

[38] ESI. (2023). *ESI Pro-SiVIC*. [Online]. Available: https://myesi.esi-group.com/downloads/software-downloads/pro-sivic-2017.0

[39] Stanford Artificial Intelligence Laboratory. (2023). *Robotic Operating System*. [Online]. Available: https://www.ros.org

[40] OSR Foundation. (2023). *Gazebo*. [Online]. Available: https://gazebosim.org/home

[41] Microsoft. (2023). *Project AirSim for Aerial Autonomy*. [Online]. Available: https://www.microsoft.com/en-us/ai/autonomous-systems-project-airsim

[42] M. Zhu, "Recall, precision and average precision," Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, Canada, Tech. Rep., 2004.

[43] OpenVINO. (2023). *Computer Vision Annotation Tool*. [Online]. Available: https://www.intel.com/content/www/us/en/developer/articles/technical /computer-vision-annotation-tool-a-universal-approach-to-data-annotation.html

[44] DLive. (2023). *Train Simulator Developer Documentation*. [Online]. Available: https://sites.google.com/a/railsimdev.com/dtgts1sdk/

4.5 HPC Platform for Railway Safety-Critical Functionalities based on Artificial Intelligence

- **Autores: Mikel Labayen** and Laura Medina and Fernando Eizaguirre and Jose Flich and Naiara Aginako
- **Revista:** Applied Sciences
- **Volumen:** 13
- **Número:** 15
- **Año:** 2023
- **Editor:** MDPI ↗

applied sciences

MDPI

Article

HPC Platform for Railway Safety-Critical Functionalities Based on Artificial Intelligence

Mikel Labayen [1,2,*,†], Laura Medina [3,†], Fernando Eizaguirre [4,†], José Flich [3] and Naiara Aginako [2]

[1] Autonomous Vehicle Department, CAF Signalling, 20018 Donostia, Spain
[2] Computer Sciences and Artificial Intelligence Department, University of the Basque Country, 20018 Donostia, Spain; naiara.aginako@ehu.eus
[3] Computer Engineering Department, Universitat Politècnica de València, 46022 Valencia, Spain; laumecha@inf.upv.es (L.M.); jflich@disca.upv.es (J.F.)
[4] Embedded Systems Department, Ikerlan Technology Research Centre, 20500 Arrasate/Mondragón, Spain; feizaguirre@ikerlan.es
[*] Correspondence: mlabayen@cafsignalling.com
[†] These authors contributed equally to this work.

Abstract: The automation of railroad operations is a rapidly growing industry. In 2023, a new European standard for the automated Grade of Automation (GoA) 2 over European Train Control System (ETCS) driving is anticipated. Meanwhile, railway stakeholders are already planning their research initiatives for driverless and unattended autonomous driving systems. As a result, the industry is particularly active in research regarding perception technologies based on Computer Vision (CV) and Artificial Intelligence (AI), with outstanding results at the application level. However, executing high-performance and safety-critical applications on embedded systems and in real-time is a challenge. There are not many commercially available solutions, since High-Performance Computing (HPC) platforms are typically seen as being beyond the business of safety-critical systems. This work proposes a novel safety-critical and high-performance computing platform for CV- and AI-enhanced technology execution used for automatic accurate stopping and safe passenger transfer railway functionalities. The resulting computing platform is compatible with the majority of widely-used AI inference methodologies, AI model architectures, and AI model formats thanks to its design, which enables process separation, redundant execution, and HW acceleration in a transparent manner. The proposed technology increases the portability of railway applications into embedded systems, isolates crucial operations, and effectively and securely maintains system resources.

Keywords: autonomous and driverless train operation; computer vision and artificial intelligence; high-performance computing; safety-critical; AI hardware accelerator

check for updates

Citation: Labayen, M.; Medina, L.; Eizaguirre, F.; Flich, J.; Aginako, N. HPC Platform for Railway Safety-Critical Functionalities Based on Artificial Intelligence. *Appl. Sci.* **2023**, *13*, 9017. https://doi.org/10.3390/app13159017

Academic Editor: Keun Ho Ryu

Received: 21 June 2023
Revised: 1 August 2023
Accepted: 2 August 2023
Published: 7 August 2023

1. Introduction

Users of the European rail industry are clamouring for a future Automatic Train Operation (ATO) system since it provides advantages such as lower operating costs, longer product life-cycles for railways and increased safety. Its definition is being worked on by the European Shift2Rail standards group [1]. For fully autonomous train operation, various rolling stock suppliers and stakeholders have already begun researching, developing and testing technologies.

Similarly to other transport sectors, different computational issues are being faced by numerous railway suppliers and stakeholders for CV- and AI-enhanced autonomous train operation. The adoption of computer equipment capable of offering the performance of high-end graphic-processor units while being able to simultaneously meet safety criteria will be necessary for the future of CV and AI advances in the railway sector. These developments will increase the size, speed, and dependability of CV and AI processing calculations.

Through the use of multi-cores, Graphic Processors Units (GPUs), and specialized accelerators, a number of HPC commercial off-the-shelf platforms provide the calculation capabilities required by autonomous systems in fields such as intelligent transportation systems, space, and robotics [2].

However, because of the challenges or barriers that HPC platforms pose to the certification process, such as support for functional and timing isolation and testability, the use of these platforms has historically been viewed as being beyond the reach of the industry of safety-critical systems (i.e., controllability and observability). Therefore, the state-of-the-art (SoA) safety-critical computing platforms cannot currently satisfy these demanding specifications.

The SELENE (Self-monitored Dependable Platform for High-Performance Safety-Critical Systems) project [3] is a European R&D initiative that develops the research provided in this article as a use case demonstration. This work proposes a high-performance platform with safety-related considerations as a main design goal in an effort to bridge this gap. The SELENE platform is an open-source Reduced Instruction Set Computer V (RISC-V) [4] multi-core processor with hardware acceleration for artificial intelligence that supports multiple types of redundancy, real-time performance monitoring, and enforcement mechanisms to ensure that the safety objectives of the applications are satisfied. Additionally, the design of this system-on-open-source chip makes it possible for it to easily adapt to other safety domains. Apart from the use case presented in this article, three other use cases from the automotive and space industries have been used to test this method.

The article is organized as follows: Section 2 gives an overview of some relevant related works and highlights the main differences with our approach. Section 3 describes the railway domain use case in which the approach is being tested. Section 4 presents the use case deployment details analysing the new HW and SW modules included in the platform and the platform architecture, taking into account the safety-related analysis of our use case. Sections 5 and 6 define the test which was carried out and presents the obtained results in order to demonstrate the performance of the solution. Finally, Section 7 presents the conclusions and future work.

2. Related Work

The use of artificial sense (in real-time and via onboard embedded hardware) has been presented via a number of demonstrations in the railway industry.The Siemens autonomous tramway pilot case [5] was one of the first demonstrations that continues to inspire researchers today [6]. With this in mind, vision-based on-board obstacle detection and distance estimation in railroads have become the most pertinent scientific approaches [7]. To test and validate an autonomous obstacle detection system, various experiments and actual pilot cases have been implemented throughout the past few years, [8–13]. Other applications have also been worked on, such as vehicle localization on light trains [14] or railway lateral signalling detection on mainline trains [15,16].

The majority of these demos, similarly to those for self-driving cars, concentrate more on the computing capability of those systems rather than how these platforms may be certified. However, certification of sophisticated computing systems is an active area of study, such as those necessary for fully automated train systems [3]. Major chip suppliers, such as Nvidia and Intel, are also creating particular platforms that support that purpose [17], by including built-in fault-tolerant mechanisms such as lockstep execution or error correction codes in memory structures. Finally, to certify complex systems affordably, a more comprehensive and cutting-edge safety certification technique is required [18].

There are currently no commercial solutions that guarantee high-performance equipment with the safety requirements to be met in the railway sector. This work aims to be the first on-board HW prototype to execute AI functions safely (and in real time) in railway operations.

3. Use Case Definition

The use case presented in this paper, which is intended to validate the HPC SELENE platform, has been titled as automatic accurate stopping and safe passenger transfer, and it consists of automatic functionality collection based on CV- and AI-enhanced techniques. Figure 1 shows the graphical representation of it. The use case specifically highlights the following three features:

- Data collection and synchronization: this captures data from stereo vision-capable cameras in real time and it synchronizes and rectifies data of both video stream signals.
- Automatic station detection and accurate stop aligning the vehicle and platform: this detects the station platform and it accomplishes precise localization inside the platform area by detecting, recognizing and tracking visual patterns. The visual landmarks have been chosen to maximize the results of the detection and identification process in any possible lighting conditions. Visual stereo sensors that have been properly calibrated assess the physical distance.
- Safe passenger transfer: this captures data from rear cameras and it manages automatic safe door functionality preventing (a) door opening operation if the train and platform are not precisely aligned and (b) door shutting operations if any passengers are entering or exiting.

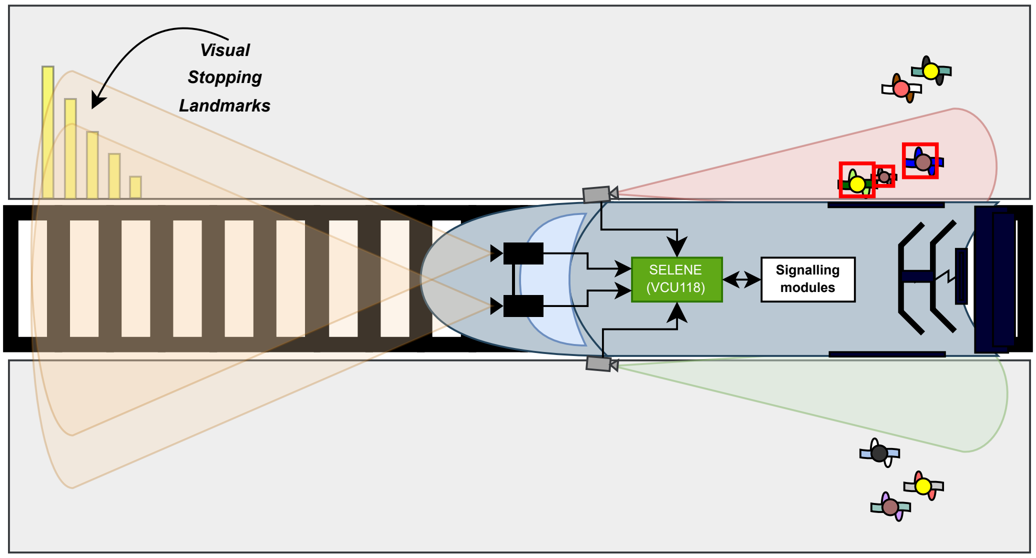


Figure 1. Physical set-up of the solution. Equipment distribution on the train.

Apart from the functional requirements, there are two key requirements to be met in this use case. The first is to keep the processing time as low as possible, since the accurate stop with a moving train (and its inertia) does not allow latencies that could turn the results of the visual analysis into obsolete data, as these would not be useful for an accurate control of the vehicle. On the other hand, the passenger detection functionality has to be secured by combinations of redundant executions over isolated resources.

3.1. System Set-Up

As shown in Figure 1, the set-up consists of two cameras (located in the train cabin) in properly calibrated stereo vision configuration, another two rear cameras (pointing at the passenger doors) and the Xilinx VCU118 board which incorporates the SELENE platform. Each camera sends a real time video stream into the system and all of these streams are analysed using the CV- and AI-enhanced algorithms to extract valuable data and send information to the next signalling equipment (decision making and actuators modules).

3.2. System Workflow

The solution architecture contains three main logical modules. The first one captures data coming from cameras and it synchronizes them in time. If the data comes from stereo cameras, it also rectifies them. The second one performs a real time data analysis using CV and AI techniques. The third collects the results of the analysis, and for those safety functionalities the 2oo2 (two-out-of-two) RootVoter (RV) logic is applied.

The most demanding computer resource functionalities are concentrated in the second logical module which is fully executed in the VCU118 board and SELENE platform:

- Platform landmark detection and identification: this detects the start/end landmarks of the platform by a pre-trained AI model (YOLOv4 [19] architecture) inference process, determining if the train is on the platform and establishing a reference point in the approximation phase for the ultimate accurate stop.
- Distance estimation: this support the precise stop process in the platform area. The distance to station stopping landmarks is calculated updating the predicted remaining distance of ATO. This calculus is based on a dense disparity map calculated by the Semi-Global Block Matching (SGBM) method [20].
- Passenger detection: using the same techniques but a different AI model, it detects passengers when they are boarding or exiting the train, managing the door opening and closing commands.

4. Deployment on SELENE Platform

This section describes the use case deployment details and it focuses specifically on four main contributions of this work: the HW accelerator, acceleration runtime, hypervisor and rootvoter.

The SELENE platform builds upon a combination of a multi-core and accelerators, which are prototyped on a FPGA System On a Chip (SoC), based on the non-proprietary RISC-V instruction set architecture (ISA). Due to their open nature, the use of an open ISA with a Linux OS and Jailhouse hypervisor [21] offers flexibility and an extension at the SW level. All these features are made compliant with the highest safety integrity levels across domains by building adequate safety measures such as monitoring, fault containment, diverse redundancy (RV availability), ease for testability, etc., in the HW and SW layers. The architecture of our railway use case and how it is implemented on top of this platform is described below.

In the use case presented in this work, safety considerations are different for each functionality. Automatic accurate stop is not a safety-related function since if the train stops beyond the platform the doors are not opened. On the contrary, a safe passenger transfer has safety implications since closing doors when passengers are still getting in/out of the train might endanger their physical health.

Due to the need for high-performance (based on parallel executions) and function separation, each task execution should rely on distinct RISC-V cores and isolated cells (except NoSafety functions which can share the same cell but should be isolated from the rest of the safety-related executions). This may be accomplished by utilizing the SELENE platform HW/SW isolation features based on the Jailhouse hypervisor.

Figure 2 shows the SELENE platform architecture in the HW and SW domains incorporating the functionalities that need to be executed in the presented use case. In the HW domain, the separation into different RISC-V CPU cores of the three main tasks can be seen. The passenger detection function, being a safety function, is redundant and its two outputs are managed by the RV of the SELENE platform. Three of the processes (two for passenger detection and one for platform landmark detection) also use hardware acceleration for their inferences. Finally, the accurate stop function requires two separate RISC-V CPU, one for AI-model inference used for landmark detection and another for distance estimation based on stereo matching algorithms. Both of them are very resource-consuming.

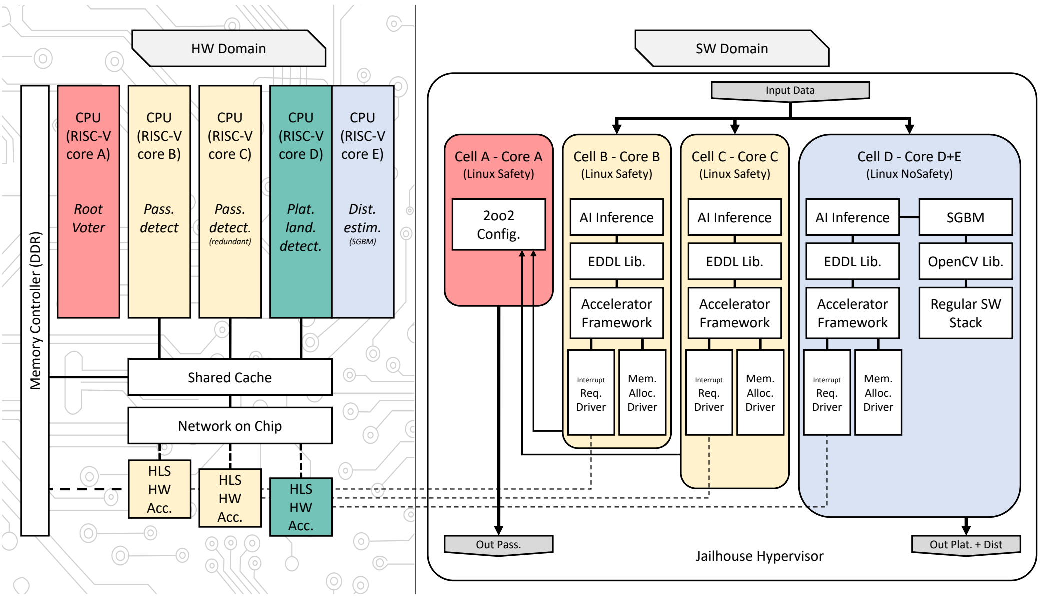


Figure 2. HW and SW architecture of the proposed solution.

In the SW domain we can see the control of the different executions of the functionalities through the creation of cells controlled by the Jailhouse hypervisor. Moreover, we can also appreciate the different SW stack executed in each core depending on the task under execution as well as the interrupts required to make use of HW acceleration.

The safe passenger transfer function, based on AI-enhanced passenger detection, is developed in one core with a replica (in the second core) to build a 2oo2 redundant system, such as those required to achieve high criticality in the railway sector [22]. A comparison of the function results is carried out by the RV HW modules incorporated in the SoC. SELENE hardware monitors make sure that safety properties are preserved.

In order to deploy this entire HW and SW architecture on the XIllinx VCU118 board, some research and development was required beyond the SoA. The exact contributions of this work with respect to the SoA are as follows:

- HLSinf HW Accelerator extension: the creation of new layers to support the YOLOv4 architecture: a Support Tensor Machine (STM) layer which is a grouping of the three different layers (softmax, hyperbolic tangent and element-to-element multiplication), and an ADD layer (element-to-element addition).
- New Acceleration Runtime: enabling HLSinf HW accelerator and Linux OS communications, accelerators control, memory allocation, and interruption manager.
- AI-inference SW library extension: a new compute service, called SELENE, to port and extend the inference library, making the platform compatible with most known AI architectures.
- Hypervisor new extension: a porting solution to RISC-V CPU and enabling process isolation.
- RootVoter new extension: a porting solution to RISC-V CPU and enabling safety-related executions on the SELENE platform based on redundant execution.

The AI hardware accelerator, AI-Inference library, and an acceleration runtime method created in this work comprise the SELENE Accelerator Framework (SAF). It works as follows: first, the European Distributed Deep Learning (EDDL) [23] inference library initializes the HLSInf [24] HW accelerator using the generated JSON configuration file. Next, the inference input data (i.e., the data to be processed) is loaded in the main memory shared with the accelerator. For this purpose, a dedicated input buffer is allocated in the memory using the Memory Allocation Driver. As soon as the input is loaded, the inference library runs the accelerator and blocks the process until the accelerator has finished (or until the timeout has been reached). The final step for the EDDL is to read the output buffer to retrieve the inference output data (i.e., the data processed by the accelerator).

4.1. SELENE AI HW Accelerator

The HLSinf accelerator is a high-level synthesis open-source FPGA accelerator which creates an efficient hardware IP for ASIC or FPGA targets and is used for inference processes of AI models based on convolutions. The central characteristic of this accelerator is flexibility, as it allows a specific AI hardware accelerator to be designed and implemented to the particular use case.

It is designed using the channel slicing concept, where a set of input channels are processed in parallel, and a set of output channels are produced in parallel. This allows the programmer to select the degree of parallelism at the design time, where a bigger parallelism implies a bigger accelerator size and more FPGA resources. This speed-up flexibility allows the user to define the best well-suited parallelism considering the available FPGA resources and the degree of parallelism desired.

This accelerator has been integrated into the SELENE SoC and interconnected with memory and the RISC-V cores using an AXI interconnect. This HLSinf accelerator on the SELENE platform can be customized to support specific data formats and Neuronal Network (NN) layers and currently supports several well-known AI models such as YOLOv3/4, Tiny-YOLO, or VGG16. The accelerator and the CPU cores share the same memory, which minimizes the cost of data movement and allows fine-grain HW/SW co-designs of the AI algorithms between the RISC-V cores and the AI accelerators to be performed. In addition, HLSinf has been designed to run in the EDDL library, providing the support needed to run offloaded AI model layers on the FPGA. It can configure and compile a given subset of network layers for use in an inference process running with EDDL. HLSinf and EDDL allow a perfectly coupled HW/SW co-design approach where some parts of the model run in the FPGA, whereas the rest run in the CPU or GPU when available.

Figure 3 shows the design of the accelerator in the SELENE Platform. This has been defined around the dataflow model using modules interconnected by data streams. This dataflow model accelerates the overall throughput of the design as it enables task-level pipelining, permitting several operations to start before the previous functions have completed all of their operations.

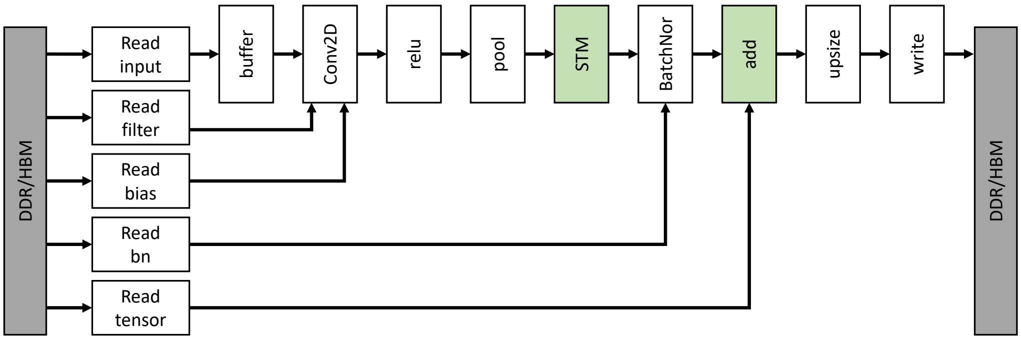


Figure 3. Design of the HLSinf accelerator (new contributions in green color).

4.2. Acceleration Runtime

This work also presents a new low-level runtime that allows the HW accelerators included in the SELENE platform to communicate with the Linux operating system. The SELENE Acceleration Runtime (SAR) is the lowest software level and it controls the accelerators, ensures memory allocation, and manages interruptions. The runtime also interacts with the EDDL inference library using an Open-CL-like Application Program Interface (API). The EDDL and the low-level runtime are included in the SELENE Linux Image deployed on the NOEL-V-based [25] platform. Thus, the final application running on the NOEL-V infers the AI algorithms and makes the deployment of the HW accelerators transparent for the use case. The SAR can handle multiple kernels and is designed to easily configure the control of all the kernels with a parametric register from a JSON file.

The accelerators require a contiguous physical memory block for data input and output. As we are using the Linux OS and we cannot directly write to the main memory (it would end in an OS crash), we use a kernel driver, called the Memory Allocation Driver, to ensure the contiguous memory allocations. An API has been designed for interfacing the SAR with the upper software level. This API contains a light OpenCL C++ compatibility layer for easier operations.

4.3. AI-Inference SW Library

The EDDL library is a general-purpose, open-source, deep-learning library used for the training and inference processes of NN models. One of the key features of EDDL is its ability to work with a wide variety of hardware, including CPUs, GPUs or FPGA. This allows users to take advantage of the best hardware for their specific use case and makes it easy to switch between different hardware platforms. Interoperability is provided with the EDDL Open Neural Network Exchange (ONNX) format support, as it allows pre-trained models in ONNX format to be loaded and it ensures compatibility with other frameworks.

The SELENE platform relies on Linux as the default operating system. In particular, it uses a Debian-based RISC-V Linux adaptation to NOEL-V. The AI software toolchain integration is built on top of this Linux distribution. To deploy artificial intelligence models on the SELENE platform, the EDDL library has been extended by including the SELENE platform as a new computing target. The EDDL library is then deployed on top of the Linux OS running on the NOEL-V processor. This allows for the inference process to be executed entirely within the NOEL-V multi-core system. Additionally, SAF is used to offload heavy computations to the SELENE AI hardware accelerators to speed up the inference process.

4.4. Jailhouse Hypervisor

Jailhouse is a partitioning hypervisor based on Linux. It configures CPU and device virtualisation features of the hardware platform in such a way that none of the resulting domains, called cells, can interfere with each other in an unexpected way. Jailhouse currently officially supports the x86-64, ARMv7 32-bit and ARMv8 64-bit architectures. For the SELENE SoC, which uses the NOEL-V processor, the code has been ported to the open RISC-V ISA. The main implementation challenge has been the decoding of transformed/pseudo instructions stored in dedicated system registers on processor exceptions (e.g., memory access violations).

As shown in Figure 2, different cells have been created for different processes (even for redundant ones) to isolate some functionality from the rest, avoiding the sharing of resources and interruptions between them.

4.5. RootVoter

The current version of SELENE SoC includes four RV cells, each has a maximum of 16 datasets to vote. The voting scheme MooN and the timeout interval are configured during cell initialization. The RV driver for Linux enables (a) resetting and initializing each RV cell, (b) loading the datasets to the dedicated RV registers, (c) polling the voting results from each RV cell, (d) parsing voting results and diagnosing the errors (if any).

The voting logic assumes that the datasets are loaded during T clock cycles (configured by software) after the configuration command. The voting starts when all N datasets are loaded to the set registers, or when at least M datasets are loaded by the end of time interval T. If at the end of this time interval less than M datasets are available, then the RV cell reports a timeout.

Once the voting is completed the, RV reports an agreement status that indicates whether at least M datasets (out of N) match among themselves and validity flags that indicate whether each particular dataset matches with the rest. The use case of this work uses only one cell for RV and the chosen voting scheme is 2oo2.

5. Test Description

Several tests have been defined to validate the platform. The tests are centered on critical requirements of the performance, process isolation and redundant execution, and also on the integration of third party libraries such as OpenCV [26]. The tests have also been found suitable to evaluate SELENE outputs in comparison to available market solutions that have higher TRL in inference execution but lack safety in-built mechanisms such as Alveo from Xillinx [27] and Jetson AGX Xavier from Nvidia [28].

In order to run the tests, part of a private *CAF* dataset [29] containing stereo images of an urban railway environment has been used. The dataset contains 19 sequences on the railway track. A sequence defines a record that starts at one station and spans the next station or two until the train stops. The frames are rectified RGB colour images coming from a stereo camera stored with lossless compression using 8-bit PNG files. The size of the images is 1280 × 720 (HD). Only a sub-part of this database has been used, those frames where landmarks and passengers are present. In total, 200 stereo image pairs from ten different sequences were used.

- System workflow validation test: the entire workflow of the system has been validated using dataset images. Apart from the correct functioning of the main functionalities, special attention has been paid to the following two points:
  - Back support libraries for the distance calculus: distance calculus requires an available implementation for the SGBM algorithm [26]. This implementation is ready in the OpenCV library but must be validated to ensure that the libraries that compute stereo SGBM matching can be cross compiled and executed in the SELENE platform with RISC-V architecture.
  - Model parsing compatibility: the models used for the use case are trained using the Darknet framework [30]. The Darknet output is not compatible with other frameworks and, for that reason, ONNX has been chosen as the sharing format. As ONNX establishes a standard format, but there are no standard parsers or exporters, the compatibility of exported models with the EDDL ONNX parser must be validated. Inference tests were used to validate this compatibility.
- AI models (passenger and landmark detectors) inference performance test: this test focuses on the performance of the machine learning algorithm in the platform. In the test, the Tiny-YOLOv4 inferences for landmark and passenger detection are executed with different computing precision. The models are 608 × 608 RGB image input models that were trained using transfer learning with a database labelled with railway traffic signals, platform landmarks and people/passengers. In addition, the goal has also been to compare the performance of the accelerator against SoA existing hardware such as Xilinx Alveo and Nvidia AGX Xavier after normalising inference time with respect to frequency.
  - VCU118: this test aims to compare the performance executing the Tiny-YOLOv4 use-case model in the VCU118 SELENE platform using different accelerator configurations (different NN layer distribution on CPU and HW accelerator and different bit number precision). It also compares the performance of the accelerator against the CPU on inference tasks to calculate the impact of implementing the accelerator over the whole platform performance.
  - Xilinx Alveo: this test is based on an inference benchmark (a technology-agnostic evaluation) evaluating HLSinf in a Xilinx Alveo Board in order to validate and evaluate the accelerator in an existing environment to isolate the results from the custom SW stack that is required for VCU118 board. Tiny-YOLOv4 for railway signalling detection was evaluated on a Xilinx Alveo with external Intel CPUs facilitating the evaluation of the accelerator isolated from CPU performance.

  - Nvidia Jetson AGX Xavier: the same image inference test is executed in the GPU of the SoA edge computing platforms.
- Distance calculus performance test: an evaluation on SGBM performance is also a target for the test. The performance of the SGBM algorithm also allows a CPU speed evaluation.
- Process isolation test: unfortunately, the process of porting the Jailhouse hypervisor to the SELENE platform could not be completed in time before the end of the project and is still ongoing. However, within this work, the correct functioning of hypervisor has been tested over RISC-V architecture using a QEMU [31] machine emulator and virtualizer. This consists of concurrently executing multiple applications of the use case on a single RICS-V SoC allowing it to evaluate non-interference properties. This also allows any impact on application precision to be evaluated as well as the performance impact of shared/contended resources. First of all, each process has been executed separately to obtain the performance data without interference from other processes. Then, in a second cell, a workload is introduced incrementally based on micro-benchmarks in interference analysis [32]. All combinations to two of the four cells have been tested. The result of these tests has been compared to the initial evaluation performed in isolation to check that performance degradation is bounded and functional behavior remains unaffected. With this configuration, the impact of several types of interference (shared memory, shared cache, shared buses) on each selected algorithm has been studied.
- Redundancy and RV test: two different tests have been carried out for RV evaluation. The first one at use-case level where *PassengerDetector* functionality is executed redundantly on the SELENE platform. The RV is configured for a 2oo2 scheme. The PC is used for interaction with the processes on the SELENE platform. Instead of the real door-closing command system, a stub is running on the PC to receive the command from the *PassengerDetector*.

  Each *PassengerDetector* process sends a vote containing the command value to the RV. In order to simulate the failure, a script has been developed enabling it to be injected in order to vote failure, send a wrong vote and test the system. The RV checks whether both of the two processes send the same vote. The master process checks the result of the RV. If the check was successful, the master process sends a command with the door-closing signal. If the check is not successful, it will send an order to keep the doors opened.

  The second one is related to low-level platform validation, where the RV subsystem has been validated by means of FPGA-based Fault Injection (FFI). This application performs a staggered redundant execution of a matrix multiplication kernel with two replicated processes. At the end of the kernel execution, each redundant process calculates the digest (CRC32) for the output results. These digests (from each process) are loaded to the dedicated dataset registers of the RV cell. For the sake of simplicity, this application uses only one RV cell. The voting scheme configured for the RV cell is 2oo2, and the configured timeout (maximum time to wait for the datasets) is 1 ms.

  FFI experiments have been carried out using a customized version of DAVOS [33] fault injection tool. Faults have been injected into the CPU cores: Cell C (which executes one of the kernel replicas), and Cell B (which executes the monitoring process). The considered faultload comprises single bit-flips in those cells of FPGA configuration memory that configure targeted SoC components (CPU cores). A total of ten thousand faults have been injected during FFI experiments (5000 faults per each targeted CPU core).

  The outcome of each individual injection run (fault effect) is described in terms of failure modes. The fault is masked when it produces no effect on the system. The fault leads to Replica fail when the RV raises the validity flag for one of the replicas. The fault leads to replica timeout when the RV raises the timeout flag for one of the replicas. Finally, the fault effect is double the modular redundancy fail when the RV is unable to establish an agreement, and the kernel result does not match the fault-free run.

At the end of the experiment, DAVOS calculates the percentage of each failure mode as the ratio between the number of registered failure modes of each type and the total number of injected faults.

6. Test Results

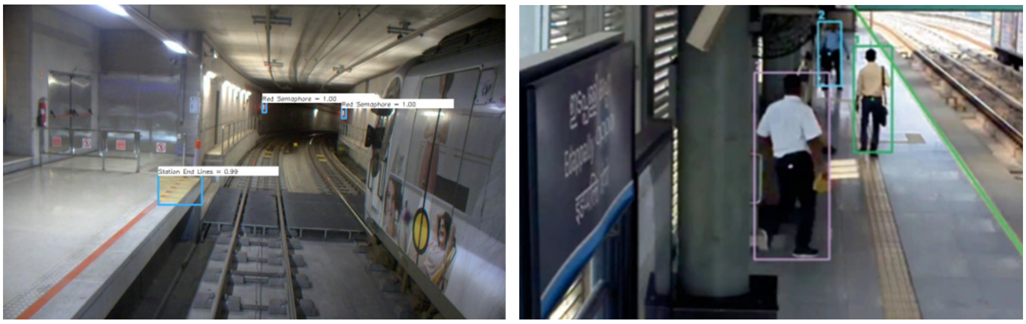
This section shows the results for the SELENE platform and compares the results with SoA platforms.

6.1. AI Model Inference Performance Results

The results for the evaluation can be seen in Table 1, together with the evaluation results printed on the input image in Figure 4. In the figure, we can see several columns of execution times of the inference of one image using the Tiny-YOLOv4 model at different platforms.

Table 1. Tiny-YOLOv4 inference times on SELENE's VCU118 board and the comparison with (a) the inference when SELENE's HW accelerator is executed at Xilinx Alveo (also using EDDL) (b) the inference when the use case is executed on the GPU of AGX Xavier and (c) the inference at CPU frequency downscaled AGX Xavier (in order to be able to set a same level comparison). "Transform" and "Others" layers are executed in CPUs. All measurements are given in milliseconds (ms).

	Tiny YOLOv4 Layers	VCU118 (100 MHz) (SELENE)	Alveo (250 MHz) (HW-Acc of SELENE)	AGX Xavier GPU (1.377 MHz)	AGX Xavier GPU (250 MHz)
	All executed on CPU	45,685,922	-	-	-
FP32	HLSinf Transform Others	1799 479 4	364 7 0	17	94
	Total	2282	371		
FP16	HLSinf Transform Others	1548 726 4	119 6 2	13	72
	Total	2278	127		
INT8	HLSinf Transform Others	796 1135 4	66 15 1	N/A	N/A
	Total	1935	82		



(**a**) (**b**)

Figure 4. Platform landmark (**a**) and Passenger (**b**) detectionsfor automatic accurate stopping and safe passenger use case. Note: (**b**) image corresponds to a platform shot. It is an example that emulates the images captured from the rear-view mirrors of the train, as the images captured for these tests are not publishable due to GDPR issues.

The VCU118 (SELENE) corresponds to the SELENE platform, using the HLSinf accelerator as the AI hardware accelerator of the platform, the EDDL library as an inference library and the SAF as the interface between the HLSinf accelerator (100 MHz) and the EDDL library. The Alveo corresponds to the inference time of the inference outside the

SELENE platform, using an Intel i7-7800-X (3.45 GHz) for the non-supported HLSinf layers and the HLSinf accelerator (200–250 MHz) deployed on the Alveo U200 board for the supported HLSinf layers. The AGX Xavier GPU corresponds to inference on the GPU of the Jetson family AGX board. Finally, the last column corresponds to the GPU downscaled because the results needed to be adjusted for the frequency of the GPU (1377 MHz) to equal the frequency of the Xilinx Alveo (200–250 MHz) in order to compare the performance of the accelerator isolated from the underlying physical technology, which limits the operation frequency.

The SELENE VCU118 platform results in Table 1 include different layer execution configuration and bit precision levels. Note that the time for one forward operation in the CPU is 4,568,592 ms, while in FP32 using the HLSinf accelerator the time lowers to 2282 ms. This result means an acceleration factor of ×2002 that rises to ×2361 when running on INT8 precision.

Comparing the GPU downscaled and the Alveo EDDL columns, the GPU behaves better while using FP32 precision. On the FP16, the HLSinf accelerator achieves 119 ms inference time per image. Taking into account that not all the layers are embedded in the accelerator, it produces slightly more inference time than downscaled. When the precision falls back to INT8, the inference time for HLSinf is 66 ms per image, together with the CPU preprocessing time required, the time to execute a forward pass on one image is 82 ms. Unfortunately, the comparison at INT8 precision is not possible as the GPU available drivers do not handle fewer than 16 bits per parameter.

6.2. Distance Calculus Performance Results

The results represented in Table 2 show that the actual inference time in the SELENE platform (two cores RISC-V CPU) is much higher than in the AGX Xavier (eight cores ARM CPU) but a direct comparison is not representative. Cumulative processing time over all cores must be calculated to obtain the computing time for all processes. In the full process time, the results show that the RISC-V CPU performance is 6.66% of the ARM performance, however operation frequency is not the same in both platforms, so frequency normalization shall be applied to obtain the actual performance for the CPU. SELENE platform CPUs run at 100 MHz and the ARM CPU runs at 2.2 GHz. The results normalizing the frequency show that the RISC-V CPU performance is actually greater than the ARM CPU.

Table 2. Depth map calculus execution times (OpenCV SGBM function in CPU). SELENE's VCU118 and AGX Xavier. The last two columns show the comparison between both platform: Raw core number normalization and Frequency Agnostic (F.A.) (100 MHz vs. 2.2 GHz) downscaled. All measurements are given in seconds (s).

					VCU118 w.r.t AGX Xavier	
	VCU118 (2 Core) SELENE		**AGX Xavier (ARM 8 Core)**		**Raw Comp.**	**F.A. Comp.**
OpenCV SGBM	Total time	Time Using Single Core	Total time	Time Using Single Core	%	%
Matching Time	48	96	0.8	6.4	6.66	146.67
Filtering Time	20	40	0.3	2.4	6	132

This test has also been used to check OpenCV compatibility and performance of an algorithm to estimate the distance from the train cabin to a stop signal on the platform. After compiling and installing OpenCV for RISC-V 64-bit architecture, a performance test consisting of the execution of the Semi-Global Block Matching (SGBM) function of OpenCV was carried out. As shown in Figure 5, the OpenCV SGBM function takes two stereo images (taken with a stereo camera, producing a left and right image) and tries to match the images creating, as a result, a disparity map which represents the distance between the detected landmarks or people to the cameras on the train.

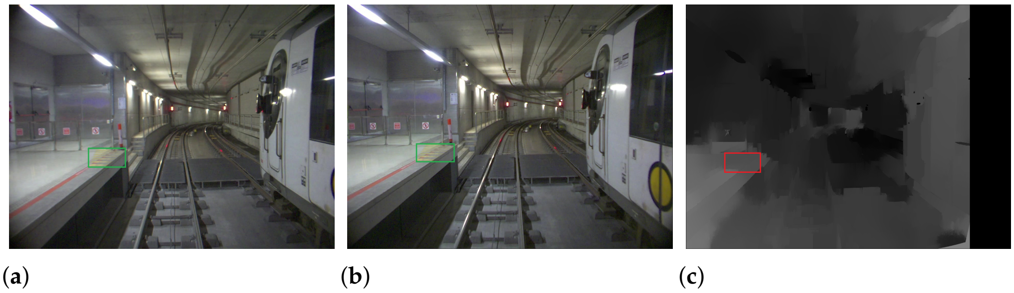


Figure 5. Distance calculusto platform landmark using stereo vision camera (left (**a**) and right (**b**) images) and extracted depth map (**c**). Green frames represent detected landmark bounding box and the red frame the landmark's corresponding area in depth map. This are is used for distance calculus.

Because OpenCV acceleration was not implemented in SELENE platform (it is planned as future work), the SGBM algorithm is just executed in CPU. Table 2 shows the execution times of the SGBM function (divided in Matching Time and Filtering Time) and the comparison with AGX Xavier board execution in its ARM CPU cores.

As expected, the Nvidia AGX Xavier with its eight cores at 2.2 GHz is much faster than the two core SELENE VCU118 at 100 MHz, however, as previously mentioned in this work, this direct comparison is not valid as the SELENE platform is an evaluation HW FPGA board with a frequency much lower than an ASIC implementation. Therefore, the comparison must be normalized to be agnostic of the frequency and the number of cores. After normalizing the cores, the SELENE platform reaches just 6.66% of the performance of the Nvidia AGX Xavier. However, after normalizing the frequencies, the RSIC-V CPU in the SELENE outperforms the ARM, demonstrating that FPGAs can be a valid option from the performance point of view.

6.3. Process Isolation Results

This Jailhouse hypervisor version was successfully executed on the QEMU, with execution of different Linux root (Safety and NoSafety) cells. The use of resources has been monitored validating the isolation capabilities of SELENE platform (shared memory, shared cache, shared buses).

Additionally, we created a simple inmate trying to escape its cell by accessing outside of its allocated memory. This attempt was correctly caught by the hypervisor that sanctioned the faulty access by a page fault exception.

6.4. Redundancy and RootVoter Results

At the use case level, all tests regarding the RV were successful. If the two scripts sent the same vote, the door enable signal is activated. If the two scripts sent a different vote, this failure is successfully detected by the RV and the doors remains closed and blocked.

At the platform level, the results of FFI show that the system has tolerated all faults injected into the kernel replica (Cell C), i.e., 0.00% of 2002 failures. The replicas themselves are quite sensitive to the injected faults: in 0.88% of cases the RV has reported a replica fail, and in 0.16% other cases the RV has reported a replica timeout. The faults injected into the monitoring process (Cell B) have not affected the behaviour of the kernel replicas, and only one 2oo2 failure per 5000 faults has been detected (0.02%).

In such a way, the described experiment has shown that the RV meets its specified functionality, i.e., it detects the errors and timeouts of replicated processes, and it establishes an agreement following the configured voting scheme. Usage of RV in the redundant applications efficiently protects the system against the faults of the replicated processes.

7. Discussion and Future Work

In this work, a new safety-critical and high-performance computing application for real-time AI-enhanced railway use has been introduced. Its design allows process isolation, redundant execution, HW acceleration and abstraction making the platform compatible for most widely used AI inference techniques and AI model architecture and formats (including open standards such as ONNX).

It is worth highlighting the implementation of specific HW and SW modules for the SELENE platform. A HW accelerator module, which can be customized to support specific data formats and neural network layers, has been deployed. HLSinf accelerator shows great performance on frequency agnostic evaluation. Using quantization and other AI model optimisation methods, the performance improves SoA. Its implementation presents a very high acceleration factor with respect to CPU execution for the Tiny-YOLOv4 algorithm. This work also presents the performance and accuracy evaluation of the use case functionalities over the SELENE platform, comparing it with executions carried out in the most widely used commercial HPC platforms, such as Nvidia's Jetson family boards or Xillinx's FPGAs.

In addition, a custom AI runtime and adapted inference SW, which abstracts the user application layer from platform specific HW configuration, has been carried out.

Finally, the rootover and Jailhouse hypervisor implementations for RISC-V based system compatibility have also been successfully validated, making it possible to execute safety-related functionalities on the platform. This solution guarantees isolating executions of the different functionalities and allows the evaluation of redundant executions with voting system when needed.

Regarding the use case, this work has demonstrated to be a valid HW platform for equipping autonomous trains that require real-time execution of safety (precision stop functionality) and non-safety (precision stop) functions based on CV and AI. With higher maturity, ASIC implementation and railway certification, the SELENE platform could suit railway industry requirements for both non-safety and safety level applications.

The next steps of the investigation will focus on improving real time execution performance (reaching lower inference times) while keeping/increasing the detection accuracy. New NN architectures will be taken into account as candidates to port them into the SELENE board. OpenCV acceleration by HW should be implemented in order to speed up basic computer vision algorithms such as SGBM. On the other hand, they will also focus on more in-depth testing and validating the platform's possibilities for redundant execution (followed by different voting systems such as 2oo3) in order to increase the safety level.

Author Contributions: Conceptualization, M.L. and J.F.; Methodology, M.L., J.F. and N.A.; Software, M.L., L.M. and F.E.; Validation, M.L., L.M. and F.E.; Formal analysis, M.L., L.M. and F.E.; Investigation, M.L., L.M., F.E., J.F. and N.A.; Writing—original draft, M.L.; Writing—review & editing, L.M., F.E. and N.A.; Supervision, J.F. and N.A.; Project administration, M.L. and J.F. All authors have read and agreed to the published version of this manuscript.

Funding: The novel approach presented in this work is being developed as a specific railway use case for autonomous train operation into SELENE European research project. This project has received funding from RIA—Research and Innovation action under grant agreement No. 871467.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: There is no available data.

Acknowledgments: The author like to thank the whole PERception (PER) development team of CAF Signalling for their help and support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shift2Rail —Home. Available online: https://shift2rail.org/ (accessed on 5 May 2023).
2. Reddi, V.J.e.a. MLPerf Inference Benchmark. *arXiv* **2019**, arXiv:1911.02549. [CrossRef]
3. CORDIS—SELENE. Available online: https://cordis.europa.eu/project/id/871467/en (accessed on 5 May 2023).
4. Waterman, A.; Lee, Y.; Patterson, D.A.; Asanović, K. *The RISC-V Instruction Set Manual, Volume I: User-Level ISA, Version 2.0.*; Technical Report UCB/EECS-2014-54; EECS Department, University of California: Berkeley, CA, USA, 2014.

5. Palmer, A.W.; Sema, A.; Martens, W.; Rudolph, P.; Waizenegger, W. The Autonomous Siemens Tram. In Proceedings of the IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]
6. Guerrieri, M.; Parla, G. Smart Tramway Systems for Smart Cities: A Deep Learning Application in ADAS Systems. *Int. J. Intell. Transp. Syst. Res.* **2022**, *20*, 745–758. [CrossRef]
7. Ristić-Durrant, D.; Franke, M.; Michels, K. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways. *Sensors* **2021**, *21*, 3452. [CrossRef] [PubMed]
8. Alstom Demonstrates Fully Autonomous Driving of a Shunting Locomotive in the Netherlands. Available online: https://www.alstom.com/press-releases-news/2022/11/alstom-demonstrates-fully-autonomous-driving-shunting-locomotive-netherlands (accessed on 5 May 2023).
9. Train Autonome Service Voyageurs: Essais Réussis. Available online: https://www.youtube.com/watch?v=vlEy7GYe684&ab_channel=GroupeSNCF (accessed on 5 May 2023).
10. Autonomous Train Tests Were Carried Out Succesfully in Finland. Available online: https://www.proxion.fi/en/autonomous-train-tests-were-carried-out-succesfully-in-finland/ (accessed on 5 May 2023).
11. Railtech—DB Cargo Automates Shunting to Boost Single Wagon Load Traffic. Available online: https://www.railfreight.com/railfreight/2021/10/27/db-cargo-automates-shunting-to-boost-single-wagon-load-traffic/?gdpr=accept (accessed on 5 May 2023).
12. Cognitive Pilot—Tram Automation Software Contract Awarded in Shanghai. Available online: https://en.cognitivepilot.com/breaking-news/fitsco-tram-english/ (accessed on 5 May 2023).
13. RailTech—Remote-Controlled Shunting. Available online: https://www.railtech.com/digitalisation/2020/09/25/remote-controlled-shunting-on-tests-in-switzerland/ (accessed on 5 May 2023).
14. Digitale Schiene—Fourteen Eyes on the Road Ahead: Second Sensors4Rail Test Project Successful. Available online: https://digitale-schiene-deutschland.de/en/Sensors4Rail-test-project (accessed on 5 May 2023).
15. Youtube—Train Autonome: Automatisation de la Lecture de la Signalisation Latérale. Available online: https://www.youtube.com/watch?v=WiYavvqh7Bk&ab_channel=GroupeSNCF (accessed on 5 May 2023).
16. La Reconnaissance Faciale des Signaux, le Projet ARTE D'alstom. Available online: https://mediarail.wordpress.com/2022/10/23/alstom-projet-arte-basse-saxe/ (accessed on 5 May 2023).
17. Perez-Cerrolaza, J.; Obermaisser, R.; Abella, J.; Cazorla, F.; Grüttner, K.; Agirre, I.; Ahmadian, H.; Allende, I. Multi-Core Devices for Safety-Critical Systems: A Survey. *ACM Comput. Surv.* **2020**, *53*, 1–38. [CrossRef]
18. Mc Guire, N.; Allende, I. Approaching certification of complex systems. In Proceedings of the 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), Valencia, Spain, 29 June–2 July 2020; pp. 70–71. [CrossRef]
19. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef].
20. Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
21. Siemens. JAILHOUSE. Available online: https://github.com/siemens/jailhouse (accessed on 5 May 2023).
22. Gerstinger, A.; Kantz, H.; Scherrer, C. TAS Control Platform: A Platform for Safety-Critical Railway Applications. *ERCIM News* **2008**, *2008*.
23. Cancilla, M.; Canalini, L.; Bolelli, F.; Allegretti, S.; Carrión, S.; Paredes, R.; Gómez, J.A.; Leo, S.; Piras, M.E.; Pireddu, L.; et al. The DeepHealth Toolkit: A Unified Framework to Boost Biomedical Applications. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 9881–9888. [CrossRef]
24. Flich, J.; Medina, L.; Catalán, I.; Hernández, C.; Bragagnolo, A.; Auzanneau, F.; Briand, D. Efficient Inference Of Image-Based Neural Network Models In Reconfigurable Systems With Pruning And Quantization. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 2491–2495. [CrossRef]
25. NOEL-V. Available online: https://www.gaisler.com/index.php/products/processors/noel-v (accessed on 5 May 2023).
26. OpenCV. Available online: https://opencv.org/ (accessed on 5 May 2023).
27. Accelerating DNNs with Xilinx Alveo Accelerator Cards. Available online: https://docs.xilinx.com/v/u/en-US/wp504-accel-dnns (accessed on 5 May 2023).
28. Jetson AGX Xavier and the New Era of Autonomous Machines. Available online: https://info.nvidia.com/rs/156-OFN-742/images/Jetson_AGX_Xavier_New_Era_Autonomous_Machines.pdf (accessed on 5 May 2023).
29. Etxeberria-Garcia, M.; Zamalloa, M.; Arana-Arexolaleiba, N.; Labayen, M. Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case. *IEEE Access* **2022**, *10*, 69200–69215. [CrossRef]
30. Biddle, P.; England, P.; Peinado, M.; Willman, B. The Darknet and the Future of Content Protection. In Proceedings of the ACM Workshop on Digital Rights Management, Washington, DC, USA, 18 November 2002; Volume 2696, pp. 155–176. [CrossRef]
31. QEMU—A Generic and Open Source Machine Emulator and Virtualizer. Available online: https://www.qemu.org/ (accessed on 5 May 2023).

32. Ensuring Software Timing Behavior in Critical Multicore-Based Embedded Systems. Available online: https://www.embedded.com/ensuring-software-timing-behavior-in-critical-multicore-based-embedded-systems/ (accessed on 5 May 2023).
33. DAVOS—A Fault Injection Toolkit for Dependability Assessment, Verification, Optimization and Selection of Hardware Desings. Available online: https://github.com/IlyaTuzov/DAVOS (accessed on 5 May 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

4.6 European Common Data Management Platform Definition for Railway AI Function Development

- **ESTADO:** Bajo revisión, pendiente de aceptación.
- **Autores: Mikel Labayen** and Daniel Ochoa de Eribe and Ander Aramburu and Marcos Nieto and Naiara Aginako
- **Revista:** Transportation Research Part C: Emerging Technologies
- **Editor:** Elsevier - Science Direct ↗

European Common Data Management Platform Definition for Railway AI Function Development[⋆]

Mr. M. Labayen[a,d,*] (Railway Engineer and Researcher), Mr. D. Ochoa de Eribe[a] (Railway Engineer), Phd. A. Aramburu[b] (Researcher), Phd. M. Nieto[c] (Researcher) and Phd. N. Aginako[d] (Researcher)

[a] *CAF Signalling, Donostia, Spain*

[b] *CAF R&D, Beasain, Spain*

[c] *Vicomtech Technology Research Centre, Basque Research and Technology Alliance, Donostia, Spain*

[d] *Computer Sciences and Artificial Intelligence Department, University of Basque Country, Donostia, Spain*

ARTICLE INFO

Keywords:
Common Data Management Platform
Artificial Intelligence
AI Training and Testing
Autonomous Vehicle
Railway

ABSTRACT

Digitalisation and automation of operations in the railway industry includes the use of Automatic Train Operation systems that provide automated functions to reach different levels of automation, known as the Grade of Automation (GoA) levels. Artificial Intelligence has emerged as technology that can substitute humans in certain driving tasks, in GoA3 (driverless) and GoA4 (unattended) modes. AI capabilities include perception, decision-making, precise positioning, or optimization of communications. The success of AI models depends on the quality and diversity of the data used for training, along with the set-up of a data life-cycle framework that covers creation, training, testing, deployment and monitorisation. The management of training datasets implies both expensive and time-consuming data gathering, labelling, curation and formatting efforts, potentially hindering the development of reliable AI systems. This paper presents a Common Data Management Platform developed by a consortium of European railway stakeholders, devised to efficiently manage data for AI training, and which is demonstrated in two different Proofs of Concept.

1. Introduction

Nowadays, autonomous driving functions in railway operation are based on developments in the automotive sector. For example, Advanced Driver Assistance Systems (ADAS), which come directly from the automotive sector, have been implemented in several demonstrators and showcases in the railway domain. These solutions have something in common: they are all based on artificial sensing.

Artificial sensing permits gathering information from the environment, which becomes a key factor when talking about enabling autonomous operation for railway transport optimization. Furthermore, autonomous driving requires the implementation of new on-board functions that complement the Automatic Train Protection (ATP). Some of these functions rely on the perception of both indoor and outdoor environments. Sensing of the outdoor environment offers driving clearance (no obstructions, signal status on green...), speed supervision (signals for speed limits), or vehicle localization. Moreover, sensing of the indoor environment will be primarily necessary in GoA4, where automated event detection and quicker reaction times will increase operational safety, on-board security, and overall service quality (maintenance).

In railway scenarios, the perception layer based on Computer vision (CV) and Artificial Intelligence (AI) technologies and including sensor fusion provides the required environment understanding. CV and AI technologies are essential for providing situational awareness, or the assessment of events, objects, and their relevance around the vehicle. However, in order to achieve AI solutions adapted to the railway domain, massive volumes of high-quality data are required. AI training and testing needs pre-recorded and synthetic scenes, data processing tools that imply complex computations and storage infrastructures. Needless to say, this huge task requires the collaboration of all the stakeholders in the rail sector.

Therefore, the creation of a complete and comprehensive Common Data Management Platform (CDMP) will benefit the whole railway ecosystem: from suppliers, who will reduce needs for initial investments, to clients, that will benefit from shorter time-to market and better products. A platform of this magnitude would avoid AI caused drift, making it possible to move towards the goal of achieving the sufficient operational maturity to enable autonomous operation. This maturity serves as the foundation for obtaining the needed certification procedures as well as to guarantee predictable behaviour.

The work presented in this article faces this challenge, and it proposes a stakeholder-agreed (European main railway players; operators, infra-managers, suppliers...) solution to the problem. This new contribution focuses on the definition of the Common Data Management Platform for artificial sense training, testing and certification and its validation through different Proofs of Concept (PoC).

This paper reports the results of that work: Section 2 resumes some pertinent previously published works from the automotive industry, first database attempts of the railway world and an analysis about reusability/exportability of automotive experiences to railway. Section 3 provides the identified new and most relevant use cases to be covered by the CDMP. Section 4 describes the high-level platform

[⋆] Declarations of interest: none.

✉ mlabayen@cafsignalling.com (M. Labayen)
ORCID(s): 0000-0001-8136-5324 (M. Labayen)

definition based on identified requirements and the description of the main modules of the platform (data acquisition/generation, data labelling and training/testing). Section 5 focuses on data management (including the functional architecture that is suggested) to guarantee the system's scalability, modularity, and interoperability. Section 6 summarises the criteria of data protection and anonymisation and section 7 the criteria related to platform access and contribution policies. Finally, the PoCs, which are a reference guide describing how the data management platform was created for a particular use-case, is presented in Section 8. Section 9 drawns the conclusions of this works and adds some suggestions for future work.

2. Related work

2.1. MLOps platforms

The convergence of advances in Deep Learning (DL), Big Data (BD) and High Performance Computing (HPC) technologies has created a fruitful technology ecosystem that leverages the creation of effective AI systems in many sectors. In the context of smart mobility, AI has proven a technology that enables advanced functions such as sensorial perception, decision making, route optimization, and, eventually, automated driving functions.

The fuel of AI is data, and, as a consequence, many efforts are devoted to creating technologies to support data creation, management, processing and monitoring. As technology develops, a vast amount of libraries, platforms, applications, standards and initiatives are emerging and thriving in the Machine Learning Operations (MLOps) landscape. The MLOps concept extends the DevOps (Development and Operations) from the SW industry by adding data and ML-specific applications MLO (2023), and therefore includes a wide range of tools and perspectives. Some examples are: versioning (DVC, Liquidata), labeling (Scale, OpenLabel), processing (Spark, Dagster), Exploration (dbt, Rapids, pandas), data lakes/warehouse (snowflake, databricks), sources (S3, Parquet, Postgres), training (Pytorch, fast.ai, RAY, Hugging face), resource management (slurm, Docker), SW management (git, visual code), experiment management (mlflow, tensorboard, neptune, comet), hyperparameter tuning (sigopt, tune), monitoring (fiddler, grafana), edge (TensorRT, Onnx, TensorFlow Lite), Web (Kubernetes, Lambda, Seldom), CI/testing (Jenkins, circleci, buildkite), etc.

All-in-one MLOps solutions with integrated services already exist, mostly promoted by large cloud vendors, such as FBLearner by Facebook, Google Cloud AI Platform, or AWS SageMaker. Other options are FloydHub, Paperspace, Gradient, Neptune or Domino Data Lab, to name a few. These offers include managed Platform-as-a-Service solutions that simplify technology choices and accelerate time-to-market development of ML solutions, at the cost of limiting the portability of the project, adhering to private formats, and elevated cloud infrastructure costs.

2.2. Data sharing

Data sharing has become one of the main pillars of the EU Digital Strategy, with the publication of the European Strategy for Data, which includes the EU Data Act EU (2023a), that joins other regulatory initiatives that cover privacy such as General Data Protection Regulation (GDPR) or Artificial Intelligence development (EU AI Act).

As a response, Open Data initiatives, such as the International Data Space Association (IDSA) IDS (2023), Gaia-XGaia-X (2023), or the EU Open Data PortalEU (2023b) have been created. They establish principles, guidelines, reference architectures and guidance to standardisation, industry, academia, legal entities and national regulatory bodies.

European projects funded by the Horizon Europe programme are requested to produce Data Management Plans (DMP) that address FAIR data (Findable, Accesible, Interoperable and Reusable) strategies. These include identification of data and metadata types, mechanisms for publication, interoperability, clear licensing options, security, ethics and privacy preservation.

2.3. Automotive/railway databases

In parallel to the MLOps frameworks and the European-level regulatory framework, a large number of datasets are being released and made openly available (mostly for research purpose) for AI training and testing, containing millions of images, point clouds and other data from a variety of sensors. These datasets can be seen as practical exercises to effectively structure and distribute data with the purpose of being used for benchmarking purposes (e.g., KITTI Geiger, Lenz, Stiller and Urtasun (2013), Virtual KITTI Cabon, Murray and Humenberger (2020), nuScenes Caesar, Bankiti, Lang, Vora, Liong, Xu, Krishnan, Pan, Baldan and Beijbom (2020), Apollo Wang, Huang, Cheng, Zhou, Geng and Yang (2019)), to gain prestige (e.g., Audi A2D2 Geyer, Kassahun, Mahmudi, Ricou, Durgesh, Chung, Hauswald, Pham, Mühlegg, Dorn, Fernandez, Jänicke, Mirashi, Savani, Sturm, Vorobiov, Oelker, Garreis and Schuberth (2020), Waymo Sun, Kretzschmar, Dotiwalla, Chouard, Patnaik, Tsui, Guo, Zhou, Chai, Caine, Vasudevan, Han, Ngiam, Zhao, Timofeev, Ettinger, Krivokon, Gao, Joshi, Zhang, Shlens, Chen and Anguelov (2020), Ford Agarwal, Vora, Pandey, Williams, Kourous and McBride (2020), Lyft5 Kesten, Usman, Houston, Pandya, Nadhamuni, Ferreira, Yuan, Low, Jain, Ondruska, Omari, Shah, Kulkarni, Kazakova, Tao, Platinsky, Jiang and Shet (2019)), or pioneering in specific application domains (e.g., Woodscape Yogamani, Hughes, Horgan, Sistu, Chennupati, Uricar, Milz, Simon, Amende, Witt, Rashed, Nayak, Mansoor, Varley, Perrotton, Odea and Pérez (2019) on fisheye segmentation, DMD Ortega, Kose, Cañas, Chao, Unnervik, Nieto, Otaegui and Salgado (2020) on driver monitoring). Their estimated volume (at Q1 2023) sums up to more than 12TB of data, and more than 3800 hours of driving.

However, the lack of standardized data formats, the heterogeneity of the purpose-specific vehicle set-ups, and customized annotation models implies that different data parsers must be developed in order to test, train or validate algorithms or models for each data source.

Apart from perception-related datasets, platforms for scenario-based testing have been created (but not openly), such as Safety PoolSafety Pool (2023), Streetwise Streetwise (2023), Scenius AVL (2023), ADSCENE VVM (2023), or PEGASUSPEGASUS (2023), containing scenario descriptions and tools to run virtual testing.

In the railway domain, the number of datasets is way more limited, and, to the best of our knowledge, no scenario-based initiative exists so far.Some datasets for camera-based AI training exist, such as RailSem19 Zendel, Murschitz, Zeilinger, Steininger, Abbasi and Beleznai (2019), a dataset with more than 8500 images of rail traffic semantic annotations on rail scenes, and FRSign Harb, Rébéna, Chosidow, Roblin, Potarusov and Hajri (2020), which contains labeled images of French railway traffic lights. Datasets for semantic segmentation use LiDAR-based set-ups to produce point-clouds of railroad environments Cserep (2022)Lamas, Soilán, Grandío and Riveiro (2021)Yu, He, Qian, Yang, Zhang and Ou (2022) or thermal images Yuan, Mei, Chen, Niu and Wu (2022).

2.4. Analysis of the reusability/exportability of AI-related technologies to railway

The automotive sector has led AI-based pioneering advances in autonomous mobility with solutions for perception, decision-making, or navigation functions. In other domains, such as agriculture or railway mobility, the industry doesn't aim to reinvent the wheel and adapt such advances into their own specificities.

In particular, railway and automotive operations have strong similarities. Use cases are often aligned: obstacle detection, traffic sign recognition, or vehicle localisation. Enabling technologies and scopes are also equivalent: (a) perception based on camera and range (RADAR, LiDAR) sensors, (b) functions for detection, identification, tracking and distance estimation, (c) digital map infrastructures and services, (d) vehicle-to-anything wireless communications, and (e) similar validation and verification technologies (data-driven, virtual testing, scenario-based evaluation, etc.).

Nevertheless, the gap exists, and differences need to be highlighted to specialize the AI-related technological choices. In some cases, differences impose additional challenges: (a) larger vehicle sizes imply more complex sensor set-ups, or (b) heavy dynamics imply longer sensing distances to actuate preventive braking manuevers. However, in general, railway operation simplifies some of the road-level dimensions: (c) simpler motion dynamics (longitudinal paths), (d) higher levels of automation operation, (e) limited or pre-fixed driving tracks (more controlled infrastructure monitoring, pre-identified risk areas such as level crossings), (f) more restricted environments and behaviour of other actors, or (g) less power/size limitations for AI equipment.

3. Use cases

This section identifies use cases to be considered in the development of a common data platform. They contain key aspects enabling an effective usage of the CDMP to search, share and combine data from different sources where all participants can benefit from each other. Moreover, the cases suggest a framework to mutually improve the datasets and models through the possibility of reporting data/model analysis and completeness issues or detecting errors to the owners. In addition, it provides the developers with the ability to evaluate their models on predefined datasets for specific tasks with several use cases. For a clear overview, the various use cases are divided into five different groups:

- **Platform workflow:** The use case describes the life cycle of the data in the platform; from the collection and upload to the deployment of an individual dataset until its discard.

- **Platform management:** These use cases describe the process of, a) receiving access to the platform with specific rights (access and rights management), b) uploading new raw data to the data management platform and the ability to update each dataset (upload and update a dataset), c) ensuring the safety of the platform (ensuring data protection) and finally d) requesting specific additional data samples e.g., snowy environment, within the platform (request of additional data).

- **Data quality:** These use cases describe the process of, a) evaluating the completeness of a given dataset (analysis of dataset completeness), b) evaluating the quality of the dataset considering the accuracy and correctness of a dataset by examining different aspects (analysis of dataset accuracy and correctness) and finally c) an uploaded raw dataset and its further supplementation which requires a careful pre-analysis (Data preparation and supplementation).

- **Data traceability:** These use cases describe the process of, a) querying the CDMP for desired task-related datasets (searching the dataset in the platform) and b) discovering issues in a given dataset from the platform and reporting them to the dataset owner (detecting issues in a dataset).

- **AI model development:** These uses case describe the process of, a) an AI model by taking or requesting desired datasets from the platform and splitting them into training, validation, and testing sets (training of AI model and model registry), b) evaluating a trained AI model in the CDMP for a defined task (testing a trained AI model) and finally c) specifying testing procedures and datasets for AI models solving defined tasks in railway (testing data management and specification).

4. Common data management platform definition

This section details the most relevant points when defining a data management platform: the requirements that it has to meet, both functional and operational, and the main core modules description that comprise the CDMP (data acquisition/generation, labelling and training/testing).

4.1. High-level requirements overview

This section overviews the fundamental requirements that establish the scope, functionality and expected methods to utilize the platform. Functional requirements determine what the platform is supposed to do and the operational ones define how to build and/or run the system. In general, operational requirements can be also understood as those non-functional requirements that determine other aspects such as performance expectations or standards to be used. The identified high-level general requirements are:

- **General functional:** The platform shall be a container of data intended for its use in AI-related processes (training, re-training and testing) and editable by multiple users simultaneously. It shall enable Create, Read, Update, Delete (CRUD) operations, permit metadata to be contained, guarantee traceability, allow back-up/archiving options and also options to categorize/organize content according to use cases, domains or relevant tags.
- **General operational:** As a general rule, the platform shall enforce the utilisation of standard file formats for sensor data, metadata and annotations. In addition, the platform shall be deployable in any local or cloud environment, be accessible via programmatic interfaces and expose callback entry points for networking protocols. It shall also include authorisation mechanisms to determine the level of access of the users.

Moreover, application-related requirements focus on AI-related applications (training, re-training and testing applications that provides the platform) must be considered. Finally, the specific utilization of the platform for AI applications mandates the definition of content-related requirements that specify characteristics of the content itself.

- **Application-Related:** The platform shall contain all data and metadata needed to feed an AI-related application:
  - **Training:** The platform shall enable a neural network to be trained using a prepared dataset (data plus labels) to produce a model that can later on be used to predict labels on new data.
  - **re-Training:** The platform shall enable the model to be re-trained (or produce an updated version of the model) using existing models and new prepared datasets (using techniques such as data augmentation, filtering, grouping...) and with the ability to measure the gain in performance, quality or any other key Performance Indicator (KPI). All of this, making use of incremental learning strategies (without repeat the training processes of previous steps).
  - **Testing:** The platform shall provide mechanisms and tools to be able to design, define and execute AI model tests as well as to save and compare the results of each test for continuous testing.
  - **Visualization:** The platform shall provide test visualization mechanism and inspection routines to analyze information about the dataset or model, e.g., balance of labels, subset KPIs, statistics of re-trained models, analyze extracted features...
- **Content-Related:** The platform shall contain training datasets in the form of data (multi-sensor recordings) plus annotations, trained models and hyper-parameters to configure all application processes (e.g., learning rate, initial weights, batch size, etc.). The platform shall also contain functional scenario descriptions for scenario-based testing, logical and specific scenario descriptions and real-world routes which can be matched with scenario tags to perform real-world tests.

4.2. Main modules

Figure 2 shows a representation of all the logic modules that can be found in the platform proposed in this work. The following ones are considered as the most important core modules inside the whole platform.

4.2.1. Data acquisition and generation

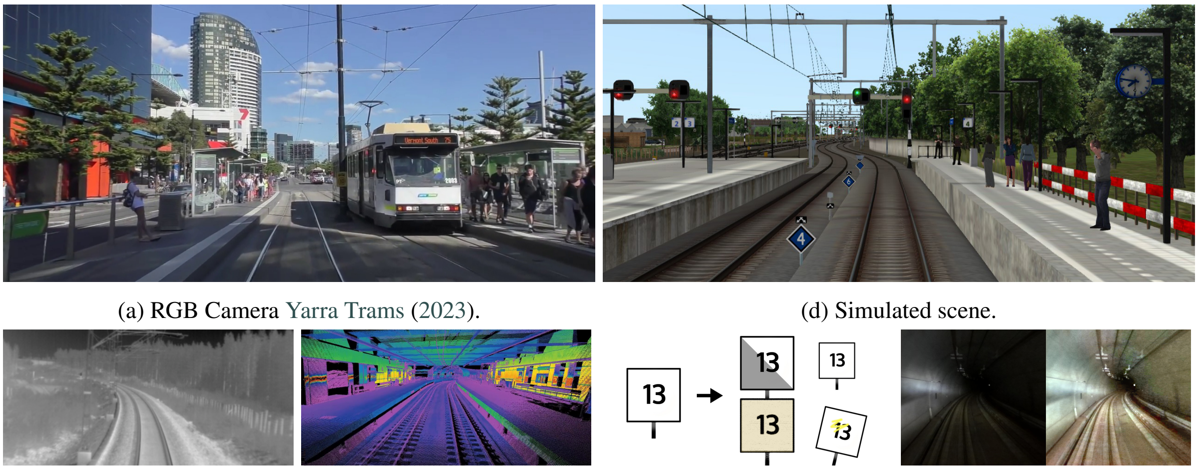
The data that will populate the datasets on the platform can be captured in a real environment or can also be generated in a synthetic environment. A common understanding in the AI community is that AI models are as good as the dataset used to train them. For this reason, data representing the expected Operational Design Domain (ODD) with high-fidelity, the situations under which the model is assumed to operate correctly, is key factor.

In order to create the most realistic dataset of the environment, the most traditional method is to acquire data adding sensing capacity to the train or infrastructure. The most traditional options use different types of sensors (Cameras, RADAR, LiDAR, IMU...). According to the functionality that the perception system wants to cover, there shall be a sensor or group of sensors that fulfil the requirements. Table 1 makes the relation between the correct sensor and the most relevant use-cases of the future autonomous train.

However, creating a dataset with real images and covering all operational conditions might be unaffordable, extremely expensive or very difficult to manage. In addition, assuming the recordings can be created, labelling them is usually the main bottleneck.

Table 1
Railway use-case Vs needed sensors.

Use-cases/sensor	RGB cam	IR cam	RADAR	LiDAR	Audio	Odom.
Obstacle detection	×	×	×	×		
Sign/signal recognition	×			×		
Switch & path monitoring	×		×	×		×
Train localization	×		×	×		
Ext. environment monitoring	×	×	×	×	×	
Infrastructure supervision	×			×		
Platform monitoring	×	×		×		
Rolling stock monitoring	×	×			×	×
Passenger's supervision	×	×		×	×	



(a) RGB Camera Yarra Trams (2023).

(d) Simulated scene.

(b) Thermal Ristic-Durrant et al. (2021).

(c) LiDAR.

(e) Augmentation.

(f) E-GAN.

Figure 1: Different data examples; a), b) and c) represents data acquired in real world using different sensor and at different railway environment; d), e) and f) are generated by simulators, data augmentation techniques and deep learning algorithms (E-GAN).

These limitations focus the attention on synthetic data generation, either by creating limited discrete samples (created from canonical images or from generative deep learning) or sequences creating a completely new scenario from a virtual simulator (from simulator engines). The first approach implies the utilization of data augmentation and DL techniques such as EnlightenGAN Jiang, Gong, Liu, Cheng, Fang, Shen, Yang, Zhou and Wang (2019) to increase the variability of the resulting dataset modifying seed examples and creating modifications (shadow, color, size, rotation, lightning...). The second method simulates an entire recording campaign, running a simulator engine which creates a virtual world where the elements of interest are represented naturally and thus shows all the required variability. Modern simulation engines (CARLA Dosovitskiy, Ros, Codevilla, López and Koltun (2017), LGSVL LGSVL (2023), Prescan Siemens (2023)) can be used via programmatic interfaces to produce virtual scenes with high-fidelity sensors, environments and behavioural models, etc. From these simulators, data can be gathered, manipulated and batch processed as desired.

4.2.2. Data semi-automatic labelling

The annotation files describe the organised rich description of the scene. These files contain labels for the scene's objects as well as other information such us sensor metadata, encoding schemes for various geometries, connections to ontologies, knowledge repositories, and other external resources. In order to build databases that are shareable and inter-operable, these labels, containing the relevant information, should be stored and organized using a common/standard format. On the other hand, the annotation criteria which comprise the guidelines to be followed while making the annotations in order to prevent the annotator's personal interpretations, should be also agreed in order to achieve the objective.

Although there are previous formats, such as JSON schema or Google Protocol Buffers, that allow comprehensible annotation both for computers and people, the automotive sector has been inclined to propose and define a standard (unique international standard for multi sensor labelling by now) for the raw content labelling for the training and testing of AI models. ASAM OpenLABELASAM (2023) proposes a univocal procedure to for classifying and describing the many elements/objects of the driving environment. Furthermore, it may be adapted to fit into the taxonomy requirements of a particular user or company as it does not define how to describe the real world (taxonomy). Because of this, OpenLABEL may be a reliable standard that applies to the railway industry.

4.2.3. Model training and testing

According to the specifications, the platform contains tools for training and testing AI models. Regarding training operations, the user will be able to submit his own code, establishing a specific and private architecture, or to choose the various well known state-of-the-art training methods available in the platform. On the other hand, regarding testing processes, the validation and testing module of the platform will be able to execute inference batches and compare, in an automatic way, the output data using the standard AI validation metrics or custom metrics established by each user. For both case the previously labelled data will be used as input data for training, and the ground truth for testing. The generated models (well tagged) could be stored in the same platform.

It is quite challenging to keep tracked of all this processes, especially if the user wants to compare different training and testing sessions and manage the built-in models throughout the different deployment phases. This platform will be able to warrant model traceability and continuous monitoring of the AI model performance in order to enable continuous integration and deployment (CI/CD) pipeline.

5. Data Management

Once the high-level requirements were gathered, it was possible to define the functional architecture of the CDMP, as shown in Figure 2. This architecture not only gives an idea about what components are the building blocks of the envisioned CDMP, but also about how data is expected to flow through them.

First of all, data are acquired by different kind of sensors (e.g., cameras) in the Data Acquisition Unit (DAU) (also compatible with the necessary parameters to boost synthetic-data-generation processes) and sent to the Data Anonymisation and Provisional Data Storage Unit (DA-PDSU), where they are made complaint with GDPR requirements and safely stored until they are requested and sent to the Data Labelling Unit (DLU). In the DLU, users can make annotations on objects and enrich the scenes by giving context or adding information about actions. The data labelling stage shall be made compliant with the latest ASAM initiatives, such as ASAM OpenLABEL and ASAM OpenScenario. After this, both the anonymised data and the created labels are sent to the Main Data Storage Unit (MDSU), where they are stored and assessed by means of the Data Analysis Unit (DAnU). From this point, authorised users have access to validated data through the Data Downloading Unit (DDU) under reasonable request.

Table 2
Open-source technologies proposed for each pipeline stage.

Process	**Technology**
Raw data ingestion	HDFS, HDF5
Data cleaning + enrichment	Spark, Hadoop
Model training and tracking	MLFlow, Comet, TensorBoard
Scene Detection (ASAM)	MongoDB + Elasticsearch,
Data Catalog	Neo4j, GraphDB
CI/CD	Gitlab CI/CD, Jenkins

Given the mechanisms available nowadays to automatise the whole AI development pipeline, it was decided to provide the platform with an additional, model-oriented functionality. This functionality starts in the Data Pre-Processing Unit (DPPU), where data are prepared for training according to the selected use case. This unit also splits the data into the training and validation datasets, which flow to the Model Development Unit (MDU), and a good, sterile testing dataset which flows to the Validation and Verification Unit (V&VU). After AI models are trained and optimised by means of the validation dataset, the best performing ones are sent to the V&VU, where they are evaluated against the test dataset. Models which are compliant with the V&V requirements are eventually transferred to the Model Registry (MR) and stored there so that authorised users can retrieve and utilise them.

Even if the certification process and the eventual model deployment and monitoring are not within the scope of this research work, they are represented in Figure 2 so that the CI/CD pipeline can be fully conceived. Table 2 contains some of the open-source technologies which could be utilised for some of the described steps.

6. Data protection and anonymisation

The collected data stem from different sensor types. Gathering data using cameras, audio or laser sensors is defined as a controlled activity by different governments and, consequently, these data are submitted to country-specific GDPR, especially where data collection includes personal information. In particular, the GDPR restricts transfer of personal data to countries outside of the EU that do not have an equivalent level of protection. For this reason, the common data management platform must provide specific tools for data protection and anonymisation depending on the specific country laws. Detecting and blurring faces or texts (i.e., car number plates) are some of the particular scenarios. Moreover, customers and contributors shall comply with the local and global data protection standards and all parties should follow secure methods of data transfer to contribute into the CDMP. Finally, the data management platform shall permit storage of data in geo-specific locations to comply with GDPR policies.

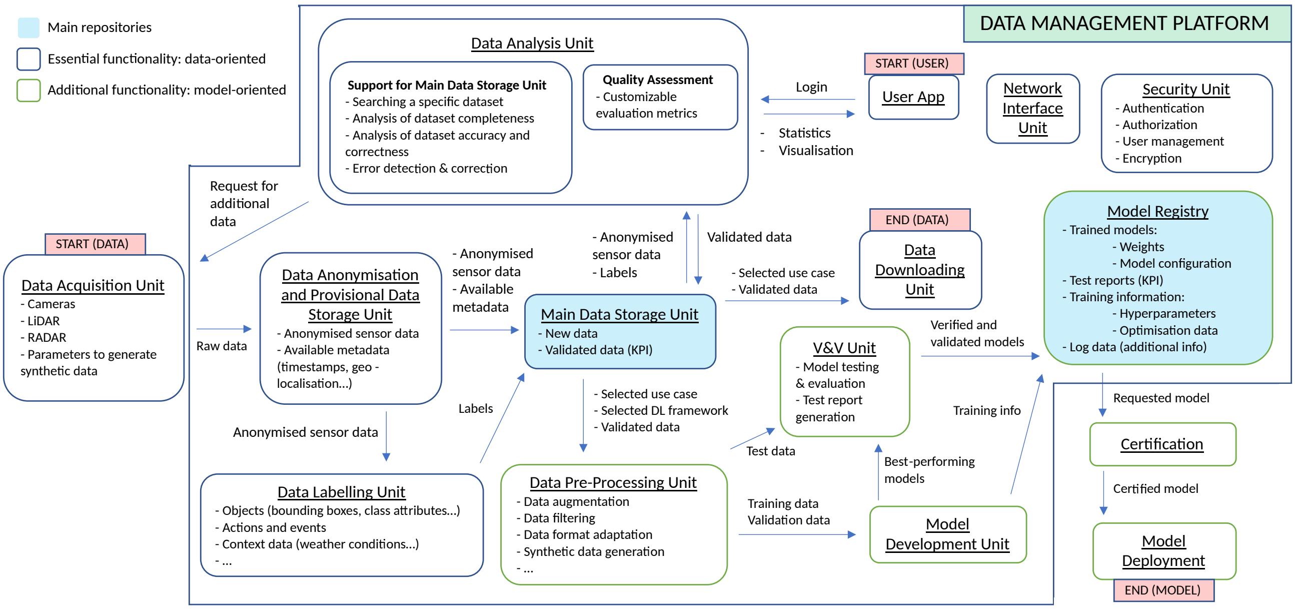


Figure 2: Functional architecture of the CDMP.

There are two main issues that should be taken into account when analyzing privacy when data is stored in the cloud. Data collected by sensors that may have contained personal information is one of them. This data must be managed according to GDPR regulations, inaccessible to unauthorised users, and adequately safeguarded against data theft. The second issue is about user privacy using the cloud. It must also be protected with solutions such as the one proposed by Malina, Hajny, Dzurenda and Zeman (2015), which is based on a non-bilinear group signature system, and can be used to provide anonymous authentication, where personal attributes can be proven without revealing the identity of the users.

7. Access and contribution policies

There is increasing awareness that without sharing of data, scientific research will become increasingly wasteful of resources. Research funding will be used for unnecessary duplication of work or the gathering of new data when existing data could be just as useful. This has led to a need for more effective sharing of data and samples.

At the same time, data must be treated, shared and transferred carefully. This section provides a template for access and contribution policy. This is a practical guideline, without intending to impose policy and practice. The template concerns access to data that have already been collected, and contribution of the data to the existing common dataset which is founded and contributed by multiple institutions.

The purpose of this policy is to preserve the confidentiality, integrity, and availability of the common dataset by restricting access to the parties which contributed into the creation of the common dataset. However further access can be granted if all contributed parties agree. These are the highlights of the Access and Contribution Policies:

- In general, the access should be made as widely available as is consistent with the consent. However, there may be occasions where it is necessary to limit access to certain groups.
- To ensure that best use is made of a limited resource, access to a collection may be limited to requestors affiliated to a recognized research institution; those with a satisfactory record in the field or those willing to pursue the research in collaboration with the contributor' group.
- Eligibility of the accessors should be examined and access to the data should be only provided to the eligible requestors.
- All contributors should have permission to read the data from CDMP.
- Only the authorized members should have permission to update the content of the platform.
- Permission to permanently remove any information should be only given to the founding group.
- The content should include authorization mechanisms to determine the level of access to certain users.
- Data access and contribution policies should ensure that data protection requirements are followed keeping personally identifiable information in data as confidential from unauthorized users.

8. PoC: Implementation and Case study

The proposed data management platform can be implemented with different flavors. On the one hand, the traditional on-premise solution has been joined by cloud and

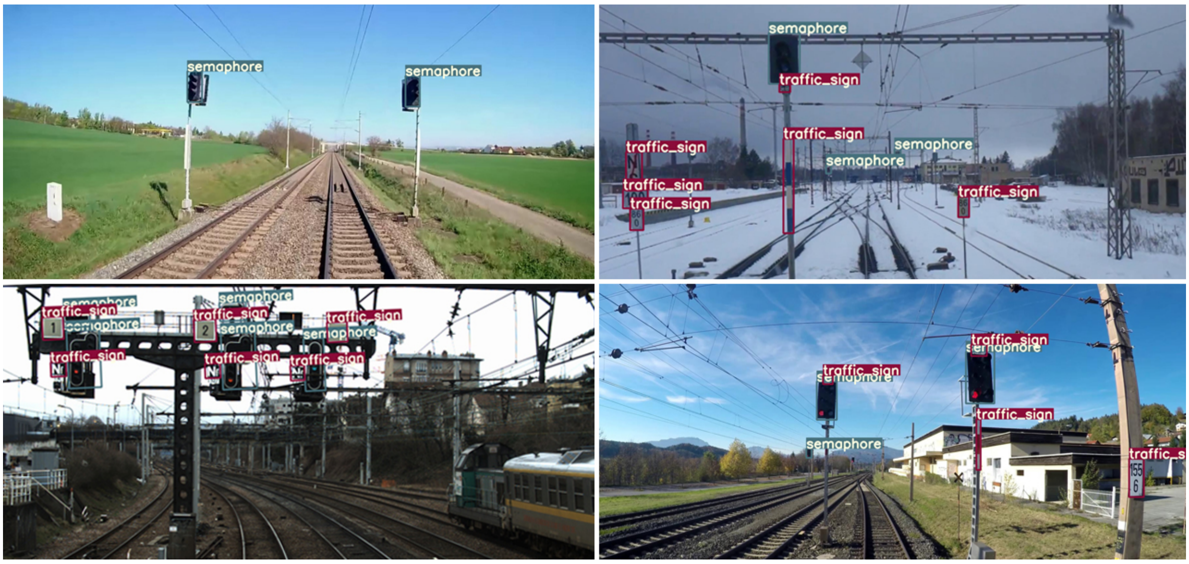


Figure 3: Railway signal & signs detection examples (FRSign database).

also hybrid options. On the other hand, the entire solution can be developed using different level implementations. For instance, low level implementations offer a deeper understanding of processes, greater control, and lower costs. However, they also require a certain level of knowledge and expertise in different areas, and their configuration can be time-consuming. In contrast, high-level solutions, such as fully managed ML services, are available for those who are not capable of building a proprietary methodology.

In order to validate the proposed solution, two real-use cases were addressed of detecting 1) railway traffic lights and signals and 2) switches using YOLO models. For both cases, the data are annotated RGB images and a specific sub-pipeline was built in batch mode using microservice technology (from data gathering to model training and testing stages). However, the implementation is different depending on the use case. Cloud and on-site infrastructures are used for the former and latter cases, respectively.

8.1. Use case 1 (cloud): railway traffic light and signal detection

In the first use case, the entire AI pipeline is performed on the AWS cloud. First, the training, validation and testing datasets were uploaded to the MDSU (S3 bucket). Secondly, the images were pre-processed in the DPPU: 1) filtered in order to obtain homogeneous traffic signal classes among the training, validation and test sets and 2) resized to fit the You Only Look Once (YOLO) Bochkovskiy, Wang and Liao (2020) model. A few thousand images were manually annotated using Video Content Description (VCD) format Nieto, Senderos and Otaegui (2021) in the DLU. Next, the DL training and inference phase models were packaged as docker images allocated on Amazon Elastic Container Registry (ECR) for posterior training and inference tasks in the MDU and V&VU (Amazon Sagemaker), achieving a mAP@0.5 of 34.9% (AP@0.5 of 44.6% for traffic lights and AP@0.5 of 25.2% for traffic signs). Finally, the trained model was stored in the MR (S3 bucket). Figure 3 shows inferences made by trained AI models.

The total cost for the training process (100 epochs) on the most expensive tested option (ml.p3.2xlarge) was less than 1.59$ for a total computation time of 1874 seconds. Depending on the necessity, the solution is easily scalable to the required instance. During the training process, we incurred a fixed cost of 60$month in terms of availability and maintenance of the AWS account and in additional costs of storage (docker images and image datasets), Virtual Private Cloud (VPC), maintenance and other costs that were negligible.

8.2. Use case 2 (on-site): railway switch detection

In the second use case, the entire AI pipeline was performed on-site. The objective of this use case was to detect railway switches and to determine whether they were open to the left or to the right based on the ego-perspective of a rail vehicle. Given that part of Railsem19's dataset is aligned with this purpose, this dataset was stored on a local MDSU. To complement Railsem19 and to showcase a local implementation of the DLU, some additional frames were extracted from a private dataset belonging to CAF Signalling. Switches appearing on them were labelled by means of the Computer Vision Annotation Tool (CVAT). The labelling process consisted on drawing bounding boxes around the identified switches and naming them according to the classes they belonged to: "switch-left", "switch-right" and "switch-unknown". After that, all these data (images and labels) were downloaded in YOLO format and stored locally in the MDSU.

It is worth highlighting here that it was intended to split this use case in two consecutive functionalities. The first of them would locate all possible switches on an image under the general "switch-all" class and the second one would

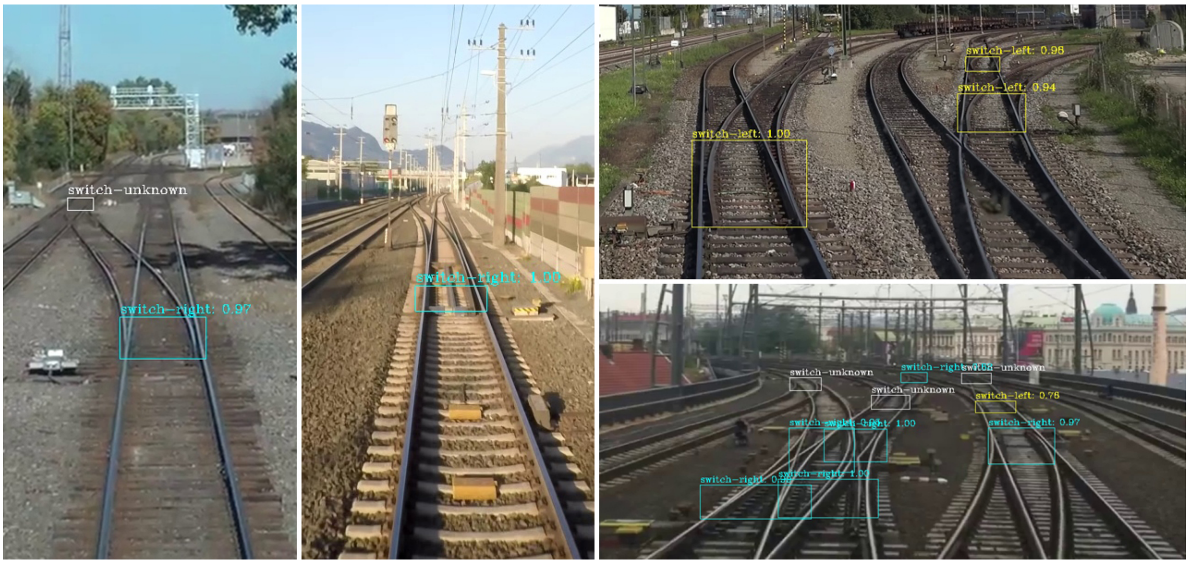


Figure 4: Railway switch detection examples (Railsem19 database).

classify the status of these switches. The main reason behind this decision was if the different switch classes were very similar, it would actually make possible to utilise visual features learnt from "switch-left" objects to locate "switch-right" objects and vice versa. Afterwards, a simpler DL model could be trained to differentiate between these two classes.

Considering this strategy, several data-preparing operations were made in the DPPU. First, Railsem19's data was filtered so that the switch-related data could be used for training. Then, Railsem19's labels were converted to YOLO format so that they could be merged with the self-annotated ones. After that, the data was prepared to train the two different kind of models needed: a YOLO model for the switch location functionality and a custom Convolutional Neural Network (CNN) for the switch classification functionality. For both kinds of models, some data augmentation operations were also performed.

Finally, the AI models were trained on a NVIDIA GeForce RTX 2080 Ti and optimised over PyTorch (switch detector) and TensorFlow (switch classifier) in the MDU and evaluated in the V&V Unit, achieving the switch detector a mAP of 38.8% and the switch classifier an accuracy of 76% on a well-balanced test set. Both models were stored locally in the MR. Figure 4 shows inferences made by trained AI models.

Comparison

After performing both experiments, we highlight the benefits of having an entire solution hosted and managed in the cloud compared with an on-site solution in terms of availability, shareability, scalability and security. In addition, a microservices-based solution encourages versatility, efficiency, and low maintenance and finally high-level frameworks offer managed tools for boosting traceability (i.e., tracking datasets and model lineage). A summary of the proposed pipelines are shown in Figure 5, considering both the cloud and the on-site approaches.

9. Conclusions and Future work

This work presents a description of a common data management platform designed for the European railway sector based on agreed requirements and specifications among the main stakeholders. This platform enables developing perception systems which are robust and safe enough to make autonomous rail operation possible. To this end, the main tasks reported in this work are: an analysis of the state-of-the-art databases from the automotive sector as a very valuable input; identification of the most relevant platform use cases; definition of the envisioned CDMP; consideration of other aspects of data management; the establishment of access and contribution policies, and the development of the first instances of the data management platform as PoC based on the addressed case studies, and has complied with the most critical requirements. As a result, this platform has shown the capability of performing the key steps of the entire AI pipeline: data ingestion, data filtering, data labeling and AI model training and testing phases.

Considering the results obtained in the PoC, the cloud solution can be concluded to be a better alternative compared to an on-site solution for the data management platform construction in terms of availability, shareability, scalability and maintainability. In addition, the microservices strategy leads to a language agnostic pipeline that accelerates deploying AI models and building the CI/CD pipelines.

Furthermore, although approaches like the one described in this work are becoming de-facto standard in the automotive sector, the railway sector is still evolving to adopt AI-centered methodologies. At the time of writing this article, there are only few remarkable open datasets related to AI for the railway domain. This work tries to bridge this gap,

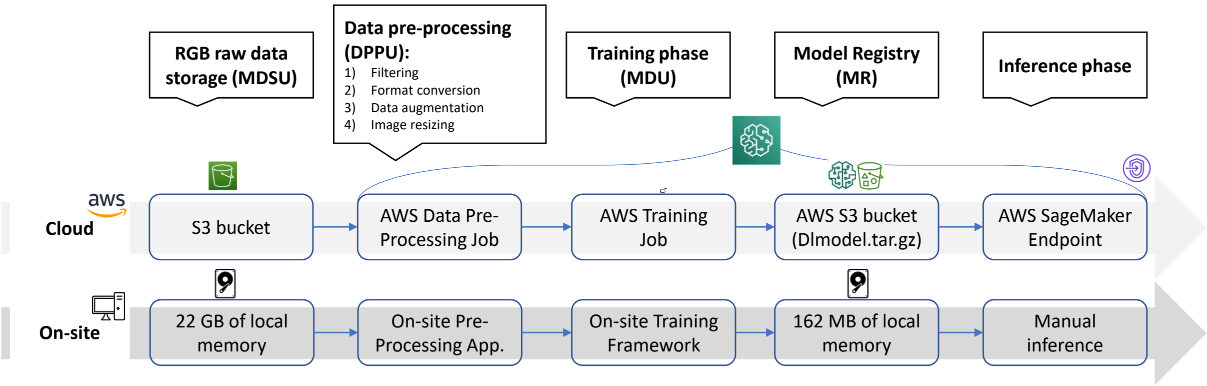


Figure 5: Proposed pipelines.

by designing and defining a common data management platform which could be better known as a common data platform.

The future iterations will eventually converge to the large-scale platform implementation and a scenario where all the interested parties in the EU share data and benefit from the data shared by others. They will also meet all the requirements and further discuss who will manage and host the platform, who will contribute, and who will be able to use it and under what conditions.

10. Acknowledgment

This work was developed as a specific task within the TAURO research project EU (2023c) inside the Shift2Rail framework. This project has received funding under grant agreement No 101014984.

11. Bibliography

References

Agarwal, S., Vora, A., Pandey, G., Williams, W., Kourous, H., McBride, J., 2020. Ford multi-AV seasonal dataset. The International Journal of Robotics Research 39, 1367–1376.

ASAM, 2023. ASAM OpenLABEL. https://www.asam.net/project-detail/asam-openlabel-v100/. Accessed 10-Aug-2023.

AVL, 2023. AVL SCENIUS. https://www.avl.com/-/scenius. Accessed 10-Aug-2023.

Bochkovskiy, A., Wang, C., Liao, H.M., 2020. Yolov4: Optimal speed and accuracy of object detection. CoRR abs/2004.10934. URL: https://arxiv.org/abs/2004.10934, arXiv:2004.10934.

Cabon, Y., Murray, N., Humenberger, M., 2020. Virtual kitti 2. arXiv:2001.10773.

Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuscenes: A multimodal dataset for autonomous driving, in: CVPR.

Cserep, M., 2022. Hungarian mls point clouds of railroad environment and annotated ground truth data. doi:10.17632/ccxpzhx9dj.1.

Dosovitskiy, A., Ros, G., Codevilla, F., López, A.M., Koltun, V., 2017. CARLA: an open urban driving simulator. CoRR abs/1711.03938. arXiv:1711.03938.

EU, 2023a. Data Act: Commission proposes measures for a fair and innovative data economy. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113. Accessed 10-Aug-2023.

EU, 2023b. data.europa.eu - The official portal for European data. https://data.europa.eu/. Accessed 10-Aug-2023.

EU, 2023c. TAURO. https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro. Accessed 10-Aug-2023.

Gaia-X, 2023. Gaia-X. https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html. Accessed 10-Aug-2023.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The kitti dataset. International Journal of Robotics Research (IJRR) .

Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A.S., Hauswald, L., Pham, V.H., Mühlegg, M., Dorn, S., Fernandez, T., Jänicke, M., Mirashi, S., Savani, C., Sturm, M., Vorobiov, O., Oelker, M., Garreis, S., Schuberth, P., 2020. A2d2: Audi autonomous driving dataset. doi:10.48550/ARXIV.2004.06320.

Harb, J., Rébéna, N., Chosidow, R., Roblin, G., Potarusov, R., Hajri, H., 2020. FRSign: A Large-Scale Traffic Light Dataset for Autonomous Trains. arXiv e-prints arXiv:2002.05665.

IDS, 2023. International Data Spaces. https://internationaldataspaces.org/. Accessed 10-Aug-2023.

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z., 2019. Enlightengan: Deep light enhancement without paired supervision. CoRR abs/1906.06972. URL: http://arxiv.org/abs/1906.06972, arXiv:1906.06972.

Kesten, R., Usman, M., Houston, J., Pandya, T., Nadhamuni, K., Ferreira, A., Yuan, M., Low, B., Jain, A., Ondruska, P., Omari, S., Shah, S., Kulkarni, A., Kazakova, A., Tao, C., Platinsky, L., Jiang, W., Shet, V., 2019. Level 5 perception dataset 2020.

Lamas, D., Soilán, M., Grandío, J., Riveiro, B., 2021. Automatic point cloud semantic segmentation of complex railway environments. Remote Sensing 13. doi:10.3390/rs13122332.

LGSVL, 2023. SVL Simulator by LG. https://www.svlsimulator.com/. Accessed 10-Aug-2023.

Malina, L., Hajny, J., Dzurenda, P., Zeman, V., 2015. Privacy-preserving security solution for cloud services. Journal of Applied Research and Technology. JART .

MLO, 2023. Machine Learning Operations. https://ml-ops.org/. Accessed 10-Aug-2023.

Nieto, M., Senderos, O., Otaegui, O., 2021. Boosting ai applications: Labeling format for complex datasets. SoftwareX 13, 100653.

Ortega, J.D., Kose, N., Cañas, P., Chao, M.A., Unnervik, A., Nieto, M., Otaegui, O., Salgado, L., 2020. Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis, in: Bartoli, A., Fusiello, A. (Eds.), Computer Vision – ECCV 2020 Workshops, Springer International Publishing. pp. 387–405. doi:10.1007/978-3-030-66823-5_23.

PEGASUS, 2023. PEGASUS. https://www.pegasusprojekt.de. Accessed 10-Aug-2023.

Ristic-Durrant, D., Franke, M., Michels, K., 2021. A review of vision-based on-board obstacle detection and distance estimation in railways. Sensors

21, 3452. doi:10.3390/s21103452.

Safety Pool, 2023. The global Initiative for certifiable AV Safety. https://www.safetypool.ai/. Accessed 10-Aug-2023.

Siemens, 2023. Simcenter Prescan. https://www.plm.automation.siemens.com/global/en/products/simcenter/prescan.html. Accessed 10-Aug-2023.

Streetwise, 2023. Accelerating Automated Driving with advanced scenario-based safety validation. https://www.tno.nl/en/digital/smart-traffic-transport/smart-vehicles/streetwise/. Accessed 10-Aug-2023.

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, Z., Anguelov, D., 2020. Scalability in perception for autonomous driving: Waymo open dataset, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA. pp. 2443–2451. doi:10.1109/CVPR42600.2020.00252.

VVM, 2023. ADScene, towards an industrial scenarios plateform for Driving Assistance Systems design & validation. https://www.vvm-projekt.de/fileadmin/user_upload/Mid-Term/Presentations/VVM_HZE_EmmanuelArnoux.pdf. Accessed 10-Aug-2023.

Wang, P., Huang, X., Cheng, X., Zhou, D., Geng, Q., Yang, R., 2019. The apolloscape open dataset for autonomous driving and its application. IEEE transactions on pattern analysis and machine intelligence .

Yarra Trams, 2023. MELBOURNE TRAM DRIVERS VIEW. https://www.youtube.com/watch?v=lMx1Bx2Ei08&ab_channel=Schony747. Accessed 10-Aug-2023.

Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Chennupati, S., Uricar, M., Milz, S., Simon, M., Amende, K., Witt, C., Rashed, H., Nayak, S., Mansoor, S., Varley, P., Perrotton, X., Odea, D., Pérez, P., 2019. Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9307–9317. doi:10.1109/ICCV.2019.00940.

Yu, X., He, W., Qian, X., Yang, Y., Zhang, T., Ou, L., 2022. Real-time rail recognition based on 3d point clouds. Measurement Science and Technology 33, 105207. doi:10.1088/1361-6501/ac750c.

Yuan, H., Mei, Z., Chen, Y., Niu, W., Wu, C., 2022. Railvid: A dataset for rail environment semantic, in: 17th International Conference on Systems, ICONS.

Zendel, O., Murschitz, M., Zeilinger, M., Steininger, D., Abbasi, S., Beleznai, C., 2019. Railsem19: A dataset for semantic rail scene understanding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

4.7 Otras publicaciones

A continuación, se listan otros artículos de menor relevancia, que aunque con aportaciones más nobles han contribuido a la consecución de los objetivos que recoge esta tesis.

4.7.1 Calibration Accuracy Measurement in Railway Overlapping Multi-Camera Systems

Martí Sánchez and Nerea Aranjuelo and Jon Ander Iñiguez de Gordoa and Pablo Alonso and Mikel García and Marcos Nieto and **Mikel Labayen** (Bajo corrección después de la primera revisión)

Relación con los objetivos de la tesis: Obj3.1

4.7.2 Common Data Management Platform for Artificial Sense Training and Testing for Railway Applications

Mikel Labayen and Daniel Ochoa de Eribe and Ander Aramburu and Marcos Nieto and Naiara Aginako, *Transport Research Procedia, 2023, Elsevier*

Relación con los objetivos de la tesis: Obj3.2

4.7.3 Safety-Critical High-Performance Computing Platforms for CV&AI-enhanced Autonomous Train Operation

Mikel Labayen and Carles Hernández and Fernando Eizaguirre and Naiara Aginako, *World Congress on Railway Research, 2022, SPARK - RSSB*

Relación con los objetivos de la tesis: Obj3.4

4.7.4 The VALU3S ECSEL project: Verification and Validation of Automated Systems Safety and Security

J.A. Agirre and L. Etxeberria and R. Barbosa and S. Basagiannis and G. Giantamidis and T. Bauer and E. Ferrari and **M. Labayen Esnaola** and V. Orani and J. Öberg and D. Pereira and J. Proença and R. Schlick and A. Smrcka and W. Tiberti and S. Tonetta and M. Bozzano and A. Yazici and B. Sangchoolie, *Microprocessors and Microsystems, 87, 104349, 2021, Elsevier*

Journal Quartile: Q2
Relación con los objetivos de la tesis: Obj3.3

4.7.5 Monocular Visual Odometry for Underground Railway Scenarios

Mikel Etxeberria-Garcia and **Mikel Labayen** and Fernando Eizaguirre and Maider Zamalloa and Nestor Arana-Arexolaleiba, *Fifteenth International Conference on Quality Control by Artificial Vision, 11794, 1-8, 2021, SPIE*

Relación con los objetivos de la tesis: Obj3.1

4.7.6 SELENE: Self-Monitored Dependable Platform for High-Performance Safety-Critical Systems

Carles Hernàndez and Jose Flieh and Roberto Paredes and Charles-Alexis Lefebvre and Imanol Allende and Jaume Abella and David Trillin and Martin Rönnbäck and Johan Klockars and Nicholas Mc Guire and Franz Rammerstorfer and Christian Schwarzl and Franck Wartet and Dierk Lüdemann and **Mikel Labayen**, *2020 23rd Euromicro Conference on Digital System Design (DSD), 370-377, 2020, IEEE*

Relación con los objetivos de la tesis: Obj3.4

4.7.7 The VALU3S ECSEL Project: Verification and Validation of Automated Systems Safety and Security

R. Barbosa and S. Basagiannis and G. Giantamidis and H. Becker and E. Ferrari and J. Jahic and A. Kanak and **M. Labayen Esnaola** and V. Orani and D. Pereira and L. Pomante and R. Schlick and A. Smrcka and A. Yazici and P. Folkesson and B. Sangchoolie, *2020 23rd Euromicro Conference on Digital System Design (DSD), 352-359, 2020, IEEE*

Relación con los objetivos de la tesis: Obj3.3

4.7.8 The ECSEL FRACTAL Project: A Cognitive Fractal and Secure EDGE based on a unique Open-Safe-Reliable-Low Power Hardware Platform Node

Aizea Lojo and Leire Rubio and Jesus Miguel Ruano and Tania Di Mascio and Luigi Pomante and Enrico Ferrari and Ignacio Garcìa Vega and Frank K. Gürkaynak and **Mikel Labayen Esnaola** and Vanessa Orani and Jaume Abella, *2020 23rd Euromicro Conference on Digital System Design (DSD), 393-400, 2020, IEEE*

Relación con objetivos de la tesis: Obj3.4

4.7.9 Application of Computer Vision and Deep Learning in the Railway Domain for Autonomous Train Stop Operation

M. Etxeberria-Garcia and **Mikel Labayen** and M. Zamalloa and N. Arana-Arexolaleiba and, *2020 IEEE/SICE International Symposium on System Integration (SII), 943-948, 2020, IEEE*

Relación con los objetivos de la tesis: Obj3.1

4.7.10 e-proctoring en e-learning: el sistema SMOWL

Mikel Labayen and Manu Fraile and Alfonso Giménez, *Tecnología en las aulas, 140, 2018, Novática*

Relación con los objetivos de la tesis: Obj2.1, Obj2.2

4.7.11 Machine Learning for Video Action Recognition: a Computer Vision Approach

Mikel Labayen and Naiara Aginako and Basilio Sierra and Igor García and Julián Flórez, *SITIS 2018 – The 14th International Conference on Signal Image Technology & Internet based Systems, 683-690, 2019, IEEE*

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

4.7.12 La Biometría como Respuesta a la Necesidad de Autenticación del Alumno Online

Ricardo Vea and **Mikel Labayen**, *La innovación educativa como agente de transformación digital en la Educación Superior. Acciones para el cambio, 240, 111-116, 2017, DYKINSON*

Relación con los objetivos de la tesis: Obj2.1, Obj2.2

4.7.13 A new verification system for “VERIFIED CERTIFICATE MOOCS”

Mikel Labayen and Ricardo Vea and Julián Flórez, *9th International Technology, Education and Development Conference, 1469-1472, 2015, IATED*

Relación con los objetivos de la tesis: Obj2.1, Obj2.2

4.7.14 SMOWL: a Tool for Continuous Student Validation based on Face Recognition for Online Learning

Mikel Labayen and Ricardo Vea and Julián Flórez and Francisco D. Guillén-Gámez and Iván García-Magariño, *6th International Conference on Education and New Learning Technologies, 5354-5359, 2014, IATED*

Relación con los objetivos de la tesis: Obj2.1, Obj2.2

4.7.15 Depth Map based Object Tracking and 3D Positioning for Non-Static Camera

Mikel Labayen and Julen García and Aritz Legarretaetxebarria and Maider Laka, *authors, 10th Signal Processing, Pattern Recognitio and Applications, 798, 2013, IATED - ACTA Press*

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

4.7.16 Image Analysis Platform for Data Management in the Meteorological Domain

Igor García and Naiara Aginako and **Mikel Labayen**, *9th International Workshop on Semantic Media Adaptation and Personalization, 89-94, 2009, IEEE*

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

4.7.17 Visual Processing of Geographic and Environmental Information in the Basque Country: Two Basque Case Studies

Álvaro Segura and Aitor Moreno and Igor García and Naiara Aginako and **Mikel Labayen** and Jorge Posada and Jose Antonio Aranda and Rubén García De Andoin, *GeoSpatial Visual Analytics - Part of the NATO Science for Peace and Security Series C: Environmental Security book series (NAPSC), 199-207, 2009, Springer Netherlands*

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

4.7.18 Weather Analysis System Based on Sky Images Taken from the Earth

Mikel Labayen and Naiara Aginako and Igor García, *The fifth International Conference on Visual Information Engineering, 146-151, 2008, IET*

Relación con los objetivos de la tesis: Obj1.1, Obj1.2

5 Patentes

5.1 Method of Detection and Recognition of Logos in a Video Data Stream

- **Inventores:** García Olaizola Igor and Aginako Bengoa Naiara and **Labayen Esnaola Mikel**
- **Número de patente:** EP2259207 (B1)
- **Fecha de expedición:** 12.02.2013
- **Estado:** Concedida
- **Tipo:** Internacional
- **Descripción:** The invention relates to a method of detection and recognition of logos in a video data stream comprising the steps of sampling frames of said video data stream; segmenting regular shapes such as, for example, circles, ellipses and rectangles; generating a vector of feature parameters of an image contained in each of said shapes; and comparing said feature parameters with a database for determining whether the images correspond to logos. The frames are captured preferably using a sampling frequency which is dynamically adapted to processing times, for the purpose of allowing the system to work in real time.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 2 259 207 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
17.10.2012 Bulletin 2012/42

(51) Int Cl.:
G06K 9/32 *(2006.01)*

(21) Application number: **09382086.8**

(22) Date of filing: **02.06.2009**

(54) **Method of detection and recognition of logos in a video data stream**

Verfahren zur Erfassung und Erkennung von Logos in einem Videodatenstrom

Procédé de détection et de reconnaissance de logos dans un flux de données vidéo

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK TR

(43) Date of publication of application:
08.12.2010 Bulletin 2010/49

(73) Proprietor: **Vicomtech-Visual Interaction and Communication Technologies Center**
20009 San Sebastian (ES)

(72) Inventors:
- **Garcia Olaizola, Igor**
  20009, SAN SEBASTIAN (ES)
- **Aginako Bengoa, Naiara**
  20009, SAN SEBASTIAN (ES)
- **Labayen Esnaola, Mikel**
  20009, SAN SEBASTIAN (ES)

(74) Representative: **Carpintero Lopez, Francisco et al**
Herrero & Asociados, S.L.
Alcalá 35
28014 Madrid (ES)

(56) References cited:
WO-A1-03/043311 WO-A2-2008/107112
US-A1- 2006 034 484

- **BALLAN L ET AL: "Automatic detection of advertising trademarks in sport video" 4TH ITALIAN RESEARCH CONFERENCE ON DIGITAL LIBRARY SYSTEMS, XX, [Online] 24 January 2008 (2008-01-24), pages 83-88, XP008128257 Retrieved from the Internet: URL:http://canto.cab.unipd.it:8090/plone/i rcdl2008/papers/Automatic_detection.pdf> [retrieved on 2008-01-24]**
- **BOHUMIL KOVAR AND ALAN HANJALIC: "Logo detection and classification in a sport video: video indexing for sponsorship revenue control" STORAGE AND RETRIEVAL FOR MEDIA DATABASES 2002, vol. 4676, no. 1, 2001, pages 183-193, XP002606037 San Jose, CA, USA DOI: 10.1117/12.451090**
- **ARJUN JAIN ET AL: "A Novel Dynamic Rate Based System for Matching and Retrieval in Images and Video Sequences" COMPUTER SCIENCE AND SOFTWARE ENGINEERING, 2008 INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 12 December 2008 (2008-12-12), pages 556-560, XP031377794 ISBN: 978-0-7695-3336-0**
- **ALBIOL A ET AL: "Detection of tv commercials" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2004. PROCEEDINGS. (ICASSP ' 04). IEEE INTERNATIONAL CONFERENCE ON MONTREAL, QUEBEC, CANADA 17-21 MAY 2004, PISCATAWAY, NJ, USA,IEEE, PISCATAWAY, NJ, USA LNKD- DOI:10.1109/ICASSP.2004.1326601, vol. 3, 17 May 2004 (2004-05-17), pages 541-544, XP010718246 ISBN: 978-0-7803-8484-2**
- **DOERMANN D ET AL: "APPLYING ALGEBRAIC AND DIFFERENTIAL INVARIANTS FOR LOGO RECOGNITION" MACHINE VISION AND APPLICATIONS, SPRINGER VERLAG, DE, vol. 9, no. 2, 1 January 1996 (1996-01-01) , pages 73-86, XP000199795 ISSN: 0932-8092**
- **LIENHART R ED - ROSENFELD A ET AL: "VIDEO OCR: A SURVEY AND PRACTITIONER'S GUIDE" 1 January 2003 (2003-01-01), VIDEO MINING; [KLUWER INTERNATIONAL SERIES IN VIDEO VIDEO COUMPUTING], NORWELL, MA : KLUWER ACADEMIC PUBL, US, PAGE(S) 155 - 184 , XP009046500 ISBN: 978-1-4020-7549-0 * Section 6.1.3; page 167 - page 171; figure 6.6 ***

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

5.2 Method for Detecting the Point of Impact of a Ball in Sports Events

- **Inventores:** García Olaizola Igor and Flórez Esnal Julián and San Román Otegui Juan Carlos and Aginako Bengoa Naiara and **Labayen Esnaola Mikel**
- **Número de patente:** EP2455911 (B1)
- **Fecha de expedición:** 08.05.2013
- **Estado:** Concedida
- **Tipo:** Internacional
- **Descripción:** The invention relates to a method for determining the point of impact of a ball in a playing field during a controversial piece of play in a sports event and comprises the steps of: recording the contentious area during the game by means of a single camera, extracting the images corresponding to the controversial piece of play, selecting the area corresponding to the ball, calculating the coordinates of the ball in pixels in each image, determining the point of intersection of the two straight lines joining the previous points and transforming the point of intersection into real coordinates. As a result of these steps it is possible to resolve the controversial piece of play with a single camera and without the aid of accessories such as radar signals, electric signals, etc.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 455 911 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
13.03.2013 Bulletin 2013/11

(51) Int Cl.:
G06T 7/00 (2006.01) *A63B 71/06* (2006.01)

(21) Application number: **10382310.0**

(22) Date of filing: **23.11.2010**

(54) Method for detecting the point of impact of a ball in sports events

Verfahren zur Erkennung der Aufprallstelle eines Balls bei Sportveranstaltungen

Procédé de détection de point d'impact d'un ballon dans les événements sportifs

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(43) Date of publication of application:
23.05.2012 Bulletin 2012/21

(73) Proprietor: **FUNDACION CENTRO DE TECNOLOGIAS DE INTERACCION VISUAL Y COMUNICACIONES VICOMTECH
20009 San Sebastian (ES)**

(72) Inventors:
- **García Olaizola, Igor**
  20009, San Sebastian (ES)
- **Flórez Esnal, Julián**
  20009, San Sebastian (ES)
- **San Román Otegui, Juan, Carlos**
  20600 Eibar (Guipúzcoa) (ES)
- **Aginako Bengoa, Naiara**
  20009, San Sebastian (ES)
- **Labayen Esnaola, Mikel**
  20009, San Sebastian (ES)

(74) Representative: **Carpintero Lopez, Francisco et al
Herrero & Asociados, S.L.
Alcalá 35
28014 Madrid (ES)**

(56) References cited:

DE-A1- 19 954 504	**GB-A- 2 403 362**
US-A- 5 489 886	**US-A1- 2009 067 670**

Printed by Jouve, 75001 PARIS (FR)

5.3 Method and System for Verifying the Identity of a User of an Online Service

- **Inventores:** Vea Orte Ricardo and **Labayen Esnaola Mikel** and Flórez Esnal Julián and Marcos Ortego Gorka
- **Número de patente:** EP3005639 (B1) ↗
- **Fecha de expedición:** 11.12.2017
- **Estado:** Concedida
- **Tipo:** Internacional
- **Descripción:** The invention relates to a method for verifying the identity of a user of an online service, with the steps of: when a user is connected to an online service, sending an IP address of an authentication server; connecting to said IP address and downloading one application for taking photos with the webcam of the user terminal; taking a photo; sending said photo and associated metadata to a management unit; storing it in a data base; automatically extracting one set of biometrical parameters per each face which appears in said photo; comparing said set of biometrical parameters with a reference biometrical model of the user to which said user ID belongs; if the result of said comparison does not unequivocally match the person in the photo with the user to which said user ID belongs, either informing the web service provider or sending said photo to a manual recognition unit for manual validation of the photo; continuously verifying the identity of the user connected to the online service through said user terminal. System and computer program product.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 3 005 639 B1**

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
09.08.2017 Bulletin 2017/32

(21) Application number: **13730502.5**

(22) Date of filing: **04.06.2013**

(51) Int Cl.:
H04L 29/06 (2006.01) ***G06F 21/32*** (2013.01)
H04L 9/32 (2006.01) ***G06K 9/00*** (2006.01)
G06Q 20/40 (2012.01) ***G07C 9/00*** (2006.01)

(86) International application number:
PCT/EP2013/061521

(87) International publication number:
WO 2014/194939 (11.12.2014 Gazette 2014/50)

(54) METHOD AND SYSTEM FOR VERIFYING THE IDENTITY OF A USER OF AN ONLINE SERVICE

VERFAHREN UND SYSTEM ZUM PRÜFEN DER IDENTITÄT EINES BENUTZERS EINES ONLINE-DIENSTES

PROCÉDÉ ET SYSTÈME DE VÉRIFICATION DE L'IDENTITÉ D'UN UTILISATEUR D'UN SERVICE EN LIGNE

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(43) Date of publication of application:
13.04.2016 Bulletin 2016/15

(73) Proprietor: **Smiley Owl Tech S.L.**
20018 Donostia (San Sebastián) (ES)

(72) Inventors:
- **VEA ORTE, Ricardo**
  E-20018 Donostia (San Sebastián) (ES)
- **LABAYEN ESNAOLA, Mikel**
  E-20018 Donostia (San Sebastián) (ES)
- **FLOREZ ESNAL, Julián**
  E-20009 Donostia (San Sebastián) (ES)
- **MARCOS ORTEGO, Gorka**
  E-20009 Donostia (San Sebastián) (ES)

(74) Representative: **Balder IP Law, S.L.**
Paseo de la Castellana 93
5a planta
28046 Madrid (ES)

(56) References cited:
US-A1- 2012 106 805 US-B1- 7 991 388

- **MOINI A ET AL: "Leveraging Biometrics for User Authentication in Online Learning: A Systems Perspective", IEEE SYSTEMS JOURNAL, IEEE, US, vol. 3, no. 4, 31 December 2009 (2009-12-31), pages 469-476, XP011327886, ISSN: 1932-8184, DOI: 10.1109/JSYST.2009.2038957**

Printed by Jouve, 75001 PARIS (FR)

5.4 Method and System for Verifying the Identity of a User of An Online Service Using Multi-Biometric Data

- **Inventores: Labayen Esnaola Mikel** and Vea Orte Ricardo and Fraile Yarza Manuel
- **Número de patente:** EP3572961 (B1)
- **Fecha de publicación:** 15.06.2022
- **Estado:** Concedida
- **Tipo:** Internacional
- **Descripción:** The invention relates to a method for verifying the identity of a user of an online service, comprising: when a user is connected to an online service from a user terminal, establishing a connection with a biometric data collecting module; downloading one application for taking photos with the webcam of the user terminal and at least one application for capturing audio or for extracting keystroke pattern; while a session with the online service is active: taking a photo of the user terminal; sending each photo and associated metadata to a management module; automatically extracting features of each face which appears in said photo; comparing said features with a biometrical model of a reference photo of the user; repeating the former steps, thus continuously verifying the identity of the user connected to the online service; using a second biometrical technique for verifying the user identity, said second biometric technique being either sound recognition or keystroke pattern analysis. If voice is detected or a keystroke pattern is extracted, voice/keystroke features are extracted and compared with a reference model. If the result of a comparison does not unequivocally match the person in the photo/audio clip/keystroke pattern with the registered user, either informing the service provider or sending the photo/audio clip for manual validation. Besides, if sound recognition is used and the detected sound is voice, another photo may be automatically taken and sent to the management module for features extraction and comparison with a biometrical model of a reference photo of the user. Similarly, if keystroke pattern recognition is used, another photo may be taken every time a keystroke pattern is detected.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 3 572 961 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
27.04.2022 Bulletin 2022/17

(21) Application number: **18382363.2**

(22) Date of filing: **25.05.2018**

(51) International Patent Classification (IPC):
G06F 21/31 $^{(2013.01)}$ ***G06F 21/32*** $^{(2013.01)}$
H04L 29/06 $^{(2006.01)}$

(52) Cooperative Patent Classification (CPC):
H04L 63/0861; G06F 21/316; G06F 21/32;
G06F 2221/2139

(54) METHOD AND SYSTEM FOR CONTINUOUS VERIFICATION OF USER IDENTITY IN AN ONLINE SERVICE USING MULTI-BIOMETRIC DATA

VERFAHREN UND SYSTEM ZUR KONTINUIERLICHEN ÜBERPRÜFUNG DER BENUTZERIDENTITÄT IN EINEM ONLINE-DIENST UNTER VERWENDUNG VON MULTIBIOMETRISCHEN DATEN

PROCÉDÉ ET SYSTÈME DE VÉRIFICATION CONTINUE DE L'IDENTITÉ D'UTILISATEUR DANS UN SERVICE EN LIGNE À L'AIDE DE DONNÉES MULTIBIOMÉTRIQUES

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(43) Date of publication of application:
27.11.2019 Bulletin 2019/48

(73) Proprietor: **Smiley Owl Tech S.L.**
20018 Donostia (San Sebastián) (ES)

(72) Inventors:
- **LABAYEN ESNAOLA, Mikel**
  20018 Donostia (Gipuzkoa) (ES)
- **VEA ORTE, Ricardo**
  20018 Donostia (Gipuzkoa) (ES)
- **FRAILE YARZA, Manuel**
  20018 Donostia (Gipuzkoa) (ES)

(74) Representative: **Balder IP Law, S.L.**
Paseo de la Castellana 93
5ª planta
28046 Madrid (ES)

(56) References cited:

EP-A1- 2 645 275	**WO-A1-2016/065405**
US-A1- 2013 036 458	**US-A1- 2014 282 868**
US-A1- 2016 127 363	

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Processed by Luminess, 75001 PARIS (FR)

Bibliografía

[AB19] H. S. G. Asep e Y. Bandung. “A Design of Continuous User Verification for Online Exam Proctoring on M-Learning”. En: *2019 International Conference on Electrical Engineering and Informatics (ICEEI)*. 2019, págs. 284-289 (vid. pág. 40).

[ABD10] Hatem Alismail, Brett Browning y M Bernardine Dias. “Evaluating pose estimation methods for stereo visual odometry on robots”. En: *the 11th Int’l Conf. on Intelligent Autonomous Systems (IAS-11)*. Vol. 3. 2010, pág. 2 (vid. pág. 56).

[Agh+21] Ali Agha, Kyohei Otsu, Benjamin Morrell et al. “Nebula: Quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge”. En: “arXiv preprint arXiv:2103.11470” (2021) (vid. pág. 47).

[Ahn04] S.J. Ahn. *Least squares orthogonal distance fitting of curves and surfaces in space*. Lecture notes in computer science. Springer, 2004 (vid. pág. 36).

[Ali+23] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh et al. “Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence”. En: “Information Fusion” 99 (2023), pág. 101805 (vid. pág. 7).

[AMA20] S. Abd Razak, N. H. Mohd Nazari y A. Al-Dhaqm. “Data Anonymization Using Pseudonym System to Preserve Data Privacy”. En: “IEEE Access” 8 (2020), págs. 43256-43264 (vid. pág. 42).

[Ans+21] Jon Zubieta Ansorregi, Mikel Etxeberria Garcia, Maider Zamalloa Akizu y Nestor Arana Arexolaleiba. “Image Enhancement using GANs for Monocular Visual Odometry”. En: *2021 IEEE International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics (ECMSM)*. IEEE. 2021, págs. 1-6 (vid. pág. 48).

[Atk+18] Christopher G Atkeson, PW Babu Benzun, Nandan Banerjee et al. „Achieving reliable humanoid robot operations in the DARPA robotics challenge: team WPI-CMU’s approach“. En: *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*. Springer, 2018, págs. 271-307 (vid. pág. 47).

[Ato+17] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu y X. Liu. “Automated Online Exam Proctoring”. En: “IEEE Transactions on Multimedia” 19.7 (2017), págs. 1609-1624 (vid. pág. 40).

[Aut23] IPG Automotive. *CarMaker*. 2023. URL: `https://ipg-automotive.com/en/products-solutions/software/carmaker/` (visitado 1 de ene. de 2023) (vid. pág. 59).

[AZA19] Abdullah Alshbtat, Dr.Nabeel Zanoon y Mohammad Alfraheed. “A Novel Secure Fingerprint-based Authentication System for Student’s Examination System”. En: “International Journal of Advanced Computer Science and Applications” 10 (ene. de 2019) (vid. pág. 40).

[BMG14] José-Luis Blanco, Francisco-Angel Moreno y Javier Gonzalez-Jimenez. “The Málaga Urban Dataset: High-rate Stereo and Lidars in a realistic urban scenario”. En: “International Journal of Robotics Research” 33.2 (2014), págs. 207-214 (vid. págs. 49, 56).

[BSC23] BSC. *Ensuring software timing behavior in critical multicore-based embedded systems*. https://www.embedded.com/ensuring-software-timing-behavior-in-critical-multicore-based-embedded-systems/. 2023 (vid. pág. 74).

[BT23] Amin Biglari y Wei Tang. “A Review of Embedded Machine Learning Based on Hardware, Application, and Sensing Scheme”. En: “Sensors” 23.4 (2023) (vid. pág. 7).

[Bur+16] Michael Burri, Janosch Nikolic, Pascal Gohl et al. “The EuRoC micro aerial vehicle datasets”. En: “The International Journal of Robotics Research” 35.10 (2016), págs. 1157-1163 (vid. págs. 47, 56).

[Cai+21] Yingfeng Cai, Tianyu Luan, Hongbo Gao et al. “YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving”. En: “IEEE Transactions on Instrumentation and Measurement” 70 (2021), págs. 1-13 (vid. pág. 58).

[Can+21] Michele Cancilla, Laura Canalini, Federico Bolelli et al. “The DeepHealth Toolkit: A Unified Framework to Boost Biomedical Applications”. En: *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021, págs. 9881-9888 (vid. pág. 69).

[Cer+09] Simone Ceriani, Giulio Fontana, Alessandro Giusti et al. “Rawseeds ground truth collection systems for indoor self-localization and mapping.” En: “Auton. Robots” 27.4 (17 de dic. de 2009), págs. 353-371 (vid. pág. 56).

[Cor+18] Santiago Cortés, Arno Solin, Esa Rahtu y Juho Kannala. “ADVIO: An Authentic Dataset for Visual-Inertial Odometry”. En: *Computer Vision – ECCV 2018*. Ed. por Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu y Yair Weiss. Cham: Springer International Publishing, 2018, págs. 425-440 (vid. pág. 56).

[Cse22] Mate Cserep. *Hungarian MLS point clouds of railroad environment and annotated ground truth data*. 2022 (vid. pág. 77).

[CUE15] Nicholas Carlevaris-Bianco, Arash K. Ushani y Ryan M. Eustice. “University of Michigan North Campus long-term vision and lidar dataset”. En: “International Journal of Robotics Research” 35.9 (2015), págs. 1023-1035 (vid. pág. 56).

[CZR05] Erik Cueva, Daniel Zaldivar y Raul Rojas. “Kalman filter for vision tracking”. En: (2005) (vid. pág. 34).

[Dai+22] Yuan Dai, Weiming Liu, Heng Wang, Wei Xie y Kejun Long. “YOLO-Former: Marrying YOLO and Transformer for Foreign Object Detection”. En: “IEEE Transactions on Instrumentation and Measurement” 71 (2022), págs. 1-14 (vid. pág. 58).

[DAV23] DAVOS. *DAVOS - a fault Injection toolkit for dependability assessment, verification, optimization and selection of hardware desings*. https://github.com/IlyaTuzov/DAVOS. 2023 (vid. pág. 74).

[DBM23] DBMR. "Industry Trends and Forecast to 2029 - Semiconductors and Electronics". En: (2023) (vid. pág. 2).

[Dos+17] Alexey Dosovitskiy, Germán Ros, Felipe Codevilla, Antonio M. López y Vladlen Koltun. "CARLA: An Open Urban Driving Simulator". En: "CoRR" abs/1711.03938 (2017). arXiv: 1711.03938 (vid. pág. 59).

[Dua23] Jianyu Duan. "Study on Multi-Heterogeneous Sensor Data Fusion Method Based on Millimeter-Wave Radar and Camera". En: "Sensors" 23.13 (2023) (vid. pág. 7).

[Eba+20] Kamak Ebadi, Yun Chang, Matteo Palieri et al. "LAMP: Large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments". En: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020, págs. 80-86 (vid. pág. 47).

[EIT23] EITB. *EITB Euskal Irrati Telebista*. 2023. URL: https://www.eitb.eus/ (visitado 30 de mar. de 2023) (vid. pág. 38).

[ESI23] ESI. *ESI Pro-SiVIC*. 2023. URL: https://myesi.esi-group.com/downloads/software-downloads/pro-sivic-2017.0 (visitado 1 de ene. de 2023) (vid. pág. 59).

[Etx+22a] Mikel Etxeberria-Garcia, Maider Zamalloa, Nestor Arana-Arexolaleiba y Mikel Labayen. "Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case". En: "IEEE Access" 10 (2022), págs. 69200-69215 (vid. págs. 15, 18, 58).

[Etx+22b] Mikel Etxeberria-Garcia, Maider Zamalloa, Nestor Arana-Arexolaleiba y Mikel Labayen. "Visual Odometry in Challenging Environments: An Urban Underground Railway Scenario Case". En: "IEEE Access" 10 (2022), págs. 69200-69215 (vid. pág. 58).

[EUC16] J. Engel, V. Usenko y D. Cremers. "A Photometrically Calibrated Benchmark For Monocular Visual Odometry". En: *arXiv:1607.02555*. 2016 (vid. pág. 56).

[EUC23] EUComission. *Europe's Rail*. https://rail-research.europa.eu/. Accessed 10-Mar-2023. 2023 (vid. pág. 91).

[Fal+13] Maurice Fallon, Hordur Johannsson, Michael Kaess y John J Leonard. "The mit stata center dataset". En: "The International Journal of Robotics Research" 32.14 (2013), págs. 1695-1699 (vid. pág. 56).

[FK10] Eric Flior y Kazimierz Kowalski. "Continuous Biometric User Authentication in Online Examinations". En: ene. de 2010, págs. 488-492 (vid. pág. 40).

[FMB17] Gianni Fenu, Mirko Marras y Ludovico Boratto. "A multi-biometric system for continuous student authentication in e-learning platforms". En: "Pattern Recognition Letters" (abr. de 2017) (vid. pág. 40).

[Fou23] Open Source Robotics Foundation. *Gazebo*. 2023. URL: https://gazebosim.org/home (visitado 1 de ene. de 2023) (vid. pág. 59).

[Gei+13] Andreas Geiger, Philip Lenz, Christoph Stiller y Raquel Urtasun. “Vision meets robotics: The kitti dataset”. En: “The International Journal of Robotics Research” 32.11 (2013), págs. 1231-1237 (vid. págs. 47-49, 56).

[GGP18] Francisco Guillen-Gamez, Iván García-Magariño y Guillermo Palacios. „Comparative Analysis Between Different Facial Authentication Tools for Assessing Their Integration in m-Health Mobile Applications“. En: mar. de 2018, págs. 1153-1161 (vid. pág. 40).

[GKS08] Andreas Gerstinger, Heinz Kantz y Christoph Scherrer. “TAS Control Platform: A Platform for Safety-Critical Railway Applications.” En: “ERCIM News” 2008 (ene. de 2008) (vid. pág. 72).

[Glo+13] Ben Glocker, Shahram Izadi, Jamie Shotton y Antonio Criminisi. “Real-time RGB-D camera relocalization”. En: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2013, págs. 173-179 (vid. pág. 56).

[GLU12] Andreas Geiger, Philip Lenz y Raquel Urtasun. “Are we ready for autonomous driving? the kitti vision benchmark suite”. En: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, págs. 3354-3361 (vid. págs. 47, 48, 50, 56).

[Gru17] Michael Grupp. *evo: Python package for the evaluation of odometry and SLAM.* `https://github.com/MichaelGrupp/evo`. 2017 (vid. pág. 50).

[Gua+22] Ling Guan, Limin Jia, Zhengyu Xie y Chaoying Yin. “A Lightweight Framework for Obstacle Detection in the Railway Image Based on Fast Region Proposal and Improved YOLO-Tiny Network”. En: “IEEE Transactions on Instrumentation and Measurement” 71 (2022), págs. 1-16 (vid. pág. 58).

[Had+22] Mohamed Hadded, Ankur Mahtani, Sebastien Ambellouis, Jacques Boonaert y Hazem Wannous. „Application of Rail Segmentation in the Monitoring of Autonomous Train’s Frontal Environment“. En: ene. de 2022, págs. 185-197 (vid. pág. 58).

[Han+14] A. Handa, T. Whelan, J.B. McDonald y A.J. Davison. “A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM”. En: *IEEE Intl. Conf. on Robotics and Automation, ICRA*. Hong Kong, China, 2014 (vid. pág. 56).

[Har+20] Jeanine Harb, Nicolas Rébéna, Raphaël Chosidow et al. “FRSign: A Large-Scale Traffic Light Dataset for Autonomous Trains”. En: “arXiv e-prints” (feb. de 2020). arXiv: `2002.05665 [cs.CY]` (vid. pág. 77).

[He+21] Deqiang He, Zhiheng Zou, Yanjun Chen, Bin Liu y Jian Miao. “Rail Transit Obstacle Detection Based on Improved CNN”. En: “IEEE Transactions on Instrumentation and Measurement” 70 (2021), págs. 1-14 (vid. pág. 58).

[How+17] Andrew G Howard, Menglong Zhu, Bo Chen et al. “MobileNets”. En: “arXiv preprint arXiv:1704.04861” (2017). arXiv: `1704.04861` (vid. pág. 43).

[IB97] Michael Isard y Andrew Blake. “Condensation - conditional density propagation for visual tracking”. En: (1997) (vid. pág. 34).

[Igo+13] García Olaizola Igor, Flórez Esnal Julián, San Román Otegui Juan Carlos, Aginako Bengoa Naiara y Labayen Esnaola Mikel. „Method for Detecting the Point of Impact of a Ball in Sports Events“. Pat. estadounidense EP20180382363 20180525. 15 de mar. de 2013 (vid. págs. 18, 19).

[Jeo+19] Jinyong Jeong, Younggun Cho, Young-Sik Shin, Hyunchul Roh y Ayoung Kim. “Complex Urban Dataset with Multi-level Sensors from Highly Diverse Urban Environments”. En: “International Journal of Robotics Research” 38.6 (2019), págs. 642-657 (vid. págs. 49, 56).

[Jia+19] Yifan Jiang, Xinyu Gong, Ding Liu et al. “EnlightenGAN: Deep Light Enhancement without Paired Supervision”. En: “CoRR” abs/1906.06972 (2019). arXiv: 1906.06972 (vid. pág. 44).

[Jia+21] Yifan Jiang, Xinyu Gong, Ding Liu et al. “Enlightengan: Deep light enhancement without paired supervision”. En: “IEEE Transactions on Image Processing” 30 (2021), págs. 2340-2349 (vid. pág. 51).

[KGC15] A. Kendall, M. Grimes y R. Cipolla. “PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization”. En: *Proceedings of the IEEE international conference on computer vision*. 2015, págs. 2938-2946 (vid. pág. 56).

[Kiy+20] A. T. Kiyani, A. Lasebae, K. Ali, M. U. Rehman y B. Haq. “Continuous User Authentication Featuring Keystroke Dynamics Based on Robust Recurrent Confidence Model and Ensemble Learning Approach”. En: “IEEE Access” 8 (2020), págs. 156177-156189 (vid. pág. 45).

[Kle+21] S Klenk, J Chui, N Demmel y D Cremers. “TUM-VIE: The TUM Stereo Visual-Inertial Event Dataset”. En: *International Conference on Intelligent Robots and Systems (IROS)*. 2021. arXiv: 2108.07329 [cs.CV] (vid. pág. 56).

[Lab+14] Mikel Labayen, Igor García, Naiara Aginako y Julian Florez. “Multimedia Tools and Applications”. En: (ago. de 2014). Ed. por Borko Furht, págs. 199-208 (vid. págs. 14, 18).

[Lab+21] Mikel Labayen, Ricardo Vea, Julián Flórez, Naiara Aginako y Basilio Sierra. “Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology”. En: “IEEE Access” 9 (2021), págs. 72398-72411 (vid. págs. 15, 18).

[Lab+23a] Mikel Labayen, Daniel Ochoa de Eribe, Ander Aramburu, Marcos Nieto y Naiara Aginako. “European Common Data Management Platform Definition for Railway AI Function Development”. En: “XXX” XXX (2023), pág. XXX (vid. págs. 17, 18).

[Lab+23b] Mikel Labayen, Laura Medina, Fernando Eizaguirre, Jose Flich y Naiara Aginako. “HPC Platform for Railway Safety-Critical Functionalities based on Artificial Intelligence”. En: “XXX” XXX (2023), pág. XXX (vid. págs. 16, 18).

[Lab+23c] Mikel Labayen, Xabier Mendialdua, Naiara Aginako y Basilio Sierra. “Semi-Automatic Validation and Verification Framework for CV & AI enhanced Railway Signalling and Landmark Detector”. En: “XXX” XXX (2023), pág. XXX (vid. págs. 16, 18).

[Lab23a] LG Electronics America R&D Lab. *SVL Simulator by LG - Autonomous and Robotics real-time sensor Simulation, LiDAR, Camera simulation for ROS1, ROS2, Autoware, Baidu Apollo. Perception, Planning, Localization, SIL and HIL Simulation, Open Source and Free.* 2023. URL: https://www.svlsimulator.com/ (visitado 1 de ene. de 2023) (vid. pág. 59).

[Lab23b] Stanford Artificial Intelligence Laboratory. *Robotic Operating System*. 2023. URL: https://www.ros.org (visitado 1 de ene. de 2023) (vid. pág. 59).

[Lam+21] Daniel Lamas, Mario Soilán, Javier Grandío y Belén Riveiro. “Automatic Point Cloud Semantic Segmentation of Complex Railway Environments”. En: “Remote Sensing” 13.12 (2021) (vid. pág. 77).

[Li+20] L. Li, X. Mu, S. Li y H. Peng. “A Review of Face Recognition Technology”. En: “IEEE Access” 8 (2020), págs. 139110-139120 (vid. pág. 43).

[Li+23] Rui Li, Mingquan Zhou, Dan Zhang, Yuhuan Yan y Qingsong Huo. “A survey of multi-source image fusion”. En: “Multimedia Tools and Applications” (jul. de 2023) (vid. pág. 7).

[Liu+16] Wei Liu, Dragomir Anguelov, Dumitru Erhan et al. “SSD: Single shot multibox detector”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2016. arXiv: 1512.02325 (vid. pág. 43).

[Liu+18] Xiaokai Liu, Xiaorui Ma, Jie Wang y Hongyu Wang. “M3L: Multi-modality mining for metric learning in person re-Identification”. En: “Pattern Recognition” 76 (2018), págs. 650-661 (vid. pág. 43).

[Liv23] Dovetail Live. *Train Simulator*. 2023. URL: https://live.dovetailgames.com/live/train-simulator (visitado 1 de ene. de 2023) (vid. págs. 59, 62).

[LSH10] Yunpeng Li, Noah Snavely y Daniel P Huttenlocher. “Location recognition using prioritized feature matching”. En: *European conference on computer vision*. Springer. 2010, págs. 791-804 (vid. pág. 56).

[MA20] Nicholas Mc Guire e Imanol Allende. “Approaching certification of complex systems”. En: jun. de 2020, págs. 70-71 (vid. pág. 68).

[Mad+17] Will Maddern, Geoffrey Pascoe, Chris Linegar y Paul Newman. “1 year, 1000 km: The Oxford RobotCar dataset”. En: “The International Journal of Robotics Research” 36.1 (2017), págs. 3-15 (vid. págs. 49, 56).

[Mal+15] L. Malina, J. Hajny, P. Dzurenda y V. Zeman. “Privacy-preserving security solution for cloud services”. En: “Journal of Applied Research and Technology. JART” (2015) (vid. pág. 82).

[MGL23] Iraj Moghaddasi, Saeid Gorgin y Jeong-A. Lee. “Dependable DNN Accelerator for Safety-critical Systems: A Review on the Aging Perspective”. En: “IEEE Access” (2023), págs. 1-1 (vid. pág. 7).

[Mic23] Microsoft. *Project AirSim for aerial autonomy*. 2023. URL: https://www.microsoft.com/en-us/ai/autonomous-systems-project-airsim (visitado 1 de ene. de 2023) (vid. pág. 59).

[Mon+13] J. V. Monaco, J. C. Stewart, S. Cha y C. C. Tappert. “Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works”. En: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Sep. de 2013, págs. 1-8 (vid. pág. 40).

[MP18] Lubasi Musambo y Jackson Phiri. “Student Facial Authentication Model based on OpenCV’s Object Detection Method and QR Code for Zambian Higher Institutions of Learning”. En: “International Journal of Advanced Computer Science and Applications” 9 (ene. de 2018) (vid. pág. 40).

[MRM21] Labayen Esnaola Mikel, Vea Orte Ricardo y Fraile Yarza Manuel. „Method and System for Verifying the Identity of a User of An Online Service Using Multi-Biometric Data“. Pat. europea EP20180382363 20180525. 22 de nov. de 2021 (vid. págs. 18, 19).

[MT17] Raul Mur-Artal y Juan D Tardós. “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras”. En: “IEEE Transactions on Robotics” 33.5 (2017), págs. 1255-1262 (vid. págs. 50, 52).

[MTS17] András L Majdik, Charles Till y Davide Scaramuzza. “The Zurich urban micro aerial vehicle dataset”. En: “The International Journal of Robotics Research” 36.3 (2017), págs. 269-273. eprint: https://doi.org/10.1177/0278364917702237 (vid. pág. 56).

[Nah+14] A. F.M.Nazmul Haque Nahin, Jawad Mohammad Alam, Hasan Mahmud y Kamrul Hasan. “Identifying emotion by keystroke dynamics and text pattern analysis”. En: “Behaviour and Information Technology” (2014) (vid. pág. 45).

[NSO21] Marcos Nieto, Orti Senderos y Oihana Otaegui. “Boosting AI applications: Labeling format for complex datasets”. En: “SoftwareX” 13 (2021), pág. 100653 (vid. pág. 83).

[NT06] Wayne Naidoo y Jules-Raymond Tapamo. “Soccer video analysis by ball, player and referee tracking”. En: (ene. de 2006), págs. 51-60 (vid. pág. 34).

[Nvi23] Nvidia. *Jetson agx xavier and the new era of autonomous machines*. https://info.nvidia.com/rs/156-OFN-742/images/Jetson_AGX_Xavier_New_Era_Autonomous_Machines.pdf. Accessed: 2023-02-01. 2023 (vid. pág. 73).

[OFC18] Daniel Olid, José M. Fácil y Javier Civera. “Single-View Place Recognition under Seasonal Changes”. En: *PPNIV Workshop at IROS 2018*. 2018 (vid. págs. 48, 56).

[OHS03] N. Owens, C. Harris y C. Stennett. “Hawk-eye tennis system”. En: *2003 International Conference on Visual Information Engineering VIE 2003*. 2003, págs. 182-185 (vid. pág. 34).

[Oka+19] Alexandra Okada, Ingrid Noguera, Lyubka Aleksieva et al. “Pedagogical approaches for e-assessment with authentication and authorship verification in Higher Education”. En: “British Journal of Educational Technology” (feb. de 2019) (vid. pág. 40).

[Ope23] OpenCV. *OpenCV*. https://opencv.org/. Accessed: 2023-02-01. 2023 (vid. pág. 72).

[Per+20] Jon Perez-Cerrolaza, Roman Obermaisser, Jaume Abella et al. “Multi-Core Devices for Safety-Critical Systems: A Survey”. En: “ACM Computing Surveys” 53 (mayo de 2020) (vid. pág. 67).

[PME11] Gaurav Pandey, James R McBride y Ryan M Eustice. “Ford campus vision and lidar data set”. En: “The International Journal of Robotics Research” 30.13 (2011), págs. 1543-1552 (vid. págs. 49, 56).

[POJ00] G. Pingali, A. Opalach e Y. Jean. “Ball tracking and virtual replays for innovative tennis broadcasts”. En: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. Vol. 4. 2000, 152-156 vol.4 (vid. pág. 34).

[Pov+11] Daniel Povey, Arnab Ghoshal, Gilles Boulianne et al. “The Kaldi Speech Recognition Toolkit”. En: *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Catalog No.: CFP11SRW-USB. Hilton Waikoloa Village, Big Island, Hawaii, US: IEEE Signal Processing Society, dic. de 2011 (vid. pág. 44).

[QEM23] QEMU. *QEMU - A generic and open source machine emulator and virtualizer*. `https://www.qemu.org/`. 2023 (vid. pág. 73).

[Ray18] Charan Singh Rayat. „Measures of Dispersion“. En: *Statistical Methods in Medical Research*. Springer, 2018, págs. 47-60 (vid. pág. 54).

[Red19] Vijay Janapa et. al. Reddi. *MLPerf Inference Benchmark*. 2019 (vid. pág. 67).

[RFM21] Danijela Ristic-Durrant, Marten Franke y Kai Michels. “A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways”. En: “Sensors” 21 (mayo de 2021), pág. 3452 (vid. pág. 78).

[Ric+17] Vea Orte Ricardo, Labayen Esnaola Mikel, Flórez Esnal Julián y Marcos Ortego Gorka. „Method and System for Verifying the Identity of a User of an Online Service“. Pat. europea WO2013EP61521 20130604. 1 de ago. de 2017 (vid. págs. 18, 19).

[S2R23] S2R. *TAURO*. `https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro`. Accessed 10-Mar-2023. 2023 (vid. pág. 58).

[Saw+19] S. Sawhney, K. Kacker, S. Jain, S. N. Singh y R. Garg. “Real-Time Smart Attendance System using Face Recognition Techniques”. En: *2019 9th International Conference on Cloud Computing, Data Science Engineering*. Ene. de 2019, págs. 522-525 (vid. pág. 40).

[Sch+17] Thomas Schops, Johannes L Schonberger, Silvano Galliani et al. “A multi-view stereo benchmark with high-resolution images and multi-camera videos”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 3260-3269 (vid. pág. 56).

[Sie23] Siemens. *JAILHOUSE*. `https://github.com/siemens/jailhouse`. 2023 (vid. pág. 70).

[SKP15] Florian Schroff, Dmitry Kalenichenko y James Philbin. “FaceNet: A unified embedding for face recognition and clustering”. En: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2015. arXiv: 1503.03832 (vid. pág. 43).

[SLS23] Fei Su, Chunsheng Liu y Haralampos-G. Stratigopoulos. “Special Issue on Testability and Dependability of Artificial Intelligence Hardware”. En: “IEEE Design & Test” 40.2 (2023), págs. 5-7 (vid. pág. 7).

[Sof23] Siemens Software. *Simcenter Prescan*. 2023. URL: `https://www.plm.automation.siemens.com/global/en/products/simcenter/%20prescan.html` (visitado 1 de ene. de 2023) (vid. pág. 59).

[spo23] sportvision. *SMT (SportsMEDIA Technology)*. 2023. URL: `https://www.smt.com/` (visitado 30 de mar. de 2023) (vid. pág. 34).

[Sta+22] Andrea Staino, Akshat Suwalka, Pabitra Mitra y Biswajit Basu. “Real-Time Detection and Recognition of Railway Traffic Signals Using Deep Learning”. En: “Journal of Big Data Analytics in Transportation” 4.1 (abr. de 2022), págs. 57-71 (vid. pág. 58).

[Sta23] University of Stanford. “Measuring trends in Artificial Intelligence”. En: (2023) (vid. pág. 2).

[Stu+12a] J. Sturm, N. Engelhard, F. Endres, W. Burgard y D. Cremers. “A Benchmark for the Evaluation of RGB-D SLAM Systems”. En: *Proc. of the International Conference on Intelligent Robot Systems (IROS)*. 2012 (vid. pág. 56).

[Stu+12b] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard y Daniel Cremers. “A benchmark for the evaluation of RGB-D SLAM systems”. En: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, págs. 573-580 (vid. págs. 50, 56).

[Ume91] Shinji Umeyama. “Least-squares estimation of transformation parameters between two point patterns”. En: “IEEE Transactions on Pattern Analysis & Machine Intelligence” 13.04 (1991), págs. 376-380 (vid. pág. 50).

[Umu+17] Yaman Umuroglu, Nicholas J. Fraser, Giulio Gambardella et al. “FINN: A Framework for Fast, Scalable Binarized Neural Network Inference”. En: *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. FPGA ’17. ACM, 2017, págs. 65-74 (vid. pág. 67).

[Unz+14] Luis Unzueta, Waldir Pimenta, Jon Goenetxea, Luís Paulo Santos y Fadi Dornaika. “Efficient generic face model fitting to images and videos”. En: “Image and Vision Computing” (2014) (vid. pág. 43).

[Vir23] 4D Virtualiz. *4D Virtualiz*. 2023. URL: https://www.4d-virtualiz.com/ (visitado 1 de ene. de 2023) (vid. pág. 59).

[Wal+17] F Walch, C Hazirbas L Leal-taix, T Sattler y S Hilsenbeck D Cremers. “Image-based localization using LSTMs for structured feature correlation”. En: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, págs. 627-637 (vid. pág. 56).

[Wan+17] Sen Wang, Ronald Clark, Hongkai Wen y Niki Trigoni. “Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks”. En: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2017, págs. 2043-2050 (vid. pág. 48).

[Wan+18a] Li Wan, Quan Wang, Alan Papir e Ignacio Lopez Moreno. “Generalized end-to-end loss for speaker verification”. En: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. 2018. arXiv: 1710.10467 (vid. pág. 44).

[Wan+18b] Quan Wang, Carlton Downey, Li Wan, Philip Andrew Mansfield e Ignacio Lopz Moreno. “Speaker diarization with LSTM”. En: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. 2018. arXiv: 1710.10468 (vid. pág. 44).

[Wan+23] Lucai Wang, Hongda Qin, Xuanyu Zhou, Xiao Lu y Fengting Zhang. “R-YOLO: A Robust Object Detector in Adverse Weather”. En: “IEEE Transactions on Instrumentation and Measurement” 72 (2023), págs. 1-11 (vid. pág. 58).

[Wat+14] Andrew Waterman, Yunsup Lee, David A. Patterson y Krste Asanović. *The RISC-V Instruction Set Manual, Volume I: User-Level ISA, Version 2.0*. Inf. téc. UCB/EECS-2014-54. EECS Department, University of California, Berkeley, mayo de 2014 (vid. pág. 68).

[Wu09] Weixin Wu. “Tennis Touching Point Detection based on High Speed Camera and Kalman Filter”. En: (2009) (vid. pág. 34).

[Xia+23] Wei Xian, Kan Yu, Fengling Han et al. “Advanced Manufacturing in Industry 5.0: A Survey of Key Enabling Technologies and Future Trends”. En: “IEEE Transactions on Industrial Informatics” (2023), págs. 1-15 (vid. pág. 7).

[Xil23a] Xillinx. *Accelerating DNNs with Xilinx Alveo Accelerator Cards*. `https://docs.xilinx.com/v/u/en-US/wp504-accel-dnns`. Accessed: 2023-02-01. 2023 (vid. pág. 73).

[Xil23b] Xillinx. *AMD Virtex UltraScale+ FPGA VCU118 Evaluation Kit*. `https://www.xilinx.com/products/boards-and-kits/vcu118.html`. Accessed: 2023-02-01. 2023 (vid. pág. 68).

[XOT13] Jianxiong Xiao, Andrew Owens y Antonio Torralba. “Sun3d: A database of big spaces reconstructed using sfm and object labels”. En: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, págs. 1625-1632 (vid. pág. 56).

[XSS19] Lu Xiaofeng, Zhang Shengfei y Yi Shengwei. “Continuous authentication by free-text keystroke based on CNN plus RNN”. En: *Procedia Computer Science*. 2019 (vid. pág. 45).

[Yan+18] Nan Yang, Rui Wang, Jorg Stuckler y Daniel Cremers. “Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry”. En: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, págs. 817-833 (vid. pág. 48).

[Yan+20] Nan Yang, Lukas von Stumberg, Rui Wang y Daniel Cremers. “D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, págs. 1281-1292 (vid. pág. 48).

[YBH15] Khalid Yousif, Alireza Bab-Hadiashar y Reza Hoseinnezhad. “An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics”. En: “Intelligent Industrial Systems” 1.4 (2015), págs. 289-311 (vid. pág. 47).

[YCK05] Fei Yan, William Christmas y Josef Kittler. “A Tennis Ball Tracking Algorithm for Automatic Annotation of Tennis Match.” En: ene. de 2005 (vid. pág. 34).

[Ye+21] Tao Ye, Xi Zhang, Yi Zhang y Jie Liu. “Railway Traffic Object Detection Using Differential Feature Fusion Convolution Neural Network”. En: “IEEE Transactions on Intelligent Transportation Systems” 22.3 (2021), págs. 1375-1387 (vid. pág. 58).

[Yeh+17] Raymond A. Yeh, Chen Chen, Teck Yian Lim et al. “Semantic image inpainting with deep generative models”. En: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 2017. arXiv: 1607.07539 (vid. pág. 44).

[YTA07] Xinguo Yu, Xiaoying Tu y Ee Luang Ang. “Trajectory-Based Ball Detection and Tracking in Broadcast Soccer Video with the Aid of Camera Motion Recovery”. En: *2007 IEEE International Conference on Multimedia and Expo*. 2007, págs. 1543-1546 (vid. pág. 34).

[Yu+03] Xinguo Yu, Changsheng Xu, Hon Leong et al. “Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video”. En: vol. 3. Nov. de 2003, págs. 11-20 (vid. pág. 34).

[Yu+22] Xinyi Yu, Weiqi He, Xuecheng Qian et al. “Real-time rail recognition based on 3D point clouds”. En: “Measurement Science and Technology” 33.10 (jul. de 2022), pág. 105207 (vid. pág. 77).

[Yua+22] H. Yuan, Z. Mei, Y. Chen, W. Niu y C. Wu. “RailVID: A Dataset for Rail Environment Semantic”. En: *17th International Conference on Systems, ICONS*. 2022 (vid. pág. 77).

[ZDJ12] Yu Zhong, Yunbin Deng y Anil K. Jain. “Keystroke dynamics for user authentication”. En: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 2012 (vid. pág. 45).

[Zen+19] Oliver Zendel, Markus Murschitz, Marcel Zeilinger et al. “RailSem19: A Dataset for Semantic Rail Scene Understanding”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Jun. de 2019 (vid. pág. 77).

[Zha+16] Z. Zhang, Mingshao Zhang, Yizhe Chang, S. Esche y C. Chassapis. “A Virtual Laboratory System with Biometric Authentication and Remote Proctoring Based on Facial Recognition”. En: 2016 (vid. pág. 40).

[Zha+18a] Huangying Zhan, Ravi Garg, Chamara Saroj Weerasekera et al. “Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, págs. 340-349 (vid. pág. 48).

[Zha+18b] Shifeng Zhang, Xiangyu Zhu, Zhen Lei et al. “FaceBoxes: A CPU real-time face detector with high accuracy”. En: *IEEE International Joint Conference on Biometrics, IJCB 2017*. 2018. arXiv: `1708.05234` (vid. pág. 43).

[Zha+18c] Zhou Zhang, El-Sayed Aziz, Sven Esche y Constantin Chassapis. “A Virtual Proctor with Biometric Authentication for Facilitating Distance Education”. En: *Online Engineering & Internet of Things*. Ed. por Michael E. Auer y Danilo G. Zutin. Cham: Springer International Publishing, 2018, págs. 110-124 (vid. pág. 40).

[Zha+19] Huangying Zhan, Chamara Saroj Weerasekera, Jiawang Bian y Ian Reid. “Visual Odometry Revisited: What Should Be Learnt?” En: “arXiv preprint arXiv:1909.09803” (2019) (vid. pág. 51).

[Zha+21] Huangying Zhan, Chamara Saroj Weerasekera, Jia-Wang Bian, Ravi Garg y Ian Reid. “DF-VO: What Should Be Learnt for Visual Odometry?” En: “arXiv preprint arXiv:2103.00933” (2021) (vid. págs. 48, 50, 52).

[Zha+23] Yan Zhang, Kefeng Li, Guangyuan Zhang, Zhenfang Zhu y Peng Wang. “DFA-UNet: Efficient Railroad Image Segmentation”. En: “Applied Sciences” 13.1 (2023) (vid. pág. 58).

[Zho+17] Tinghui Zhou, Matthew Brown, Noah Snavely y David G Lowe. “Unsupervised learning of depth and ego-motion from video”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, págs. 1851-1858 (vid. pág. 48).

[Zhu+18] Alex Zihao Zhu, Dinesh Thakur, Tolga Ozaslan et al. “The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception”. En: “IEEE Robotics and Automation Letters” 3.3 (2018), págs. 2032-2039 (vid. pág. 56).

[ZL23] Zhaoyun Zhang y Jingpeng Li. “A Review of Artificial Intelligence in Embedded Systems”. En: “Micromachines” 14.5 (2023) (vid. pág. 7).

[Zou+23] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo y Jieping Ye. “Object Detection in 20 Years: A Survey”. En: “Proceedings of the IEEE” 111.3 (2023), págs. 257-276 (vid. pág. 6).

[Zui94] Karel Zuiderveld. „Contrast Limited Adaptive Histogram Equalization“. En: *Graphics Gems*. 1994 (vid. pág. 43).

[Zuñ+20] David Zuñiga-Noël, Alberto Jaenal, Ruben Gomez-Ojeda y Javier Gonzalez-Jimenez. “The UMA-VI dataset: Visual–inertial odometry in low-textured and dynamic illumination environments”. En: “The International Journal of Robotics Research” 39.9 (2020), págs. 1052-1060. eprint: `https://doi.org/10.1177/0278364920938439` (vid. pág. 56).

Índice de figuras

Índice de tablas