eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

APPLICATION-DRIVEN VISUAL COMPUTING TOWARDS INDUSTRY 4.0

Author: Helen V. Diez McCready

Supervised by:

Ph.D. Alejandro García-Alonso Montoya

&

Ph.D. Jorge Posada Velásquez

PhD Thesis

San Sebastian, July, 2018

To my parents.

Abstract

This research work focuses in three fields within the Visual Computing research area: Interactive Virtual Agents (IVA), Immersive VR/AR Environments and Interactive 3D Model Management. The contributions come from applied research projects in different areas: e-learning, marketing, health, manufacturing (safety and simulation). This knowledge provides proof of concept for some challenges emerging from Operator 4.0: human factors, simulation, image generation, and model integration. Operator 4.0 is a key element within Industry 4.0.

The thesis studies the development of autonomous, modular, scalable, ubiquitous and engaging IVA. IVA that can interact with users in a human-like way thanks to their attributes: realistic appearance, verbal and non-verbal communication and natural behaviour skills.

Immersive VR/AR Environments are analyzed in the context of production planning, product design, process simulation, testing and verification. The research analyzes solutions for cost-efficient applications that reduce risks for the operators, provide real-time feedback and show relevant information to the worker.

Interactive 3D Model Management improves multimedia CAD model management and visualization, via automatic model conversion from CAD data into web interactive display of 3D models.

Finally, this knowledge is presented in the context of Industry 4.0 visual computing challenges.

Agradecimientos

Es difícil saber hasta dónde se tienen que remontar estos agradecimientos, pero para no aburrir al lector voy a limitarlos a la etapa que atañe a esta tesis.

Gracias a Vicomtech por darme la oportunidad de desarrollar mi carrera profesional durante estos ocho años y por permitirme culminar ese trabajo a través de esta tesis. Gracias a los diferentes jefes que he tenido dentro de Vicomtech: Igor G. Olaizola, Iñigo Barandiaran y por supuesto David Oyarzun, bajo tu dirección creamos un departamento envidiable e inolvidable.

Agradezco su dedicación a Jorge Posada, sus consejos han contribuido a enmarcar y enriquecer esta tesis.

Sin duda esta memoria no habría sido posible sin el profesor Alex García-Alonso. Gracias por tu apoyo, tus consejos, tus ánimos, tu comprensión, tu paciencia y sobretodo por entender y respetar cúales eran mis prioridades vitales más allá de la tesis.

Como afortunadamente en la vida no todo es trabajo, gracias a todos los compañeros y amigos que me habéis acompañado durante estos años. Sin vosotros no creo que hubiese llegado hasta aquí. Os voy a echar de menos.

Gracias a mi familia, con la que sé que siempre puedo contar y sin la cual nada tendría sentido.

Por último, gracias a mis padres, porque empiezo a entender el esfuerzo y sacrificio que habéis hecho y seguís haciendo. Solo espero que sintáis que ha merecido la pena.

Contents

List of Figures

List of Tables

INTRODUCTION

CHAPTER 1

Introduction

This chapter serves as an introduction to the motivation and objectives of this thesis. The chapter is structured as follows. Section 1.1 introduces the global trend Industry 4.0 and exposes how Visual Computing technologies can be applied to support the Smart Factories. Section 1.2 outlines the objectives followed in this thesis and Section 1.3 presents how the document is structured.

1.1 Motivation

The Industry 4.0 initiative, described as the Fourth Industrial Revolution, started in Germany. It was defined by Kagermann et al. (2013) as the union of three key components in manufacturing environments: Internet of Things (IoT), Cyber-Physical Systems (CPS) and Smart Factories. Hermann et al. (2016) noted that while lot of academic and practical discussion on the topic of Industry 4.0 was made, the term was not concrete enough and different practitioners used slightly different concepts. They reviewed the literature to find the actual key technological components from the most frequent terms in related publications. Among the top technologies required to address this challenges are *Industrial Internet of Things (IIoT)*, *Cloud Computing*, *Human Robot collaboration*, *Big Data* and also *Visual Computing* (Posada et al., 2015; Stork, 2015).

Beyond the technological aspects, one key element, in present and future manufacturing, is the *human operator* (Vallés, 2013). The Industry 4.0 concept does identify the human workforce as a relevant and active factor (Posada et al., 2015). Nevertheless, the integration of the human operator in the new *Smart Factories* poses ethical and societal challenges with scientific, technological and political implications (Manyika, 2016).

The technologies related with Visual Computing (common to some of the main topics of Computer Graphics and Computer Vision, and covering the full cycle acquisition, analysis, synthesis and interaction with visual data) are key elements to cope with some of the challenges of Industry 4.0. To cite some of the technologies that could fit under the Visual Computing term; virtual reality, augmented reality, computer vision and visual analytics could be highlighted.

The following two sections deal with Industry 4.0 and Visual Computing concepts and relations between them.

1.1.1 Industry 4.0

Posada et al. (2015) presented how Visual Computing can contribute as a key enabling technology to Industry 4.0. In this work they based their study on three main criteria: i) the integration dimensions, ii) the product and production and iii) the human factors.

This work identified critical intersections between the priorities of Industry 4.0 and the Visual Computing technologies (see *Figure 5* in Posada et al. (2015)). They summarized these challenges and applications in three tables:

- Table 1.1 showed how Visual Computing technologies are related with the *integration* dimensions.
- Table 1.2 analyzed *product and production* priorities in Industry 4.0.
- Table 1.3 described the third criterion related to the *human factors*.

Table 1.1: Industry 4.0 future factory visual computing challenges in three integration dimensions. (Posada et al., 2015)

Table 1. Industrie 4.0 future factory visual computing challenges in three integration dimensions.

Industrie 4.0 integration dimension	Visual computing enabling technologies and challenges
Vertical integration (networked manufactured systems and autonomous cyber-physical production systems)	*Virtual environments*. Visually empowered 3D simulation scenarios for new ways of planning production, especially suitable for dynamic and fast changes. Scenarios for testing different configurations. *Real-time representation of production*. Visualizing flows of information, material, and knowledge in the factory, not only physical representation. *3D scanning and 3D reconstruction of factories*. Adapting old factories to new paradigms. *End user interfaces*. Editing configurations in demanding work conditions, such as production lines.
End-to-end digital engineering integration (holistic life-cycle management)	*Natural flow of a persistent and interactive digital model*. Product life-cycle management involving large industrial 3D CAD/CAM models with full access to semantic/dynamic integration data in Web3D. *3D real-time simulations for CPE production.* *New paradigms of 3D geometric representation*. New processes (such as laser-based manufacturing, fast-speed material removal, and micro- and nano-manufacturing) and new materials (such as biomaterials and metallic powders for 3D metal printing). *Computer vision "closing the loop" in 3D production planning*. Real-time coupling of production process and 3D models. Geometry adaptation to physical conditions.
Horizontal integration through value networks (value chain integration)	*Augmented reality (AR) for service-based actions with providers and clients*. Ergonomic aspects of the solutions going from the lab to real factories and the integration with the information systems. Intelligent media streaming/search to improve service (as in teleoperation). *3D model automatic simplification*. Preserving critical features for service tasks while allowing interaction/visualization in mobile low-power client devices.

Table 1.2: Industry 4.0 future factory visual computing challenges in product and production. (Posada et al., 2015)

Table 2. Industrie 4.0 future factory visual computing challenges in product and production.

Industrie 4.0 product and production	Visual computing enabling technologies and challenges
Product self-awareness (history, status, location, delivery strategy, and service)	*Integration of GIS (outdoor) with in-factory (indoor) localization-visualization systems*. Individualized product tracking and as underlying connection layer between factories and products when delivered. *Cyber-physical 3D equivalence*. At all times linking product digital model and situational status.
Personalization and flexibility (flexible adaptation to individual customer requirements)	*3D interactive tools*. Empowering end users in the final product configuration. *Automatic generation of options catalogs*. Accounting for production parameters and user preferences. *3D shape automatic adaptation*. Fitting production and manufacturing restrictions. *Linking of 3D changes with resource impacts*. Accounting for time and cost.
Optimized decision making (with access to real-time production and design data)	*Visual analytics of production big data*. Trillions (or more) samples per year (such as GE Industrial Big Data[19]). *Real-time mixing of production big data with 3D digital engineering design data.* *User interface dynamic adaptation of information to user profile, devices, and context*. Visual analytics system for the engineer and the worker.
Emergence of new services and business models	*Digital coexistence or alter-ego of the physical product*. Enabled by Web3D, localization and mobile interaction technologies, allowing new services (such as social networks of users for the same product line).
Resource/energy efficiency and sustainable production	*Dynamic resource visualization at the factory level*. Including sustainability footprint (such as CO_2 consumption), energy distribution in a plant, and material waste. Can be mixed with VR and in some cases AR.

Table 1.3: Industry 4.0 future factory visual computing challenges, human factors. (Posada et al., 2015)

Table 3. Industrie 4.0 future factory visual computing challenges, human factors.

Industrie 4.0 relation with human factors	Visual computing enabling technologies and challenges
Work organization and design (productivity enhancement though improved human intervention)	*New HMI modalities.* Allowing new modes of interaction adapted to workers' job restrictions (such as voice-based interaction and gesture recognition) *Post-WIMP interfaces.* Adapting to the future trend of mobile devices in the factory. *Advanced manufacturing and production-planning visualization.* Linking SCADA systems with VR paradigms for interaction in the planning and understanding of the production plan. *Interfaces with manufacturing execution systems.* Allowing different configuration capabilities.
Foster creativity in skilled workers	*End-user tools for visualization of flexible production plans.* Including the location of different machines and persons (with complementary skills) in alternative production scenarios. Tools for discussions between engineers and workers.
Training and continuing professional development (capture and systematic reuse of the knowledge of the worker)	*Multimedia capture and intelligent retrieval of worker knowledge.* Capturing and transferring knowledge between the workers. *3D authoring tools and end-user UI.* Operational training of complex machines and in some cases in virtual and augmented reality setups.
Safety and security	*Cognitive computer vision systems.* Detecting and contextualizing events occurring in the factory to improve safety and security. Focusing on hazardous area exposition and collision detection with massive objects. *Visual simulation for emergency response in the factories.*
Sociotechnical interaction (co-working with configurable robots)	*Visual programming of robot interactions.* Imitating human motion based on computer vision for anthropomorphic robots. Easing the use and control of the robot by the worker, not necessarily the engineer. *Virtual environments simulations.* For human and robot coexistence in production and for different configurations and parameters.

This human-centred factories have led to a new concept: **Operator 4.0**. Romero et al. (2016a) presented the concept of Operator 4.0 to describe the worker in a Human Cyber-Physical System scenario, as an evolution of previous generations of operators. An Operator 4.0 is a “smart and skilled operator who performs ‘work aided’ by machines if and as needed”. In Romero et al. (2016b) the authors go on to present different typologies of Operator 4.0 based on the technologies they work with.

They define the concept of Human Cyber-Physical Systems (H-CPS) as systems engineered to improve *human abilities* and improve *human physical sensing* and *cognitive capabilities* by means of various technologies, including Visual Computing.

1.1.2 Visual Computing for Industry 4.0

The following large quote from Posada et al. (2015) provides an appropiate foundation for this section:

"*Diverse technologies such as big data, advanced Human-Machine Interaction (HMI), 3D models and simulations, cloud computing, CPS, Internet of Things and Services, machine to machine, and 'smartization' can be applied in Industry 4.0 solutions.*

To achieve CPS for industry, the virtual simulation of products and processes, before and during operation, is a key aspect for achieving critical goals for product configuration and production flexibility. The modeling and simulation of processes covering the full product life-cycle (from design to disposal) is a relevant aspect, especially with the emergence of the cyber-physical equivalence (CPE), which refers to the fact that the virtual and physical dimensions coexist and are synchronized in time.

Visual analytics could help link otherwise separated technologies such as industrial big data, IoT and cloud services, intelligent devices, and semantic technologies. The manufacturing industry is one of the most demanding and challenging scenarios for visual analytics, and earlier research pointed to the many billions and even trillions of individual products that are produced per year.

A new generation of HMI applied to industry is needed to optimize the configuration of manufacturing jobs, including the operation of machines and production lines as well as aspects related to extended training and qualification. These are intelligent and multimodal assistance systems that put the person in the center of production."

All visual computing related technologies have a relevant aspect related with HMI. For example, virtual and augmented reality are both advanced visualization techniques oriented towards showing information to the operators while they are interacting either with virtual or real objects.

Soon, mobile technology will shift towards what is now known as contextual computing, defined in large in (Abowd et al., 1998). In this incoming scenario technologies such as augmented reality will play a key role. Augmented Reality enables showing additional information to the operators adapted to the production status, operators role or preferences, adding valuable information allowing operators to make better-informed shop floor decisions.

As pointed out in several publications (Evans and Annunziata, 2012; Kagermann et al., 2013; Posada et al., 2015) Industry 4.0 paradigms are not only focused on the integration of new technologies for improving the efficiency of production of machines and production lines but also to the empowering of the operators. All recognize the strategic importance of skilled and empowered operators in order to fulfil the smart factory or the factories of the future vision.

Other technologies related to visual computing, such as computer vision have been traditionally applied to dimensional or surface defects identification inside a manufacturing line. Nowadays this technology is being extended to monitoring operators while operating in scenarios or environments such as collaborative robotics or *co-bots* (Mohammed et al., 2016). Most industrial robots today operate behind safety fences. With the introduction of the so-called co-bots these barriers have started to be removed. However, these types of robots are still not fully safe for the operator and additional actions need to be taken such as adding external or on-robot cameras to monitor operator movements and avoid collision. The apparition of new sensors able to deliver high-speed full 3D information along with the development of new 3D real-time image processing are fostering the adoption of these types of solutions in the shop-floors.

Virtual Reality is an appropriate technology to deploy continuous training procedures in the shop-floor, allowing the operators to learn beforehand unseen production situations. Also, virtual reality is a very suitable technology for training safety protocols or simulating dangerous operations (Matsas and Vosniakos, 2017).

1.2 Objectives and Methodology

This thesis proposes different Visual Computing contributions which are relevant in three main areas: *Interactive Virtual Agents*, *Interactive VR/AR Environments* and *Interactive 3D Model Management*. The thesis follows an application-driven approach: starting from an initial focus on e-learning, training and education, towards Industry 4.0 development.

The research considered in this thesis falls within the following classification of Visual Computing technologies applied to the Industry 4.0 defined by Segura & Diez et al. (2018). The three main areas described above are directy related to these technologies:

A. Visual Analytics. Provides easy to understand depictions of large amounts of data and relations that are not immediately seen.

B. Collaborative Robotics Interaction. Co-bots enhance workers manual abilities enabling more precise or force-requiring operations to be performed.

C. Human-Machine Interaction (HMI). Act as the main interface between the worker and the supporting automation systems.

D. Media/Social Network. Industrial social networks allow communication between workers or between them and management increasing worker satisfaction and transfer of experience.

E. Virtual Reality (VR). Permits the vision and virtual use of elements out of reach, enabling safe use of hazardous equipment and enhanced learning of procedures.

F. Augmented Reality (AR). Augments the workplace with relevant information not normally visible and useful for the work required at the moment.

These six technologies cover a broad amount of research topics. This thesis compiles research work carried out since 2010 in several application-driven research projects that can be classified within them.

The roadmap for this thesis could not be foreseen when the research work that led to this thesis started. The research path has been driven by the application research projects that will be presented on the following chapters. Practical problems - challenges - defined the route to follow in the research work along the initial years. On 2014-2015 a new light illuminated this candidate's research work: the formal recognition by the scientific community of the relevance of Visual Computing technologies for the field of Industry 4.0. The challenges already addressed provided experience and motivation to join the scientific community discussion towards Industry 4.0.

The two goals derived from this work are presented next.

Thesis Goal I (Contributions to Visual Computing)

The initial goal pursued in this research work was to contribute in Visual Computing technologies in the following areas within the above mentioned taxonomy:

- **Interactive Virtual Agents.** (deals with technology C)

- **Interactive VR/AR Environments.** (B)(E)(F)
- **Interactive 3D Model Management.** (A)

Thesis Goal II (Projecting the contributions towards Industry 4.0)

The results obtained from this application-driven research have been related to Industry 4.0 challenges. The challenges resolved in these applications have a direct correlation with those that have to be faced in the Smart Factories. The knowledge acquired in the three research areas from this thesis has highlighted the path towards Industry 4.0.

The thesis (see Chapters 2, 3 and 4) contextualizes and describes the contributions provided by the research work within Goal I. Chapter 5 relates these contributions with the Industry 4.0 challenges (Goal II).

1.3 Document Structure and Background

This thesis presents application-driven research where Interactive Virtual Agents, Immersive VR/AR Environments and Interactive 3D Model Management are integrated into several case studies.

Each Chapter (2, 3 and 4) provides background, a summarized description of some applied research projects and their contributions. An effort has been made in order to classify these projects into the Visual Computing areas contemplated from the Industry 4.0 point of view. This classification eases the final correlation presented in Section 5.2 regarding Thesis Goal II.

Chapter 2, Interactive Virtual Agents, deals with the challenges related to the third criteria described by Posada et al. (2015), *human factors*. This chapter analyzes three characteristics of Interactive Virtual Agents, which are, Realistic Appearance, Verbal and Non-Verbal Communication and Natural Behaviour Skills. The *human factor* criteria has other challenges, such as production planning, flexible production plan visualization, etc. Chapter 2 will define its domain.

Virtual Reality and Virtual Environments are Visual Computing technologies relevant for Industry 4.0 which affect several dimensions within the three criteria mentioned in Section 1.1.1 (Posada et al., 2015). This is why **Chapter 3**, addresses Virtual Reality and analyzes three characteristics which the scientific community has stated as fundamental: Virtual Environments, Interactivity and Immersion.

3D CAD/CAM models are essential in various industrial sectors, such as automotive, aeronautic, and machine-tool. The representation of 3D geometries for CAD/CAM is a challenge for computer graphics researchers, because, as radically new fabrication processes emerge, they require new 3D representations and algorithms. **Chapter 4**, Interactive 3D Model Management, focuses on improving the multimedia management of complex 3D CAD models.

As this thesis covers three complementary areas, each technical chapter will start with its specific background review. The structure of the technical chapters is described next, followed by Chapter 5 (Conclusions and Further Work).

- Chapter 2. Interactive Virtual Agents:
  - Section 2.1: Realistic Appearance.
  - Section 2.2: Verbal and Non-Verbal Communication.
  - Section 2.3: Natural Behaviour Skills.
- Chapter 3. Interactive VR/AR Environments:
  - Section 3.1: Virtual Environments.
  - Section 3.2: Interactivity.
  - Section 3.3: Immersion.
- Chapter 4. Interactive 3D Model Management:
  - Section 4.3: Automatic Model Conversion.

APPLICATION-DRIVEN VISUAL COMPUTING TOWARDS INDUSTRY 4.0

CHAPTER 2

Interactive Virtual Agents

Human-centred manufacturing will be needed in Industry 4.0 in order to increase flexibility, agility, and competitiveness (Vallés, 2013). Skilled workers and engineers for a competitive smart factory are of strategic importance. From now on, we will refer to these persons as Operator 4.0.

In the Smart Factories the work of the Operator 4.0 will be supported by smart assistance systems with multimodal, user-friendly interfaces (Henning, 2013).

Nowadays, Interactive Virtual Agents (IVA) are frequently applied in diverse fields, such as education, training, entertainment, etc. The role of the virtual agent depends on the application; it can act for example as a guide, tutor, instructor, assistant or as an information presenter.

This thesis supports the idea that Industry 4.0 will be another field of application for IVA. These IVA will increase the Operator 4.0's motivation and will engage them on the task at hand. These agents endowed with certain artificial intelligence and human-like physical appearance will react to the Operator 4.0's interaction by means of non-verbal gestures, natural behaviour and verbal communication. This new kind of collaboration between humans, robots and virtual agents has been defined as Hybrid Team (Schwartz et al., 2016a, b).

Although IVA are commonly used, their design, development, animation and integration are not easy tasks. Research has outlined a number of desirable attributes for these kind of agents (Ahmed, 2005). This chapter presents some prove of concept to achieve IVA capable of communicating in a realistic way with humans. To accomplish this, certain aspects have been taken into account, mainly; realistic appearance, verbal and non-verbal communication and natural behaviour skills. The background of these three aspects is described in sections 2.1, 2.2 and 2.3, respectively.

Section 2.4 describes research achievements provided by application-driven research projects carried out within the field of IVA. These results prove the usability, scalability and integrability of such agents in diverse fields. Finally, Section 4.5 summarizes and highlights the contributions on this topic and enumerates several publications.

2.1 Realistic Appearance

Physical appearance of virtual agents (VA) can influence the way users behave (Merola and Peña, 2009; Martey and Consalvo, 2011; Martey et al., 2014).

An important attribute when defining VA's appearance is anthropomorphism, this is, the extent to which a character has either the appearance or behavioural attributes of humans (Koda and Maes, 1996; Nowak, 2004; Nowak and Biocca, 2003). In some studies, highly anthropomorphic VA have been rated as more credible, engaging, and likeable than less anthropomorphic images (Koda and Maes, 1996; Wexelblat, 1998). Moreover, Nowak (2004) found that the strongest predictor of these variables, was the degree of masculinity or femininity of VA.

Therefore, gender has also been a matter of study (Zanbaka et al., 2006; Sloan, 2015). Gender stereotypes that are common in social interaction studies (Carli, 1990; Eagly, 1978) replicate with virtual characters. Male participants are more persuaded by female speakers than male speakers, and female participants are more persuaded by male speakers than female speakers (Zanbaka et al., 2006).

In virtual learning environments, students feel higher positive affect and self-efficacy after interacting with female VA (Plant et al., 2009). The study carried out by Feng et al. (2017) demonstrate evidence that female characters should be considered in favour of male characters when designing games and interventions that simulate negatively-valued and/or emotionally-charged social situations such as a conflict-resolution.

Yee and Bailenson (2007) found that height of VA had an influence on the users. Users with tall VA were more confident with offers in a bargaining task, while those assigned short VA were more likely to accept an unfavourable decision.

Physically attractive avatars are rated higher on social competence, social adjustment, and intellectual competence. Research suggests that attractive VA are significantly more persuasive in changing the user's opinion (Khan and De Angeli, 2009; Khan and Sutcliffe, 2014). These findings show that the physical appearance of VA has an influence on the user in an interaction-based context.

Finally, the colour of clothing can also influence the user's behaviour. For example, the experiment carried out by Peña et al. (2009), proved that participants using black-cloaked VA developed more aggressive intentions and attitudes but less group cohesion than those using white-cloaked VA. Another experiment compared the performance of red and blue teams in a popular multi-player first-person-shooter (FPS) computer game. Results outlined that red teams offered an advantage over blue teams in virtual competitions (Ilie et al., 2008).

Regarding realism, McCloud (McCloud, 1994) stated that individuals see themselves as iconic images but see others in a more detailed form, that is, as realistic images. Gulz and Haake (Gulz and Haake, 2006) extended this idea to the role of animated pedagogical agents and stated that if the agent is acting as a teacher, the student will see it as "the other person" and therefore it is better to represent it in a human form. However, the risk of falling into the "uncanny valley" (Mori, 1970) also exists. This term refers to the unsettling feeling humans experience when observing humanoid objects which appear almost, but not exactly, like real human beings.

Moreover, realistic appearance goes hand in hand with behavioural realism. Higher realism in the appearance of VA may lead to heightened expectations for behavioural realism (Slater and Steed, 2002; Garau et al., 2003).

2.2 Verbal and Non-Verbal Communication

Communication is simply the act of transferring information from one place to another. There are various categories of communication and more than one may occur at any time. The different categories of communication include (Tubbs, 2012; Knapp et al., 2013; Littlejohn and Foss, 2010):

- Verbal Communication: face-to-face, telephone, radio or television and other media.
- Non-Verbal Communication: body language, gestures, how we dress or act.
- Written Communication: letters, e-mails, books, magazines, the Internet or via other media.
- Visual Communication: graphs and charts, maps, logos and other visualizations can communicate messages.

Written and visual communication are used by humans constantly. They can be included in applications as text or images. However, they are not intrinsic to the capabilities of IVA. Therefore, written communication has not been considered. This work has focused on allowing IVA to communicate with verbal and non-verbal skills.

Verbal communication, as in speech, and non-verbal communication, as in, facial gestures, both imply facial animation. One of the key moments in the history of facial animation is the definition of the MPEG-4 standard (Bauer et al., 1999).

The standard sets three key features for facial animation:

- Facial Definition Parameters (FDPs), define the shape and texture of the human face.
- Facial Animation Parameters (FAPs), represent a complete set of basic facial actions, and allow for the representation of most facial expressions.

- Facial Animation Parameter Units (FAPUs), are defined in order to allow interpretation of the FAPs on any facial model in a consistent way.

Most studies regarding facial animation are based in these parameters (Ali et al., 2018; Yong, 2014; Leandro Parreira Duarte et al., 2014).

According to Radovan and Pretorius (2006) the different methods for facial animation can be classified in three groups: i) based in geometry, ii) based in images and iii) based in real movements.

The pioneer work of Parke (1972) in the field of facial animation can be classified in the first group. This and other primitive works transform the vertices one by one in each frame of the animation. Later on, new works based on pseudo-muscles, which simulate geometrically the effects of muscles or directly simulate the physical behaviour of the muscles and the skin were developed (DeRose et al., 1998; Wang, 2010).

Other methods, usually focused on the film industry, are based in images instead of geometry. Facial animation based in image manipulation can be categorized into the following techniques:

- morphing, different key images are interpolated to obtain the animation(Su and Liu, 2001).

- texture manipulation, the changes in the texture of the facial mesh create the animation (Fei, 2001).

- blend-shape interpolation, the intermediate frames between two modeled faces are computed (Huang et al., 2011).

Facial animation based in real movements has also been widely studied. Beeler et al. (2011) analyze the captured data to recognize predefined motions and launch these predefined movements in the virtual face. In the work presented by Arghinenti (2011) the captured data is only a layer of the facial animation engine. Other layers cope with expressions, phonemes and muscles. Deng et al. (2006) combine a motion capture system with a face scanning system in order to obtain a realistic animation with low-level details, such as wrinkles. Once they obtain the minimum number of key expressions, blend-shape interpolation is used to generate the final animation.

In visual speech synthesis, apart from the facial animation technique, one of the key aspects to obtain a realistic animation is the interpolation between the phonemes that have to be synthesized. One of the first works in this field and one of the main references is the one presented by Cohen et al. (1993). They create exponential functions for facial parameters that rise until the phoneme's exact time and fall afterwards. The so called co-articulation is obtained by the combination of different phonemes and the functions for their associated facial parameters.

Since the first works in co-articulation were published, several methods have been presented that take speech unities and convert them in a fluent and realistic facial animation: rule-based, Markov models (Yamamoto et al., 1998), neural networks (Massaro et al., 1999), to name but a few.

Regarding non-verbal communication, McNeill (1992) explains that in some speech contexts about three-quarters of all sentences are accompanied by hand gestures. The scientific community has established four type of hand gestures (Cassell et al., 1994):

- Iconic: or illustrators, they are descriptive gestures often used to illustrate speech.
- Metaphoric: or representational gestures, represent an abstract feature concurrently spoken about.
- Deictic: indicate a point in the space.
- Beats: small formless waves of the hand that occur to emphasize words.

Hence, body animation is also required for non-verbal communication.

MPEG-4 standard (Bauer et al., 1999) also defined the Body Definition Parameters (BDPs). MPEG-4 defines 68 FAPs and 186 Body Animation Parameters (BAPs) which are control parameters of animation. Detailed character movements such as hip twisting are defined in BAPs. However, the number of the BAPs is too large to be implemented efficiently. For this reason, recent character animation methods do not use the Face and Body Animation object defined in MPEG-4. Most methods of 3D character body animation can be categorized into Skeleton-based and Physics-based approach.

The Skeleton-based approach animates the character body by characterizing the relation between underlying skeletal structure and motion. These methods are called Skeletal Subspace Deformation (SSD) or Skinning. Magnenat-Thalmann et al. (1988) proposed a Joint dependent Local Deformation (JLD) operator which is able to generate object grasping animation of hand through the binding between joint and local deformation. Nowadays, this method is referred as Linear Blend Skinning (LBS). With LBS the mesh is deformed according to skinning weights which associate each bone with a set of the mesh vertices Song et al. (2015).

2.3 Natural Behaviour Skills

Agents capable of performing facial expressions, hand-arm gestures, eye-gaze and conversational abilities are known as Embodied Conversational Agents (ECA) (Cassell, 2000). Research has outlined that users find IVA with autonomous conversational behaviours to be more natural than IVA whose behaviours are controlled (Cassell and Vilhjálmsson, 1999).

Providing natural behaviour skills to IVA is a widely studied issue. However, defining natural behaviour is not an easy task. The literature and theory of affective computing imply several conditions for synthesized motion to appear natural (Abrilian et al., 2005). Speed of interaction and emotion/speech-correlated believable body motion are among the most important functionalities (Mlakar and Rojc, 2011).

Diverse XML-based languages have been developed to define the behaviour of ECA.

The Behavior Expression Animation Toolkit (BEAT) was the first toolkit to automatically suggest appropriate gestures, communicative facial expressions, pauses, and intonational contours for an input text. It also provided the synchronization information required to animate the behaviours in conjunction with a character's speech (Cassell et al., 2001).

Multimodal Utterance Representation Markup Language (MURML) defines a notation for gesture specification and synchronization with co-verbal speech. MURML also focused on specifying the actual form of body or face behaviours. (Kopp et al., 2003)

The Affective Presentation Markup Language (APML) specifies the agent's behaviour at the meaning level. APML represents, among others, communicative intent, emotions, interaction and cultural aspects (De Carolis et al., 2004).

The Rich Representation Language (RRL) was developed in the NECA project (Krenn, 2003). RRL is a special purpose markup language which represents a wide range of expert knowledge required at the interfaces between the different components in the NECA architecture (Piwek et al., 2004).

Finally, BML (Kopp et al., 2006) is a markup language oriented to describe verbal and non-verbal behaviour of IVA. BML defines elements like gestures and facial expressions and also allows specifying their temporal alignment. BML has become a standard and other markup languages allow compatibility with it.

For example, the aim of PML (Scherer et al., 2012) is to specify knowledge about the environment an non-verbal behaviour to IVA and it includes compatibility with BML. Moreover, as the authors explain in Čereković et al. (2010) BML has some advantages, it is intuitive and simple to implement.

In addition, there are several research projects based on BML whose main objective is to generate non-verbal behaviour for virtual characters. There are several interpreters like SmartBody (Thiebaux et al., 2008), BMLRealizer (Árnason and Porsteinsson, 2008) or Elckerlyc (Welbergen et al., 2009) that animate virtual characters receiving a BML archive as input. However, to generate that input, there is only one behaviour planning developed (Lee and Marsella, 2006). This non-verbal generator analyzes the syntactic and semantic of a text and the affective state of a character to decide the behaviour.

An interesting tool for the creation of interactive applications with multiple virtual characters is Visual SceneMaker (Gebhard et al., 2012). It is planed to add BML and other markup languages to integrate non-verbal behaviour generation.

2.4 Challenge Analysis based on Case Studies

The previous sections have outlined a number of desirable features that IVA should have in order to seem natural and appealing for a human interlocutor. This thesis work, is based on the author's research developed to obtain realistic IVA with those features. Furthermore, these IVA have been integrated into R&D Projects of diverse nature which have proven their usability.

The research fields studied in this section include: E-learning, Marketing and Health Care. Section 4.5 discusses the results obtained and extrapolates them to the Operator 4.0 context.

2.4.1 E-Learning

The use of online learning or e-learning has increased significantly in the past years and it is expected to keep growing in the future. According to the latest survey by Stratistics MRC, the Global E-Learning Market was accounted for $165.21 billion in 2015 and is expected to reach $275.10 billion by 2022 growing at a Compound Annual Growth Rate (CAGR) of 7.5% during the forecast period (Costello, 2017).

Distance learning offers a series of benefits that traditional learning does not, for example in terms of mobility, affordability or flexibility e-learning happens to be much more suitable for nowadays lifestyle, moreover this technology has enabled increasingly dynamic and engaging learning experiences.

Future manufacturing enterprises will use interactive e-learning tools to facilitate students', apprentices', and new workers' gain in understanding of advanced manufacturing operations, also involving the use of new Information and Communication Technology (ICT) paradigms (Vallés, 2013).

Below two R&D Projects developed in the field of E-learning are presented. Their main contributions and experiences are summarized.

SPEEP

The main objective of the SPEEP project was the development of a Computer Assisted Language Learning (CALL) platform to help students improve their pronunciation of a foreign language through the use of speech, natural language processing and virtual reality technologies.

The platform had to deal with the following requirements:

- Input. The solution had to allow voice and/or text as standard inputs. In the case of voice, it had to be processed by a speech recognition engine to obtain the transcription automatically.

- Languages. The platform had to operate for Basque and English.
- 3D Reconstruction. The animation had to be done at two levels; the first one employing the natural videos and audio recorded from the webcam and the microphone, and the second one using a Text-to-Speech engine and IVA. This way, through the use of advanced techniques of co-articulation and 3D graphics, the platform had to generate 3D virtual reconstructions considering both the correct way to pronounce the input and the way in which the user had pronounced it.

As seen in Figure 2.1, the interaction platform designed in this project showed the head of the agent, not the body. The objective was to visualize the pronunciation of the words in a very realistic way. Therefore, only the facial animation of the virtual agent was taken into account.

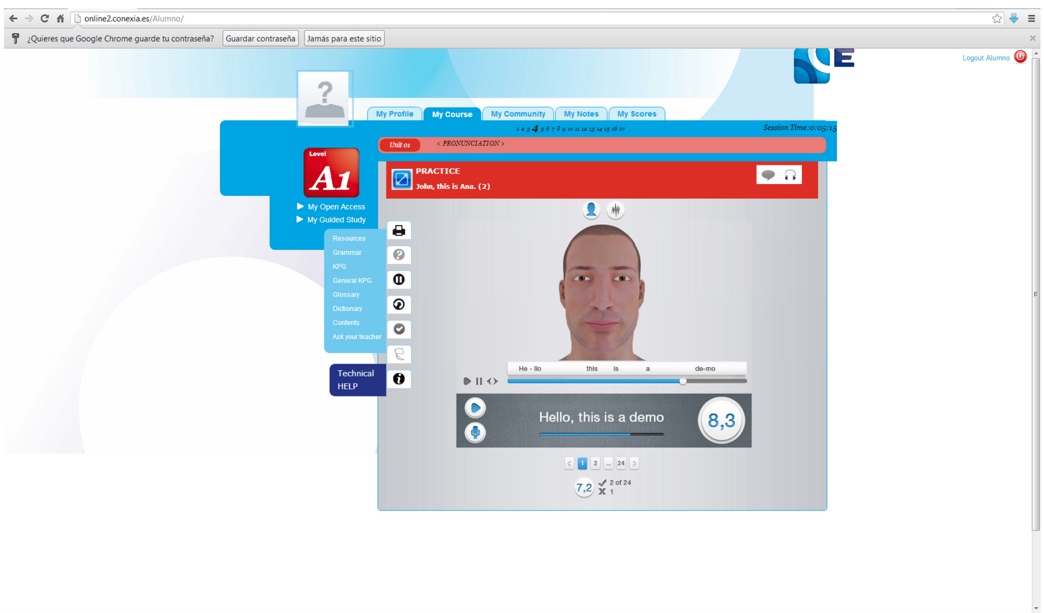


Figure 2.1: SPEEP R&D Project.

As explained in Section 2.2, there are three main groups of facial animation techniques.

In this R&D project, morphing was discarded. Although being a fast resource for modeling, it is limited to the number of variants made for each expression. Another problem is that morphing is based on the transformation of a big mass of vertices, this implies having very little control over any modification, so the risk of corrupting the geometry increases with the number of vertices.

The solution developed was based in pseudo-muscles. As the approach was based in the geometry, the use of bones as modifiers made the tool compatible with any humanoid agent. The configuration of face and body was valid for any model used, and no changes were made per agent as when using morphing. Once a configuration was modeled, it could be applied to any agent.

In order to have greater control over the different expressions, a facial bone structure composed by 64 bones was implemented. To achieve this, the movements of the different muscles of the face were studied and a bone hierarchy was applied to allow the movements of the facial bio-mechanical structure (Figure 2.2).

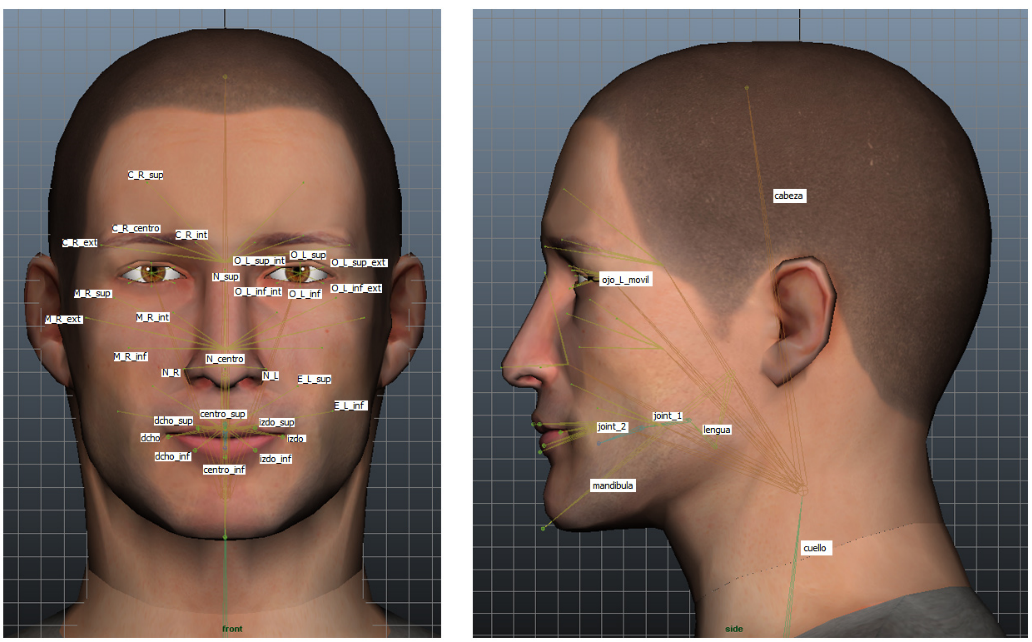


Figure 2.2: Face bone structure.

Once the animation control was defined, there was a need to synchronize speech and face animation. The lip-sync module developed was as follows.

The sentence to pronounce was generated using a text-to-speech synthesizer. The text introduced for each sentence was used to extract the corresponding phonemes, which were aligned to the synthesized audio file using an acoustic model obtained by training .

To configure the face, the FDPs of the MPEG-4 standard (Pandzic and Forchheimer, 2003) were used. During the rendering of IVA, in each frame the phonemes that surrounded the actual time were interpolated with the values of the corresponding FAPs following the method presented by Cohen et al. (1993).

A study of the computational cost of the different methods that are called in each frame during the facial animation showed that the most time consuming function was the skinning i.e. the computation of the new position of the vertices according to the transformations of their neighbour bones. In order to overcome this situation, the animation engine implemented a hardware skinning algorithm that delegated all bone transformations to the GPU. It was implemented as a GLSL vertex shader that takes the original mesh and the transformation matrices of the bones as inputs, and calculates the new vertex positions as output (Mujika & Diez et al., 2013; Mujika et al., 2014).

3DTUTOR

The aim of 3DTUTOR was the design, development and implementation of a virtual and smart 3D Tutorial system interoperable with multiple platforms and systems. It had to coordinate the immersive and personalised virtual teaching processes.

3DTUTOR was represented by IVA which role was to guide each student through the different lecture units. The objective was to allow the student to obtain knowledge in an autonomous and customised way, without the presence of a real tutor. The system should help the student through three different channels:

- Motivation through emotional intelligence.
- Evaluation of the student.
- Personalised teaching.

IVA were included in an immersive e-Learning system to improve the quality of asynchronous and distance education.

The real tutor was able to make a remote monitoring of the student's evolution thanks to the evolution profile generated automatically by the system. In addition, this evolution profile was useful to modify the answers of the virtual tutor and to adapt the teaching method.

Figure 2.3 shows the integration of a virtual tutor into Moodle Learning Management System.

Figure 2.3: 3DTUTOR R&D Project.

Unlike SPEEP project, in 3DTUTOR the whole body of the IVA was visualized. The virtual tutor behaved as real tutor, so it had to perform several body movements. Some examples of animations that needed to be implemented were; pointing out elements, spontaneous gestures with the hands, walking cycles, arm gestures or head movements.

Therefore, the facial animations were developed as described previously (2.4.1). For the body, a simplified structure of the human skeleton was applied, discarding all the bones that hardly reverted in modifications when making the movements needed for the required objectives. Altogether, 66 bones controlled the body (Figure 2.4).

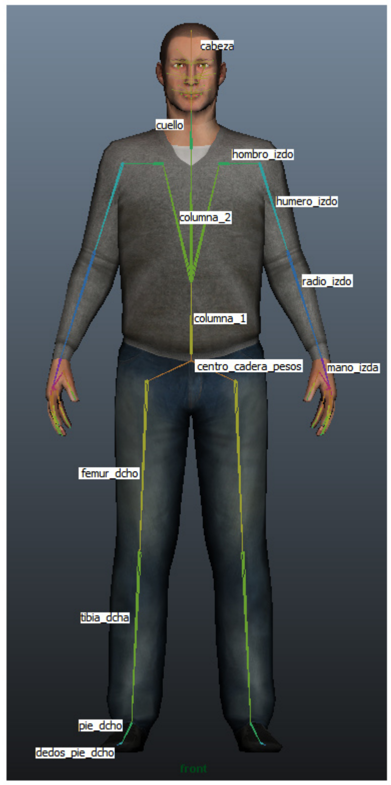


Figure 2.4: Body bone structure.

To achieve an autonomous virtual tutor, a behaviour module was developed (Diez et al., 2014b). Depending on the interaction of the user the virtual tutor reacted appropriately.

The student's interaction with the e-learning platform can be analyzed in different ways, for example, controlling the speed in which the student answers questions or the duration of each session. If a student takes too long to answer a question, the behaviour module will send an impulse to the IVA and it will react to this delay by asking some questions to the student; "is the exercise too difficult?", "are you tired?". The duration of each session can also be verified in order to evaluate the student's learning progress. Very short or long sessions might suggest the exercise does not match the student's knowledge level, in this case the IVA will suggest the student different exercises or will repeat the test to determine the student's level.

This kind of communication makes the students feel accompanied and they appreciate the concern showed by the IVA. This attention on the one hand makes them feel that their work is being valued and on the other hand infers some pressure on them, so it is more difficult for them to abandon the course.

2.4.2 Marketing

Recent studies have highlighted how business-to-business (B2B) firms have started to use digital marketing tools, especially the social media marketing, in the same way of business-to-consumer (B2C) firms (Wang et al., 2016). This new marketing approach of B2B companies is connected with the growing international competitiveness of industrial markets.

Next a case study developed for marketing is presented.

AVATARCO

The main objective of AVATARCO was to develop a web platform to easily generate interactive multimedia content for marketing actions. The introduction of IVA endowed with verbal and non-verbal behaviour was a key functionality of the platform.

The project proposed innovative solutions:

- Simple online edition of multimedia content including clips selection, spatial and temporal composition, the inclusion of effects like transitions.
- Create and edit IVA allowing the edition of their physical appearance, the voice and the behaviour (pose, gestures) according to the scene.
- Create HTML5 interactive content. Possibility to create traditional videos or create new interactive content to improve the experience of the audience.

- Manipulate online/offline contents, allowing the possibility of integrating contents through social media.

An authoring tool (Figure 2.5) was developed to allow users to easily define several aspects of the IVA:

- physical appearance (cartoon or realistic)
- automatic behaviour animation (body, limbs, face)
- clothing and complements
- gender, ethnic, age

The facial and body animation techniques used were based on previous work (Section 2.4.1). Which proves the modularity and re-usability of these developments.

The results from this R&D project are framed under the affective component of hybrid agents.

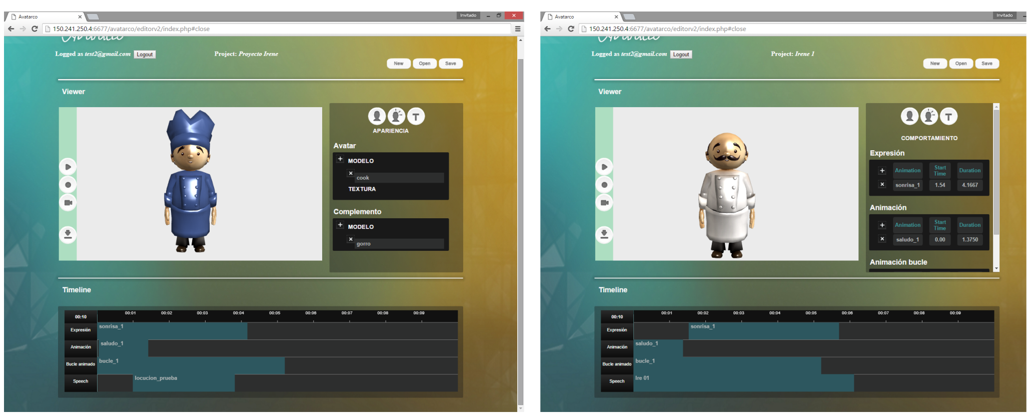


Figure 2.5: AVATARCO R&D Project.

2.4.3 Health Care

Safety and health are crucial aspects in Industry 4.0 (Romero et al., 2016a; Vallés, 2013). Health 4.0 concept has derived from the Industry 4.0 concept. It refers to digitally connected health care systems that will allow personalized and individualized health care support (Thuemmler and Bai, 2017). Health 4.0 will also reduce costs, improve management, optimize processes, increase efficiency and enhance the security planning.

Demographic changes will also have to be taken into account in a human-centred industry. Assistance systems will have to be introduced to meet the higher requirements of elderly or impaired workers and the increasing need for continuous worker qualification (Wolf et al., 2017).

Two case studies developed in the context of health care are presented in the following sections.

Health Assistant

The aim of Health Assistant was to develop intelligent and proactive IVA oriented to the health sector with the possibility of integrating them in data bases which contained information, diagnosis, treatments, prevention treatments, etc. The aim was to motivate patients in their therapies, specifically chronic patients.

The role of this assistant was to guide patients in rehabilitation therapies. Thus, the patient was guided by the IVA who motivated them and safeguarded the treatments.

The characteristics of the IVA were:

- Integrability: The goal was to integrate the IVA in different types of databases, such as wikis, CRMs (Customer Relationship Management), etc. and in any type of device or platform.
- Proactivity: Based on the information of databases (patient's profile, historic, diagnosis, etc.) the assistant suggested activities to the patient in a proactive way.
- Customization: The client was able to customize the appearance of the IVA in order to adapt them to the patient's preferences. In addition, it was possible to customize the behaviour and the way of communication adapting them to the patient. Besides, the IVA used the patient's preferences to suggest healthy activities or new habits.

- Expressiveness: The IVA could speak and had facial and corporal expressiveness. Thus, the assistant was capable of verbal and non-verbal communication.

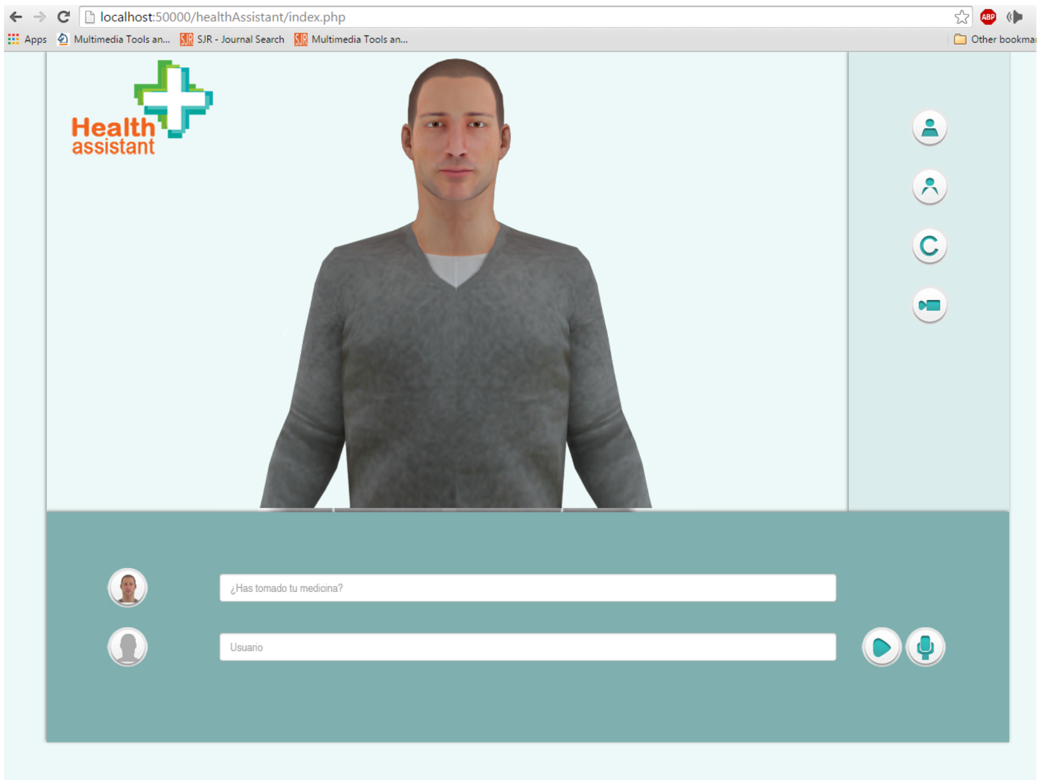


Figure 2.6: Health Assistant R&D Project.

Most of the achievements in this project are consequence of the modularity and re-usability of the IVA. The novelty of this research is the knowledge managed by the IVA. In the previous projects the IVA reacted to affective stimuli. In this case, the IVA made decisions and provided advice gathered from databases (Palmaz et al., 2016).

AVAC

The aim of AVAC project was to create a new product to ease the production of audiovisual content aimed at deaf people. The application used stenotype techniques (phonetic coding), voice recognition and IVA to achieve this goal.

Through the stenotype and voice recognition, the system was able to generate subtitles that were interpreted by IVA who performed Sing Language. The project was focused on audiovisual contents in Basque and Spanish. However the system was designed to integrate support for other languages.

The innovation of AVAC was based on:

- The creation of a Basque stenotype system.
- Translation from Basque to Spanish Sign Language (LSE).
- Adaptation during the execution of the signs, depending on the virtual interpreter's mood.

AransGi (Association of families of Deaf people of Gipuzkoa) evaluated the prototype with deaf people. The association had a huge interest on successful results and offered its support in tasks of advising and evaluation. The suggestions provided by the evaluators are summarized as follows:

- The sentences interpreted by the IVA were easy to understand. However, they suggested that the transition between signs should be smoother.
- The expression of the IVA had to be exaggerated. The emotions were recognized, but when people use sign language they accentuate their facial expression.
- Eyebrows, gaze and shoulders are also important for the correct interpretation of sign language. For example, eyebrows are fundamental to express questions or exclamations.
- The behaviour and attitude during the waiting state was correct.
- The IVA emulated the behaviour of a real person.

This project is the oldest among those mentioned in this section. Its results provided huge experience and founded the architecture and algorithms required by more advanced projects.

2.5 Chapter Summary and Related Publications

The case studies presented in this chapter show the suitability of introducing IVA to enhance the sensing and cognitive capabilities of the Operator 4.0.

The IVA presented act as Intelligent Personal Assistants (IPAs). IPAs are mobile, autonomous and software agents capable of performing tasks or services on behalf of humans (Garrido et al., 2010).

One of the operator typologies defined by Romero et al. (2016b) is the *smarter operator*, this is, an operator aided by an IPA.

This chapter has proven that IVA can help the *smarter operator* in interfacing with machines, computers, databases and other information systems in a human-likely way.

The IVA developed and presented in this chapter allow voice interaction with the user. To achieve this a co-articulation system based on pseudo-muscles was implemented (Mujika & Diez et al., 2013).

In several of the case studies, the IVA assess the user in an autonomous way; as a tutor guiding the user through the lectures, as a medical counsellor solving the patient's doubts (Diez et al., 2013).

Several behaviour rules have been defined to endow IVA with natural behaviour skills. Most of these rules are related to hand and arm movements or facial expressions, but some are also related to gaze, blinking or posture. These rules intend to make the IVA more lively and human-like by emulating the behaviour of real people (Carretero & Diez et al., 2012; Diez et al., 2014b).

The IVA's appearance can be entirely defined to encounter the users needs. AVATARCO authoring tool has proven to allow modeling of the IVA's appearance. It provides the designer with a user-friendly tool to define the IVA appearance (realistic/cartoonish), gender, ethnic group, clothing, etc (Carretero & Diez et al., 2016).

Moreover, the case studies presented demonstrate the modularity an reusability of the IVA developed as they have been successfully integrated in applications of diverse nature (e-learning, marketing and health care).

Related Publications

- (Carretero & Diez et al., 2012) Improving Gestural Communication in Virtual Characters. del Puy Carretero, M., Ardanza, A., García, S., Diez, H. V., Oyarzun, D., & Ruiz, N. In International Conference on Articulated Motion and Deformable Objects (pp. 69-81). Springer, Berlin, Heidelberg. (2012).

- (Diez et al., 2013) 3D Animated Agent for Tutoring based on WebGL. Diez, H. V., García, S., Sánchez, J. R., & del Puy Carretero, M. In Proceedings of the 18th International Conference on 3D Web Technology (pp. 129-134). ACM. (2013).

- (Mujika & Diez et al., 2013) Realistic Visual Speech Synthesis in WebGL. Mujika, A., Diez, H. V., Alvarez, A., Urteaga, M., & Oyarzun, D. In Proceedings of the 18th International Conference on 3D Web Technology (pp. 207-207). ACM. (2013).

- (Diez et al., 2014b) IAAN: Intelligent Animated Agent with Natural Behaviour for Online Tutoring Platforms. Diez, H. V., García, S., Sánchez, J. R. , del Puy Carretero, M., & Oyarzun, D. In Proceedings of the 6th International Conference on Agents and Artificial Intelligence-Volume 2 (pp. 123-130). SCITEPRESS-Science and Technology Publications, Lda. (2014).

- (Carretero & Diez et al., 2016) Providing Physical Appearance and Behaviour to Virtual Characters. del Puy Carretero, M., Diez, H. V., García, S., & Oyarzun, D. In International Conference on Articulated Motion and Deformable Objects (pp. 98-107). Springer International Publishing. (2016).

CHAPTER 3

Immersive VR/AR Environments

Virtual Reality (VR) can be applied in various engineering applications, such as product design, modeling, shop floor controls, process simulation, manufacturing planning, training, testing and verification (Mujber et al., 2004). Optimized decision-making in Industry 4.0 will be a combination of interactive virtual reality and advanced simulations of realistic scenarios.

Augmented Reality (AR) can also be applied to Industry 4.0 through several tasks, such as robot path planning, AR collaborative design, plant layout or simulation (Nee et al., 2012). They also present the main challenges for the future AR systems on the industrial scope, focusing on the human interaction with the environment. The present trend of manufacturing focuses on the modern ways of human interaction with the workstation in order to help on common tasks giving extra information about the workstation and the manufacture.

Although, research on VR is large, there is little consensus on pointing out its principal characteristics. For example, Brey (2008) defined it as: “A three-dimensional interactive computer-generated environment that incorporates a first-person perspective.” Søraker (2011) reviewed the definitions given to VR by various researchers (Steuer, 1992; Stanovsky, 2004; Brey, 2008) to extract the criteria to form a consensual definition. He stated that all the definitions discussed included some form of “computer simulation” and “interactivity”.

More precisely, Sherman and Craig (2002) listed four essential elements in VR: virtual worlds, sensory feedback (responding to user input), immersion and interactivity.

Zhao (2009) performs a thorough survey on the state of the art of VR. This work discusses the major contents and current situation of VR research in the aspects of VR modeling, VR representation technology, human-machine interaction and equipment, VR development suite and supporting environment, as well as VR applications, and indicate several problems of the theory and technology.

Taking these definitions into account, this chapter has focused its research on three characteristics of VR: Virtual Environments, Interactivity and Immersion. Sections 3.1, 3.2 and 3.3 present a general overview of these characteristics, which are also necessary in AR. Next, 3.4 analyzes each of these characteristics through the results obtained in diverse R&D projects in which virtual environments, interaction devices and immersion devices were integrated. Finally, Section 3.5 summarizes the chapter and lists a number of publications.

3.1 Virtual Environments

Virtual Environments (VEs) are the first challenge stated by Posada et al. (2015) (Table 1.1) within the integration dimension of Industry 4.0. Among other cases, there is a need for visually empowered 3D simulation scenarios for testing different configurations. In the relation of Industry 4.0 with human factors, the simulation of Virtual Environments is also crucial for the interaction between humans and robots. Visual simulation for emergency response in factories is also a challenge in Industry 4.0 and requires Virtual Environments.

Overall, Virtual Environments for Industry 4.0 have to cover many different challenges, such as 3D modeling and representation, real-time simulation, animation, interaction and rendering. This section analyzes the requirements needed in Virtual Environments to overcome these challenges.

Bell (2008) analyzes the different definitions given to the term virtual world (Bartle, 2004; Castronova, 2008; Combs, 2004) and provides a common definition. This combined definition is as follows: “A synchronous, persistent network of people, represented as avatars, facilitated by networked computers".

According to Schroeder (2008), "virtual worlds are persistent virtual environments in which people experience others as being there with them". The sense of "being there" depends on the accuracy in obtaining an immersive and presence feeling in those virtual environments.

Another study of the concept was held by Girvan (2013), who studied 88 peer-review articles which offered a definition of virtual world. The analysis resulted on the following definition for virtual environment:

"A persistent, simulated and immersive environment, facilitated by networked computers, providing multiple users with avatars and communication tools with which to act and interact in-world and in real-time" (Girvan, 2013).

Akchelov and Galanina (2016) also performed a review of 44 peer-review articles to form the following definition:

"Computer-based, three or two dimensional environment or space, that can simulate real world where users represented by avatars are able to communicate or interact simultaneously or synchronically" (Akchelov and Galanina, 2016).

More precisely, for the Industry 4.0, the virtual environment for manufacturing should include (Mujber et al., 2004) :

- Functionality: the virtual prototype should be clearly defined and realistically simulated to address product functionality and dynamic behaviour.
- Human interaction: the human functions involved must be realistically simulated, or the human must be included in the simulation.
- Environment: an off-line computer simulation of the functions can be carried out, or a combination of computer off-line and real time simulation can be carried out.

3.2 Interactivity

One of the key aspects addressed by Industry 4.0 is the improvement of human-machine interaction (HMI) paradigms. Cyber-physical systems (CPS) are central to the Industry 4.0 vision. The term CPS refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities (Baheti and Gill, 2011).

Posada et al. (2015) point out a number of interaction possibilities that open with the use of emerging information and communication technologies (ICT) (*Figure 3, page 29*):

- Enhanced human-machine cooperation: including human interaction with robots and intelligent machines.
- Improved human-in-the-loop interaction between the cyber and physical worlds.

The coexistence of humans and robots or the appearance of radically new ways of interaction in factories (gestures, mimics, haptics, etc.) entails new challenges. This section analyzes the interaction challenges involved in VR.

Interactivity refers to the ability to navigate virtual worlds and to interact with objects, characters and places (Stanovsky, 2004; Brey, 2008).

Defining interaction techniques in virtual environments has been broadly studied(Hand, 1997; Bowman et al., 2001, 2004). Jankowski and Hachet (2013) present a review of the state of the art of non-immersive interaction techniques based on the following interaction tasks:

- Navigation: is related to the motor task of moving the viewpoint through an environment. If it includes a cognitive component, it is referred to as wayfinding.
- Selection and Manipulation: Another typical task performed in a 3D virtual environment is object selection and its direct manipulation. Interaction techniques for the 3D manipulation include three fundamental tasks: object translation, object rotation, and object scaling.
- System Control: Related to the communication between user and system which is not part of the VE. Techniques that support system control tasks in 3D are classified as: graphical menus, voice commands, gestural interaction, virtual tools with specific functions.

3.3 Immersion

Immersion and interactivity are terms hard to distinguish. Figure 3.1 shows levels from low to high of these concepts and some technologies which support them.

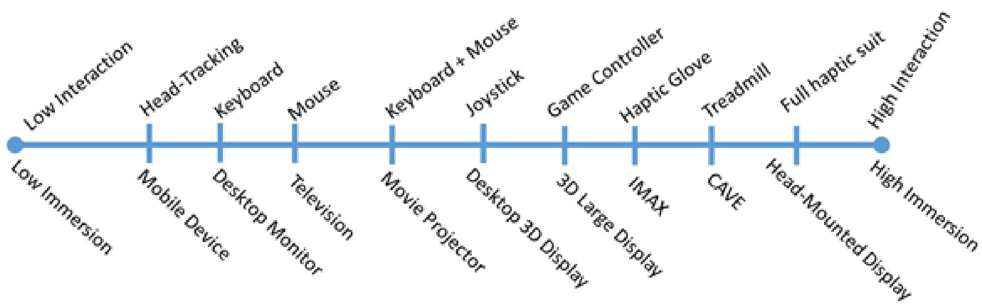


Figure 3.1: Immersion and interactivity from low to high with example technologies (Klippel et al., 2015).

This section studies the influence of immersion in VR experiences and analyzes its applicability in Industry 4.0.

Regarding immersion itself, Slater and Wilbur (1997) stated the degree of immersion could be objectively assessed depending on several characteristics such as "the extent to which a display system can deliver an inclusive, extensive, surrounding and vivid illusion of virtual environment to a participant" and "the extent of body matching, and the extent to which there is a self-contained plot in which the participant can act and in which there is an autonomous response".

Brown and Cairns (2004) identify three levels of immersion: engagement, engrossment and total immersion. Total immersion rarely occurs, normally users experience engrossment, in which the game controls become invisible and the users feel more involved in the story, or engagement, which might be influenced by interactivity (Choi and Baek, 2011).

However, Brey (2008) does not consider immersion to be fully necessary to experience virtual reality. One of the reasons he argues is that total immersion depends on certain devices which are not available for most users. In this way, Brey prefers the concept "first-person perspective" to "immersion".

Regarding Industry 4.0, Posada et al. (2015) do not explicitly categorize immersion into any of their three criteria challenges. However, the R&D project named MACHS presented in this work (Posada et al., 2015) is an example of how visual computing technologies can address some human factors requirements in Industry 4.0.

The goal of this project was to design an immersive experience to train machine-tool maintenance staff (Mujika et al., 2011). MACHS project was championed by the Danobat industrial group (part of Mondragon Corporation, one of Europe's largest industrial groups). The project contains two main tools. The first one: an authoring tool where an expert in machine-tool manufacturing can design a course for students that are learning how to use a machine or how to do its maintenance. The second tool, receives the output of the first one and runs a course in a 3D virtual environment where the student must follow the steps designed by the expert.

Similarly, Schuster et al. (2016) tested a collaborative virtual learning environment (VLE) to prepare engineering students to meet the demands of Industry 4.0. This work shows how students like to be immersed by the virtual environment and to interact with it intuitively and naturally.

In the last years, the new generation of Head-Mounted Displays (HMDs) have become popular devices to provide highly immersive experiences which are very difficult to achieve by other means. They are excellent tools to view virtual environments in a realistic way. However, these devices are still reserved for few users. Moreover, these devices do not offer interactive capabilities with the objects in the VE. Additional devices are attached to the HMD set-up to overcome this situation.

3.4 Challenge Analysis based on Case Studies

The previous sections have outlined a number of characteristics for VR; Virtual Environments, Interactivity and Immersion. These characteristics have been integrated into different applications in the following fields: safety, marketing and manufacturing.

These applications are briefly described in the following lines. Section 3.4.1 describes de different virtual environments designed, Section 3.4.2 describes the interaction devices integrated and Section 3.4.3 shows he immersion devices used.

Each section concludes with a summary of the challenges and contributions.

PREST Project

The research context in PREST project was interactive training on occupational hazard prevention and more precisely in fire safety in buildings. The fire warden trainee had to deal with the fire by a) finding the safest evacuation route or b) putting the fire out.

Two virtual environments were designed for this project; a warehouse and an office building, these are described in section 3.4.1. A gamepad and the Leap Motion Controller were used as interaction devices, as explained in section 3.4.2. Oculus Rift Head-Mounted Display (HMD) was used to provide a fully immersive experience (see section 3.4.3).

The results and challenges encountered in this project were published in Diez et al. (2016) and Diez et al. (2017a).

IDISS Project

The aim of IDISS project was to improve the interaction between Digital Signage systems and potential customers within marketing research context. A 3D virtual agent was integrated into the platform to offer a more natural interaction between the device and the audience, assessing them in their requests and adapting the products to their needs.

The target of the platform were Mall visitors, so a virtual environment reproducing this scenario was designed (see section 3.4.1). The audience could interact with the Digital Signage system through hand gestures that were tracked using a Kinect (see section 3.4.2). Large screen monitors were used offering a semi-immersive VR experience.

Futurad company introduced this solution in their Digital Signage marketing channels throughout Panama (Futurad, 2012).

Virtual Operator Case Study

This case study presents a Virtual Operator, understood as the collaboration of an Operator with VR (Romero et al., 2016b).

The concept here is as follows, a collaborative robot picks up parts from the working environment. Then it places them in the platform of a 3D scanner or in a conveyor. From time to time, an operator replaces that part for another. In these cases, the robotic arm must detect the presence of the arm or hand of the operator and stop its movement to avoid the collision with the operator. Trying out this situation with humans could be dangerous. Thanks to the developed virtual training program, the operator can perform its work in a safe way.

The virtual environment designed in this case study is explained in section 3.4.1. Leap Motion Controller was used as interaction device (see section 3.4.2) and a fully immersive experience was achieved thanks to the Oculus Rift HMD (see section 3.4.3).

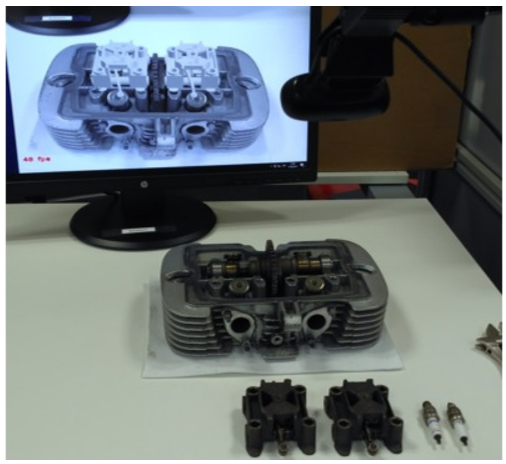
The results from this case study are under a second review publication process for the International Journal: Computers & Industrial Engineering on the special issue: “The Operator 4.0: Towards Socially Sustainable Factories of the Future” (Segura & Diez et al., 2018).

Augmented Operator Case Study

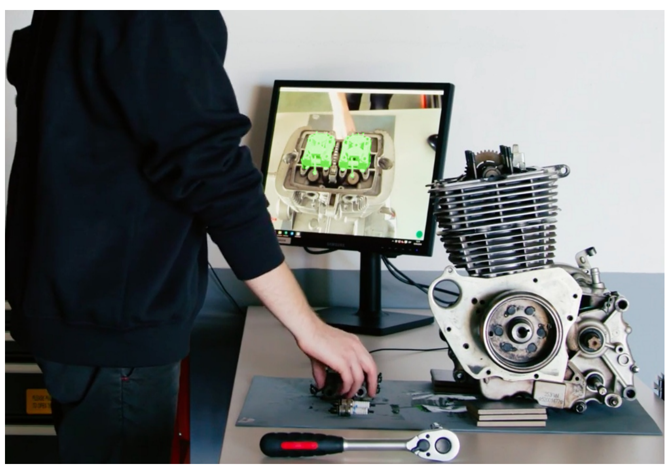
This case study was developed for the automotive industry dealing with assembling or maintaining engines (see Figure 3.2 (a)). The idea was to facilitate the training of newcomers or to assist operators through the assembly process, avoiding the need of physical manuals. Cognitive load is reduced by placing the elements of the assembly process directly in the operators perspective of the surrounding environment. This way the performance is not dependent on the operator remembering all operations required for a particular product.

The AR system was specially designed using markerless technologies (Teichrieb et al., 2007) in order to avoid polluting the station with artificial markers and interfering with industrial critical processes. Any kind of media content was displayed to the operator with stable and real time rendering. The system was conceived as an easy-to-use and cost-efficient tool. Thus, the elements involved were a web camera and a computer with a monitor. With those few elements the system performed the main processes required for the application.

The results from this case study were published in Ugarte et al. (2016).



(a) Preparation of augmented instructions.



(b) Augmented operator in the workshop.

Figure 3.2: Augmented Operator - Assisted assembly of an engine using AR.

3.4.1 Virtual Environment Description

The following section describes the diverse VEs that were designed in the previously summarized projects.

Warehouse

This VE was designed to train experts on occupational hazard prevention and more precisely in fire safety in buildings. Thus, a two floor building with a warehouse unit on the ground floor and an office unit on the first floor was chosen as a suitable scenario.

Both floors are connected by a lift and two staircases can also be found at the front and rear of the warehouse. Exit from the first floor is also possible by an external staircase.

The content of the building was also taken into account when designing the VE. Hazardous or flammable substances were placed to improve the simulation of fire propagation. Figure 3.3 shows how the ground floor of the warehouse was packed with wooden pallets and cardboard boxes. The possible interaction of the fire with these objects was considered as it could influence the evacuation routes.

Figure 3.3: Warehouse - Virtual Environment.

Office Building

In this case, an office building was designed to train wardens in evacuation tasks. Vicomtech Research Center was used to model this VE to make it as realistic as possible.

As seen in Figure 3.4, this environment had 16 office units, 2 bathrooms and 3 common areas. The content in this environment was mainly compound by office material and furniture.

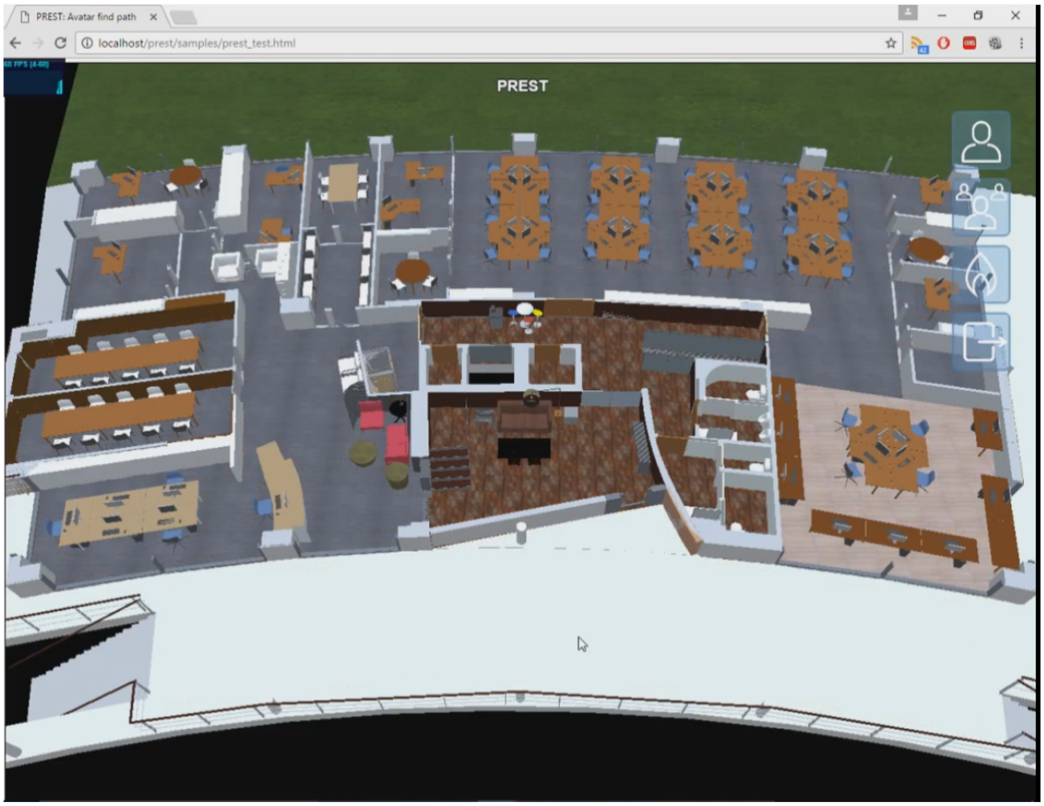


Figure 3.4: Office Building - Virtual Environment.

Mall

This VE was designed for marketing purposes. The aim of this R&D project was to improve the interaction between Digital Signage systems and the audience using 3D Interactive Virtual Agents (IVA) to provide a more natural communication and by adapting contents according to the audience.

The VE designed for this purpose was a Mall, with shop windows in which the audience could view the available products (Figure 3.5).

The IVA was modeled following the characteristics described in Chapter 2. The appearance of the IVA was designed to resemble the audience to which it was addressed.

Figure 3.5: Mall - Virtual Environment.

Virtual Operator

Vicomtech's laboratory has the UR10 robotic arm from Universal Robots (Universal-Robots, 2018) and the 3-finger adaptive robot from ROBOTIQ (RobotiQ, 2018). Therefore, a simple scene representing this set-up was modeled for this case study (see Figure 3.6).

The collaborative robotic arm and the gripper device were modeled with the same characteristics as their real counterparts.

UR10 has six rotating joints with 360-degree rotation freedom. Its maximum speed is 120-degrees per second for the base and shoulder joints and 180-degrees per second for the elbow and the three remaining wrist joints.

The 3-finger Gripper can almost pick anything with its 4 grip modes: pinch, wide, scissor and basic. Fingers can also be controlled separately and get feedback from each of them.

Their Digital Twins were linked in real time with their real counterparts, providing real physics and behaviour to the Digital Twins.

A virtual operator was also modeled and represented by an Interactive Virtual Agent (IVA), which was designed using the characteristics described in Chapter 2. This virtual operator was capable of handling the 3D objects of the scene.

The 3D objects were easily imported into the platform using the 3D model conversion and management systems explained in Chapter 4.

Figure 3.6: Virtual Operator - Virtual Environment.

Conclusions

Four different VEs have been presented in this section. Different rendering platforms were used; Unity Engine and Web browser.

Modeling a realistic VE is not an easy task, it involves physics, environmental effects, photo-realistic graphics, lightning, sound effects, and so on.

The level of detail that each rendering system can support is different. Thus, the VEs should be designed taking this into account.

However, the state of the art is constantly changing and it is hard to predict what capabilities future devices will support. As a consequence, any new application requires to find an equilibrium among different factors: geometric model detail, textures, physics, environment, etc. This balance will also depend on current HW performance and availability.

3.4.2 Interaction Devices

Several interaction devices were integrated in the previously mentioned projects, this section explains each of them.

Leap Motion Controller

Leap Motion Controller is a small USB device that scans an area of roughly 8-cubic feet above it. It tracks both hands and all ten fingers. With the release of the Orion software, the Leap Motion Controller can be used with VR headsets such as the Oculus Rift and HTC Vive.

This setup was used in PREST R&D project. Leap Motion Controller was attached to the Oculus Rift HMD, to detect the trainee's gestures: fist, pinch and thumbs up gestures were detected (Figure 3.7).

Fist gesture detection was used to evaluate whether the trainee was opening or closing doors. The "fist" gesture was chosen for this action as it is the one performed in real life to grab door handles.

Pinch gesture was used to detect the user grabbing objects in the scene, fire extinguishers in particular. The full action process implied the user getting close to the extinguisher, reaching an arm towards it and pinching it by holding the index and thumb fingers together.

Finally, to simulate the squeezing of the nozzle, the following gesture was tracked; one hand of the user performing a "thumbs up" gesture, closing the hand with the thumb heading up. This gesture was chosen because it is similar to the one users should perform in real life and to differentiate it from the "fist" gesture, in which the thumb is not visible.

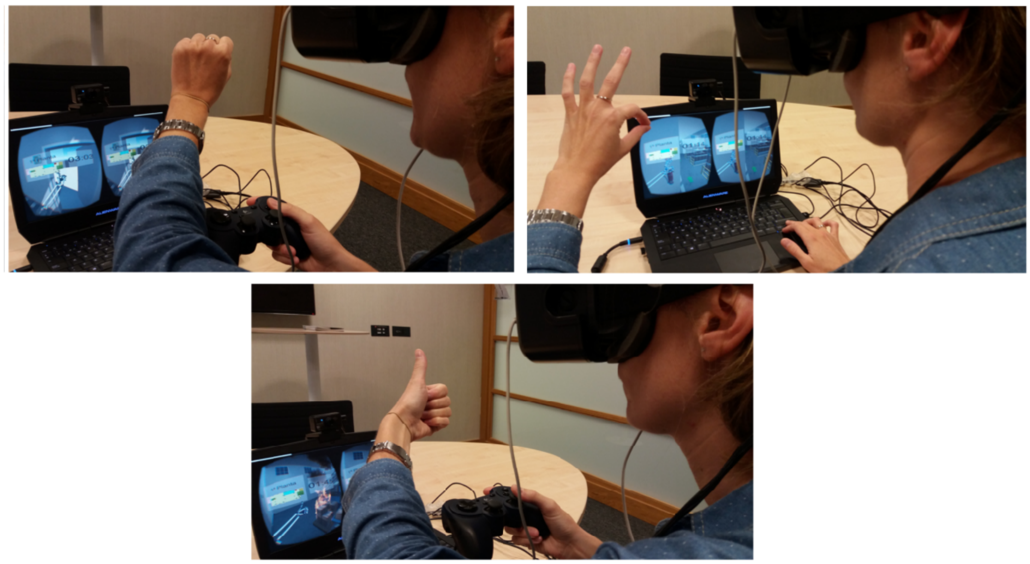


Figure 3.7: Leap Motion Controller used to detect hand gestures in PREST R&D project.

Leap Motion Controller was also used in the Virtual Operator case study (Figure 3.8). The operator's hands were visualized in the virtual scene and their movements were tracked in real time. If the Digital Twin of the collaborative robotic arm detected a possible collision with the virtual hands, it automatically stopped preventing any dangerous situation. The safety for the trainee is guaranteed since the interaction is purely virtual, but with real physical behaviour.

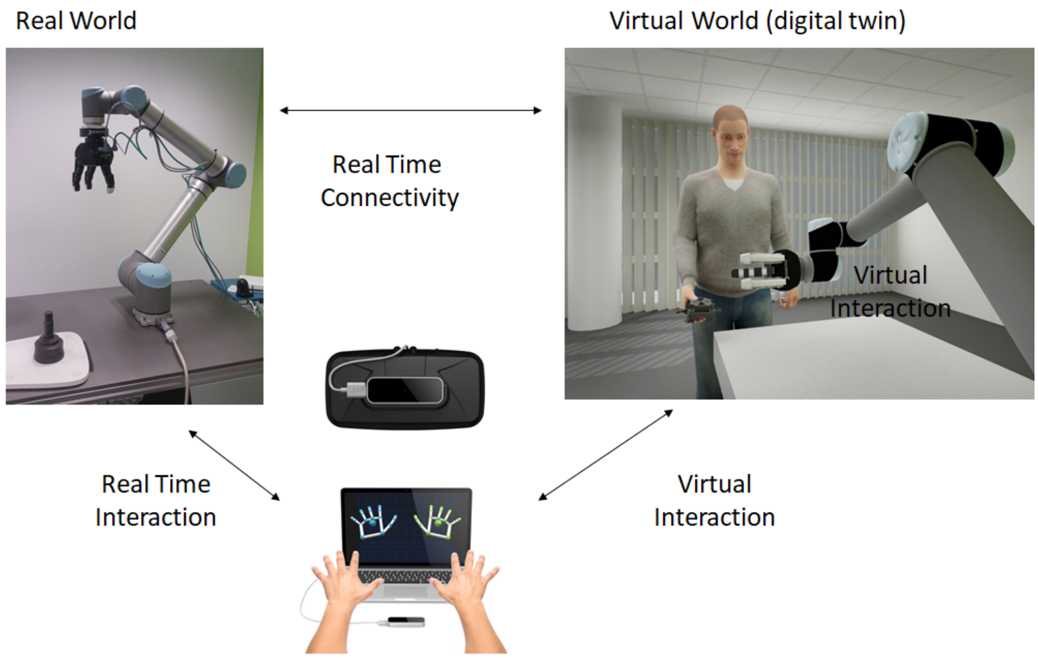


Figure 3.8: Setup for the Virtual Operator case study using Leap Motion Controller.

Gamepad

Gamepads are familiar interaction devices mainly used in video games. The Logitech Rumblepad2 gamepad was used for navigation purposes into PREST R&D project.

Instead of using the keyboard to navigate a gamepad was used. In this case both hands were needed to hold the gamepad and the user had to let go of one hand to perform the hand gestures for selecting and manipulating objects of the VE.

In most applications, the right stick of the gamepad is used for navigating (forward, backward, left and right) and the left stick for turning the camera view. This left stick was not necessary in this case as the user could look around by moving the head with the 360-degree tracking system offered by the HMD. So, in the gamepad the user only had to use the right stick, none of the other buttons were needed. This made it simple for the user, despite wearing the HMD.

Kinect

A method for detecting and tracking the user's hand gestures was implemented in IDISS R&D Project. For this purpose, Kinect interaction device was used.

The Kinect device captures video with an integrated camera and a depth map using infrared sensors. With the help of OpenNI and Nite APIs the system is able to detect and track human body parts and perform gesture recognition.

The system developed detected the human body and the head and the right and left hand positions. The 3D position was projected into 2D screen coordinates to obtain the distance between the user and the camera. This way the interaction was restricted to users that were facing the camera and close enough to it, avoiding interaction with people passing by, as a Mall is a very busy space.

In order to perform gesture recognition the user's hands were segmented from the rest of the body and tracked. The Nite API allowed to track and detect the click gesture, the waving gesture and the rising hand gesture. The coordinates of the hand were also converted to screen coordinates so the user could use the hand as a mouse for selecting or clicking objects in the Digital Signage system.

Conclusions

Interaction presents many challenges in VR. It has to be intuitive in order to result natural for the user. Hand gestures used to interact with the VE must be both natural to perform and easy to understand. The user must be able to easily correlate the hand gesture needed in real life with the one performed for the virtual experience.

Furthermore, the tracking systems must be very robust or the experience will be disappointing for the user. Having to repeat the same gesture, without getting the expected interaction feedback can result extremely frustrating.

3.4.3 Immersion Devices

The immersion devices tested are explained next.

Oculus Rift

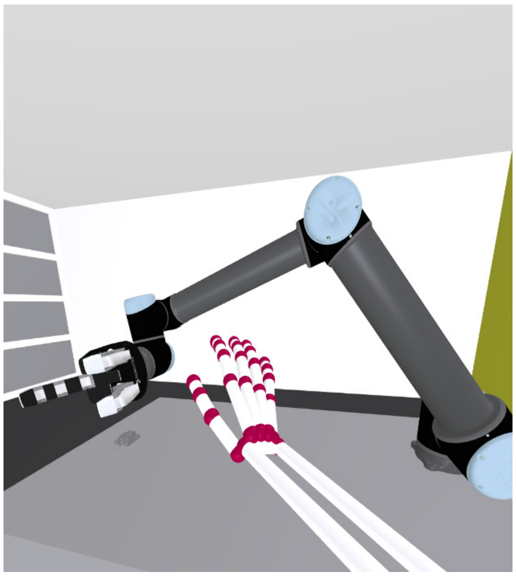
Oculus Rift HMD was used in PREST project and in the Virtual Operator case study. More specific, Development Kit 2 (DK2) was used. The HW specifications of DK2 are as follows:

- High Resolution and Refresh Rate: 1920x1080 (960x1080 per eye) resolution and a maximum refresh of 75Hz.
- Low Persistence OLED Display: to help reduce motion blur and judder, significantly improving image quality and reducing simulator sickness.
- Positional Tracking: precise low latency positional tracker.
- Built-in Latency Tester: it constantly measures system latency to optimize motion prediction and reduce perceived latency.

The case studies using HMD were developed based on *WebVR*. The users could visualize the virtual scene from two points of view. Next, these points of view for the Virtual Operator case study are described (see Figure 3.9).

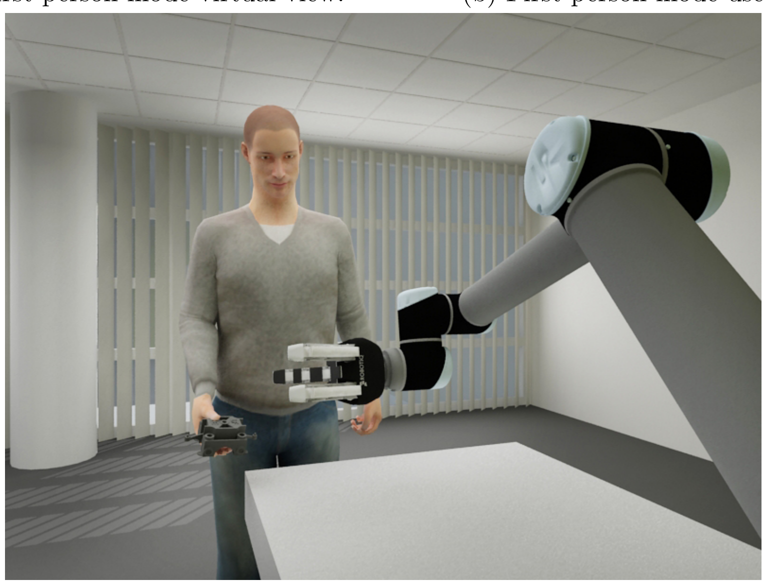
In the first person point of view mode, the virtual operator could visualize the real movements of the robot as the robotic arm and its Digital Twin were connected. The operator viewed the scene in an immersive way through the HMD. The Leap Motion gesture sensor attached to the HMD allows real time tracking of the hands of the operator.

In the third person point of view mode, the operator could view a agent interacting with the Digital Twin. This animated agent was endowed with natural behaviour, performed realistic movements and reproduced facial expressions, using the knowledge from Chapter 2.



(a) First person mode virtual view.

(b) First person mode user view.



(c) Third person mode virtual character interacting with Digital Twin.

Figure 3.9: Virtual Operator using Oculus Rift and Leap Motion Controller.

Conclusions

Interaction can be tiring, this is why, the seated position was defined in PREST project. The stand-up position poses additional problems like i) problems with the cabling, ii) collisions with the real-world furniture and iii) anxiety regarding the cognitive disassociation of the virtual position and the real position, specially when stairs are included in the scenarios.

A conflict between looking and walking control was found in some cases. Looking is controlled by head orientation while walking is controlled by interaction devices: some users immersion experience is confused by this interaction duality.

Immersive walking experiences within VE might be ruined when frequent collisions with the environment happens. So VE with many obstacles or narrow free paths may impair the immersive experience.

3.5 Chapter Summary and Related Publications

The case studies presented in this chapter show how VR can improve human abilities by means of intelligent Human-Machine Interaction (HMI) interfaces using Human-Computer Interaction (HCI) techniques.

This chapter summarizes research focused on three characteristics of VR: virtual environments, interactivity and immersion. Diverse VEs have been modeled, various interaction devices have been tested and different immersion displays have been used. Case studies in the fields of safety, marketing and manufacturing highlight the knowledge acquired in this field.

This chapter has proven that the Virtual Operator concept presented by Romero et al. (2016a) is possible. VR can replicate a manufacturing environment and allow the operator to interact with it with reduced risk and real-time feedback. The Virtual Operator case study presented in this chapter demonstrates that a safe working environment can be achieved combining VR and Digital Twins, which offer feedback in real-time.

VEs are one of the challenges relevant to Industry 4.0. This chapter shows various VEs modeled for various CPS tools. Several aspects must be taken into account when modeling 3D VEs; the final audience, the degree of realism, the rendering platform, the interaction devices and the visualization platform (Diez et al., 2016, 2017a).

The three CPS tools presented in this chapter have provided a wide range of study as their fields of application have been diverse.

HMI is a crucial aspect in Industry 4.0. Furthermore, new ways of interaction are constantly being developed (walking platforms, specific hand input devices, haptic gloves, etc.). The CPS tools developed in the context of this thesis have served to study challenges regarding HMI:

- The interactive multimodal platform in IDISS (Kinect) provided natural interaction tracking the users' hand gestures.
- The IVA integrated into the platform enhanced the HCI by allowing natural communication (Diez et al., 2014a).
- Finger gestures (Leap Motion Controller) proved to be useful in training fire wardens in the PREST project (Diez et al., 2016).

The three immersive experiences reported in this chapter provided a satisfactory experience (Diez et al., 2016, 2017a; Segura & Diez et al., 2018). However, HMD, as many other research works have pointed out, still have a limited usability time range. Our experiences (Diez et al., 2016, 2017a) show that:

- conflict between looking and walking interaction must be solved.
- systems must be designed in a way that user's movements are automatically improved to avoid collisions with the VE.

Related Publications

- (Diez et al., 2014a) Interactive Multimodal Platform for Digital Signage. Diez, H. V., Barbadillo, J., García, S., del Puy Carretero, M., Alvarez, A., Sánchez, J. R., & Oyarzun, D. In International Conference on Articulated Motion and Deformable Objects (pp. 128-137). Springer, Cham. (2014).

- (Diez et al., 2016) Virtual Training of Fire Wardens through Immersive 3D Environments. Diez, H. V., García, S., Mujika, A., Moreno, A., & Oyarzun, D. In Proceedings of the 21st International Conference on Web3D Technology (pp. 43-50). ACM. (2016).

- (Ugarte et al., 2016) Augmented Reality System to Assist in Manufacturing Processes. Ugarte R., Barrena N., Diez H. V., Álvarez H., & Oyarzun D. In Proceedings of the XXVI Spanish Computer Graphics Conference (pp. 75-82). Eurographics Association. (2016).

- (Diez et al., 2017a) Interaction Challenges in the Development of Fire Warden VR Training System using a HMD. Diez, H. V. , Moreno, A. & Garcia-Alonso, A. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP, 84-91. (2017).

- (Segura & Diez et al., 2018) Visual Computing Technologies to support the Operator 4.0. Segura, Á., Diez, H. V., Arbelaiz, A., Posada, J., A. & Garcia-Alonso. Computers & Industrial Engineering. (under second review, 2018).

CHAPTER 4

Interactive 3D Model Management

The main objective of the research presented in this chapter is to improve the multimedia management of complex 3D models, such as CAD models. The management should be as automatic as possible. Providing a solution to this problem will allow content managers to drag CAD files directly into a web application. In this way, it will be possible to create useful management tools for e-commerce content, Industry 4.0 content and other application areas.

The developments and tests presented in this chapter, show that it is possible to automatically convert these models into web compatible formats. The proposed solution is a compound of algorithms and processes that perform this task requiring no extra intervention from the user. The numerical results show that sharing 3D content on the web is immediate and simple. The automatic conversion mechanism has been also validated building an e-commerce management platform.

This work and its contributions were developed in the context of e-commerce (Diez et al., 2017b). However, as stated in the publication the results could be applied in other areas.

As stated by Posada et al. (2015), new improvements in the full lifecycle of products and increasing availability through Web3D in 3D CAD/CAM modeling will be a key enabling technology for innovative Insdustry 4.0 scenarios and emerging business models. Industrial 3D CAD/CAM models with full access to semantic/product dynamic data will be needed to allow a natural flow of the persistent and interactive digital model in Product Lifecycle Management accessible by Web3D technology.

This chapter is organized as follows; Section 4.1 explains the context in which this research was developed. Section 4.2 introduces the background of the research, Section 4.3 discusses automatic CAD to web conversion, Section 4.3 presents the results obtained, Section 4.4 validates the solution within an e-commerce platform and finally Section 4.5 summarizes contributions.

4.1 Research Context

Sales agents involved in e-commerce create and manage online catalogues to offer their products and allow potential customers to view and interact with them. Operators in the Industry 4.0 need to visualize and manage manufacturing content on ubiquitous platforms. In both areas, CAD data is commonly used. However, nor sales agents nor operators need to have technical background on this data format.

Embedding CAD data into web-based repositories has several challenges in contrast to common element types such as text and images.

- CAD file formats are not included in web standards and browsers do not directly support them.
- There is no standard 3D model file format that can be embedded in web pages, unlike image or video formats.
- The WebGL 3D rendering API can render polygonal models (triangle meshes), but CAD models are usually not composed of polygons but are defined in terms of higher complexity parametric shapes.
- When parametric CAD models are transformed into polygonal meshes, the result often contains a large number of polygons and the file is too large for an efficient use by a web client.

The main challenges for using such content on the web are thus to transform CAD files into formats suitable for web-based handling and rendering, and to present this content to client viewers in an efficient and user-friendly repository. That is, from the point of view of the CAD designer it has to be easy to add content to the repository. And from the point of view of the visitor, searching, accessing and viewing content has to be easy and quick.

The main bottleneck for the last challenge is time transfer. This can be reduced with mesh simplification algorithms, but at the expense of visual quality. The system will have to find a compromise between shape quality and file size (i.e. time transfer).

To deal with the above challenges, the integration of CAD data involves solving the following issues which are discussed in the following sections:

- Convert different native CAD file formats into a unique model format.
- Polygonal conversion: Convert non-polygonal data into a polygonal model. For example, curved surfaces into triangle meshes.
- Polygonal mesh simplification: Reduce the number of polygons that describe an object maintaining its appearance and making it suitable for the Web: fast transmission rate and efficient rendering.

These issues make it difficult to create and update web catalogues. Sales agents and operators, from now on users, have to deal with technical issues that exceed their knowledge field.

The web platform proposed in this chapter allows users -with no particular knowledge- to view, manage and interact with complex 3D models. It allows users to upload CAD models into the web. So, 3D model catalogues for e-commerce or Industry 4.0 can be managed in an easy and intuitive way. It also provides a collaborative mechanism to share models.

The platform proposed in this chapter solves the issues presented above in an automatic way. It offers a device independent, interoperable, plug-in free solution. The platform serves the purpose of an online catalogue; users can easily upload their 3D models by simply dragging their native 3D formats into the online platform. The system automatically converts these file formats into a web compatible file format and stores them in a database.

The platform must also provide a solution to other problems. The platform should be able to efficiently visualize the models on the web and it should include some interaction tools. WebGL technology has been chosen to render 3D models in browsers.

A web based platform will make sharing 3D models immediate, creating a collaborative work environment in which anyone with an internet connection enabled device (PC, smartphone, tablet) can interact with the models from the online catalogue.

4.2 Background

To reach the main goal, automatic model conversion, the platform should perform a sequence of operations on the original models to make them adequate to the Web. These operations include tessellation, polygon simplification and scene graph reduction. However, this operations entail various problems.

There are many different CAD formats. Each CAD product usually makes use of its own file format. This format proliferation makes it quite difficult to solve the automatic conversion problem. Some CAD tools have their proprietary conversion mechanism. However, managing web models from different sources adds a heavy load on content managers.

Section 4.2.1 addresses the problem of handling various CAD formats. Polygon simplification is discussed in Section 4.2.2 and simplification and scene graph reduction is explained in Section 4.2.3. Finally, Section 4.2.4 analyzes some existing platforms related with the problem.

4.2.1 CAD Formats

These are some of the most widely used CAD exchanging formats:

- AutoCAD DXF (Finkelstein, 2008) (Drawing eXchange Format) is a CAD data file format developed by Autodesk for enabling data interoperability between AutoCAD and other programs.

- IGES (Association et al., 2007) (Initial Graphics Exchange Specification) defines a vendor neutral data format that allows digital exchange of information among CAD systems (version 6.0 1998-01-05).

- STEP (STandard for the Exchange of Product) was developed as the successor of IGES; it is an ISO 10303-21 (Dunn, 1992) (last reviewed in 2012) standard for the computer interpretable representation and exchange of product manufacturing information.
- VRML (Virtual Reality Modeling Language) (Carey and Bell, 1997) is a text file format where vertices and edges for a 3D polygon can be specified along with the surface colour, transparency, etc.
- In 2001, X3D (Web3D-Consortium, 2001) arrived as an XML encoding of VRML. X3D provides both the XML encoding and the Scene Authoring Interface (SAI) to enable web applications to incorporate real-time 3D, presentations and controls into non 3D content.
- COLLADA (COLLAborative Design Activity) (Khronos-Group, 2011) defines an open standard XML schema for exchanging digital assets among various graphics software applications. Unlike X3D, COLLADA does not define the semantic in the 3D scenes, it is an intermediate format whose primary goal is to represent rich data in multiple forms, to enable the transformation of assets as they journey from content tools that use higher level description paradigms to applications that require platform specific optimized descriptions (Arnaud and Parisi, 2007).

None of the above formats are directly embeddable in web pages. The challenge is to select one or more native CAD input formats and to provide a tool that converts them into a format that can be interpreted and rendered in a web page.

4.2.2 Polygonal Conversion

CAD models are normally designed as parametric models. Parametric models offer very precise and realistic models. However, interactive visualization requires converting them into a polygonal representation by approximating these surfaces into many polygons. This process is also known as tessellation.

There are essentially two methods presented by Frey and George (2000) to convert a parametric surface into its polygonal representation. However, this step is usually performed by each CAD tool.

4.2.3 Polygonal Mesh Simplification

Algorithms used to simplify meshes are based on the so called simplification operators; the most popular ones are vertex decimation, edge collapsing and vertex clustering.

- Vertex decimation operator was first proposed by Schroeder (Schroeder et al., 1992) Vertex decimation operates on a single vertex by deleting that vertex and retriangulating the resulting hole.
- After vertex decimation, edge collapsing became the most common mesh simplification algorithm, such that today nearly all iterative algorithms use some sort of edge collapsing (Hoppe et al., 1993).
- In vertex clustering the bounding box of the mesh is divided into a grid, and all of the vertices in a given cell are replaced with a single representative vertex. Faces that become degenerate are removed from the resulting simplified mesh. Lindstrom (Lindstrom, 2000) showed how Quadric Error Metric (QEM) simplification algorithm can be used to generate higher quality results.

Software tools can also be used to simplify the polygonal complexity of 3D models. Some of these tools work as plugins, as stand-alone applications or as programming libraries:

- Polygon Cruncher (Software, 2015).
- Simplygon (Simplygon-Studios, 2015).
- MeshLab (Cignoni and Ranzuglia, 2015).

Considerable research has been conducted on polygonal mesh simplification (Cignoni et al., 1998; Luebke, 2001, 2003). Shamir (2008) presents a state of the art on mesh segmentation techniques, and concludes that the key factor for choosing both the algorithm and the criteria for mesh segmentation is the application in mind.

Thakur et al. (2009) present a list of CAD model simplification techniques relevant for physics based simulation problems and characterize them based on their attributes. They state that there are many open research issues such as the lack of formal analysis of computational complexity or the lack of application specific error measures.

4.2.4 Related Platforms

As for visualization technologies are concerned, WebVR is an experimental JavaScript API that provides access to virtual reality devices. This technology has been used to visualize CAD models (Zhang et al., 2009; Wei-dong, 2010) and also to manage big city information, proving the usability of such technology for 3D city visualization (Lv et al., 2016).

As applications are concerned, CyberCAD (Tay and Roy, 2003) in 2003 aimed to establish a virtual synchronous collaborative design environment to overcome geographical constraints, shorten product development time and cost through the Internet. This project developed a proprietary framework for networked CAD.

Han et al. (2002) built a pilot real-time 3D system to promote the Internet based collaborative engineering design (modeling) using STEP standard to store database. The system provides a web-based search tool with the concept of metadata for navigating the product data.

Lu et al. (2016) proposed Open3D platform to enable the collaborative curation of large scale city models. This platform allows simultaneous city modeling allowing multiple users to work on different aspects of the same 3D model.

GrabCAD (Stratasys-Solution, 2009) was founded in 2009 as a marketplace to connect engineers with CAD related jobs. In 2011 it evolved into a community for engineers to share CAD models. It is short of social network where engineers can create a personal profile, store different projects, view, upload and download models from other members, communicate with each other and collaborate in other userś projects. In 2013, GrabCAD released Workbench, a cloud based product data management (PDM) solution. This platform offers visualization of CAD models using original format viewers. This workbench offers visualization of CAD models using original format viewers. One of its features is section cutting. They resolve this issue by applying shaders to the visualization of the 3D model. However, the platform proposed performs an accurate geometric cut of the model applying Binary Space Partitioning algorithm citepnaylor2004binary,hughes2014computer.

The platforms currently available do not provide a solution to the objectives described in this chapter. There was a need to find solutions that allowed a user-friendly connection between CAD systems and online marketing applications.

In the long run, this work does not intend to compete with commercial platforms. This research work shows that there are new challenges that should be addressed by them. Meanwhile the web platform presented in this chapter can provide a useful service. This work may push the industry towards these new challenges. As a consequence, user-friendly CAD to Multimedia web based publishing and searching tools might become common tools in the near future.

4.3 Automatic Model Conversion

CAD models contain complex modeling descriptions and structures. These models are defined by primitives, surfaces, curves, etc. In real-time, 3D rendering models must be represented as flat polygons, normally triangles, so curves, spheres, cylinders, etc. must be approximated to polygonal meshes. These polygons are normally subdivided into triangle sets so that the graphics hardware can render them onto the screen.

Real-time graphics rendering engines are libraries used to ease application development. They include functionalities such as loading models, scene and camera management, animations, visual effects, etc. These libraries support polygonal formats which are oriented to interactive visualization. Therefore, CAD formats are not usually supported natively by the graphical engines.

Hence, the proposed platform must take a CAD format as an input and convert it into a web 3D compatible file format. Figure 4.1 illustrates the whole automatic process followed since the reception of the original CAD file until the web visualization of the 3D model. Decisions taken to accomplish this process are explained bellow.

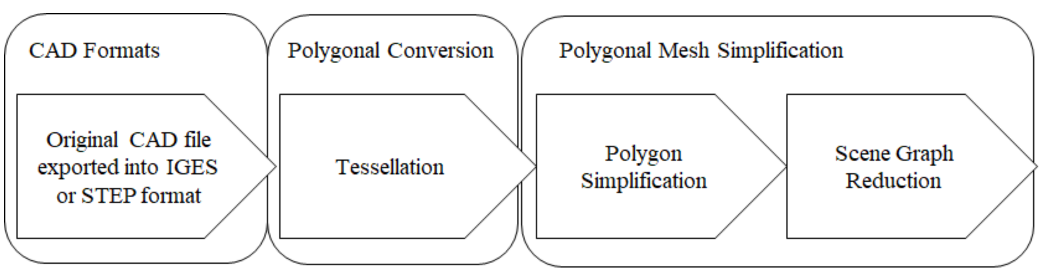


Figure 4.1: Original CAD model conversion into Web compatible format.

Many CAD design programs own proprietary formats which cannot be interpreted by other applications. However, as mentioned in Section 4.2.3, these programs also support 3D standard file formats. Table 4.1 lists the exporting file formats supported in the most common CAD applications.

STL, U3D, VRML, OBJ and COLLADA are polygonal file formats for visualization purposes, not CAD specific. But they are supported by some CAD applications. As shown in Table 4.1, the file formats supported by all applications are IGES and STEP. This fact simplifies the conversion process. Hereafter for the rest of the chapter, any CAD file will be treated as a STEP file, since there is a valid exporting option to obtain a STEP file.

However, CAD programs export IGES models not only as polygonal meshes: they also use primitives and complex curves to define the 3D model. In order to use IGES as input to the platform, a tessellation process must be performed.

Graphics processing libraries have been used to make these conversions. Open CASCADE (OpenCASCADE, 2008) and OpenSceneGraph (Martz, 2007) have been selected. Open CASCADE Technology is a software development platform applied in development of specialized CAD/CAM/CAE applications. On the other hand, OpenSceneGraph is an open source high performance 3D graphics toolkit, used by application developers in fields such as visual simulation, games, virtual reality, scientific visualization and modeling.

As shown in Table 4.2 STL format is the only exchangeable format between these two libraries. Finally, as mentioned in the introduction, WebGL technology has been chosen to render 3D models into web browsers. A higher level library must be chosen to ease the programming job. Table 4.3 shows the formats supported by some of these higher level JavaScript libraries.

Taking the afore-mentioned information into account, the process developed to convert original CAD models into web 3D compatible models is explained in the following paragraphs.

First, the original CAD models are exported into IGES or STEP files, since these two file formats are supported by most commonly used software. Then, OpenCASCADE is used to tessellate or triangulate these files. CAD complex primitives and surfaces are approximated into triangles obtaining a polygonal mesh. This process is controlled by modifying the tessellation process. Increasing the level of quality parameter, the obtained result will be more similar to the original 3D model but at the cost of having more triangles, and hence, an oversized file. Lower quality value generates smaller and easier to handle files. To finish this part of the conversion process, the tessellated results are exported to STL format, which as seen in Table 4.2 is an exchangeable file format between Open CASCADE and OpenSceneGraph.

Next, OpenSceneGraph is used to import this STL file, optimize it and convert it into a COLLADA file.

The optimization process joins equal or very close vertices. This process reduces the number of vertices obtaining a simplified mesh. Deleting too many triangles may result in a deformed model with little resemblance with the original. To prevent this undesired effect, an approximation error parameter is set. This parameter will control the balance between the simplification obtained and the shape accuracy.

This optimization process requires a high amount of memory. In order to address this issue, a batching utility has been developed. When the system has to deal with a largest STL file, it is automatically divided into smaller text files; each of which contains a section of the original geometry. Each subfile is optimized following the process previously explained. After every subfile has been optimized they are converted into COLLADA file format and merged into one single file containing the optimized model.

This COLLADA file is sent to the web client where it must be loaded and parsed by JavaScript graphic software. ASCII file formats are not particularly efficient to store huge data sets as those required to describe 3D models (space coordinates, vertices, normals, texture coordinates, colours, etc.). These ASCII files may also contain duplicated or too precise information regarding geometries or vertices. To solve these problems and speed up the interpretation of text files a Scene Graph Reduction Module (SGRM) has been developed.

This SGRM module performs, among others, the following optimizations to de COLLADA file:

- Fixed point precision. The precision of vertex coordinates, texture coordinates and normal vector coordinates can be controlled.

- Coordinate or normal vector simplification. The SGRM searches for duplicated groups of coordinates or normal vectors; it leaves only one group and references the rest.

- Geometry simplification. The SGRM searches for duplicated geometries; it leaves only one and references the rest.

- Multimaterial simplification. Some objects with multimaterials are not well interpreted by browsers, so the SGRM may simplify them.

Finally, this preprocessed COLLADA file is sent to the web application for rendering.

Table 4.1: Exportable formats in a selection of CAD software (Diez et al., 2017b).

		CAD application				
		SolidEdge	**SolidWorks**	**CATIA**	**Unigraphics**	**Pro-Engineering**
File Format	**Parasolid**	×	×			×
	JT	×			×	×
	ACIS	×	×			×
	CATIA	×		×		×
	IGES	×	×	×	×	×
	STEP	×	×	×	×	×
	ProE-assem		×			
	STL	×	×	×		
	U3D	×	×			×
	VRML		×	×	×	
	OBJ					×
	COLLADA					

Table 4.2: Import/Export file formats in Open CASCADE and OpenSceneGraph (Diez et al., 2017b).

		Graphics Library / Engine	
		OpenCASCADE	OpenSceneGraph
File Format	Parasolid	×	
	JT		
	ACIS	×	
	CATIA		
	IGES	×	
	STEP	×	
	ProE-assem		
	STL*	×	×
	U3D*		
	VRML*	×	×
	OBJ*		×
	COLLADA*		×

Table 4.3: Formats supported by JavaScript libraries (Diez et al., 2017b).

		High level Web Library			
		X3DOM	O3D	GLGE	Three.js
File Format	X3D	×			
	COLLADA		×	×	×
	OBJ			×	
	JSON		×		×
	UTF8				×

Results

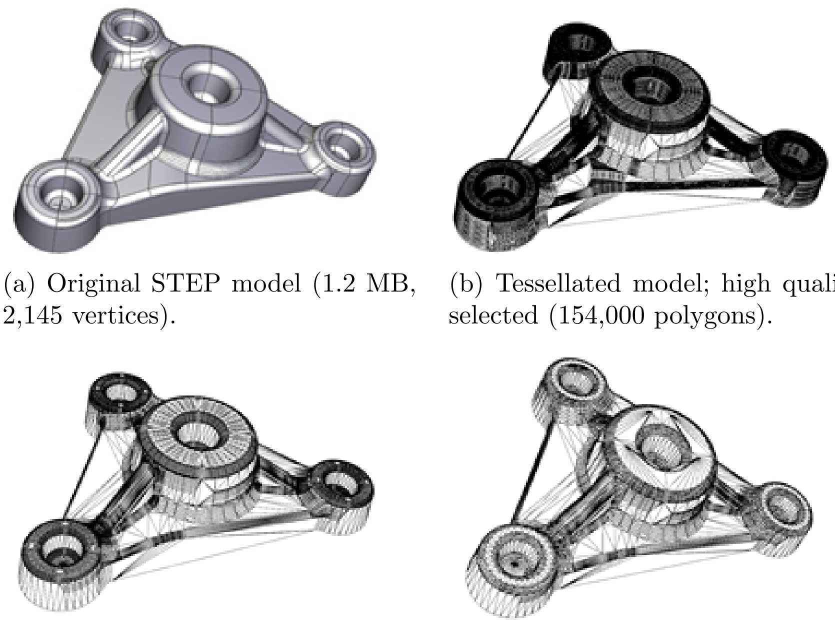
This Section presents some model conversion results.

Simple STEP model conversion

The input for this first example is a simple STEP file. As seen in Figure 4.4a this model has not been designed with a very high detail. The number of vertices used to define its surfaces is very low: 2,145 vertices. The first automatic step tessellates this file, approximating its surfaces to triangulated meshes. In Figure 4.4b a high quality tessellation is shown with 154,000 polygons. This number can be reduced selecting a lower quality level. With medium quality for this model achieved 35,700 polygons (see Figure 4.4c).

It may happen that a huge polygonal STEP file does not contain the original surfaces description. In these cases, the scene graph reduction module is applied to the tessellated model. Triangles and vertices are deleted maintaining a bounded approximation error. Using the previous model (35,700 polygons), an automatically simplification achieves a model with 10,340 polygons (see Figure 4.4d).

Table 4.4: Simple STEP model conversion example.



(a) Original STEP model (1.2 MB, 2,145 vertices).

(b) Tessellated model; high quality selected (154,000 polygons).

(c) Tessellated model; medium quality se-lected (42,069 polygons)

(d) Simplified polygonal model (10,340 polygons)

Complex STEP model conversion

The following example presents a much more complex situation. The initial input model in this example is a 165 MB STEP file. Its tessellated result, shown in Figure 4.5a, is composed by 2.2 million polygons and its STL ASCII file size is 579 MB. The model is too heavy and complex to be visualized even by a native desktop application.

Following the algorithm described in Section 4.4, the model is automatically split in various sub-models and treated separately. In this case, the model was divided into 15 sub-models. Each of these pieces was processed and finally these pieces were merged to obtain the resulting optimized file (see Figure 4.5b).

Table 4.5: Complex STEP model conversion example.

(a) Tessellated model (579 MB, 2.2 million polygons).

(b) Simplified model (852,393 polygons).

Conversion Time Rates

This section shows the time needed to convert several CAD models of diverse size and complexity into COLLADA web compatible format. Each example includes the size of the models (KB) after each conversion/optimization process and the final number of polygons (units).

A medium quality level of conversion has been applied to these models. As defined in Figure 4.1 the conversion process includes:

- Tessellation.
- Polygon simplification.
- Optimization or Scene Graph Reduction (SGR).

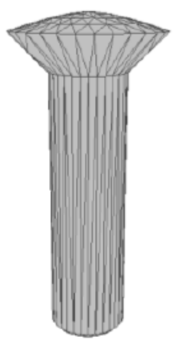
As shown in the following examples, tessellation and polygon simplification are the most time consuming steps of the process (see Table 4.6 to Table 4.12).

Next, the most attention getting data of these examples are explained.

The size of the models after the tessellation process is bigger than the initial size. Original CAD files are defined as parametric models; these curved surfaces must be converted into polygons. In order to obtain a precise approximation of the model, a vast number of polygons is needed. However, this model is used as an intermediate process of the conversion. It is not transferred at any time.

In most cases, the size of the final multimedia file is smaller than the original one. However, in Table 4.7 the final size of the converted model is slightly bigger than the initial size. This happens with models designed in high detail. In order to preserve the appearance of the original model, a bigger amount of polygons is needed and the final size might be slightly increased. However, the size of the converted models remains manageable for multimedia transmission purposes.

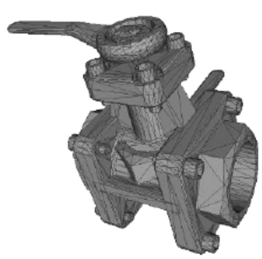
Table 4.6: Conversion Times for model #1 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	87	Transfer: 0.0.10
Tessellation	128	0.158
Polygon Simplification	30	0.143
Optimization (SGR)	24	0.199

Final Number of Polygons: 1,374

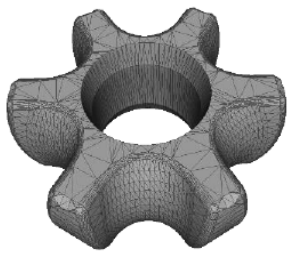
Table 4.7: Conversion Times for model #2 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	1,120	0.035
Tessellation	20,101	2.015
Polygon Simplification	1,623	3.543
Optimization (SGR)	1,325	0.566

Final Number of Polygons: 77,355

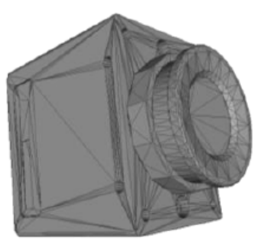
Table 4.8: Conversion Times for model #3 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	2,245	0.069
Tessellation	5,350	1.559
Polygon Simplification	557	2.512
Optimization (SGR)	459	0.084

Final Number of Polygons: 18,831

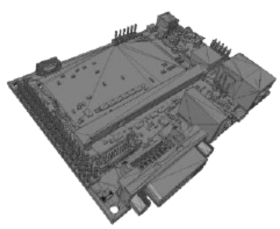
Table 4.9: Conversion Times for model #4 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	22,296	0.282
Tessellation	34,192	15.940
Polygon Simplification	4,463	5.016
Optimization (SGR)	3,748	0.789

Final Number of Polygons: 229,887

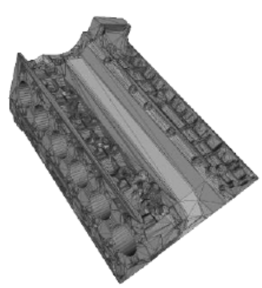
Table 4.10: Conversion Times for model #5 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	62,006	0.767
Tessellation	70,253	107.329
Polygon Simplification	3,538	67.093
Optimization (SGR)	2,888	0.445

Final Number of Polygons: 193,323

Table 4.11: Conversion Times for model #6 (Diez et al., 2017b).



	Size (KB)	Time (s)
Initial	108,808	1.196
Tessellation	530,791	149.263
Polygon Simplification	13,025	573.354
Optimization (SGR)	10,827	0.609

Final Number of Polygons: 486,579

Table 4.12: Conversion Times for model #7 (Diez et al., 2017b).

	Size (KB)	Time (s)
Initial	165,613	1.749
Tessellation	881,030	356.989
Polygon Simplification	19,934	1002.751
Optimization (SGR)	16,199	1.653

Final Number of Polygons: 852,393

4.4 Challenge Analysis based on a Case Study

In this Section, a platform design that provides a structured solution to the problems related to e-commerce catalogue is presented.

The platform is designed following a client-server model. To ensure interoperability between modules, the platform is based on Web Services technology. Specifically, SOAP messaging protocol is used.

The above mentioned functionalities will be composed by three software modules within a common platform: Authoring Tool, Web Visualization Tool and Content Management. Section 4.3 explains the geometric issues taken into account to perform the Authoring Tool, Section 4.4.1 describes the Web Engine developed for the platform, Section 4.4.2 deals with semantic issues and finally, Section 4.4.3 briefly introduces the Content Manager module (Figure 4.2).

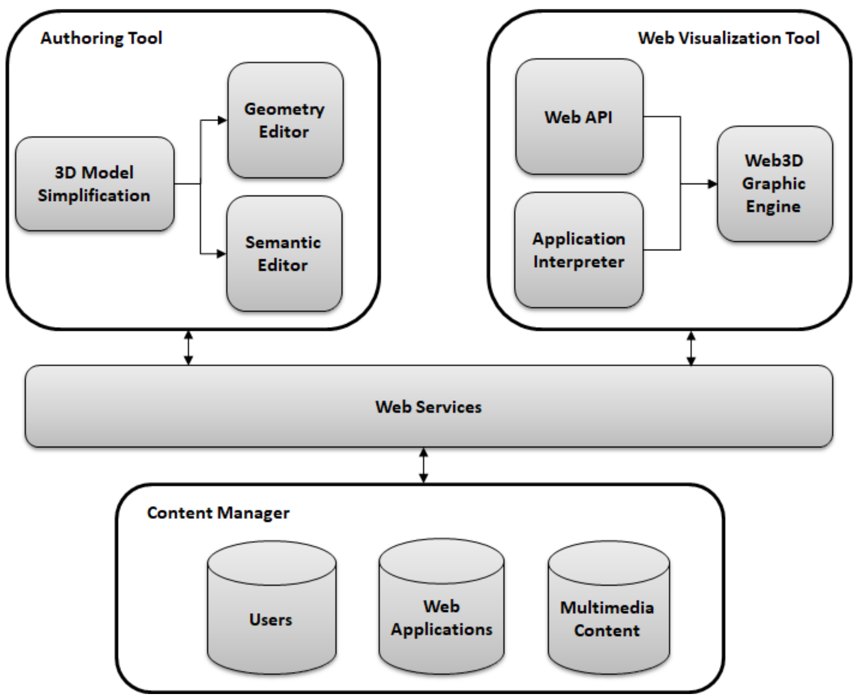


Figure 4.2: 3D Model Management Platform Architecture.

4.4.1 Web3D Visualization and Interactive Tools

The engine developed in this work uses the very well known Three.js library. This engine is used to visualize the 3D model on a web browser (Figure 3). The engine is offers the following functionalities:

- 3D Model loading. Three.js includes a function to load COLLADA models. Nevertheless, this function has been slightly modified to prevent errors reading the textures of materials.

- Real time rendering. Three.js library provides backends for Canvas, SVG, CSS and WebGL. This platform uses the WebGL backend to render the models. This allows creating complex 3D scenes in a much simpler way.

- Interactivity. Several functionalities have been designed so that users can interact with the 3D models in various ways.

  Camera controls: In order to explore the CAD model, users can interact with the model; zoom in, zoom out, view it from different perspectives, etc.

  Cut tool: With this tool the user can perform several cuts to the 3D models. When this functionality is activated, a cutting plane is visualized over the model. The user can interact with this plane with three degrees of freedom to perform and visualize different cuts on the model.

  Measure tool: Picking is used to select different faces or points of the model. This information is then used as a measuring tool, calculating the distance or the angle between the selected faces.

  Snapshot tool: The user can choose to store a thumbnail of the CAD model. This tool allows taking a snapshot of the model in the desired position and associating this image with the metadata of the model.

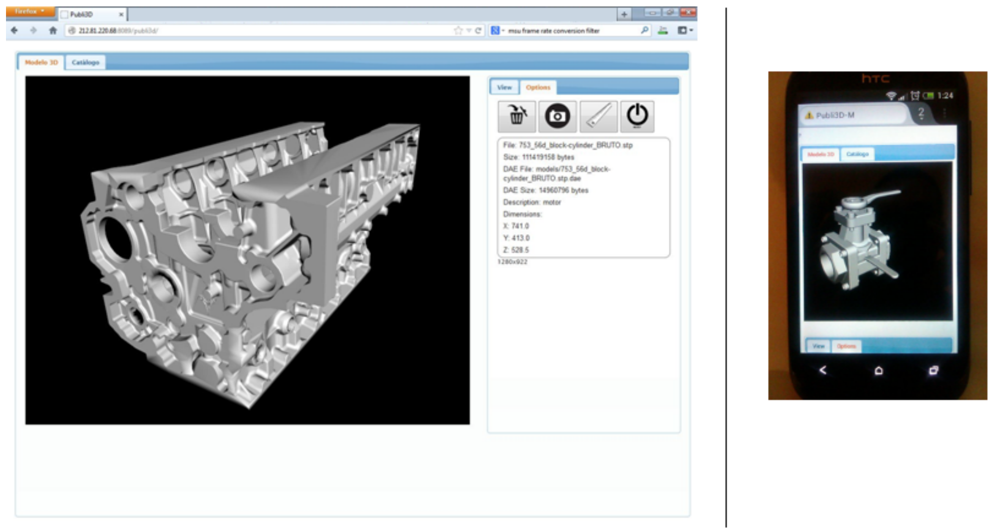


Figure 4.3: Web Visualization Tool a) Firefox web browser and b) Smartphone.

4.4.2 Semantic Issues

The following Section addresses the problem of storing and searching for the previously loaded 3D models.

CAD models contain metadata which typically provide semantic information including a brief description of the data set, the area covered by the data set, the data structure and file format, the coordinate system or projection of the data, the time when the data was collected and method of collection or the quality or accuracy of the data.

On the other hand, in Computer Vision, image descriptors describe certain elementary characteristics of the images such as shape, colour, texture or size. The advantage of these descriptors is that they can be mathematically compared: similar images will have similar descriptors. This quality is widely used by search engines for 2D image classification.

In this case, 3D projections are stored with the CAD converted model as metadata. These projections can be generated automatically. Users can easily take snapshots of specific views and store them. These images can be used by classical image classification systems. In this way, content based retrieval processes can be integrated into the platform (Quartulli and Olaizola, 2013).

When the user takes a snapshot of the loaded 3D model, this image is stored in Cloudinary. It is a proprietary cloud platform to manage large numbers of images. Each image file is associated with a unique identifier and URL. Cloudinary extracts the main metadata from the images (codec, size, etc.). Clearly, this data offered by Cloudinary is not enough to search images by similarity. The Semantic Editor Module proposed in this work adds content descriptors: colour and texture histograms and number of detected faces and their orientation.

In essence, the descriptor of each image is a vector of numbers that represents all those features. The enriched metadata is stored in a MongoDB database in the cloud. When a user initiates an image search in the platform a query with the searching criteria is sent. The platform then retrieves all the images from the database with similar features to the one in the query.

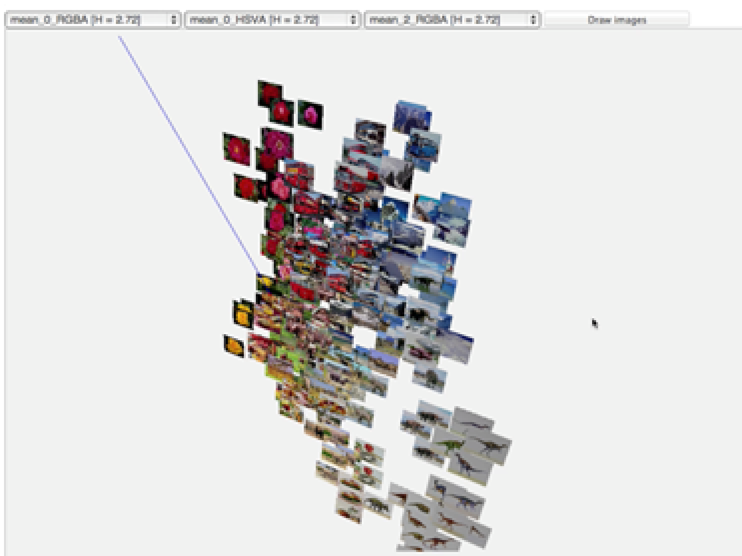


Figure 4.4: Image cloud distributed in a 3D space. Images are grouped based on 3 descriptors regarding their colour similarity.

A way to evaluate the level of similarity between the retrieved images is to calculate the Euclidean distance between the descriptors in the query and those from the result. As mentioned before, these descriptors are represented as a vector of numbers so their Euclidean distance can be easily obtained.

The Semantic Editor Module in this work calculates the distance between the image used in the query and those in the database and retrieves the 20 more similar images. This collection is done applying the methodology proposed by Silva et al. (2015) which is based on a probabilistic k-nearest neighbour supervised classification algorithm.

4.4.3 Content Management

The platform presented in this work serves as an online marketing catalogue. The contents of this catalogue are stored in a MongoDB database.

MongoDB (Martz, 2007) is a cross platform document oriented database system. Classified as a NoSQL database, MongoDB eschews the traditional table based relational database structure in favour of JSON like documents with dynamic schemas, making the integration of data in certain type of applications easier and faster. These databases hold a set of collections and these collections hold a set of documents. A document is a set of key value pairs.

Documents have dynamic schema which means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a document collection may hold different type of data.

The designed database has a collection named catalogue in which all the information regarding the 3D models is stored. When a new model is uploaded into the platform a new document is added to the catalogue collection. Initially, these documents are described only by two fields; name and description. These fields are edited by the user when uploading a new model. When the model has been processed as explained in Section 4.3, extra information regarding various aspects (optimized size, path, snapshot image, etc.) is added to this document.

4.5 Chapter Summary and Related Publications

This chapter presents a platform for CAD model management for online marketing. It eases the development of e-commerce systems that contain catalogues with 3D content. As stressed in the introduction, the goal of this work is to offer a user-friendly "CAD to multimedia" web based publishing solution and to solve some of its challenges.

The platform has been tested with diverse CAD models from different locations, devices, operating systems and browsers. The platform has been tested both through wired internet connection and through 3G mobile connection.

The import process considers automatic primitive tessellation, mesh reduction, huge model splitting, scene graph reduction and Web3D compatible model generation (COLLADA). Model uploading rates vary depending on the size of the model: light models (90 KB) take only a few milliseconds to upload (example in Table 4.6), however larger files (160 MB) can take up to 20 minutes (example in Table 4.12). Then, converted models are smoothly managed and rendered in the client side.

To test the platform various models have been dragged and dropped onto the web application. Loading rates on the client vary between a few seconds and a few minutes depending on the size of the file. Large files can be compressed to reduce the loading times (Bayardo et al., 2004).

Tests have been made on Mozilla Firefox version 45 and Google Chrome version 50 browsers. Additionally, the platform has also been tested, except the model loading functionality, on smartphones with Android operating system and Firefox Mobile or Chrome Mobile browsers. A solution for 3D model management has been considered based on virtual 2D images of the 3D models, evaluating image descriptors.

CAD systems are relevant in Industry 4.0 since they allow to redesign, simulate and monitor models to be produced without having to stop the production line. Industry 4.0 demands all machines to have a networking process in which the product is able to modify the process if needed in order to build a flexible production line. Online CAD modifications according to each specific product is an advantage for product personalization. This will allow to optimize the resources of the company and supply chain suppliers while providing the customers with a complete quality product which can be easily adapted to the needs of current society.

Related Publications

- (Diez et al., 2017b) 3D Model Management for e-commerce. Diez, H. V., Segura, A., García-Alonso, A. & Oyarzun, D. Multimedia Tools and Applications, 76(20):21011-21031. (2017).

CONCLUSIONS AND FURTHER WORK

CHAPTER 5

Conclusions and Further Work

This thesis has provided contributions for Visual Computing technologies in three main areas of study. The results obtained from these application-driven research have highlighted the path to solve further challenges presented by Industry 4.0.

The contributions of this thesis compile the research work carried out since 2010 in several applied research projects. The research is focused in three main areas within *Visual Computing*: Interactive Virtual Agents, Immersive VR/AR Environments and Interactive 3D Model Visualization.As a whole, they also contribute to another broader field: *Industry 4.0*.

This work has materialized in eleven publications included in this volume (see Annex I).

5.1 Contributions to Visual Computing (Thesis Goal I)

Chapters 2, 3 and 4 compile knowledge and experience gathered along eight years. An application-driven research through diverse fields, such as e-learning, training and e-commerce, has led to solve several challenges for Visual Computing.

This section summarizes the main contributions within the Visual Computing field. Previous chapters describe contributions in a more detailed and contextual mode (see Sections 2.1, 2.2, 2.3, 3.1, 3.2, 3.3 and 4.3).

Chapter 2 focused on *Interactive Virtual Agents (IVA)*. The case studies presented have validated the development of autonomous, modular, scalable, ubiquitous and, more importantly, engaging IVA. These IVA can interact with users in a human-like way thanks to their attributes: realistic appearance, verbal and non-verbal communication and natural behaviour skills. The IVA have been integrated in applications of diverse fields, such as e-learning, marketing, health care, safety and manufacturing. These applications validate the suitability of these agents in human-centred tasks. These agents can be applied as an interface between the worker and the supporting automation system, this is as Human-Machine Interaction (HMI) interfaces.

Chapter 3, *Immersive VR/AR Environments*, shows how VR can be applied in production planning, product design, process simulation, testing and verification. Furthermore, this can be achieved in a cost-efficient way, reducing risks for the operators and acquiring real-time feedback. Appropriate virtual environments must be designed taking into account the final user and the final rendering platform. The case studies presented in Chapter 3 show a number of interaction and immersion devices that can be used. The Virtual Operator case study shows how VR and Co-bots can work together in a safe environment, enhancing workers cognitive and physical abilities. AR can be used to show relevant information to the worker in a non-intrusive way as shown in the Augmented Operator case study.

Chapter 4, *Interactive 3D Model Management*, presents an online solution for multimedia CAD model management and visualization. An automatic model conversion was implemented to allow users to upload CAD files into the Web and create online product catalogues. Large industrial 3D CAD model management will be crucial in Industry 4.0. Web3D technology will allow the visualization and interaction of these models in mobile low-power devices.

5.2 Projecting the Contributions Towards Industry 4.0 (Thesis Goal II)

As stressed in Section 1.1 Industry 4.0 presents a number of challenges. The application-driven research followed in this thesis work has contributed to provide proof of concept for those challenges.

Chapter 2 has contributed to new modes of interaction. IVA have been implemented to allow natural Human-Machine interaction. These agents endowed with autonomous behaviour have been integrated in several applications. Voice and gesture commands have been developed to allow natural communication. Five publications have presented the results from this research.

As seen in Figure 5.1 (in blue), these IVA contribute to some of the challenges presented by Posada et al. (2015) related with human factors.

- *New HMI modalities.* Allowing new modes of interaction adapted to workers' job restrictions (such as voice-based interaction and gesture recognition).
- *Post-WIMP interfaces.* Adapting to the future trend of mobile devices in the factory.
- *3D authoring tools and end-user UI.*

Challenges regarding human factors have also been addressed in Chapter 3. VR and AR were applied to several training applications. Safety and security issues were treated in the fire warden case study developing an emergency response application. The Virtual Operator case study showed the possibilities of VR in collaborative works with robots. Figure 5.1 (in red) highlights the addressed challenges related to the third criterion from Posada et al. (2015).

- Operational training of complex machines and in some cases in *virtual and augmented reality setups.*
- *Visual simulation for emergency response in the factories.*
- *Easing the use and control of the robot by the worker.*

- *Virtual environments simulations.* For human and robot coexistence in production and for different configurations and parameters.

Table 3. Industrie 4.0 future factory visual computing challenges, human factors.

Industrie 4.0 relation with human factors	Visual computing enabling technologies and challenges
Work organization and design (productivity enhancement though improved human intervention)	*New HMI modalities.* Allowing new modes of interaction adapted to workers' job restrictions (such as voice-based interaction and gesture recognition) *Post-WIMP interfaces.* Adapting to the future trend of mobile devices in the factory. *Advanced manufacturing and production-planning visualization.* Linking SCADA systems with VR paradigms for interaction in the planning and understanding of the production plan. *Interfaces with manufacturing execution systems.* Allowing different configuration capabilities.
Foster creativity in skilled workers	*End-user tools for visualization of flexible production plans.* Including the location of different machines and persons (with complementary skills) in alternative production scenarios. Tools for discussions between engineers and workers.
Training and continuing professional development (capture and systematic reuse of the knowledge of the worker)	*Multimedia capture and intelligent retrieval of worker knowledge.* Capturing and transferring knowledge between the workers. *3D authoring tools and end-user UI.* Operational training of complex machines and in some cases in virtual and augmented reality setups.
Safety and security	*Cognitive computer vision systems.* Detecting and contextualizing events occurring in the factory to improve safety and security. Focusing on hazardous area exposition and collision detection with massive objects. *Visual simulation for emergency response in the factories.*
Sociotechnical interaction (co-working with configurable robots)	*Visual programming of robot interactions.* Imitating human motion based on computer vision for anthropomorphic robots. Easing the use and control of the robot by the worker, not necessarily the engineer. *Virtual environments simulations.* For human and robot coexistence in production and for different configurations and parameters.

Figure 5.1: Challenges faced for Industry 4.0 related with human factors. Those addressed in Chapter 2 in blue, those addressed in Chapter 3 in red.

Finally, Chapter 4 has also contributed to some challenges presented by Industry 4.0. Regarding the integration dimensions, the automatic CAD model conversion into Web3D compatible models for management and visualization presented in Chapter 4 aligns directly with the challenges highlighted in Figure 5.2.

- *Natural flow of a persistent and interactive digital model.* Product lifecycle management involving large industrial 3D CAD/CAM models with full access to semantic/dynamic integration data in Web3D.

- *3D model automatic simplification.* Preserving critical features for service tasks while allowing interaction/visualization in mobile low-power client devices.

Table 1. Industrie 4.0 future factory visual computing challenges in three integration dimensions.

Industrie 4.0 integration dimension	Visual computing enabling technologies and challenges
Vertical integration (networked manufactured systems and autonomous cyber-physical production systems)	*Virtual environments.* Visually empowered 3D simulation scenarios for new ways of planning production, especially suitable for dynamic and fast changes. Scenarios for testing different configurations.
	Real-time representation of production. Visualizing flows of information, material, and knowledge in the factory, not only physical representation.
	3D scanning and 3D reconstruction of factories. Adapting old factories to new paradigms.
	End user interfaces. Editing configurations in demanding work conditions, such as production lines.
End-to-end digital engineering integration (holistic life-cycle management)	*Natural flow of a persistent and interactive digital model.* Product life-cycle management involving large industrial 3D CAD/CAM models with full access to semantic/dynamic integration data in Web3D.
	3D real-time simulations for CPE production.
	New paradigms of 3D geometric representation. New processes (such as laser-based manufacturing, fast-speed material removal, and micro- and nano-manufacturing) and new materials (such as biomaterials and metallic powders for 3D metal printing).
	Computer vision "closing the loop" in 3D production planning. Real-time coupling of production process and 3D models. Geometry adaptation to physical conditions.
Horizontal integration through value networks (value chain integration)	*Augmented reality (AR) for service-based actions with providers and clients.* Ergonomic aspects of the solutions going from the lab to real factories and the integration with the information systems. Intelligent media streaming/search to improve service (as in teleoperation).
	3D model automatic simplification. Preserving critical features for service tasks while allowing interaction/visualization in mobile low-power client devices.

Figure 5.2: Challenges faced in Chapter 4 for Industry 4.0 related with the integration dimensions.

Regarding product and production challenges, the case study presented in Chapter 4 shows a Web3D application for CAD model visualization that integrates several interactive tools. This aligns with the following challenges seen in Figure 5.3:

- *3D interactive tools.* Empowering end users in the final product configuration.
- *Automatic generation of options catalogs.* Accounting for production parameters and user preferences.
- *Digital coexistence or alter-ego of the physical product.* Enabled by Web3D, localization and mobile interaction technologies, allowing new services (such as social networks of users for the same product line).

Table 2. Industrie 4.0 future factory visual computing challenges in product and production.

Industrie 4.0 product and production	Visual computing enabling technologies and challenges
Product self-awareness (history, status, location, delivery strategy, and service)	*Integration of GIS (outdoor) with in-factory (indoor) localization-visualization systems*. Individualized product tracking and as underlying connection layer between factories and products when delivered. *Cyber-physical 3D equivalence*. At all times linking product digital model and situational status.
Personalization and flexibility (flexible adaptation to individual customer requirements)	*3D interactive tools*. Empowering end users in the final product configuration. *Automatic generation of options catalogs*. Accounting for production parameters and user preferences. *3D shape automatic adaptation*. Fitting production and manufacturing restrictions. *Linking of 3D changes with resource impacts*. Accounting for time and cost.
Optimized decision making (with access to real-time production and design data)	*Visual analytics of production big data*. Trillions (or more) samples per year (such as GE Industrial Big Data[19]). *Real-time mixing of production big data with 3D digital engineering design data*. *User interface dynamic adaptation of information to user profile, devices, and context*. Visual analytics system for the engineer and the worker.
Emergence of new services and business models	*Digital coexistence or alter-ego of the physical product*. Enabled by Web3D, localization and mobile interaction technologies, allowing new services (such as social networks of users for the same product line).
Resource/energy efficiency and sustainable production	*Dynamic resource visualization at the factory level*. Including sustainability footprint (such as CO_2 consumption), energy distribution in a plant, and material waste. Can be mixed with VR and in some cases AR.

Figure 5.3: Challenges faced in Chapter 4 for Industry 4.0 related with product and production.

5.3 Further Work

This thesis work has presented a conceptual framework, supported by specific practical case studies, showing the application of different Visual Computing technologies. The application of these technologies can empower operators in the context of Industry 4.0 scenarios.

HMI interfaces, Virtual Reality, Augmented Reality and Collaborative Robotics Interaction are some of the most relevant technologies for the Operator 4.0, since they not only improve productivity and efficiency, but are essential to tackle the social, inclusion and interaction aspects that are central to these new sociotechnical systems. The application of Visual Computing technologies can contribute decisively to the enhancement of the operator's ability to perform traditional tasks, and to the definition of new tasks and scenarios.

Figure 5.4 illustrates this idea: the operator in the centre of the figure is surrounded by a number of tasks to perform, inner ring, which are supported by Visual Computing technologies, outer ring. The worker is thus empowered by Visual Computing to perform better or make decisions with stronger criteria.

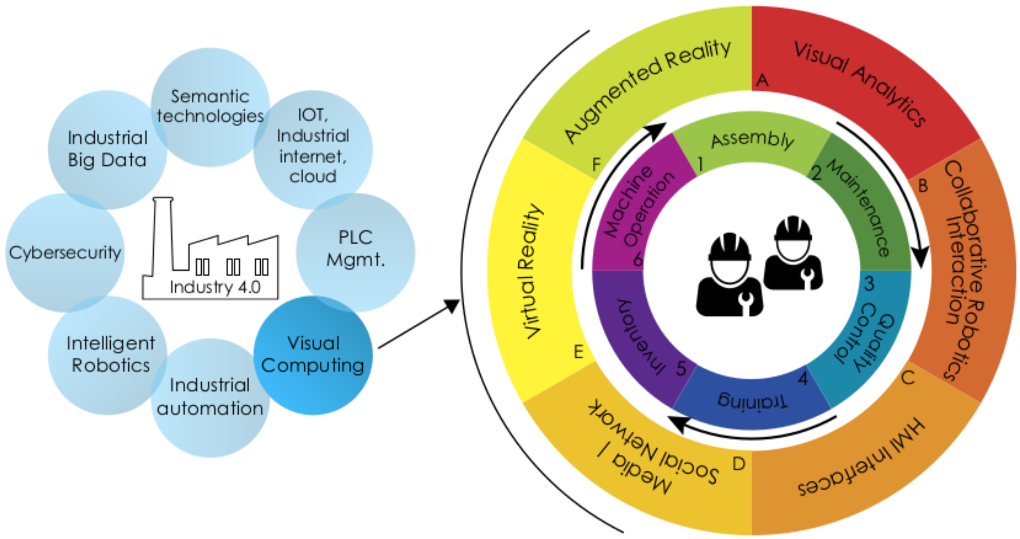


Figure 5.4: Left: Key enabling technologies in Industry 4.0 (From Posada et al. (2015)). Right: General outline of Visual Computing technologies improving Operator 4.0 tasks (From Segura & Diez et al. (2018))

The case studies presented in this thesis are good real-world examples of how a "Visual Computing enhanced" Operator 4.0 can have a central role in future industrial production scenarios.

The robotic Digital-Twin in the Virtual Operator case study described in Section 3.4 shows how an operator can be trained in collaborative robotics interaction to be prepared (technically and psychologically) for a novel interaction with machines and robots.

The Augmented Reality case study described in Section 3.4, shows the applicability of AR to support assembly operations based on robust markerless tracking.

Hence, this thesis is directly addressing the need of adaptation and balancing of new production forms with the enhancement of the skills and abilities of the operators.

Future work also includes the measurement and analysis of the actual impact of these case studies in the factory, including both productivity/efficiency and social/psychological aspects.

Further work will also have to explore other Visual Computing technologies for Industry 4.0. Visual analytic solutions will be crucial for optimized decision making, linking industrial Big Data with semantic technologies and product management technologies. Social networks will also play a role in increasing worker's performance and satisfaction. Other technologies, such as cloud computing and cibersecurity will also have a crucial role in the Industry 4.0

Bibliography

Abowd, D., Dey, A. K., Orr, R., and Brotherton, J. (1998). Context-awareness in wearable and ubiquitous computing. *Virtual Reality*, 3(3):200–211.

Abrilian, S., Devillers, L., Buisine, S., and Martin, J.-C. (2005). EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *Proceedings HCI (no pages)*.

Ahmed, K. S. (2005). Animated pedagogical agents in web-based intelligent learning environment. *https://pdfs.semanticscholar.org/bdb3/014a0568d1e5b077a89f4b6ba3cabe31bc9c.pdf*.

Akchelov, E. and Galanina, E. (2016). Virtual world of video games. In *Games and Virtual Worlds for Serious Applications (VS-Games), 2016 8th International Conference on*, pages 1–4. IEEE.

Ali, I. R., Kolivand, H., and Alkawaz, M. H. (2018). Lip syncing method for realistic expressive 3d face model. *Multimedia Tools and Applications*, 77(5):5323–5366.

Arghinenti, A. (2011). Animation Workflow in KILLZONE 3TM: A fast facial retargeting system for game characters. In *ACM SIGGRAPH 2011 Talks*, page 37. ACM.

Árnason, B. and Þorsteinsson, A. (2008). The CADIA BML realizer. *http://cadia.ru.is/projects/bmlr/*, 97.

Arnaud, R. and Parisi, T. (2007). Developing web applications with collada and x3d. *White paper*.

Association, U. P. D. et al. (2007). Initial graphics exchange specification. *US Product Data Association, Berkeley*.

Baheti, R. and Gill, H. (2011). Cyber-physical systems. *The impact of control technology*, 12:161–166.

Bartle, R. A. (2004). *Designing virtual worlds*. New Riders.

Bauer, S., Kneip, J., Mlasko, T., Schmale, B., Vollmer, J., Hutter, A., and Berekovic, M. (1999). The MPEG-4 Multimedia Coding Standard: Algorithms, Architectures and Applications. *J. VLSI Signal Process. Syst.*, 23(1):7–26.

Bayardo, R. J., Gruhl, D., Josifovski, V., and Myllymaki, J. (2004). An evaluation of binary xml encoding optimizations for fast stream based xml processing. In *Proceedings of the 13th international conference on World Wide Web*, pages 345–354. ACM.

Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R. W., and Gross, M. (2011). High-quality passive facial performance capture using anchor frames. In *ACM Transactions on Graphics (TOG)*, volume 30, page 75. ACM.

Bell, M. W. (2008). Toward a definition of "virtual worlds". *Journal For Virtual Worlds Research*, 1(1).

Bowman, D., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice, CourseSmart eTextbook*. Addison-Wesley.

Bowman, D. A., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. (2001). An introduction to 3D user interface design. *Presence: Teleoperators and virtual environments*, 10(1):96–108.

Brey, P. (2008). Virtual reality and computer simulation. *The handbook of information and computer ethics*, page 361.

Brown, E. and Cairns, P. (2004). A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1297–1300. ACM.

Carey, R. and Bell, G. (1997). *The annotated VRML 2.0 reference manual*. Addison-Wesley.

Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology*, 59(5):941.

Carretero & Diez, M., Ardanza, A., García, S., Díez, H., Oyarzun, D., and Ruiz, N. (2012). Improving gestural communication in virtual characters. In *International Conference on Articulated Motion and Deformable Objects*, pages 69–81. Springer.

Carretero & Diez, M., Diez, H. V., García, S., and Oyarzun, D. (2016). Providing Physical Appearance and Behaviour to Virtual Characters. In *International Conference on Articulated Motion and Deformable Objects*, pages 98–107. Springer.

Cassell, J. (2000). Embodied conversational agents.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 413–420. ACM.

Cassell, J. and Vilhjálmsson, H. (1999). Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous agents and multi-agent systems*, 2(1):45–64.

Cassell, J., Vilhjálmsson, H. H., and Bickmore, T. (2001). Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 477–486. ACM.

Castronova, E. (2008). *Synthetic worlds: The business and culture of online games*. University of Chicago press.

Čereković, A., Pejša, T., and Pandžić, I. S. (2010). A controller-based animation system for synchronizing and realizing human-like conversational behaviors. In *Development of Multimodal Interfaces: Active Listening and Synchrony*, pages 80–91. Springer.

Choi, B. and Baek, Y. (2011). Exploring factors of media characteristic influencing flow in learning through virtual worlds. *Computers & Education*, 57(4):2382–2394.

Cignoni, P., Montani, C., and Scopigno, R. (1998). A comparison of mesh simplification algorithms. *Computers & Graphics*, 22(1):37–54.

Cignoni, P. and Ranzuglia, G. (2015). MeshLab. *http://www.meshlab.net/*.

Cohen, M. M., Massaro, D. W., et al. (1993). Modeling coarticulation in synthetic visual speech. *Models and techniques in computer animation*, 92:139–156.

Combs, N. (2004). A virtual world by any other name. *Terra Nova weblog*.

Costello, H. (2017). Global e-learning market research report and forecast to 2017-2022. *https://www.reuters.com*.

De Carolis, B., Pelachaud, C., Poggi, I., and Steedman, M. (2004). APML, a markup language for believable behavior generation. In *Life-like characters*, pages 65–85. Springer.

Deng, Z., Chiang, P.-Y., Fox, P., and Neumann, U. (2006). Animating blendshape faces by cross-mapping motion capture data. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 43–48. ACM.

DeRose, T., Kass, M., and Truong, T. (1998). Subdivision surfaces in character animation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 85–94. ACM.

Diez, H. V., Barbadillo, J., García, S., del Puy Carretero, M., Álvarez, A., Sánchez, J. R., and Oyarzun, D. (2014a). Interactive Multimodal Platform for Digital Signage. In *International Conference on Articulated Motion and Deformable Objects*, pages 128–137. Springer.

Diez, H. V., García, S., Mujika, A., Moreno, A., and Oyarzun, D. (2016). Virtual training of fire wardens through immersive 3d environments. In *Proceedings of the 21st International Conference on Web3D Technology*, pages 43–50. ACM.

Diez, H. V., García, S., R Sánchez, J., del Puy Carretero, M., and Oyarzun, D. (2014b). IAAN: Intelligent Animated Agent with Natural Behaviour for Online Tutoring Platforms. In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence-Volume 2*, pages 123–130. SCITEPRESS-Science and Technology Publications, Lda.

Diez, H. V., García, S., Sánchez, J. R., and del Puy Carretero, M. (2013). 3D animated agent for tutoring based on WebGL. In *Proceedings of the 18th International Conference on 3D Web Technology*, pages 129–134. ACM.

Diez, H. V., Moreno, A., and García-Alonso, A. (2017a). Interaction Challenges in the Development of a Fire Warden VR Training System using a HMD. In *VISIGRAPP (2: HUCAPP)*, pages 84–91.

Diez, H. V., Segura, Á., García-Alonso, A., and Oyarzun, D. (2017b). 3D model management for e-commerce. *Multimedia Tools and Applications*, 76(20):21011–21031.

Dunn, M. (1992). Industrial automation systems and integration-product data representation and exchange. *ISO/WD*.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, 85(1):86.

Evans, P. C. and Annunziata, M. (2012). Industrial internet: Pushing the boundaries. *General Electric Reports*.

Fei, K. (2001). Expressive textures. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation*, pages 137–141. ACM.

Feng, D., Jeong, D. C., Krämer, N. C., Miller, L. C., and Marsella, S. (2017). Is it just me?: Evaluating attribution of negative feedback as a function of virtual instructor's gender and proxemics. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 810–818. International Foundation for Autonomous Agents and Multiagent Systems.

Finkelstein, E. (2008). *AUTOCAD 2009 & AUTOCAD LT 2009 BIBLE (With CD)*. John Wiley & Sons.

Frey, P. J. and George, P. L. (2000). *Mesh generation: application to finite elements*. Wiley Online Library.

Futurad (2012). El futuro de la publicidad. *MivTech Inc, http://futurad.tv/avatar.html*.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536. ACM.

Garrido, P., Martinez, F., and Guetl, C. (2010). Adding semantic web knowledge to intelligent personal assistant agents. In *Proceedings of the ISWC 2010 Workshops*, pages 1–12. Philippe Cudre-Mauroux.

Gebhard, P., Mehlmann, G., and Kipp, M. (2012). Visual scenemaker - a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces*, 6(1-2):3–11.

Girvan, C. (2013). What is a virtual world? definition and classification. Technical report, Technical Report. Trinity College, Dublin Google Scholar.

Gulz, A. and Haake, M. (2006). Design of animated pedagogical agents-A look at their look. *International Journal of Human-Computer Studies*, 64(4):322–339.

Han, S., Choi, Y., Yoo, S., and Park, N. (2002). Collaborative engineering design based on an intelligent STEP database. *Concurrent Engineering*, 10(3):239–249.

Hand, C. (1997). A survey of 3D interaction techniques. In *Computer graphics forum*, volume 16, pages 269–281. Wiley Online Library.

Henning, K. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0. *Acatech–National Academy of Science and Engineering, www.acatech.de*.

Hermann, M., Pentek, T., and Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In *49th Hawaii International Conference on System Sciences (HICSS)*, pages 3928–3937. IEEE.

Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. (1993). Mesh optimization. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 19–26. ACM.

Huang, H., Chai, J., Tong, X., and Wu, H.-T. (2011). Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. In *ACM Transactions on Graphics (TOG)*, volume 30, page 74. ACM.

Ilie, A., Ioan, S., Zagrean, L., and Moldovan, M. (2008). Better to be red than blue in virtual competition. *Cyberpsychology & Behavior*, 11(3):375–377.

Jankowski, J. and Hachet, M. (2013). A survey of interaction techniques for interactive 3D environments. In *Eurographics 2013-STAR*.

Kagermann, H., Wahlster, W., and Helbig, J. (2013). Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 – Securing the Future of German Manufacturing Industry. Final report of the industrie 4.0 working group, acatech – National Academy of Science and Engineering, München.

Khan, R. and De Angeli, A. (2009). The attractiveness stereotype in the evaluation of embodied conversational agents. In *IFIP Conference on Human-Computer Interaction*, pages 85–97. Springer.

Khan, R. F. and Sutcliffe, A. (2014). Attractive agents are more persuasive. *International Journal of Human-Computer Interaction*, 30(2):142–150.

Khronos-Group (2011). COLLADA-3D Asset Exchange Schema. *https://www.khronos.org*.

Klippel, A., Bagher, M. M., and Wallgrun, J. O. (2015). Immersion, presence, and interaction.

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Koda, T. and Maes, P. (1996). Agents with faces: The effect of personification. In *Robot and Human Communication, 1996., 5th IEEE International Workshop on*, pages 189–194. IEEE.

Kopp, S., Jung, B., Lessmann, N., and Wachsmuth, I. (2003). Max-a multimodal assistant in virtual reality construction. *KI*, 17(4):11.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Thórisson, K. R., and Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *International workshop on intelligent virtual agents*, pages 205–217. Springer.

Krenn, B. (2003). The NECA Project: Net Environments for Embodied Emotional Conversational Agents - Project Note. *KI*, 17:30.

Leandro Parreira Duarte, R., El Rhalibi, A., and Merabti, M. (2014). Coarticulation and speech synchronization in MPEG-4 based facial animation. *Kybernetes*, 43(8):1165–1182.

Lee, J. and Marsella, S. (2006). Nonverbal behavior generator for embodied conversational agents. In *International Workshop on Intelligent Virtual Agents*, pages 243–255. Springer.

Lindstrom, P. (2000). Out-of-core simplification of large polygonal models. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 259–262. ACM Press/Addison-Wesley Publishing Co.

Littlejohn, S. W. and Foss, K. A. (2010). *Theories of human communication*. Waveland Press.

Lu, Z., Guerrero, P., Mitra, N. J., and Steed, A. (2016). Open3D: crowd-sourced distributed curation of city models. In *Web3D Symposium Proceedings*.

Luebke, D. P. (2001). A developer's survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications*, 21(3):24–35.

Luebke, D. P. (2003). *Level of detail for 3D graphics*. Morgan Kaufmann.

Lv, Z., Li, X., Zhang, B., Wang, W., Zhu, Y., Hu, J., and Feng, S. (2016). Managing big city information based on WebVRGIS. *IEEE Access*, 4:407–415.

Magnenat-Thalmann, N., Laperrire, R., and Thalmann, D. (1988). Joint-dependent local deformations for hand animation and object grasping. In *In Proceedings on Graphics interface'88*. Citeseer.

Manyika, J. (2016). Technology, jobs, and the future of work. *http://www.mckinsey.com*.

Martey, R. M. and Consalvo, M. (2011). Performing the looking-glass self: Avatar appearance and group identity in Second Life. *Popular Communication*, 9(3):165–180.

Martey, R. M., Kenski, K., Folkestad, J., Feldman, L., Gordis, E., Shaw, A., Stromer-Galley, J., Clegg, B., Zhang, H., Kaufman, N., et al. (2014). Measuring game engagement: multiple methods and construct complexity. *Simulation & Gaming*, 45(4-5):528–547.

Martz, P. (2007). OpenSceneGraph quick start guide. *PMARTZ Computer Graphics Systems*.

Massaro, D. W., Beskow, J., Cohen, M. M., Fry, C. L., and Rodgriguez, T. (1999). Picture my voice: Audio to visual speech synthesis using artificial neural networks. In *AVSP'99-International Conference on Auditory-Visual Speech Processing*.

Matsas, E. and Vosniakos, G.-C. (2017). Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 11(2):139–153.

McCloud, S. (1994). *Understanding comics*. William Morrow Paperbacks.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Merola, N. and Peña, J. (2009). The effects of avatar appearance in virtual worlds. *Journal For Virtual Worlds Research*, 2(5).

Mlakar, I. and Rojc, M. (2011). Towards ECA's Animation of Expressive Complex Behaviour. *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 185–198.

Mohammed, A., Schmidt, B., and Wang, L. (2016). Active collision avoidance for human–robot collaboration driven by vision sensors. *International Journal of Computer Integrated Manufacturing*, pages 1–11.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35.

Mujber, T. S., Szecsi, T., and Hashmi, M. S. (2004). Virtual reality applications in manufacturing process simulation. *Journal of materials processing technology*, 155:1834–1838.

Mujika, A., Barrena, N., García, S., and Oyarzun, D. (2014). Elastic Bone Transformation for Realistic Facial Animation in WebGL. In *Future Vision and Trends on Shapes, Geometry and Algebra*, pages 107–116. Springer.

Mujika, A., Oyarzun, D., Iparragirre, I., Dominguez, J., and Arambarri, J. (2011). MACHS: An authoring tool to create serious games for machine-tool operator training. In *Emerging eLearning Technologies and Applications (ICETA), 2011 9th International Conference on*, pages 147–152. IEEE.

Mujika & Diez, H., Alvarez, A., Urteaga, M., and Oyarzun, D. (2013). Realistic visual speech synthesis in webgl. In *Proceedings of the 18th International Conference on 3D Web Technology*, pages 207–207. ACM.

Nee, A. Y., Ong, S., Chryssolouris, G., and Mourtzis, D. (2012). Augmented reality applications in design and manufacturing. *CIRP Annals-manufacturing technology*, 61(2):657–679.

Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, 9(2):JCMC925.

Nowak, K. L. and Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494.

OpenCASCADE (2008). Open CASCADE Technology, 3D modeling & numerical simulation. *http://www.opencascade.org*.

Palmaz, S., Cuadros, M., and Etchegoyhen, T. (2016). Statistically-guided controlled language authoring. In *International Workshop on Controlled Natural Language*, pages 37–47. Springer.

Pandzic, I. S. and Forchheimer, R. (2003). *MPEG-4 facial animation: the standard, implementation and applications*. John Wiley & Sons.

Parke, F. I. (1972). Computer generated animation of faces. In *Proceedings of the ACM annual conference-Volume 1*, pages 451–457. ACM.

Peña, J., Hancock, J. T., and Merola, N. A. (2009). The priming effects of avatars in virtual settings. *Communication Research*, 36(6):838–856.

Piwek, P., Krenn, B., Schröder, M., Grice, M., Baumann, S., and Pirker, H. (2004). RRL: A rich representation language for the description of agent behaviour in NECA. *https://arxiv.org/ftp/cs/papers/0410/0410022.pdf*.

Plant, E. A., Baylor, A. L., Doerr, C. E., and Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education*, 53(2):209–215.

Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., de Amicis, R., Pinto, E. B., Eisert, P., Döllner, J., and Vallarino, I. (2015). Visual computing as a key enabling technology for industrie 4.0 and industrial internet. *IEEE computer graphics and applications*, 35(2):26–40.

Quartulli, M. and Olaizola, I. G. (2013). A review of EO image information mining. *ISPRS journal of Photogrammetry and Remote Sensing*, 75:11–28.

Radovan, M. and Pretorius, L. (2006). Facial animation in a nutshell: past, present and future. In *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, pages 71–79. South African Institute for Computer Scientists and Information Technologists.

RobotiQ (2018). RobotiQ-Gripper. *https://www.universal-robots.com*.

Romero, D., Bernus, P., Noran, O., Stahre, J., and Fast-Berglund, Å. (2016a). The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation towards Human-Automation Symbiosis Work Systems. In *APMS (Advances in Production Management Systems)*.

Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., and Gorecky, D. (2016b). Towards an Operator 4.0 Typology: a human-centric perspective on the fourth industrial revolution technologies. In *International Conference on Computers & Industrial Engineering CIE 46*.

Scherer, S., Marsella, S., Stratou, G., Xu, Y., Morbini, F., Egan, A., Rizzo, A., and Morency, L.-P. (2012). Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In *Intelligent virtual agents*, pages 455–463. Springer.

Schroeder, R. (2008). Defining virtual worlds and virtual environments. *Journal For Virtual Worlds Research*, 1(1).

Schroeder, W. J., Zarge, J. A., and Lorensen, W. E. (1992). Decimation of triangle meshes. In *ACM Siggraph Computer Graphics*, volume 26, 2, pages 65–70. ACM.

Schuster, K., Groß, K., Vossen, R., Richert, A., and Jeschke, S. (2016). Preparing for industry 4.0–collaborative virtual learning environments in engineering education. In *Automation, Communication and Cybernetics in Science and Engineering 2015/2016*, pages 417–427. Springer.

Schwartz, T., Feld, M., Bürckert, C., Dimitrov, S., Folz, J., Hutter, D., Hevesi, P., Kiefer, B., Krieger, H.-U., Lüth, C., et al. (2016a). Hybrid Teams of Humans, Robots, and Virtual Agents in a Production Setting. In *Intelligent Environments (IE), 2016 12th International Conference on*, pages 234–237. IEEE.

Schwartz, T., Zinnikus, I., Krieger, H.-U., Bürckert, C., Folz, J., Kiefer, B., Hevesi, P., Lüth, C., Pirkl, G., Spieldenner, T., et al. (2016b). Hybrid teams: flexible collaboration between humans, robots and virtual agents. In *German Conference on Multiagent System Technologies*, pages 131–146. Springer.

Segura & Diez, H. V., Arbelaiz, A., Posada, J., and García-Alonso, A. (2018). Visual Computing Technologies to support the Operator 4.0. *Computers & Industrial Engineering, under revision.*

Shamir, A. (2008). A survey on mesh segmentation techniques. In *Computer graphics forum*, volume 27, pages 1539–1556. Wiley Online Library.

Sherman, W. R. and Craig, A. B. (2002). *Understanding virtual reality: Interface, application, and design*. Elsevier.

Silva, J. L., Bengoa, N. A., Quartulli, M., Olaizola, I. G., and Zulueta, E. (2015). Web-based supervised thematic mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5):2165–2176.

Simplygon-Studios (2015). Automagic 3D Optimization. *https://www.simplygon.com*.

Slater, M. and Steed, A. (2002). Meeting people virtually: Experiments in shared virtual environments. In *The social life of avatars*, pages 146–171. Springer.

Slater, M. and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616.

Sloan, R. J. S. (2015). *Virtual Character Design for Games and Interactive Media*. CRC Press.

Software, M. (Accessed in 2015). Polygon Cruncher: the optimization studio. *http://www.mootools.com/plugins/us/polygoncruncher/index.aspx*.

Song, H., Heo, S., Kang, J., and Lee, S. (2015). 3D Character Animation: A Brief Review. *Journal of International Society for Simulation Surgery*, 2(2):52–57.

Søraker, J. H. (2011). Virtual entities, environments, worlds and reality: Suggested definitions and taxonomy. *Trust and virtual worlds: Contemporary perspectives*, pages 44–72.

Stanovsky, D. (2004). Virtual reality. *The Blackwell Guide to the Philosophy of Computing and Information*, pages 167–177.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93.

Stork, A. (2015). Visual computing challenges of advanced manufacturing and industrie 4.0 [guest editors' introduction]. *IEEE computer graphics and applications*, 35(2):21–25.

Stratasys-Solution (2009). GrabCAD. *https://www.grabcad.com/*.

Su, M.-C. and Liu, I.-C. (2001). Application of the self-organizing feature map algorithm in facial image morphing. *Neural processing letters*, 14(1):35–47.

Tay, F. E. H. and Roy, A. (2003). CyberCAD: a collaborative approach in 3D-CAD technology in a multimedia-supported environment. *Computers in Industry*, 52(2):127–145.

Teichrieb, V., do Monte Lima, J. P. S., Apolinário, E. L., de Farias, T. S. M. C., Bueno, M. A. S., Kelner, J., and Santos, I. H. (2007). A survey of online monocular markerless augmented reality. *International Journal of Modeling and Simulation for the Petroleum Industry*, 1(1).

Thakur, A., Banerjee, A. G., and Gupta, S. K. (2009). A survey of CAD model simplification techniques for physics-based simulation applications. *Computer-Aided Design*, 41(2):65–80.

Thiebaux, M., Marsella, S., Marshall, A. N., and Kallmann, M. (2008). Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pages 151–158. International Foundation for Autonomous Agents and Multiagent Systems.

Thuemmler, C. and Bai, C. (2017). Health 4.0: Application of Industry 4.0 Design Principles in Future Asthma Management. In *Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare*, pages 23–37. Springer.

Tubbs, S. L. (2012). Human communication: Principles and contexts.

Ugarte, R., Barrena, N., Diez, H. V., Álvarez, H., and Oyarzun, D. (2016). Augmented Reality System to Assist in Manufacturing Processes. In *Spanish Computer Graphics Conference (CEIG)*, pages 75–82, Salamanca, Spain.

Universal-Robots (2018). UR10 Universal Robots. *https://www.universal-robots.com*.

Vallés, J.-L. (2013). *Factories of the Future 2020*. European Comission.

Wang, W. (2010). *Muscle-based Facial Animation*. PhD thesis, Texas A&M University.

Wang, W. Y., Pauleen, D. J., and Zhang, T. (2016). How social media applications affect B2B communication and improve business performance in SMEs. *Industrial Marketing Management*, 54:4–14.

Web3D-Consortium (2001). X3D - Extensible 3D. *http://www.web3d.org/x3d/what-x3d*.

Wei-dong, W. (2010). Research and Application of Engineering Drafting Teaching Mode based on Technology of Three Degree CAD. *Science & Technology Information*, 23:479.

Welbergen, H., Reidsma, D., Ruttkay, Z. M., and Zwiers, J. (2009). Elckerlyc-A BML realizer for continuous, multimodal interaction with a virtual human. *Journal on Multimodal User Interfaces*, 3(4):271–284.

Wexelblat, A. (1998). Don't make that face: A report on anthropomorphizing an interface. *Intelligent Environments*, 173(179):98–02.

Wolf, M., Kleindienst, M., Ramsauer, C., Zierler, C., and Winter, E. (2017). Current and future industrial challenges: Demographic change and measures for elderly workers in industry 4.0. In *Proceedings of the 9th International Sientific Conference Management of technology step to sutainiable production*. Croatian Association for PLM.

Yamamoto, E., Nakamura, S., and Shikano, K. (1998). Lip movement synthesis from speech based on hidden markov models. *Speech Communication*, 26(1):105–115.

Yee, N. and Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290.

Yong, J. (2014). Design and Implementation of 3D Facial Animation Based on MPEG-4. In *Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 4*, pages 753–759. Springer.

Zanbaka, C., Goolkasian, P., and Hodges, L. (2006). Can a virtual cat persuade you?: the role of gender and realism in speaker persuasiveness. In

Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 1153–1162. ACM.

Zhang, M., Lv, Z., Zhang, X., Chen, G., and Zhang, K. (2009). Research and Application of the 3D Virtual Community Based on WEBVR and RIA. *Computer and Information Science*, 2(1):84.

Zhao, Q. (2009). A survey on virtual reality. *Science in China Series F: Information Sciences*, 52(3):348–400.

ANNEX I

List of Publications

1. Improving Gestural Communication in Virtual Characters. del Puy Carretero, M., Ardanza, A., García, S., Diez, H. V., Oyarzun, D., & Ruiz, N. In International Conference on Articulated Motion and Deformable Objects (pp. 69-81). Springer, Berlin, Heidelberg. (2012).

2. 3D Animated Agent for Tutoring based on WebGL. Diez, H. V., García, S., Sánchez, J. R., & del Puy Carretero, M. In Proceedings of the 18th International Conference on 3D Web Technology (pp. 129-134). ACM. (2013).

3. Realistic Visual Speech Synthesis in WebGL. Mujika, A., Diez, H. V., Alvarez, A., Urteaga, M., & Oyarzun, D. In Proceedings of the 18th International Conference on 3D Web Technology (pp. 207-207). ACM. (2013).

4. IAAN: Intelligent Animated Agent with Natural Behaviour for Online Tutoring Platforms. Diez, H. V., García, S., Sánchez, J. R. , del Puy Carretero, M., & Oyarzun, D. In Proceedings of the 6th International Conference on Agents and Artificial Intelligence-Volume 2 (pp. 123-130). SCITEPRESS-Science and Technology Publications, Lda. (2014).

5. Providing Physical Appearance and Behaviour to Virtual Characters. del Puy Carretero, M., Diez, H. V., García, S., & Oyarzun, D. In International Conference on Articulated Motion and Deformable Objects (pp. 98-107). Springer International Publishing. (2016).

6. Interactive Multimodal Platform for Digital Signage. Diez, H. V., Barbadillo, J., García, S., del Puy Carretero, M., Alvarez, A., Sánchez, J.

R., & Oyarzun, D. In International Conference on Articulated Motion and Deformable Objects (pp. 128-137). Springer, Cham. (2014).

7. Virtual Training of Fire Wardens through Immersive 3D Environments. Diez, H. V., García, S., Mujika, A., Moreno, A., & Oyarzun, D. In Proceedings of the 21st International Conference on Web3D Technology (pp. 43-50). ACM. (2016).

8. Augmented Reality System to Assist in Manufacturing Processes. Ugarte R., Barrena N., Diez H. V., Álvarez H., & Oyarzun D. In Proceedings of the XXVI Spanish Computer Graphics Conference (pp. 75-82). Eurographics Association. (2016).

9. Interaction Challenges in the Development of Fire Warden VR Training System using a HMD. Diez, H. V. , Moreno, A. & Garcia-Alonso, A. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP, 84-91. (2017).

10. Visual Computing Technologies to support the Operator 4.0. Segura, Á., Diez, H. V., Arbelaiz, A., Posada, J., A. & Garcia-Alonso. Computers & Industrial Engineering. (under second review, 2018).

11. 3D Model Management for e-commerce. Diez, H. V., Segura, A., García-Alonso, A. & Oyarzun, D. Multimedia Tools and Applications-Volume 76 (pp. 21011-21031). (2017).

Improving Gestural Communication in Virtual Characters

- Authors: María del Puy Carretero, Aitor Ardanza, Sara García, Helen V. Diez, David Oyarzun and Nuria Ruiz
- Booktitle: International Conference on Articulated Motion and Deformable Objects
- Year: 2012
- Publisher: Springer

Improving Gestural Communication in Virtual Characters

María del Puy Carretero[1], Aitor Ardanza[1], Sara García[1],
Helen Díez[1], David Oyarzun[1], and Nuria Ruiz[2]

[1] 3D Animation and Interactive Virtual Environments, Vicomtech,
P. Mikeletegi 57, Donostia-San Sebastián, Spain
[2] Elixir Films S.L., C/ Portuetxe, 23B, 4ª planta, oficina 13-6,
Donostia-San Sebastián, Spain
{mcarretero,aardanza,sgarcia,hdiez,doyarzun}@ivicomtech.org,
nuriarc@elixir-films.com

Abstract. Gestural communication is part of non-verbal communication, but it is also a method that deaf people use to communicate with others through sign languages. In this paper, we present the research to improve communication in virtual characters no matter if it is spoken or signed using non-verbal communication. A brief study of hand gestures has been made to obtain some rules to improve natural behaviour in virtual characters. These rules, like the use of beats, are applied in an animation engine that emulates the behaviour of real people. The behaviour of the virtual character can also be modified by other channels such as user's input, so the character's attitude has to be adapted. This paper describes the architecture and main modules of the prototype designed to achieve natural behaviour in virtual characters and present the validation results obtained during the first phase of the development.

Keywords: 3D Virtual Character, Natural Behaviour, Communication, Gestures.

1 Introduction

Communication through hands is extended and each culture has its own movements. Although there are some gestures that can be considered as universals [1] such as "ok", "stop", "goodbye", "crazy", etc., Birdwhistell [2] concluded that there are no universal gestures, but they are acquired over time influenced by the culture.

Scientific community has established four types of hand gestures [3]:

- Iconics: represent some feature of the accompanying speech.
- Metaphorics: represent an abstract feature concurrently spoken about.
- Deictics: indicate a point in the space.
- Beats: are small formless waves of the hand that occur to emphasize words.

McNeill [4] explains that in some speech contexts about three-quarters of all sentences are accompanied by hand gestures; of these, about 40% are iconic, 40% are beats, and the remaining 20% are divided between deictic and metaphoric gestures.

F.J. Perales, R.B. Fisher, and T.B. Moeslund (Eds.): AMDO 2012, LNCS 7378, pp. 69–81, 2012.

These types of gestures accompany verbal communication and their use is to complement the speech. However there is another kind of communication that involves gestures with hands and arms, that is sign language. Thanks to sign languages, deaf people can communicate with others making gestures using their hands and arms but these signs are always accompanied by other non-verbal elements such as facial expressions, pose, etc.

This work presents the research that involves improving gestural communication in virtual characters. There are two main objectives. First, it is to improve natural behaviour in virtual characters thanks to the execution of gestural movements like beats or pointing, as well as other non-verbal elements such as blinking. Secondly, it is to achieve natural transition in the execution of the movements which makes the communication and the behaviour of the virtual character more natural.

Section below summarizes the state of the art of hand gestures. In this section we analyze typical gestures and the moment when they are generally made. Section 3 makes a brief summary about other research related to natural behaviour and methods to achieve it. Section 4 explains the architecture designed to provide natural gestural behaviour to the virtual character. In section 5 two different projects related to improving communication with virtual characters are explained. These projects validate the architecture and one of them has been tested with real users who gave us their opinion. Finally, section 6 shows our conclusions.

2 Type of Hand Gestures

In this section a study of hand gestures is presented. As Cassel et. al., [3] explained, there are four types of hand gestures. They are made depending on the purpose of the speaker. However, sometimes they are made unconsciously. Thanks to this study, behaviour rules have been obtained in order to apply them to virtual characters. Thus, natural behaviour in gestural communication is achieved.

2.1 Iconic Gestures

McNeill [4] was who named this type of gestures that are characterized by representing something concrete and accompany to what is being said. Other authors [5] had named these gestures as "illustrators" but the definition is the same: they are hand movements that are directly linked with speech and that help to illustrate what is being said verbally.

It seems that iconic gestures represent the shape and/or other physical characteristics of the object or thing that is being spoken [6]. However, some researches expose that the similarity with its referent does not explain totally all the apparitions of iconic gestures [7]. Researches reveal that iconic gestures are influenced by the context [6]. In addition humans are very different from each other. Depending on personal or cultural background, differences in gestures are obvious [8]. That means for example that gestures can be different to represent the same object.

According to McNeill [4] the most important aspect of iconic gestures is the ability to articulate, from the point of view of the orator, which are the most relevant characteristics of the context. McNeill concluded that iconic gestures allow observing the thoughts of the orator because these movements are not forced by rules or standards.

2.2 Metaphoric Gestures

Metaphoric gestures are reproduced when a concept is being explained. These gestures are made using three-dimensional space and are used to describe an idea or something specific such as “take hold of an arm” or something more general such as “shake hands”. As Cassell [9] explained, the concept metaphoric gestures represent has no physical form; instead the form of the gesture comes from a common metaphor. Cassel gave as an example it can be said "the meeting went on and on" accompanied by a hand indicating rolling motion.

Parrill and Sweetser [10] proposed that even metaphoric gestures have an iconic aspect. This iconic aspect invokes some visual or concrete situation, entity or action, by means of the hand shapes and motions. So it seems that these gestures are similar to iconic gestures. However, the difference is that metaphoric gestures represent a more abstract idea [11].

2.3 Deictic Gestures

Deictic gestures accompany to some linguistic elements of pointing. These elements can refer to people such as “you”, “he”, “she”, etc.; to a place like “here”, “there”, “up”, etc.; to time such as “today”, “yesterday”, “now”, etc.

According to Alibali [12] words like “this”, “that”, etc., need to be accompanied with deictic gestures so they can be interpreted by the interlocutor. Most of the deictic gestures are made with the index finger extended and the rest of the fingers forming a fist. In Table 1 a short list of typical deictic gestures is shown.

Table 1. Example of some deictic gestures

Meaning	Gesture
Involve to the interlocutor	Point to interlocutor
Involve to oneself	Point to oneself
Now/here	One or both index finger pointing down
There	Point to somewhere
Up/down	Put up/down one arm with the index finger extended
To the left/right	Move one arm to left/right
Fom up to down / From down to up	Move one arm from up to down / down to up

Cienki and Müller [13] argued that some uses of abstract deictic could also be considered metaphoric, if the gesture is interpreted as pointing to an event (a new scene) as an object. However as Kobsa et. al., [14] explained deictic gestures not only involves hands and fingers but also other elements such as head, eyes or even pointing with a pencil.

2.4 Beats

Beats are short and simple gestures that accompany the rhythm of the discourse, so beats accompany the prosodic peaks of the speech. These gestures are named as beats because they are similar to the movements that conductors make [15].

The function of these gestures is similar to pitch and intonation when speaking. Beats act as markers of the structure of the information associated to the expression [16]. Krahmer and Swerts [17] research shows that there is a connection between the beat gesture and prosody in the production of it as well as in its perception. Thus, it can be said that beats represents the same rhythm than prosody accents.

As Kettebekov [11] explained, these gestures are possibly the most spontaneous and the smallest of the gestures made with hands. According to McNeill [4] the role of beats is to indicate that the narrator is taking the conversation through different states.

3 Related Work

Providing natural behaviour to virtual characters is the objective of several research groups. There are different studies about the natural behaviour of virtual characters in diverse applications and cases of study. In most of them, the authors base their research on hand and arm gestures.

For example, with the aim of increasing the success of virtual characters as other method of human-computer interaction, Rieger [18] compiled a list of speech acts to give the message and the state of the system. For example to inform about the status of the system, if a warning message appears, the virtual character has to show importance, if it is busy, shows regrets, if error, shows sorry, etc. In his architecture, Rieger distinguishes between behaviour rules and gesture rules to develop the controller and renderer of the virtual character. The author explains that the behaviour is responsible for the virtual characres mimic, speech, and head-movements; and gesture rules are related to hand movements.

López et. al., [19] established some rules in order to provide realism to a virtual character during the interaction with people in a spoken dialog system. Their objective was to make the virtual character look alive. The experimental case of study was a domotic service where users “called” to home to check the state of their electrical appliance. The strategy followed by the spoken dialog system was designed to deal with some critical states. For example, at the beginning of the dialogue, the virtual character has to look at the camera, smile and wave a hand; when the system starts to speak, the virtual character has to look directly at the user and raise the eyebrows; or when the system gives some explanation to the user, the virtual character has to make beat gestures with the hands.

Bergman and Kopp [20] expose that the speech is also influenced by the gestures. The authors explain that they made some experiments and concluded that when making gestures is allowed, the vocabulary is richer than when gestures are forbidden. They mention the conclusion of Rauscher et. al [21] that prohibiting gestures in cartoon narrations made speech less fluent. In addition in these cases the speaker tends to use more words like “uhm”. To conclude, making gestures enriches the language used by people and the communication is more fluent.

Alseid and Rigas [22] studied the user’s view of facial expressions and body gestures in e-learning applications. To their experiment they used different expressions (interested, amazed, happy and smiling, neutral and thinking) and eight body gestures categorized into positive (hands clenching – front and back, open palms, pointing, chin stroking and hands steepling) and negative (arms folded and legs grossed). Users evaluated positively body animations. Their results revealed that including specific body gestures in interactive e-learning interfaces could be attractive for users.

In order to improve the transition between two different gestures during the generation of French Sign Language, Efthimiou and Kouroupetroglo [23] use coarticulation parameters to avoid using an intermediate rest posture. In their first version of their technique coarticulation was implemented as a simple interpolation between the last frame of a sequence and the first frame of its following sequence.

The related work shows several authors are researching on improving the natural gestural behaviour of virtual character including body gestures, and researching on developing techniques to make the animation more natural.

Our goal is to provide not only specific gestures in predefined states, but also improve the natural and unpredictable actions of virtual characters. The objective is to establish a randomly but coherent behaviour based in behaviour rules, that is modified by specific gestures depending on the context of the interaction. Besides, these gestures and actions vary depending on the mood of the virtual character. In addition the transition between gestures has to result natural in order to emulate movements of real people. For that purpose it is necessary to develop transition techniques between predefined gestures.

4 Architecture

This section explains our platform's global architecture. The main objective is to improve the gestural movements of virtual characters when they communicate with others. Figure 1 shows the modules of the architecture.

The main module is the Animation Engine and it is supplied by the rest of the components of the architecture and it is also in charge of rendering the final appearance and behaviour of the virtual character. This module works influenced by several factors that change the attitude of the virtual character:

- The Cognitive Module takes into account the inputs of the system and it obtains an estimated emotion to modify the behaviour of the virtual character.
- A Markup Language, such as BML (Behaviour Markup Language) [24] or VHML (Virtual Human Markup Language) [25] can be used to indicate some specific movements to the virtual character.
- The TextToSpeech (TTS) module synthesizes text to obtain audio files that form the voice of the virtual character. In addition it obtains a *.pho file that contains the information that is needed to animate the lips according to what text says.
- The Behaviour Rules Database contains some rules of behaviour that are applied randomly but coherently to provide natural behaviour to the virtual character.

4.1 Cognitive Module

The Cognitive Module is responsible for estimating the virtual character's emotion depending on the inputs of the system and it is dependent on the logic of the application. That means that it is necessary to know the objectives of the application to be able to predict with which emotion the virtual character has to react to the user's actions.

So, for an interactive system to be capable of generating emotions, it will need to be provided with a module that is automatically capable of generating emotions. The emotions will be the result of a cognitive process that evaluates the events in the environment. There are several emotional models on which this cognitive perspective have been based. Our Cognitive Module implements Roseman Models [26] due to the reasons argued by Ortiz et. al., [27]. This module is able to predict Ekman's emotions.

However, depending on the characteristics of the application it is extensible to adapt this module using another cognitive model which fits better to the objectives of the application. No matter which cognitive model is used, the important result is the estimated emotion with which the virtual character has to react. This emotion has influence on the final animation both facial and corporal.

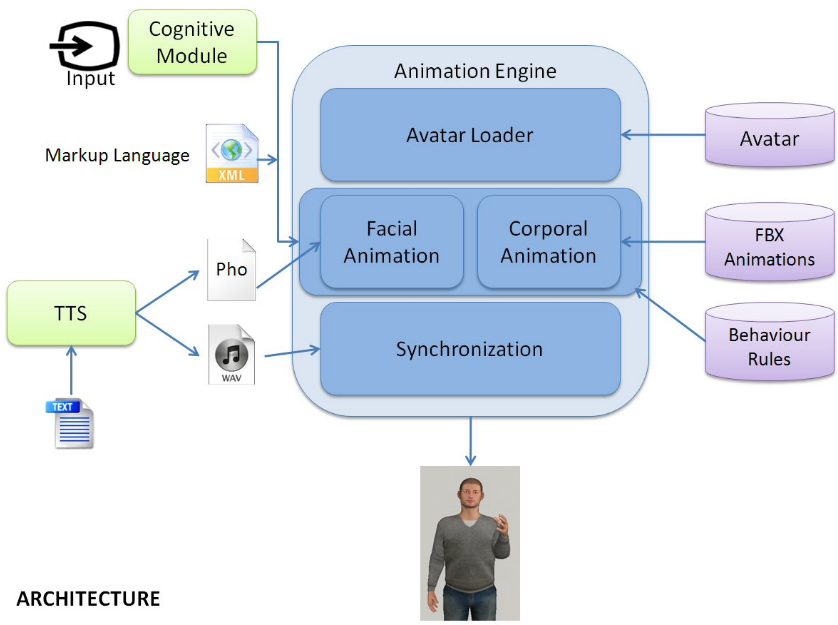


Fig. 1. Architecture

4.2 TextToSpeech (TTS) Module

The TTS Module uses a commercial synthesizer to obtain the audio file which composes the voice of the virtual character. For our prototypes we use Loquendo TTS that is speech synthesis software available with several voices in different languages, to fully meet the needs of all application environments.

In addition, this module is in charge of obtaining the *.pho file. This file contains the phonemes that are needed to animate the lips of the virtual characters as well as the duration of each phoneme. Thanks to this information, the mouth of the virtual character is animated simulating a real person talking.

4.3 Databases

The architecture uses three main databases that feed the animation engine (which will be explained later). Thanks to these databases it is possible to load the virtual character, to apply its behaviour and modify it according to external factors such us user's interaction or behaviour rules that affect the natural behaviour of the virtual character.

- *Avatar database* contains the 3D models that form the physical appearance of the virtual character. A virtual character is created using 3D design software like 3ds Max or Maya. Apart from the physical appearance of the character it is also required that the facial expressions of the virtual character are created in order to apply facial animation techniques as it will be explained in the following section. The structure of each character is stored in an xml file which contains all the 3D models required to load the virtual character and to animate it.
- *FBX animations database* is composed for the pre-defined animations that the virtual character has to execute during its animation. These animations can be obtained thanks to motion capture systems like CyberGlobe II that captures the movements of hand and fingers (http://www.cyberglovesystems.com), or Organic Motion (http://www.organicmotion.com) that captures movements of the whole body. The animations obtained thanks to these systems are applied to the skeleton of each character and stored in the database.
- *Behaviour Rules database* includes some rules to provide virtual characters with natural behaviour. These rules were obtained through the research of several authors as explained in section 2. Most of them are related to hand and arm movements while interacting with others, but they also include rules related to the gaze; for example, looking at the listener when ending a sentence [28], or to the postures, such as changing balance generally caused by fatigue and small variations in body posture due to muscle contractions. These rules pretend to make the virtual character seem more alive and to emulate the behaviour of a real person.

4.4 Animation Engine Module

The Animation Engine Module is the main module of the architecture. It is responsible for generating the virtual character's animation as well as to load it. This module is developed using Open Scene Graph (http://www.openscenegraph.org/) and loads the virtual character from the Avatar database. Once the model of the character is loaded, the animation that forms the virtual character's behaviour can be applied. For that purpose there are two different sub-modules: Speech Animation and Behaviour Animation.

The Speech Animation sub-module executes the facial animation when the virtual character speaks. The facial animation is generated using morphing techniques where the key facial expressions considered to make the animation are the Ekman's emotions as well as other expressions such as eyes closed, winks and a neutral face. For this task the phonemes of the speech are needed. As it was explained before, they are given by the *.pho file.

The Behaviour Animation sub-module deals with both the predefined rules for natural behaviour and with the emotion obtained with the cognitive module. The rules are launched automatically but they are executed randomly in order to give improvised behaviour as people usually do. The animations of these basic movements are taken from the FBX database and they are applied using inverse kinematics to concatenate several animations and to obtain realistic movements. The cognitive module changes the behaviour executed by the rules. This module guesses the attitude that the virtual character has to show depending on the inputs of the system. Thus, the speed of the gestures and movements is changed. In addition the character's expression is modified. The face shows the predicted emotion, so it implies more realism to the virtual character.

The natural behaviour of the virtual character can also be modified by other specific movements and gestures. These gestures are provided by the input of the system or by a predefined behaviour in Markup Language files (XML). This may be because it can be needed to execute certain movements at given moments. For example, as it will be explained later, when the virtual character's role is a virtual tutor, the specific movements are given by the real tutor when he/she prepares the course.

The Render module synchronizes the audio (the voice of the virtual character) and the generated animation. The result is a realistic virtual character emulating a real person and his/her natural behaviour while interacting with others.

5 Applications That Validate the Architecture

The architecture is being validated in two different projects related with virtual characters and natural behaviour in communication with people. In the first one, the main objective is to develop natural behaviour in virtual teachers in order to emulate a real one to improve e-learning courses. The second project improves the behaviour in virtual characters that "speak" Sing Language. These projects are explained in the following sections.

5.1 Natural Behaviour in e-Learning

E-learning applications that integrate virtual characters as teachers have been developed during years. However most of these researches are focused on talking heads, or predefined characters that do not have enough realism in their animations and are far from real behaviour.

The objective of our project is to improve the communication between the virtual teacher and students. The virtual teacher has two different tasks: to explain the lessons and to evaluate the students. During the phase of explanation: the virtual character emulates the behaviour of a real teacher giving explanations by speaking and making gestures that help to understand the meaning of the lecture. In addition the virtual teacher makes natural gestures that humans make unconsciously such as beats and others not related to hands, such as blinking.

The gestures that help to understand the meaning of speech are usually associated with words or sentences. It is necessary to know which gestures the virtual character has to reproduce in order to make an analysis of the sentences. That is not an objective of the project, so as a preliminary solution, and as the lessons are pre-edited by real tutors, the course editor indicates which gestures the virtual character has to reproduce while explaining the lesson. For that purpose, we use markup languages oriented to virtual characters like BML (Behaviour Markup Language) or VHML (Virtual Human Markup Language) in order to indicate which gesture goes with which word or sentence. For example, the course editor can select deictic gestures or predefined iconic or metaphoric gestures. These gestures depend on the context so they have to be designed previously.

The gestures that the character has to execute unconsciously, such as blinking or beats, are given by the rules established in the behaviour rules database and executed through the behaviour animation sub-module. However this behaviour can be modified by the markup gestures.

The second phase is the student's evaluation. In this case the natural behaviour is modified depending on the input of the student. The Cognitive Module predicts the emotion that the virtual teacher has to express depending on the student's responses. For example if the student's answers are correct the virtual teacher should be glad and act according to this emotion. In contrast, if the student's answers are wrong, the virtual character gets angry and its behaviour is ruddier. Once the emotion is predicted and it is known if the answer is correct, the character's behaviour is given by the behaviour rules and adapted with the emotion. For example, if the student answers correctly, the virtual character makes gestures with the arms and hands indicating good!, ok!, perfect!, etc. The face shows happiness and the virtual tutor says that the answer is correct and encourages the student to keep up the good work. On the other hand, if the answer is wrong the character makes negation gestures with the arms and hands, the face shows sadness and the virtual tutor replies that the answer is incorrect and advices to concentrate or to revise the lesson.

Fig. 2 shows two different virtual characters' reactions depending on the student's answers. On the left the student answered correctly, so the virtual tutor makes a satisfaction gesture with his arms and his face shows gladness. On the right the student's answer was incorrect so the virtual character rejects the answer and shows a sad expression.

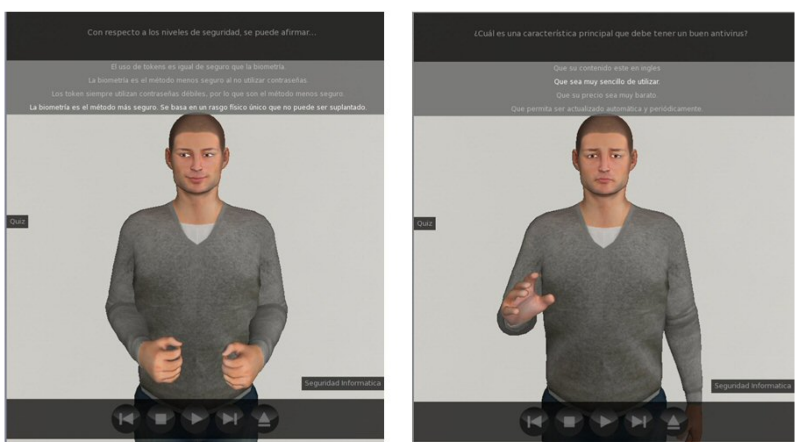


Fig. 2. Different virtual characters' reaction depending on the answer, happy if it is correct (left), sad if it is wrong (right)

5.2 Virtual Sign Language Interpreter

The main objective of the second project is to make media services accessible to people with a hearing disability through the use of sign language. For that purpose, a virtual character whose role is to interpret Spanish Sign Language (LSE) is used. What the virtual character has to interpret is given via text, which is translated to LSE and executed by the virtual interpreter. The translation from text to LSE is being developed in a given domain (meteorology) and this will be described in successive papers.

Getting a realistic and natural animation of a signing character in real time is an arduous task. There are two main points to solve. The first point is to ensure that deaf or other hear-disabled people understand the sequence of gestures the virtual character reproduces. To achieve this, a natural and realistic concatenation of gestures is necessary.

The second point is related to a correct interpretation of the gestures. Emotions and mood are reflected by people when they communicate with others. Emotions are noticed by the facial expression; however they also modify the execution of gestures. That means that the same gesture is slightly different depending on how the person feels, for example happy, sad or angry. When someone is happy the movements are a bit quicker, when a person is feeling sad they are a bit slower and when being angry they are more abrupt. So it is necessary to modify the speed of the animation execution. In addition, in Spanish Sign Language (LSE), there are some gestures that are different depending on the mood of the person who signs.

Thanks to the Animation Engine Module, our architecture realizes the animation of the signing character in LSE. The Animation Engine interprets the sequence of gestures that a Text-LSE module obtains from the translation from text to Spanish Sign Language.

The animation engine loads the required gestures from Sign Language Database (FBX database). These gestures are captured previously with a motion capture system and they need to be made by experts in LSE in order to get realistic movements.

The animation engine also loads the virtual interpreter from a database. Avatars have to be designed according to some rules. For example, all the bones and joints of the hands and arms have to be defined in order to reproduce the gestures properly.

The Animation Engine generates the animation corresponding to the sign language translation. This animation is modified by a given emotion that can be given manually or depending on the user's interaction thanks to the cognitive module. This emotion is expressed on the face of the character, but it also has an influence on the execution of the animation changing its speed. Figure 3 shows the same sign gesture executed with different emotions: neutral, angry and sad.

Fig. 3. A virtual character signing with different emotions: neutral, angry and sad

In addition the animation engine provides natural behaviour to the virtual interpreter. This means that the character does not just play the necessary gestures, but it also has other movements to improve the communication. Thus it seems that the virtual character is alive. For example, in waiting state (without talking in LSE) it makes movements like a real person does when waiting; while interpreting in sign language, the virtual character acts as a real person moving the head, eyes or body as people do, as well as expressing emotions. That provides realistic behaviour to the virtual interpreter allowing deaf and other hear-disabled people to better understand the information that the virtual interpreter offers.

5.3 Evaluation and Results

The virtual tutor application is still under development and has not been evaluated by real users (students and teachers) yet. However we have a first vision of deaf people that have evaluated the virtual interpreter. This not only tests this application, but it also allows us to know if the behaviour of the character seems natural to people. The virtual interpreter was evaluated by one deaf person, by a real interpreter and by a person who knows LSE. They tested the application and made several comments that are summarized as follows:

- Although they understood the sentences interpreted by the virtual character, they suggested that the transition between signs should be smoother.
- The expression of the virtual interpreter has to be more exaggerated because deaf people give importance to this factor. They agree that the emotions can be recognized, but when people use sign language they exaggerate their expression.
- Eyebrows, gaze and shoulders are also important for the correct interpretation of sign language. For example, eyebrows are fundamental to express questions or exclamations. So it is necessary to include some rules to this behaviour.
- The behaviour and attitude during the waiting state is correct. Their opinion was that the virtual character emulates the behaviour or a real person.

In general, the evaluation of deaf people was positive, they appreciate the efforts made to make media services more accessible and their opinion about the result was positive although it was a preliminary development. They mentioned that the behaviour during the waiting state was good emulating the behaviour of real people.

6 Conclusions

We have designed and developed an architecture that improves the natural behaviour of virtual characters when they communicate with people. It is focussed on hands and arms but it also includes other aspects such as emotions, blinking, posture, gaze, etc.

For that purpose we have analyzed the common behaviour of people when they communicate with others especially the hands and arms movements. From this behaviour we have obtained some common rules and they have been integrated in the Animation Engine that executes them by default but randomly.

The default behaviour of the virtual character is modified by other specific gestures. In addition the emotion and mood of the character vary the execution of the gestures and the behaviour in the speed of the movements.

The architecture is being tested in two different projects that integrate virtual characters to communicate with people. The first project improves the communication with students in e-Learning applications. The virtual teacher has a natural behaviour and explains the lesson accompanied by gestures that make it easier to understand the concepts. In the second project the virtual character communicates through sign language with deaf people. However the virtual interpreter not only executes the gestures, but he/she also modifies them depending on the mood or emotion. In addition, the virtual character has a default behaviour that improves the realism of the virtual interpreter during the waiting state (without using sign language).

The evaluation was satisfactory. Although the project was in an initial phase, deaf people gave us a good feedback about the result. They suggested to make the transition between gestures a bit smoother and to exaggerate the emotions and expressions because that is something they do when they communicate with others. However they admitted the emotions and the purpose of the character could be recognized.

As future work we plan to improve the way of executing specific gestures. In other words, our objective is to analyze the speech and to obtain the gestures and movements that real people would make in each case. In addition we pretend to extend the basic rules for natural behaviour based on the virtual character's personality. This is because each person has his/her own way of acting so the same idea can be applied to virtual characters.

References

1. Andersen, M.L., Taylor, H.F.: Sociology: The essentials, 4th edn. Thomson Wadsworth, Beltmon (2007)
2. Birdwhistell, R.L.: El lenguaje de la expresión corporal. Ediciones Gustavo Gili, Spain (1979)
3. Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, W., Douville, B., Prevost, S., Stone, M.: Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: Computer Graphics. Annual Conf. Series, pp. 413–420. ACM (1994)
4. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press, USA (1992)
5. Ekman, P., Friesen, W.: The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. Semiotica 1, 9–98 (1969)
6. Bergmann, K., Kopp, S.: Modeling the Production of Co-Verbal Iconic Gestures by Learning Bayesian Decision Networks. Applied Artificial Intelligence 24(6), 530–551 (2010)
7. Streeck, J.: Depicting by gesture. Gesture 8(3), 285–301 (2008)
8. Hostetter, A., Alibali, M.: Raise your hand if you're spatial–Relations between verbal and spatial skills and gesture production. Gesture 7(1), 73–95 (2007)
9. Cassell, J.: A Framework For Gesture Generation And Interpretation. In: Cipolla, R., Pentland, A. (eds.) Computer Vision in Human-Machine Interaction. Cambridge University Press (1998)
10. Parrill, F., Sweetser, E.: What we mean by meaning. Gesture 4(2), 197–219 (2004)

11. Kettebekov, S.: Exploiting Prosodic Structuring of Coverbal Gesticulation. In: Proceedings of the ACM International Conference on Multimodal Interfaces, pp. 105–112 (2004)
12. Alibali, M.W., Kita, S., Young, A.J.: Gesture and the process of speech production: We think, therefore we gesture. Language and Cognitive Processes 15(6), 593–613 (2000)
13. Cienki, A., Müller, C.: Metaphor, gesture, and thought. In: Gibbs Jr., R.W. (ed.) The Cambridge Handbook of Metaphor and Thought. Cambridge University Press, Cambridge (2008)
14. Kobsa, A., Allgayer, J., Reddig, C., Reithinger, N., Schmauks, D., Harbusch, K., Wahlster, W.: Combining deictic gestures and natural language for referent identification. In: Proceedings Eleventh COLING, Bonn, Germany, pp. 356–361 (1986)
15. Giorgolo, G.: Space and Time in Our Hands. PhD. Universiteit Utrecht (2010)
16. Pierrehumbert, J., Hirschberg, J.: The meaning of intonational contours in interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (eds.) Intentions in Communication, ch. 14. MIT Press (1990)
17. Krahmer, E., Swerts, M.: The efects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. Journal of Memory and Language 57(3), 396–414 (2007)
18. Rieger, T.: Avatar Gestures. Journal of WSCG 2003 11(2), 379–386 (2003)
19. López, B., Hernández, A., Díaz Pardo de Vera, D., Santos de la Cámara, R., Rodríguez, M.C.: ECA gesture strategies for robust SLDS. In: Convention - Communication, Interaction and Social Intelligence on the Symposium on Multimodal Output Generation, Aberdeen, Scotland (2008)
20. Bergmann, K., Kopp, S.: Multimodal content representation for speech and gesture production. In: Theune, M., van der Sluis, I., Bachvarova, Y., André, E. (eds.) Symposium at the AISB Annual Convention: Multimodal Output Generation, Aberdeen, UK, pp. 61–68 (2008)
21. Rauscher, F.H., Krauss, R.M., Chen, Y.: Gesture, speech, and lexical access: The role of lexical movements in speech production. Psychological Science 7, 226–231 (1996)
22. Alseid, M., Rigas, D.: Users' views of facial expressions and body gestures in e-learning interfaces: an empirical evaluation. World Scientific and Engineering Academy and Society, WSEAS (2009)
23. Efthimiou, E., Kouroupetroglou, G. (eds.): Proc. of the 9th International Gesture Workshop, Athens, Greece, May 25-27, 2011, pp. 25–27 (2011)
24. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjalmsson, H.: Towards a Common Framework for Multimodal Generation in ECAs: The Behavior Markup Language. In: Proceedings of the 6th International (2006)
25. Marriott, A.: VHML – Virtual Human Markup Language. In: Talking Head Technology Workshop, at OzCHI Conference (2001)
26. Roseman, I., Antoniou, A., Jose, P.: Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. Cognition and Emotion 10 (1996)
27. Ortiz, A., Oyarzun, D., Carretero, M., Posada, J., Linaza, M.T., Garay-Vitoria, N.: Analysing emotional interaction through avatars. A generic architecture. e-Minds: International Journal on Human-Computer Interaction I(5) (March 2009) ISSN: 1697-9613 (print) - 1887-3022 (online)
28. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica 26, 22–63 (1967)

3D Animated Agent for Tutoring based on WebGL

- Authors: Helen V. Diez, Sara Garía, Jairo R. Sánchez and María del Puy Carretero
- Booktitle: Proceedings of the 18th International Conference on 3D Web Technology
- Year: 2013
- Publisher: ACM

3D Animated Agent for Tutoring Based on WebGL

Helen V. Diez *, Sara García †, Jairo R. Sánchez ‡, Maria del Puy Carretero §
Vicomtech-IK4 Research Center, San Sebastián, Spain

Abstract

The goal of the work presented in this paper is to develop a 3D web based online tutoring system that enhances the motivation and cognitive development of students. To achieve this, a virtual assistant will be integrated to the e-learning platform; this 3D modeled e-tutor will evaluate each student individually, it will react to their learning progress by empathetic gestures and it will guide them through the lectures according to their personal needs. The accomplishment of these tasks will imply a thorough study of the latest techniques on artificial intelligence, multi-agent architectures and their representation by means of 3D emotional avatars.

CR Categories: I.2.11 [Distributed Artificial Intelligence]: Intelligent agents; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality; I.3.8 [Computer Graphics]: Applications;

Keywords: virtual agents, e-learning, artificial intelligence, Web3D technology

1 Introduction

The use of online learning or e-learning has increased significantly in the past years and it is expected to keep growing in the future. According to the latest survey by Ambient Insight Research [Adkins 2013] the aggregate growth rate for self-paced e-learning products and services expected for the next five year period (2011-2016) is 7.6%. Distance learning offers a series of benefits that traditional learning does not, for example in terms of mobility, affordability or flexibility e-learning happens to be much more suitable for nowadays lifestyle, moreover this technology has enabled increasingly dynamic and engaging learning experiences.

Many studies have been held to research the benefits of online learning, but this paper focuses on the benefits that the introduction of virtual agents in e-learning platforms may have in the cognitive process of students. This work introduces an animated agent with real-time humanlike responses to students' interaction, by means of non-verbal gestures, natural behaviour and verbal communication; this artificial intelligence will increase the students' motivation and will engage them to the lectures.

Studies like [Bloom 1984] demonstrated the effectiveness of one-on-one human tutoring against other methods of teaching, and [Lepper et al. 1993], defended the idea that education could be globally improved if every student was provided with a personal tutor. This is something almost impossible to achieve in traditional education but it is not so in online learning. Intelligent animated agents represent a new generation of human computer interface (HCI) design. Animated pedagogical agents [André et al. 1997] [Shaw et al. 1999] [Lester and Stone 1997] [Piesk and Trogermann 1998] [Schöch V and G 1998] are life-like autonomous agents that facilitate human learning by interacting with learners and make computer-based learning more engaging and effective [Johnson 1998].

However, most commonly used e-learning systems are not interoperable with other platforms such as Learning Management Systems, web-based virtual world platforms, Virtual Reality learning systems or simulators; this causes several compatibility problems and limits their use. The work presented in this paper solves this issue as it is entirely web based so it will run in any compatible browser. Meaning that it will run in any browser supporting WebGL [Leung and Salga 2010] technology as it is the Application Programming Interface (API) chosen in this work to render the 3D agent via web. WebGL is based on OpenGL, which is a widely used open source 3D graphics standard. Nowadays, most common browsers support this technology; Google Chrome, Mozilla Firefox, Apple Safari or Opera.

This paper is organized as follows; Section 2 analyzes the related work carried out in the last years concerning animated agents in virtual reality systems, Section 3 describes the architecture followed to accomplish the goals of this work, Section 4 shows validation results from the platform. The final section is about conclusions and future work.

2 Related work

The purpose of this work is to develop a virtual agent which will fulfill the role of a virtual tutor in various web-based e-learning systems. Such virtual tutor must react to the students' needs as if it were a real tutor, so this virtual agent must be provided with sufficient artificial intelligence to react in an autonomous way and natural behaviour to give the impression of interacting with a real teacher.

Research has outlined a number of desirable attributes for this kind of pedagogical agents [Ahmed 2005]. The agent must be an autonomous character and it must be able to perform almost every action without the direct intervention of other agents. Moreover, it must react to changes in the environment and respond to them over a certain period of time.

Many studies [Dehn and Van Mulken 2000] [Johnson et al. 2000] [Moundridou and Virvou 2002] [Baylor and Ryu 2003] have found that rendering agents with lifelike features, such as facial expressions, deictic gestures and body movements may rise the so called persona effect. A persona effect is a result of anthropomorphism derived from believing that the agent is real and authentic [Van Mulken et al. 1998] [Baylor and Ebbers 2003]. The persona effect shows that the presence of a lifelike character in an interactive learning environment can have a strong positive effect on students' perception of their learning experience [Lester et al. 1997] .

Providing natural behaviour skills to virtual agents is also a

*e-mail:hdiez@vicomtech.org
†e-mail:sgarcia@vicomtech.org
‡e-mail:jrsanchez@vicomtech.org
§e-mail:mcarretero@vicomtech.org

Web3D 2013, June 20 – 22, 2013, San Sebastian, Spain.

widely studied issue. However, defining natural behaviour is not an easy task. The literature and theory of affective computing imply several conditions for synthesized motion to appear natural [Abrilian et al. 2005]. Speed of interaction and emotion/speech-correlated believable body motion are among the most important functionalities [Mlakar and Rojc 2011]. Rieger [Rieger et al. 2003] developed a series of rules in order to increase the acceptance of virtual agents in Human-Computer Communication and established a correlation table between the message to rely and the emotion to show.

As far as the agent's appearance is concerned, research does not give a clear answer as to which design is best, whether to portray the agent with a real human look or with an iconic feature. For example, [Nass et al. 2000] proposed that the appearance of agents should be made to resemble that of the learners. However, such view is opposed by [Buisine and Martin 2007], who cited Kohar's recommendation [Kohar and Ginn 1997]; dramatized characters can display more exaggerated emotions than realistic humanlike agents.

Steve [Johnson and Rickel 1997] was one of the first pedagogical agents capable of expressing emotions; it was designed as a stereoscopic 3D character that cohabited with learners, it has been applied to naval training tasks. However, Steve was originally designed to operate in immersive virtual environments and not over the Web. Adele [Shaw et al. 1999], was the evolution of Steve into the World Wide Web, Adele's design was based on an autonomous agent paradigm able to use facial expressions and react to students' actions. More than a decade has passed since these works were presented, technology has evolved a lot since then and the internet has become accessible to almost everyone.

D'Mello's study [D'mello and Graesser 2012] presents two interactive intelligent systems that promote learning and engagement thanks to an animated teacher. These studies have motivated further work such as AutoTutor-Lite which is a simpler version of the former but keeps the most important features. AutoTutor-Lite is optimized for web-based learning environments; it uses a lightweight semantic engine that can be implemented as a small plug-in of flash movie that works on the learner's computer. It is web-served so the user can interact with AutoTutor-Lite in the web, introduce content and receive automatic tutoring back [Hu 2011].

More recently, [Benin et al. 2012] presented the implementation of a WebGL talking head for compatible browsers and iOS mobile devices based on LUCIA; a three-dimensional animated computer talking head which repeats any input text in six different emotional ways.

Although this system meets many of the requirements desired in our work (web-based, no additional plug-ins required, 3D modeled character) it does not interact with the users, neither reacts to their behaviour nor guide them throughout the learning process. The following section describes the architecture followed to achieve the goals of this work.

3 Architecture

This section introduces the architecture of the Animated Agent Engine implemented to work on e-learning platforms. The goal of this work is to develop a humanlike character to guide students in their learning process and to encourage them by showing emotional feedback. To accomplish this a virtual character has been designed following several rules regarding appearance and natural behaviour.

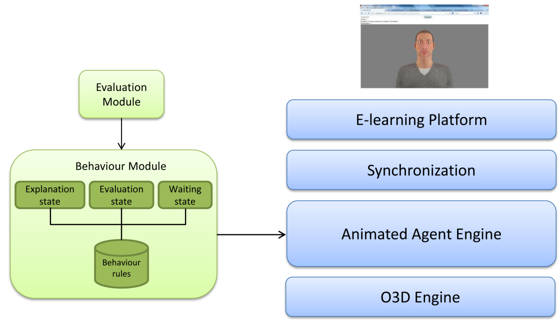


Figure 1: *Animated Agent Engine Architecture.*

The following sections describe the most relevant modules of the architecture shown in Figure 1.

3.1 Behaviour Module

In order to make the agent's behaviour more natural the agent will be able to perform several hand gestures. The scientific community has established four type of hand gestures [Cassell et al. 1994]:

- Iconic: or illustrators, they are descriptive gestures often used to illustrate speech.
- Metaphoric: or representational gestures, represent an abstract feature concurrently spoken about.
- Deictic: indicate a point in the space.
- Beats: small formless waves of the hand that occur to emphasize words.

Thus, the agent will be able to point out elements in the screen or perform spontaneous gestures with the hands. The agent will also vary the speed in which he performs these gestures depending on his mood [Carretero et al. 2012]. When people are sad their movements and gestures tend to be slower than when they are happy, in this case they execute movements in a quicker and energetic way.

The animated agent will also be able to reproduce facial expressions. Based on Ekman's six universal emotions [Ekman and Friesen 1981] the Animated Agent Engine is able to show the following facial expressions: happiness, sadness, surprise, anger and neutral expression. Disgust and fear have not been taken into account as they might lead to the opposite reaction desired from the student, the aim of this animated character is to stimulate the learner and these emotions may have a negative influence. Figure 2 shows happiness, surprise, anger and sadness from the agent in response to questions made by the student.

The role of the agent will be divided into three states:

- Explanation state: the animated character will introduce the lecture to student and will explain any necessary information. To make this explanation more appealing to the learner the agent will perform several natural behaviour movements, such as, iconic or deictic gestures. The behaviour in this stage can be edited by the real tutor who prepares the lessons. For this purpose, an authoring tool to edit the agents behaviour is provided.
- Evaluation state: the virtual agent reacts depending on the student's responses. This state is influenced by the Evaluation Module.

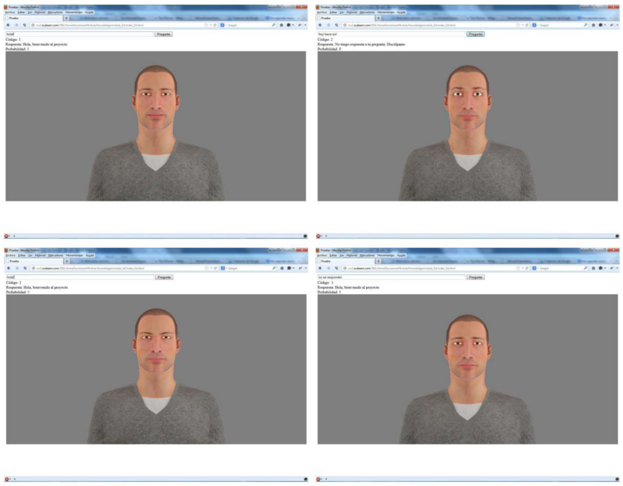


Figure 2: *Emotions from the Animated Agent reacting to student's input.*

- Waiting state: this state is executed when the virtual agent is neither explaining nor evaluating. The e-learning platform will provide several content in order to complete the course; text, video or audio may appear to clarify concepts, in this case the agent will perform waiting state movements. Humans do not stand hieratic when waiting for something to happen, we move our head, eyes, balance our body. The Animated Agent Engine takes this factor into account and allows the virtual agent to perform this kind of waiting movements when necessary.

These states and their behaviour rules influence the Animated Agent Engine and indicate how the virtual agent must react.

As mentioned before, in the Explanation state a real tutor should edit the virtual agent's behaviour so that the agent acts as similar as a real tutor would. In addition to the authoring tool provided for this purpose a Behavioural rule database will be used.

Most of these rules are related to hand and arm movements or facial expressions, but some are also related to gaze, blinking or posture.

The Evaluation Module will animate the virtual agent depending on the learner's interaction, for example, the agent will express happiness if the student answers correctly, it will shake his head when the answer is incorrect or it will appear sad if the student fails an exam. The Waiting state will use behavioural rules related to body posture, head movements, gaze, etc. For example, the virtual agent will cross his arms or will balance his body randomly in order to emulate human natural behaviour when standing for long periods of time.

These rules intend to make the virtual agent more lively and humanlike by emulating the behaviour of real people. Following the system logic, they will be executed randomly so that they do not result repetitive.

3.2 Evaluation Module

The animations reacting to the student's learning progress will be executed in the evaluation state. This state is meant to accompany the learner throughout the course so it will react in real-time to the learner's input.

The Evaluation Module will keep track of each student along the course in order to give personalized feedback to each individual. When the student types a question into the web browser, this input is evaluated taking into account several aspects and the logic inside the module determines the appropriate code. This code is used as the input for the Agent Module and depending on its value a certain emotion or movement is performed by the virtual agent.

As mentioned in the previous section, some behaviour rules have been defined for this state. In the case of facial expressions, when the student types a question and the returning code indicates the system has the answer, the virtual agent will express happiness, if the code indicates that the system does not understand the question, the agent will appear surprised or confused. The virtual agent will also execute different movements when evaluating the student; if the answer is correct it will rise up his thumb, if the answer is incorrect he will shake his head or hand (Figure 3).

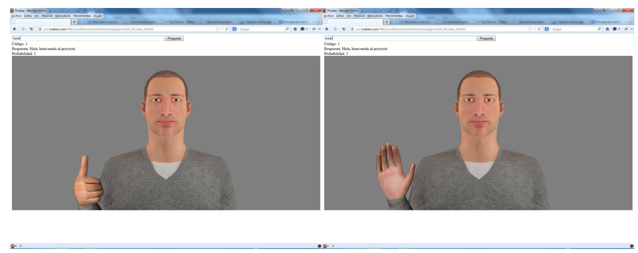


Figure 3: *Approval and Disapproval gestures.*

Furthermore, the Evaluation Module will save information regarding several aspects such as; the knowledge level of the student, the number of times a student enters a course, the frequency of interaction with the platform, the mistakes a student makes before finding the correct answer, etc. All this information will be stored in a database and will be analyzed by the Evaluation Module. Depending on the result of this evaluation the Animated Agent Engine will perform different actions. Figure 4 shows an example of this logic.

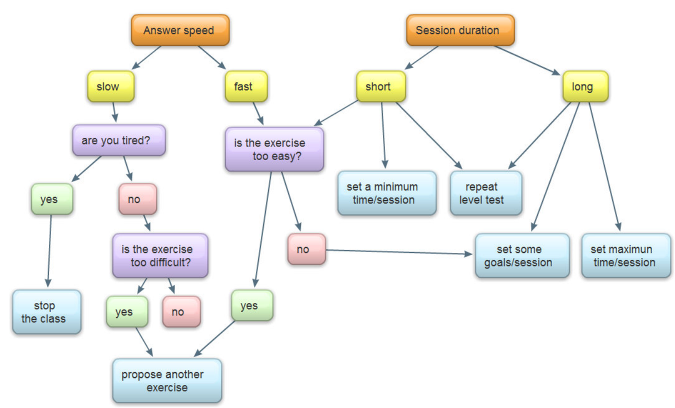


Figure 4: *Example of the Evaluation Module logic.*

The student's interaction with the e-learning platform can be analyzed in different ways, for example, controlling the speed in which the student answers questions or the duration of each session. If a student takes too long to answer a question, the Evaluation Module will send an impulse to the agent and it will react to this delay by asking some questions to the student; "is the exercise too difficult?", "are you tired?". The duration of each session can also be verified in order to evaluate the student's learning progress. Very short or long sessions might suggest the exercise does not match the student's knowledge level, in this case the virtual agent will suggest the student different exercises or will repeat the test to determine the students level.

This kind of communication makes the students feel accompanied and they appreciate the concern showed by the virtual agent. This attention on the one hand makes them feel that their work is being valued and on the other hand infers some pressure on them, so it is more difficult for them to abandon the course.

3.3 Animated Agent Engine

According to McCloud [McCloud 1994] individuals see themselves as iconic images but see others in a more detailed form, that is, as realistic images. Gulz and Haake [Gulz and Haake 2006] extended this idea to the role of animated pedagogical agents and stated that if the agent is acting as a teacher, the student will see it as "the other person" and therefore it is better to represent it in a human form. The agent of this work has been designed following these ideas; it will clearly perform the role of a teacher and not a classmate so it will appear in a realistic way.

As presented in the introduction, WebGL technology has been selected to render the 3D virtual agent into any compatible browser without the use of additional plug-ins. The Animated Agent Engine is composed by several modules developed using JavaScript programming language and following all the HTML5 and Web3D standards. These modules have been developed as an abstraction layer over O3D[1] engine which has been selected amongst other engines (GLGE, x3dom, etc.) for its benefits, as it is not a very high level API it allows great flexibility when developing new features.

The Agent Module is in charge of setting the emotions and the natural behaviour to the virtual character depending on the input received from the *Evaluation Module* and the *Behaviour rules*.

The animations executed in the platform can be either programmatically implemented or predefined by the designer in the Collada file. In this work the last option has been used, the animations are predefined in a timeline and they are coded sequentially following the next cycle; neutral expression - animation - neutral expression.

Each coded expression is reproduced when the animation engine decides which animation to launch. When several animations have to be reproduced the transition between one animation and another happens by reverting the timeline to the initial neutral state and then executing the transition from that neutral state to the following animation.

The realistic perception of the human body is mainly due to the complexity of its structure. The numerous details that compound the facial and body gestures result from simultaneous movements of muscles, bones, tendons and other fibers. This complex system cannot be reproduced as it occurs in nature when designing virtual agents in 3D virtual worlds, as muscles, fibers or tendons are not available tools. These human characteristics must be reproduced with the technology offered by software, namely, bones, morphing and bezier curves. The use of the last is not very common, normally morphing or bones methods are used to deform the geometry mesh.

Between these two possibilities morphing has been discarded, despite it being a much faster resource for modeling, it is limited to the number of variants made for each expression or phoneme. Another problem is that morphing is based on the transformation of a big mass of vertices, this implies having very little control over any modification, so the risk of corrupting the geometry increases with the number of vertices.

[1] http://code.google.com/p/o3d/

The use of bones as modifiers makes the tool compatible with any humanoid agent we want to include. The configuration of a face and a body will be valid for any model we use, and no changes will have to be made per agent as when using morphing, doing it the first time will be enough. With this method not only the designer decides how a face behaves, but leaves the door open to programmatically make adjustments or new representations.

The fact of having a clear hierarchy of the internal structure and the knowledge of the influences between bones, makes the tool very simple for anyone in the team working with the virtual agent.

In order to have greater control over the different expressions, a facial bone structure has been implemented. To achieve this, the movements of the different muscles of the face have been studied and a bone hierarchy has been applied to allow the movements of the facial biomechanical structure.

In the body a simplified structure of the human skeleton has been applied, discarding all the bones that hardly revert in modifications when making the movements required for the application. Some examples of animations that need to be implemented are; walking cycles, sitting down action, arm gestures or head movements.

To configure the face the Facial Definition Parameters of the MPEG-4 standard [Pandzic and Forchheimer 2003] has been use, but some of these points have been adapted to match the criteria required in this work. For example, simplification of the less visible areas, optimization of complex mouth zones, organization of the eyes and eyelids into a hierarchy to obtain mayor control and reduction of adjacent control points (Figure 5).

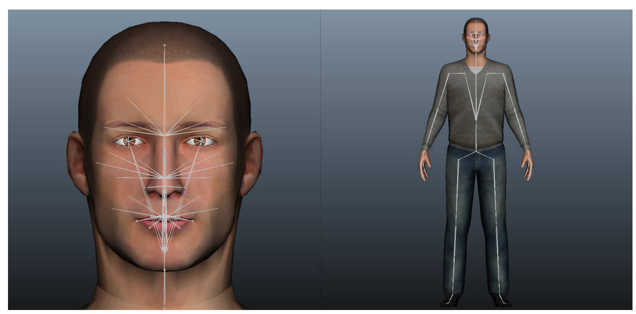


Figure 5: *Face and Body Bone Structure.*

The designer chooses the modifications for each group of bones, when the phonemes and the expressions are correctly defined by animation keyframes, the entire scene is exported in Collada format and it is ready to be loaded into the platform.

However, compared with morphing animations, bones technique implies a performance dropdown that can be particularly meaningful in Web3D environments. Computationally speaking, each vertex has to be weighted by the local transformation matrix of each affecting bone, while in the former method the final position is obtained with a simpler linear interpolation. This fact involves a serious bottleneck in the animation pipeline since the arithmetic engines of JavaScript interpreters are not able to perform the required computations in real-time.

In order to overcome this situation our animation engine implements a hardware skinning algorithm that delegates all bone transformations to the GPU. It has been implemented as a GLSL vertex shader that takes the original mesh and the transformation matrices of the bones as inputs, and calculates the new vertex positions as output. It is worth mentioning that this approach implies that the mesh information that the engine has is no longer related with the actual position of the vertices. This can be a drawback, for instance, in cases where the bounding box of the mesh is needed, like in collision detection or culling operations. Nevertheless, the performance obtained is far greater than the CPU implementation, being able to move in real time complex bone structures.

4 Validation

Although we are still working on the development of the modules that constitute the platform, some of our final users in different applications have seen the animated agent presented in this work and they have all agreed that it is more likeable and engaging than other agents, they state that it seems very natural and humanlike and this favors a more realistic way of interaction.

This validation process has been followed not only to verify the usefulness of the virtual character and its positive effect on learners but also to improve its natural behaviour and facial emotions. Next we mention some of the comments received from the users about our animated agent organized by positive aspects the users stress about the agent and some suggestions they have made in order to improve its behaviour.

Positive aspects:

- Including natural behaviour in the waiting state makes the virtual agent more lively.
- Common gestures such as “good”, “no”, “ok”, etc., give realism to the virtual agent.
- Certain gestures make the virtual agent funny, so the human-computer interaction turns out to be funny.
- The virtual agent’s emotions are easy to recognize.
- Varying the speed in which the virtual agent performs gestures depending on its mood is very interesting.

Suggestions:

- The virtual agent’s movements may end up being too repetitive.
- When the virtual agent is in an explanation stage, it would be convenient to reduce the amount of natural behaviour movements.
- The concatenation of some movements may seem unnatural.
- The virtual agent’s facial expressions should be magnified.

The positive comments suggest that our animated agent results more appealing than non-animated ones. Its natural behaviour improves the human-computer interaction. Integrating a humanlike agent in the platform approaches the users to a real classroom scenario so they react accordingly. The suggestions made by the users have also been very useful and we have taken them into account for future work in order to improve the agent’s behaviour.

5 Conclusions and Future work

A 3D animated agent based on WebGL has been presented in this paper. Thanks to the Animated Agent Engine developed in this work the system is compatible with any web browser supporting WebGL technology, thus it solves the interoperability issues that e-learning systems normally present.

The Evaluation Module implemented in this work provides the animated agent with sufficient artificial intelligence to react to the students interaction in real-time. This logic turns the animated agent into an autonomous agent that needs no exterior intervention to make its decisions.

The idea for future work is to validate the functionality of the global platform in order to confirm the positive effect that the animated character has in real students. We plan to verify that the agent’s natural behaviour and its real-time emotional response to the student’s inputs has a beneficial effect on the student’s learning engagement and final cognitive results.

Although this work has not been tested with real students, we have validated the animated agent with other applications and we have received feedback from users. They all agree the animated agent seems natural and appealing. The suggestions made by the users have been taken into account and we are working to solve them.

We agree that some of the agent’s movements may result too repetitive. In order to resolve this issue we are thinking on adding new movements to the natural behaviour module or including algorithms to modify the original movements. In addition, we plan to change the behaviour rules so that the Animated Agent Engine launches gestures with equal meanings randomly, namely, to disagree with the learner the virtual character will be able to arbitrarily decide whether to shake his head or move his hand.

We have also noticed that natural behaviour movements may sometimes distract the user from the main message, so we are going to modify the behaviour rules so that they focus on the meaning of the message and control sporadic movements.

Finally, to exaggerate the expressiveness of the virtual character the 3D model can be modified, new 3D graphic designs can be developed in order to meet the needs of each user. Nevertheless, this aspect does not require any changes from the Animated Agent Engine.

Two main animations are performed on the facial area: phonemes and expressions. These animations must be well correlated in order to achieve autonomous and natural behaviour from the virtual agent. For this purpose, a weighting of the influence of the vertices of the phonemes with their corresponding expression will be made, so that the agents mood is recognized without losing the mouth posture.

When the work is finished we plan to validate the system with real students in order to evaluate their learning progress and prove whether animated agents enhance the cognitive development.

Acknowledgements

The authors were partially supported by the Spanish Ministry of Economy and Competitiveness grant IPT-2011-0889-900000.

References

ABRILIAN, S., DEVILLERS, L., BUISINE, S., AND MARTIN, J.-C. 2005. Emotv1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *HCI International*.

ADKINS, 2013. The worldwide market for self-paced elearning products and services: 2011-2016 forecast and analysis, January.

AHMED, K. S. 2005. Animated pedagogical agents in web-based intelligent learning environment.

André, E., Rist, T., and Müller, J. 1997. WebPersona: A Life-Like Presentation Agent for Educational Applications on the World-Wide Web. In *Workshop on "Intelligent Educational Systems on the World Wide Web", 8th World Conference of the AIED Society*.

Baylor, A., and Ebbers, S. 2003. The pedagogical agent split-persona effect: When two agents are better than one. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, vol. 2003, 459–462.

Baylor, A. L., and Ryu, J. 2003. The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research 28*, 4, 373–394.

Benin, A., Leone, G. R., and Cosi, P. 2012. A 3d talking head for mobile devices based on unofficial ios webgl support. In *Proceedings of the 17th International Conference on 3D Web Technology*, ACM, 117–120.

Bloom, B. S. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher 13*, 6, 4–16.

Buisine, S., and Martin, J.-C. 2007. The effects of speech–gesture cooperation in animated agents behavior in multimedia presentations. *Interacting with Computers 19*, 4, 484–493.

Carretero, M., Ardanza, A., García, S., Díez, H., Oyarzun, D., and Ruiz, N. 2012. Improving gestural communication in virtual characters. In *Proceedings of the 7th international conference on Articulated Motion and Deformable Objects*, Springer-Verlag, Berlin, Heidelberg, AMDO'12, 69–81.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, ACM, 413–420.

Dehn, D. M., and Van Mulken, S. 2000. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies 52*, 1, 1–22.

D'mello, S., and Graesser, A. 2012. Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS) 2*, 4, 23.

Ekman, P., and Friesen, W. V. 1981. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Nonverbal communication, interaction, and gesture*, 57–106.

Gulz, A., and Haake, M. 2006. Design of animated pedagogical agentsa look at their look. *International Journal of Human-Computer Studies 64*, 4, 322–339.

Hu, X. 2011. Autotutor lite (interactive event).

Johnson, W. L., and Rickel, J. 1997. Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin 8*, 1-4, 16–21.

Johnson, W. L., Rickel, J. W., and Lester, J. C. 2000. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education 11*, 1, 47–78.

Johnson, W. L. 1998. Pedagogical agents. In *Proceedings of ICCE98*, 13–22.

Kohar, H., and Ginn, I. 1997. Mediators: guides through online tv services. In *CHI'97 extended abstracts on Human factors in computing systems: looking to the future*, ACM, 38–39.

Lepper, M. R., Woolverton, M., Mumme, D. L., and Gurtner, J. L. 1993. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In *Computers as Cognitive Tools*, S. P. Lajoie and S. J. Derry, Eds. Lawrence Erlbaum Associates, 75–105.

Lester, J. C., and Stone, B. A. 1997. Increasing believability in animated pedagogical agents. In *Proceedings of the first international conference on Autonomous agents*, ACM, New York, NY, USA, AGENTS '97, 16–21.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. 1997. The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 359–366.

Leung, C., and Salga, A. 2010. Enabling webgl. In *Proceedings of the 19th international conference on World wide web*, ACM, 1369–1370.

McCloud, S. 1994. *Understanding comics*. William Morrow Paperbacks.

Mlakar, I., and Rojc, M. 2011. Towards ecas animation of expressive complex behaviour. *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, 185–198.

Moundridou, M., and Virvou, M. 2002. Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning 18*, 3, 253–261.

Nass, C., Isbister, K., and Lee, E.-J. 2000. Truth is beauty: Researching embodied conversational agents. *Embodied conversational agents*, 374–402.

Pandzic, I. S., and Forchheimer, R., Eds. 2003. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons, Inc., New York, NY, USA.

Piesk, and Trogermann. 1998. Presenting educational contents in nonlinear narrative structures by conventional virtual actors. In *Proceedings of ED-MEDIA98*, 1885–1888.

Rieger, T., et al. 2003. Avatar gestures. *Journal of WSCG 2003 11*, 379–386.

Schöch V, S. M., and G, W. 1998. adi an empirical evaluation of a tutoring agent. In *Proceedings of ED-MEDIA98*, 1271–1282.

Shaw, E., Johnson, W. L., and Ganeshan, R. 1999. Pedagogical agents on the web. In *Proceedings of the third annual conference on Autonomous Agents*, ACM, 283–290.

Van Mulken, S., André, E., and Müller, J. 1998. The persona effect: How substantial is it? *People and Computers*, 53–66.

Realistic Visual Speech Synthesis in WebGL

- Authors: Andoni Mujika, Helen V. Diez, Aitor Álvarez, Miren Ugarte and David Oyarzun
- Booktitle: Proceedings of the 18th International Conference on 3D Web Technology
- Year: 2013
- Publisher: ACM

Realistic Visual Speech Synthesis in WebGL

Andoni Mujika⋆, Helen Diez⋆, Aitor Alvarez⋆, Miren Urteaga†, David Oyarzun⋆ *
⋆ Vicomtech-ik4
† University of Basque Country

Abstract

This paper presents the work that has been done to develop a web application that shows the face of a virtual character pronouncing the sentences the user sets. The level of realism was high and the performance was fast enough. The application makes use of WebGL, speech processing, text to speech and co-articulation technologies to obtain the virtual pronunciation.

1 Introduction

Since the pioneer works [Parke 1972] in facial animation, hundreds of methods have been presented to make a virtual character pronounce a sentence, but very few [Benin et al. 2012] are based on the emerging technology WebGL. This paper presents a project that works in this direction, SPEEP, partly funded by the Basque Government. The project creates a system for foreign language pronunciation learning where the key part is the visualization of a virtual character pronouncing the corresponding sentences in a web. Figure 1 shows the interface designed for the project integrated in the language learning system. Since the sentences that have to be synthesized in the virtual character are not predefined i.e. the user writes the desired sentence, the system cannot work with animations that have been generated in a previous phase of the project. Thus, we can define our project as a work in real-time Visual Speech Synthesis.

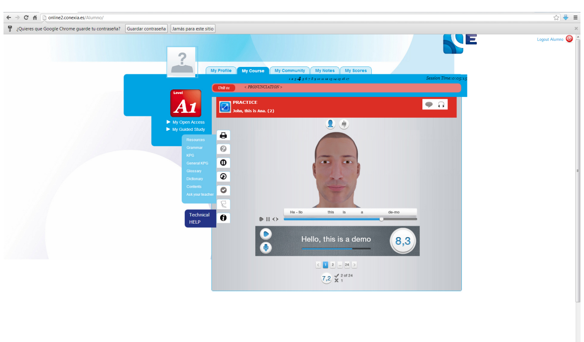


Figure 1: *Interface of the system.*

2 Implementation

In a web browser, the user of the application will find a text field to write the desired sentence and once it is written and confirmed, it is sent to the server. In the server, the text is converted to speech with the desired voice by the module for text-to-speech conversion and the module for voice transformation. On the other hand, the text is used to get the phonemes and the syllables of the sentence and align them with the audio file generated by the text-to-speech module. Once the three files needed (audio, phonemes and syllables) are generated, they are sent to the client.

The client receives the information about the phonemes that must be rendered and their exact times. Then, during the rendering of the virtual character, in each frame the co-articulation engine takes the phonemes that surround the actual time and interpolates the values of the Facial Animation Parameters following the method presented by Cohen and Massaro [Cohen et al. 1993]. Besides, several rules have been defined that change the value of some parameters and the position of the maximum point of the phoneme in its interval, depending on the type of syllable and the phonemes involved. For example, the consonant of a syllable CV (consonant-vowel) is placed in a different position of its interval comparing to a consonant of a syllable VC.

A study of the computational cost of the different methods that are called in each frame during the facial animation showed that the most time consuming function was the skinning i.e. the computation of the new position of the vertices according to the transformations of their neighbor bones. So, we decided to make use of GPU's power and we implement the skinning using shaders, obtaining a considerable increase in the number of frames per second.

For the speech generation, the server takes the text written by the user and converts it into an audio file. Then, speech and text are automatically aligned using an acoustic model obtained by training. Finally, in order to identify each syllable for a natural mouth articulation of the avatar, a syllabification module was built.

3 Conclusion

The results obtained so far in the project SPEEP are satisfactory. A big number of vertices are needed to obtain a realistic virtual face and although all the vertices are not transformed in all frames, the amount of vertices that are moved is big enough to become a problem. Specifically, the virtual face that is used in the project SPEEP is compound of 22802 vertices that form 43449 polygons. Nevertheless, the problems that slower performance of web applications can cause have been overcome. Several strategies have been implemented to make the application work faster; fast enough even in computers with commodity graphic card. Moreover, a high level of realism has been achieved in Visual Speech Synthesis.

References

Benin, A., Leone, G. R., and Cosi, P. 2012. A 3d talking head for mobile devices based on unofficial ios webgl support. In *Proceedings of the 17th International Conference on 3D Web Technology*, ACM, New York, NY, USA, Web3D '12, 117–120.

Cohen, M. M., Massaro, D. W., et al. 1993. Modeling coarticulation in synthetic visual speech. *Models and techniques in computer animation 92*.

Parke, F. I. 1972. Computer generated animation of faces. In *Proceedings of the ACM annual conference-Volume 1*, ACM, 451–457.

*e-mail:{amujika, hdiez, aalvarez, doyarzun}@vicomtech.org, miren.urteaga@ehu.es

Web3D 2013, June 20 – 22, 2013, San Sebastian, Spain.

ACM 978-1-4503-2133-4/13/06

IAAN: Intelligent Animated Agent with Natural Behaviour for Online Tutoring Platforms

- Authors: Helen V. Diez, Sara Garía, Jairo R. Sánchez and María del Puy Carretero and David Oyarzun
- Booktitle: Proceedings of the 6th International Conference on Agents and Artificial Intelligence-Volume 2
- Year: 2014
- Publisher: SCITEPRESS-Science and Technology Publications, Lda

IAAN: Intelligent Animated Agent with Natural Behaviour for online tutoring platforms

Helen V. Diez, Sara García , Jairo R. Sánchez, Maria del Puy Carretero and David Oyarzun
Vicomtech-IK4 Research Center, San Sebastián, Spain
{*hdiez, sgarcia, jrsanchez, mcarretero, doyarzun*}*@vicomtech.org*

Keywords: Agents, Artificial intelligence, Natural behaviour, Behaviour markup language.

Abstract: The goal of the work presented in this paper is to develop an Intelligent Animated Agent with Natural Behaviour (IAAN). This agent is integrated into e-learning platforms in order to perform the role of an online tutor. The system stores into a database personalized information of each student regarding their level of education, their learning progress and their interaction with the platform. This information is then used by the 3D modeled virtual agent to give personalized feedback to each student; the purpose of the agent is to guide the students throughout the lectures taking into account their personal needs and interacting with them by means of verbal and non-verbal communication. To achieve this work a thorough study of natural behaviour has been held and a complex state machine is being developed in order to provide IAAN with the sufficient artificial intelligence as to enhance the students motivation and engagement with the learning process.

1 INTRODUCTION

Online tutoring has become very popular in the past years and according to several studies such as the latest survey by Ambient Insight Research (Adkins, 2013); the aggregate growth rate for self-paced e-learning products and services expected for the next five year period (2011-2016) is 7.6%. Distance learning offers a series of benefits that traditional learning cannot compete with, for example in terms of mobility, affordability or flexibility e-learning happens to be much more suitable for nowadays lifestyle. However, online learning also has its drawbacks, the lack of supervision from a tutor in the courses may lead to demotivation, boredom and the final drop of the courses. This is why the integration of virtual agents represented as human characters into these platforms can be an effective solution to make the students feel supported and accompanied throughout the course as a real teacher would.

Michael Graham Moore (Moore, 1989) classified the possible interactions in distance education into three types:

- learner-content
- learner-instructor
- learner-learner

He acknowledged the "learner-instructor" interaction as the most important. Moreover, studies like (Bloom, 1984) demonstrated the effectiveness of a one-on-one human tutoring system against other methods of teaching, and (Lepper et al., 1993), defended the idea that education could be globally improved if every student was provided with a personal tutor. This is something almost impossible to achieve in traditional education but it is not so in online learning.

This work introduces IAAN, a virtual agent represented as a 3D modeled character endowed with intelligence and natural behaviour, designed to perform the role of a real tutor in e-learning platforms. IAAN reacts in real time to the students' interaction with the platform by means of verbal and non-verbal communication. Furthermore, in order to make IAAN as realistic as possible an Intelligent Animated Agent Editor is included into the system allowing real tutors to define the appearance and behaviour of IAAN in response to different situations. This editor is based on the Behaviour Markup Language standard (BML [1]).

Additionally, this work is entirely web based so it solves the interoperability issues presented by most commonly used e-learning systems with other platforms such as Learning Management Systems, web-based virtual world platforms, Virtual Reality learning systems or simulators.

IAAN has been partially integrated into Moodle

[1] http://www.mindmakers.org/projects/bml-1-0/wiki

(Moodle, 2013) Learning Management System.

This paper is organized as follows; Section 2 analyzes the related work carried out in the last years concerning agents and artificial intelligence, Section 3 describes the architecture followed to accomplish the goals of this work, Section 4 shows integration results from the work. The final section is about conclusions and future work.

2 RELATED WORK

The purpose of this work is to integrate an intelligent 3D virtual agent into e-learning platforms. Concerning this process, Buraga (Buraga, 2003) proposes an agent-oriented extensible framework based on XML family for building a hypermedia e-learning system available on the world-wide-web. This intelligent tutoring system is composed of four major components, the information processed by each component can be stored by XML documents. Some of the components are implemented as intelligent agents.

Angehrn et al. (Angehrn et al., 2001) suggests the use of K-InCA to provide a personalized e-learning system to help people learn and adopt new behaviours. The agent continuously analyses the actions of the user in order to build and maintain a "behavioural profile" reflecting the level of adoption of the "desired" behaviours. Using this profile, the agent provides customized guidance, mentoring, motivation and stimuli, supporting the gradual transformation of the users behaviours.

Defining natural behaviour is not an easy task either. The literature and theory of affective computing imply several conditions for synthesized motion to appear natural (Abrilian et al., 2005). Speed of interaction and emotion/speech-correlated believable body motion are among the most important functionalities (Mlakar and Rojc, 2011). Rieger (Rieger et al., 2003) developed a series of rules in order to increase the acceptance of virtual agents in Human-Computer Communication and established a correlation table between the message to rely and the emotion to show.

Steve (Johnson and Rickel, 1997) was one of the first pedagogical agents capable of expressing emotions; it was designed as a stereoscopic 3D character that cohabited with learners, it has been applied to naval training tasks. However, Steve was originally designed to operate in immersive virtual environments and not over the Web.

Project GRETA (Poggi et al., 2005) presents a multimodal Embodied Conversational Agent (ECA) capable of interpreting APML [2] mark-up language to generate synchronized speech, face, gaze and gesture animations.

More recently, (Benin et al., 2012) presented a three-dimensional animated talking head which repeats any input text in six different emotional ways.

3 INTELLIGENT ANIMATED AGENT ROLES

The final goal of the work presented is to integrate a 3D animated agent into e-learning platforms in order to asses and guide students as a real teacher would.

Many studies have been held to identify the qualities of a good teacher (Azer, 2005) (Korthagen, 2004). Besides, students and teachers do not always agree in the importance of these qualities. From the perspective of students, Brown and McIntyre (Brown and McIntyre, 1993) and Batten (Batten et al., 1993) found the two qualities with highest frequency of mention were the teachers ability to "explain clearly", and "help us with our work". On the other hand, two qualities seen by teachers as crucial, but not mentioned by students, were "planning, structuring and organising the classroom, and fostering student involvement and participation".

Considering previous research, IAAN has been designed to perform several roles throughout the course depending on the needs of the lesson, these roles have been divided into three states:

- **Explanation state**: IAAN will be able to explain new concepts to the user. Humans tend to gesticulate when introducing an idea to others; performing arm movements or pointing out objects, IAAN will act alike.
- **Evaluation state**: One of the most important roles that IAAN must perform is the one of a real tutor. IAAN will guide the students through the lectures and will interact with them giving them personal feedback, responding in real-time to their interaction with the platform.
- **Waiting state**: When IAAN is not in any of the previous states, for example, when self-explanatory audiovisual content is being displayed to the student, IAAN will enter a waiting state mode. IAAN will not interact directly with the student but he will be animated. Humans do not stand hieratical when waiting for something to happen; we balance our body, gaze, cross our arms, etc. As IAAN is endowed with natural be-

[2] http://apml.areyoupayingattention.com

haviour, he will also perform these kind of movements in this state.

The following section introduces the Intelligent Animated Agent Editor, this tool is being developed to ease the job of defining the virtual agent's appearance and behaviour. With the help of this tool the course editor will be able to define the Explanation and Waiting state. This tool is the core of the Behaviour Module. The Evaluation state will be discussed in more detail in Section 5.

3.1 Intelligent Animated Agent Editor

This tool is being implemented to allow real tutors to design IAAN's appearance and behaviour as they deem most appropriate in each case. To ensure quality education it is important to leverage the knowledge and expertise provided by real teachers.

Regarding IAAN's appearance, McCloud (McCloud, 1994) stated that individuals see themselves as iconic images but see others in a more detailed form, that is, as realistic images. Gulz and Haake (Gulz and Haake, 2006) extended this idea to the role of animated pedagogical agents and stated that if the agent is acting as a teacher, the student will see it as "the other person" and therefore it is better to represent it in a human form. However, the risk of falling into the "uncanny valley" (Mori, 1970) also exists and furthermore depending on the target of the course the editor may consider to represent IAAN in a more cartoonish shape, for example, if the course is targeted for kids. So the decision of IAAN's appearance is entirely left in the hands and teaching experience of the real tutor. The agent's appearance is selected from a list of predefined 3D models and imported into the IAA Editor.

With respect to facial expression, the IAA editor is based on Ekman's six universal emotions (Ekman and Friesen, 1981), so IAAN will be able to show the following facial expressions: happiness, surprise, anger, sadness, disgust and fear. An example of IAAN performing these facial expressions can be found in Figure 1.

IAAN performs several hand gestures, the scientific community has established four type of hand gestures (Cassell et al., 1994):

- Iconic: or illustrators, they are descriptive gestures often used to illustrate speech.
- Metaphoric: or representational gestures, represent an abstract feature concurrently spoken about.
- Deictic: indicate a point in the space.

Figure 1: IAAN represented as a 3D virtual agent expressing facial emotions (happiness, surprise, anger, sadness).

- Beats: small formless waves of the hand that occur to emphasize words.

IAAN communicates in a verbal way with the students, for this purpose the IAA Editor includes a text editor. This text is then transformed into speech by a Text-to-Speech synthesizer.

As shown in Figure 2 the IAA Editor is composed of three main areas; the animations regarding natural behaviour can be found in the *Options Area*, the real tutor selects the desired animation from one of the available menus and drags it onto the *Timeline Area*, this action is repeated as many times as necessary until the desired behaviour is achieved. Then the final result is visualized in the *Viewer Area*.

The composition created in the timeline is translated into a BML file. In the following BML example created with the IAA Editor, a welcoming message has been designed. This BML file is then used as an input to the Animation Engine.

```
<bml xmlns=
   "http://www.bml-initiative.org/bml/bml-1.0"
    character="Iaan"
    id="bml1">
 <gesture id="behavior1" lexeme="hello-waving"
     start="2" end="3"/>
 <faceLexeme  id="behavior2" lexeme="happy"
      amount="0.8" start="2" end="3"/>
 <speech id="speech1" start="4">
    <text>Wellcome to the first lesson!</text>
 </speech>
</bml>
```

4 IMPLEMENTATION

The architecture chosen to accomplish the goals presented in the previous section is shown in Figure 3.

Figure 2: Intelligent Animated Agent Editor.

The main modules involved in the design of IAAN are the Behaviour Module, the Evaluation Module and the Animated Engine.

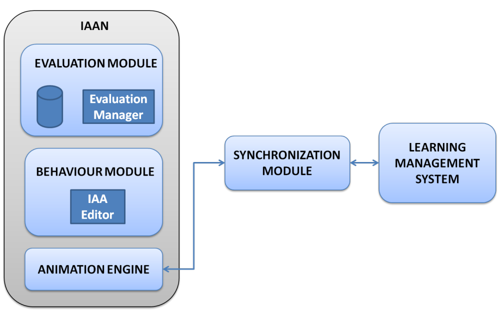


Figure 3: Platform Modules.

The Behaviour Module has been described in Section 3. This module is in charge of defining the behaviour of IAAN in each situation, to make IAAN as realistic as possible a real tutor is in charge of describing IAAN's interaction with the students. To ease this job the IAA Editor has been developed.

Next the Evaluation Module and the Animated Engine will be explained in more detail.

4.1 Evaluation Module

IAAN must accompany the students throughout the course and interact with them whenever is necessary. To accomplish this IAAN will keep track of each student along the course in order to give personalized feedback to each individual.

This is achieved by storing multiple information into a database and by defining a complex state machine that will inform IAAN when and how to interact with each student.

4.1.1 User Profile Database

It is crucial to know as much as possible about each student in order to assess them according to their personal needs. Table 1 shows the information stored into the database for further analyses, regarding personal information as well as the student's interaction with the e-learning platform.

Table 1: User Profile Information.

Personal Information	Platform Usage
Name	Logins
Age	Session Duration
Address	Interaction Speed
Knowledge Level	Mistake Frequency

Taking this information into account the course editor is able to design specific evaluation rules for each student.

The first thing the course editor must decide is the `EVALUATION_HARSHNESS`, initially this parameter is set depending on the student's age and knowledge level, but as the course progresses and the student acquires more knowledge this scale can be modified.

The Platform Usage information is designed to establish a BEHAVIOUR_PATTERN for each student, for example; the amount of logins per week, the session duration, the frequency in answering questions, the mistaken answers. Bearing this pattern in mind, IAAN will interact with the student whenever a disorder in the pattern takes place.

4.1.2 Evaluation Manager

The Evaluation Manager is designed as a finite state machine (FSM). As seen in Figure 4 the Evaluation Manager takes as inputs the EVALUATION_HARSHNESS, the BEHAVIOUR_PATTERN and the BML_FILES that define IAAN's natural behaviour which have been previously designed with the IAA editor .

Taking these inputs into account the Evaluation Manager waits for behaviour pattern alerts and if they occur the manager orders IAAN to interact with the student by means of the previously designed BML files. Regarding the User Profile Information the possible alerts have been defined as follows:

- LOGIN_ALERT: The course editor is in charge of establishing an amount of logins per week (or month) for each student in order to fulfil his assignments. If this recommendation is altered in any way the platform is notified with an alert message and IAAN is launched to interact with the student.
- INTERACTION_ALERT: Once the student has logged in he interacts with the platform in a determined frequency. Modifying this frequency may mean several things, for example, if the interaction speed increases it may indicate the student finds the lesson too easy, on the contrary if the speed decreases the lesson might be too difficult or it may simply mean the student is taking a break. IAAN is able to interact with the student to find out what is happening.
- MISTAKE_ALERT: If the student commits more mistakes than usual the course level might not be appropriate or the student might not be paying attention. IAAN might enter the Explanation state in order to clarify concepts or draw the student's attention by introducing a multimedia effect.
- SESSION_ALERT: A recommended minimum and maximum time is set to perform each session, if this time is altered IAAN shows up to check if the student has finished his assignment.

The messages in round boxes from Figure 4 represent examples of IAAN communicating with the student in each alert situation. For example, if the Evaluation Manager detects an INTERACTION_ALERT IAAN launches the BML file describing the behaviour to adopt under this circumstance, for instance, IAAN will ask the student: "Are you taking a break?". Another example would be in the event of a MISTAKE_ALERT, in this case IAAN will express concern by asking: "Is the lesson too difficult?".

With these queries IAAN seeks to encourage the student to continue working, letting him know he is not alone and that he is being supervised.

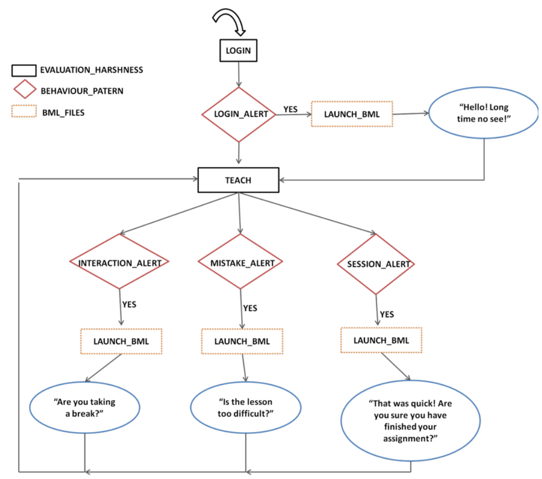


Figure 4: Evaluation FSM Example.

4.2 Animation Engine

The Animation Engine is composed by several modules developed using JavaScript programming language and following all the HTML5 and Web3D standards. These modules have been developed as an abstraction layer over O3D[3] engine which has been selected amongst other engines (GLGE, x3dom, etc.) for its benefits, as it is not a very high level API it allows great flexibility when developing new features.

WebGL (Leung and Salga, 2010) technology has been used to render IAAN via the web. WebGL is based on OpenGL, which is a widely used open source 3D graphics standard. Nowadays, most common browsers support this technology; Google Chrome, Mozilla Firefox, Apple Safari or Opera.

The modules that compose the Animation Engine are in charge, amongst other features, of rendering 3D modeled characters into web browsers, parsing BML files which define the character's behaviour and ani-

[3]http://code.google.com/p/o3d/

mations and communicating with the e-learning platform.

5 INTEGRATION

The work presented in this paper is entirely web based, this fact makes the platform compatible with most widely-used learning management systems (LMS)[4]; Edmodo (Edmodo, 2013), Moodle (Moodle, 2013), Blackboard (Blackboard, 2013), SumTotal Systems (SumTotal, 2013), etc.

In the work presented IAAN has been partially integrated into Moodle. Moodle is based on a Model-View-Controller coding design pattern. To integrate IAAN into Moodle the modules from this work must be added to the configuration file of the platform.

This platform offers a very interesting feature for our work; the *Configurable Reports*. This block is a Moodle custom reports builder designed in a modular way to allow developers to create new plugins. The types of reports available are:

- Courses reports, with information regarding courses.
- Categories reports, with information regarding categories.
- User reports, with information regarding users and their activity in a course.
- Timeline reports, this is a special type of report that displays a timeline showing data depending on the start and end time of the current row.
- Custom SQL Reports, custom SQL queries.

Taking advantage of this feature a Synchronization Module is being developed as a communication bridge between Moodle and the Animation Engine (Figure 3).

For this work a very simple English Course has been created in Moodle with multiple choice quizzes for the student to answer. Figure 5 shows the integration of IAAN into the created course.

Validation results regarding the entire platform integration have not yet been performed as it is still work in progress. However, some of our final users in different applications have been able to interact with IAAN and they have pointed out its natural behaviour and communication as an engaging and realistic way of interaction.

[4]http://edudemic.com/wp-content/uploads/2012/10/top-20-lms-software-solutions.png

6 CONCLUSIONS AND FUTURE WORK

An intelligent virtual agent represented as a 3D modeled character has been presented in this paper. Thanks to the IAA editor the agent is gifted with natural behaviour, allowing real tutors to use their experience to define the agent's reactions in different circumstances.

The Evaluation Module described in this work is still being developed. The main goal of this module is to turn the animated agent into an autonomous agent that needs no exterior intervention to respond to students' interaction in real-time.

The modules of this work have been developed following a Model-View-Controller coding pattern to ease the integration with the selected e-learning platform. IAAN has been successfully integrated into Moodle, though only partially functional as the Evaluation Module is still under development.

The Animation Engine introduced in this work is constantly improved to suit new needs. The final goal is to develop an animation engine capable of reproducing human behaviour as realistic as possible.

A synchronization module is being developed in order to optimize the communication between IAAN and Moodle. Furthermore, we are studying other LMS in order to develop a general synchronization module turning IAAN into a multiplatform assistant.

Finally, we are defining the validation phase to test the work presented in this paper with real students in order to confirm IAAN's positive effect. The results of these evaluations will verify whether IAAN's natural behaviour and real-time emotional response has a beneficial effect on the students learning engagement and final cognitive results.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness grant IPT-2011-0889-900000.

Figure 5: Integration of IAAN into Moodle.

REFERENCES

Abrilian, S., Devillers, L., Buisine, S., and Martin, J.-C. (2005). Emotv1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *HCI International*.

Adkins (2013). The worldwide market for self-paced elearning products and services: 2011-2016 forecast and analysis.

Angehrn, A. A., Nabeth, T., Razmerita, L., and Roda, C. (2001). K-inca: Using artificial agents to help people learn and adopt new behaviours. In *ICALT*, pages 225–226.

Azer, S. A. (2005). The qualities of a good teacher: how can they be acquired and sustained? *Journal of the Royal Society of Medicine*, 98(2):67–69.

Batten, M., Marland, P., Khamis, M., and for Educational Research., A. C. (1993). *Knowing how to teach well : teachers reflect on their classroom practice / Margaret Batten, Perc Marland, Mon Khamis*. ACER Hawthorn, Vic.

Benin, A., Leone, G. R., and Cosi, P. (2012). A 3d talking head for mobile devices based on unofficial ios webgl support. In *Proceedings of the 17th International Conference on 3D Web Technology*, pages 117–120. ACM.

Blackboard (2013). Wherever your learning takes place. `http://uki.blackboard.com`. Accessed November, 2013.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16.

Brown, S. and McIntyre, D. (1993). *Making Sense of Teaching*. Open University Press.

Buraga, S.-C. (2003). Developing agent-oriented e-learning systems. In *in Proceedings of The 14th International Conference on Control Systems And Computer Science vol. II, I. Dumitrache and C. Buiu, Eds, Politehnica*. Press.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 413–420. ACM.

Edmodo (2013). Where learning happens. `https://www.edmodo.com`. Accessed November, 2013.

Ekman, P. and Friesen, W. V. (1981). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Nonverbal communication, interaction, and gesture*, pages 57–106.

Gulz, A. and Haake, M. (2006). Design of animated pedagogical agentsa look at their look. *International Journal of Human-Computer Studies*, 64(4):322–339.

Johnson, W. L. and Rickel, J. (1997). Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin*, 8(1-4):16–21.

Korthagen, F. A. (2004). In search of the essence of a good teacher: towards a more holistic approach in teacher education. *Teaching and Teacher Education*, 20(1):77 – 97.

Lepper, M. R., Woolverton, M., Mumme, D. L., and Gurtner, J. L. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In Lajoie, S. P. and Derry, S. J., editors, *Computers as Cognitive Tools*, pages 75–105. Lawrence Erlbaum Associates.

Leung, C. and Salga, A. (2010). Enabling webgl. In *Proceedings of the 19th international conference on World wide web*, pages 1369–1370. ACM.

McCloud, S. (1994). *Understanding comics*. William Morrow Paperbacks.

Mlakar, I. and Rojc, M. (2011). Towards ecas animation of expressive complex behaviour. *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 185–198.

Moodle (2013). Open-source community-based tools for learning. `https://moodle.org`. Accessed November, 2013.

Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education*, 3:1–6.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., and de Carolis, B. (2005). GRETA. A Believable Embodied Conversational Agent. pages 27–45.

Rieger, T. et al. (2003). Avatar gestures. *Journal of WSCG 2003*, 11:379–386.

SumTotal, S. (2013). Learning management and talent management solutions. `http://www.sumtotalsystems.com`. Accessed November, 2013.

Providing Physical Appearance and Behaviour to Virtual Characters

- Authors: María del Puy Carretero, Helen V. Diez, Sara Garía and David Oyarzun
- Booktitle: International Conference on Articulated Motion and Deformable Objects
- Year: 2016
- Publisher: Springer

Providing physical appearance and behaviour to virtual characters

María del Puy Carretero, Helen V. Diez, Sara García and David Oyarzun

Vicomteh-IK4, Paseo Mikeletegi 57, 20009 Donostia-San Sebastián, Spain
{mcarretero,hdiez,sgarcia,doyarzun}@vicomtech.org
http://www.vicomtech.org

Abstract. The goal of the work presented in this paper is to develop a web based authoring tool to create the physical appearance of a virtual character and to provide its behaviour. One of the main characteristics of this tool is its simplicity to create a virtual character, provide its behaviour and the way to integrate it anywhere. The appearance of each virtual character and its behaviour are stored in a database. In order to specify the behaviour, the standard BML (Behaviour Markup Language) is used. It is interpreted by a BML Parser module that indicates to the animation engine how the virtual character has to act. The authoring tool is tested creating different virtual characters and integrating them in two unrelated applications with different roles.

Keywords: 3D Virtual Character, Authoring Tool, Animation

1 Introduction

Nowadays virtual characters are frequently used in several kind of applications. The role of the virtual characters depends on the application; it can act for example as a guide, assistant or as an information presenter. Although they are commonly used, their design, development, animation and integration are not easy tasks. These functionalities are commonly done by 3D designers and animation experts. The goal of the work presented in this paper is to develop an authoring tool to create the physical appearance of a virtual character and to provide its behaviour in order to make it lively. In addition, other objective is to provide different options to integrate and use the created virtual characters in different applications.

One of the most typical method of giving the virtual character's behaviour is by using a mark-up language, which contains different labels to indicate the text to be spoken, emotions, facial and body gestures, and the exact moments when these have to be reproduced. The authoring tool presented in this paper is based on Behavior Markup Language (BML) standard because of the reasons given in Related Work section [1].

The paper is organized as follows; Section 2 analyses the related work regarding different methods and possibilities of creating virtual characters and giving them behaviour. Section 3 explains the method followed to accomplish the goals of this work, describing in detail each authoring tool. Section 4 explains the global architecture to create and integrate virtual characters. Section 5 shows different use cases and applications of the presented work. And the final section is about conclusions and future work.

2 Related Work

The idea of virtual characters editor is not new. Virtual worlds like Seconf Life[1], or games such as The Sims[2] or World of Warcraft[3] allow the user to configure the appearance of the virtual character who will represent them. However the use of these editors is exclusively of the application (the virtual world or the game) and the virtual character only can be used on them. In addition, the actions of the virtual character are defined by the actions of the user. That is, the behaviour is defined by what wants to do the end user at each moment.

In other kind of applications, behaviour definition usually is done by means of authoring tools and by using scripting and markup languages [2]. During the last years, several markup languages have appeared in order to specify the behaviour of virtual characters such as VHML [3], AML [4] or CML [5]. A comparison between them and other markup languages oriented to virtual characters can be found in [6].

BML [1] is a markup language, developed several years ago, oriented to describe verbal and non-verbal behaviour of virtual characters. BML defines elements like gestures and facial expressions and also allows specifying their temporal alignment. BML has become a standard and other markup languages allow compatibility with it. For example, the aim of PML [7] is to specify knowledge about the environment an non-verbal behaviour to virtual characters and it includes compatibility with BML. Moreover as the authors explain in [8] BML has some advantages such as it is intuitive and simple to implement.

In addition there are several projects based on BML whose main objective is to generate non-verbal behaviour for virtual characters. There are several interpreters like SmartBody [9], BMLRealizer [10] or Elckerlyc [11] that animate virtual character receiving a BML archive as input. However, to generate that input, there is only one behaviour planning developed [12]. This non-verbal generator analyzes the syntactic and semantic of a text and the affective state of a character to decide the behaviour, so it is not oriented to be decided by end users.

An interesting tool for the creation of interactive applications with multiple virtual characters is Visual SceneMaker [13]. It is planed to add BML and other markup languages to integrate non-verbal behaviour generation, so it does not include our main objective.

Taking all this into consideration, a tool based on BML with the following features has not been found:

[1] http://secondlife.com/
[2] https://www.thesims.com/
[3] http://eu.battle.net/wow/

– it has to be Web based in order to be accessible from any device and at any time,
– it has to be easy to use,
– it has to be modular in order to add new functionalities,
– it has to allow appearance and behaviour edition,
– the appearance and behaviour have to be easily integrable in other applications.

3 Authoring Tool: Intelligent Animated Agent Editor

The authoring tool is divided in two different functionalities. The first one is to create the physical appearance of the virtual character. The second one is to provide behaviour to an already defined virtual character. The proposed tool is a web based application in order to make it accessible from any device.

Previous to using the Editor the user must have the 3D models to work on. These models have to be created and maintained by 3D graphic designer experts in order to design and create the different options to create the virtual characters.

To change the appearance of the 3D models the texture of their attributes must be changed. To change the behaviour different animations designed for that model have to be loaded. So for each model it is needed to have stored various files; COLLADA files for the 3D models and animations and .jpeg or .png files for the textures.

To handle the application an SQL database is used. This database stores the user profiles, projects, 3D models, textures, animations and Timeline sequence. To store the textures and animations, the route in which each file is found is stored.

3.1 Appearance Editor

The appearance editor has been designed to configure the physical appearance of virtual characters in a simple way. End users do not have to have prior knowledge in design of 3D computer graphics. The appearance of the virtual character is configured by selecting different options in hair, face, eyes, body, clothing and other complements.

Figure 1 shows the interface of the Appearance Editor. As it can be seen, there are three main work areas:

– **3D Visor:** this area shows in real time the appearance of the virtual character. When the end user selects an option from the menu, this change is automatically shown here.
– **Options menu:** it is composed by a series of libraries that facilitate the virtual character editing task. There are three different libraries:
  • *Virtual Characters:* this library stores the 3D models of each virtual character. The end user only has to choose one to customize it.

- *Textures:* each virtual character has several textures associated. The end user can select one of them to customize it and change the appearance according to his/her preferences.
- *Complements:* for each virtual character it is necessary to have a series of accessories or complements. These complements can be hair styles, hats, moustaches, different kind of jewels, etc. The texture can also be selected to change its appearance.

– **Timeline:** although this area can be seen from the Appearance Editor, this functionality is to define the virtual character's behaviour as explained in the next section.

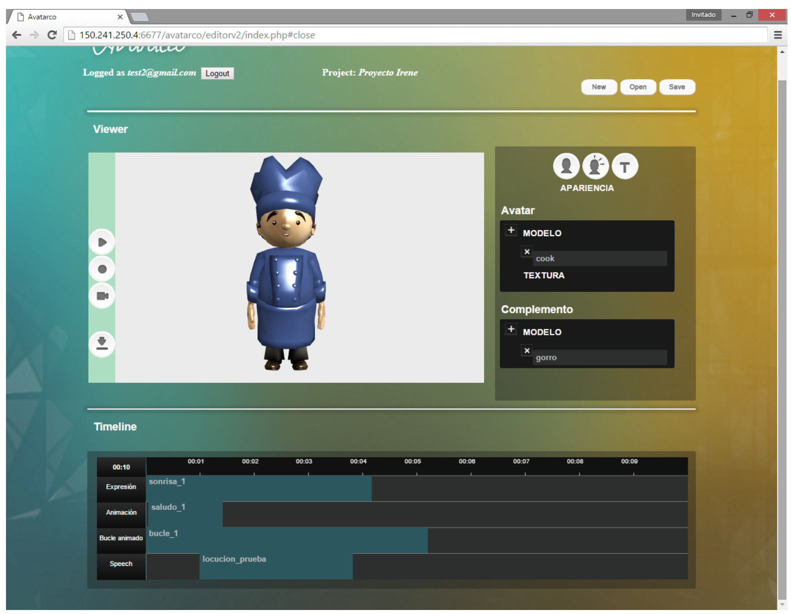


Fig. 1. Appearance Editor Interface

Thus the end user can select the basic appearance with the virtual characters menu, customize it through selecting different textures and have the possibility to add several complements to personalise the physical appearance to his/her liking. Once the end user has finished the appearance configuration of the virtual character, the project containing this character can be saved in the database with a given name.

3.2 Behaviour Editor

The objective of the Behaviour Editor is to animate the virtual character and to provide it with the ability to speak in an easy way without the need of having technical knowledge in animation of virtual characters.

Figure 2 shows the interface of the Behavior Editor. As it can be seen, there are three different areas:

- **Viewer Area:** in this area, end users can watch the defined behaviour by pressing play button.
- **Options Menu:** as in the Appearance Editor, it is composed by a series of libraries that facilitate the task of editing the behaviour. As it was mentioned before, these libraries have to be developed by experts. There are three different libraries:
  - *Expressions:* it contains predefined facial animations to express emotions such as: happiness, sadness, surprise, etc. The user can select the intensity of the expression.
  - *Animations:* it contains different animations to personalise the behaviour such as: greeting with hand, walking, etc.
  - *Loops:* These animations include all those movements that people do unconsciously and they are applied to virtual characters to make their behaviour more natural. For example: blinking, swaying body, etc.
- **Timeline:** it displays the list of selected options in chronological order. In order to specify the behaviour, the end user has to drag&drop expressions, animations and loops from the Options Menu.

As the aim of virtual character is to emulate the behaviour of a real person, it will be able to communicate in a verbal way. For this purpose there is also an option to introduce the text that the virtual character has to speak. This text is transformed into speech by a Text-to-Speech synthesizer. iSpeech [14] has been used for this purpose. To enter the text there is a button with the 'T' letter above Options Menu that opens a pop-up menu to write the text.

Thus, once the appearance of the virtual character has been defined with the Appearance Editor, this character can be doted with behaviour with this tool. The end user only has to write the text for the virtual character to reproduce. This text is converted into speech and an audio file is stored in the project folder. Then the user has to drag&drop this feature onto the Timeline, and add the animations that the virtual character has to do while acting. The user selects the start time of each action as well as the duration of each one.

The result can be visualized at any moment in the Viewer Area pressing the play button of the interface and the end user can modify it as many time as needed.

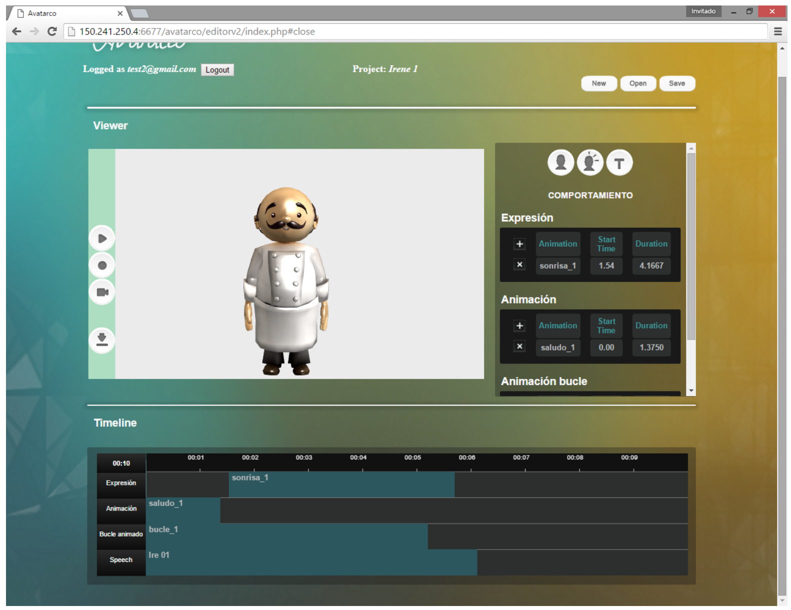


Fig. 2. Behavior Editor Interface

The composition created in the Timeline is translated and stored into a BML file. In the following BML example designed with the Behavior Editor, an introduction message has been created.

```
<bml xmlns="http://www.bml-initiative.org/bml/bml-1.0"
      character="Iaan" id="bml1">
    <gesture id="behaviour1" lexeme="hello-waving"
    start="2" end="5"/>
    <faceLexeme id="behaviour2" lexeme="happy" amount="0.8"
     start="2" end="5"/>
    <speech id="speech1" start="4">
        <text>Hello! My name is Iaan
        and I am a virtual character</text>
    </speech>
</bml>
```

This BML file will be the input of the Animation Engine. The following section explains all the process to obtain a virtual character acting.

4 Global Arquitecture

Figure 3 shows the global architecture in order to define the appearance and behaviour of virtual characters. As it can be seen, both editors are accessible from the same URL, explained in the previous section.

Once the behaviour of the virtual character is defined, it is stored as BML format, which is interpreted by the BML Parser module. This module extracts the text that has to be reproduced by the virtual character and converts it into audio by means of a text-to-speech module. Besides, the BML Parser extracts the behaviour of the virtual character of the BML labels. This behaviour is represented by the Animation Engine Module.

The Animation Engine is composed by several modules developed using JavaScript programming language and following all the HTML5 and Web3D standards. WebGL [15] technology has been used to render the virtual character via web. WebGL is based on OpenGL, which is a widely used open source 3D graphics standard. Besides nowadays, most common browsers support this technology without the need of plug-ins; Google Chrome, Mozilla Firefox, Apple Safari or Opera[16].

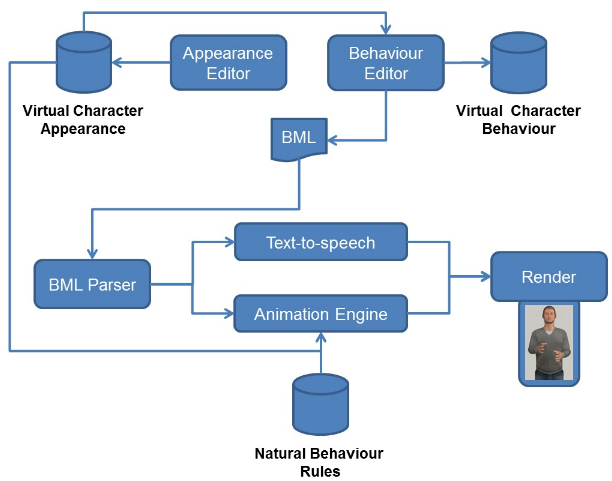


Fig. 3. Global Architecure

The Animation Engine runs as follows: the appearance of the virtual character is loaded from the Virtual Character database. It has natural behaviour, involving blinking, looking sideways, changing the weight of the body between both feet, crossing arms, etc. With this natural behaviour the virtual character emulates the actions of a real person in a waiting state. The natural behaviour rules are the same ones for each character and they are applied when there is no other behaviour running. In order to change the behaviour, the user has to edit one with the web based Behavior Editor.

When the Animation Engine receives the BML input with the designed behaviour, the animation engine stops any default (natural) animation except blinking. Then it starts the sequence of animations, emotions and other defined behaviour.

Thanks to both tools, the creation of virtual characters is very intuitive and fast. Its integration with other contents and applications can be done by two different means:

- By generating a video from the Behavior Editor itself. In order to generate the video FFmpeg [17] is used.
- By integrating the Animation Engine in the application where the virtual character has to appear.

How to integrate the virtual character depends on the characteristics and specifications of the final application. Through a video, the virtual character will only give information. Whereas if the animation engine is integrated, more interactive characters can be created. The following section explains different use cases.

5 Application and Use Cases

As it has been explained before, there are two different ways of integrating the result of the authoring tool. It depends on the final application as well as the role of the virtual character.

The authoring tool has been tested integrating the virtual character with the following roles: a presenter and a virtual teacher.

5.1 Virtual character as a presenter

One of the use cases of the Appearance and Behavior Editors has been to create a virtual presenter. The role of the virtual character is to give or present different kind of information. In this case, the main objective was to mix the virtual presenter with other kind of information as pictures, text, videos, etc. The idea was to obtain a video of the virtual character acting in order to use it in other editor to mix it with the rest of the content.

Thanks to the authoring tools, it was possible to create the introduction made by the virtual character in an easy way. The designer only had to select the appearance, write the speech and select the behaviour of the character.

As the virtual character had to be mixed with other contents, a chroma video was obtained in order to integrate it properly. Thus this video of the virtual character acting can be integrated with further multimedia content to create the final video of the presentation.

5.2 Virtual character as virtual teacher

E-learning applications are very commonly used in the academic area. Increasingly frequent use of virtual characters who exercise the role of teachers. The objective is to guide students in their learning process as a real teacher does.

One of the tasks of the virtual character is to introduce the lessons. For this task the authoring tool can be used to create the lessons. Moreover, teachers evaluate students so also the behaviour editor can be used to create these responses.

The difference with the previous application is that one of the advantages of the e-learning is that the student can decide his/her scheduling, so the application is interactive. In addition learning management systems (LMS) are normally used for this purpose. In this case the selected LMS was Moodle, so the animation engine was integrated into Moodle in order to show the virtual teacher. Lessons and possible responses of the teacher were created with the authoring tool and integrated into the Moodle course. Students could see the virtual tutor teaching the lesson and also received feedback from virtual tutor in the exercise phase.

6 Conclusions and Future Work

This paper presents an innovative tool to define the appearance and behaviour of virtual characters. Its design has been thought for users who are not familiar with the design of computer graphics, animation and programming. Thus, the user interface is specifically designed to be intuitive and very easy to use.

The integration of the result can be easily done by two different means: by generating a video and playing it; and by integrating the animation engine into the final application. We plan to evaluate the usability of the system with real users of the application.

As future work, we are working on reusing the defined behaviour in other virtual characters. As the authoring tool presented in this paper works now, the behaviour is defined for a particular virtual character. This is because each virtual character has its own defined animations. These animations cannot be used in other characters because each one has its particular characteristics. Although human virtual characters are normally used, and all of them have similar physiognomy, the length of the skeleton bones is not equal, so during the animations can be non detected collisions. Reusing the defined behaviour is an interesting functionality in order to reuse scripts, or to have the opportunity to change the selected virtual character, etc.

Another useful functionality is to define behaviour to several virtual characters at the same time. Thus, it would be possible to have several characters in the same scene with different or similar roles. For example to emulate news presenters or other kind of simulations.

References

1. Vilhjlmsson, H., Cantelmo, N., Cassell, J., Chafai, N.E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C. and Ruttkay, Z., (2007). The behavior markup language: Recent developments and challenges. In Intelligent virtual agents (pp. 99-111). Springer Berlin Heidelberg.
2. Jung, Y., Wagner, S., Jung, C., Behr, J., Fellner, D. (2010). Storyboarding and pre-visualization with x3d. In Proceedings of the 15th International Conference on Web 3D Technology (pp. 73-82). ACM.
3. Marriot, A., 2001. VHML. http://www.vhml.org/
4. Kshirsagar, S., Magnenat-Thalmann, N., Guye-Vuillme, A., Thalmann, D., Kamyab, K., Mamdani, E. (2002, May). Avatar markup language. In ACM International Conference Proceeding Series (Vol. 23, pp. 169-177).
5. Arafa, Y., Mamdani, A. (2003, January). Scripting embodied agents behaviour with CML: character markup language. In Proceedings of the 8th international conference on Intelligent user interfaces (pp. 313-316). ACM.
6. Oyarzun, D., Ortiz, A., Carretero, M.P., Gelissen, J., Garcia-Alonso, A., Sivan, Y. (2009). ADML: a framework for representing inhabitants in 3D virtual worlds. In Proceedings of the 14th international Conference on 3D Web Technology (pp. 83-90). ACM
7. Scherer, S., Marsella, S., Stratou, G., Xu, Y., Morbini, F., Egan, A., Morency, L. P. (2012). Perception markup language: towards a standardized representation of perceived nonverbal behaviors. In Intelligent virtual agents (pp. 455-463). Springer Berlin Heidelberg.
8. erekovi, A., Peja, T., Pandi, I. S. (2010). A controller-based animation system for synchronizing and realizing human-like conversational behaviors. In Development of Multimodal Interfaces: Active Listening and Synchrony (pp. 80-91). Springer Berlin Heidelberg.
9. Thiebaux, M., Marsella, S., Marshall, A. N., Kallmann, M. (2008, May). Smartbody: Behavior realization for embodied conversational agents. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1 (pp. 151-158). International Foundation for Autonomous Agents and Multiagent Systems.
10. Árnason, B.P., Porsteinsson, A., The CADIA BML realizer. http://cadia.ru.is/projects/bmlr/
11. Welbergen, Herwin van and Reidsma, Dennis and Ruttkay, Zsfia M. and Zwiers, Job (2010) Elckerlyc - A BML Realizer for continuous, multimodal interaction with a Virtual Human. Journal on Multimodal User Interfaces, 3 (4). pp. 271-284. ISSN 1783-7677
12. Lee, Jina, and Stacy Marsella. Nonverbal behavior generator for embodied conversational agents. Intelligent virtual agents. Springer Berlin Heidelberg, 2006.
13. Gebhard, P., Mehlmann, G., Kipp, M. (2011) Visual SceneMaker A tool for authoring interactive virtual characters. Journal on Multimodal User Interfaces July 2012, Volume 6, Issue 1-2, pp 3-11
14. iSpeech (accessed May 2016), http://www.ispeech.org/
15. Leung, C. and Salga, A. (2010). Enabling webgl. In Proceedings of the 19th international conference on World wide web, pages 13691370. ACM
16. WebGL - 3D Canvas graphics,(accessed April 2016), http://caniuse.com/
17. FFmpeg (accessed May 2016) https://ffmpeg.org/

Interactive Multimodal Platform for Digital Signage

- Authors: Helen V. Diez, Javier Barbadillo, Sara Garía, María del Puy Carretero, Aitor Álvarez, Jairo R. Sánchez and David Oyarzun
- Booktitle: International Conference on Articulated Motion and Deformable Objects
- Year: 2014
- Publisher: Springer

Interactive Multimodal Platform for Digital Signage

Helen V. Diez, Javier Barbadillo, Sara García,
Maria del Puy Carretero, Aitor Álvarez,
Jairo R. Sánchez, and David Oyarzun

Vicomteh-IK4, Paseo Mikeletegi 57, 20009 Donostia-San Sebastián, Spain
{hdiez,jbarbadillo,sgarcia,mcarretero,aalvarez,jrsanchez,doyarzun}@vicomtech.org
http://www.vicomtech.org

Abstract. The main objective of the platform presented in this paper is the integration of various modules into Web3D technology for Digital Signage systems. The innovation of the platform consists on the development and integration of the following technologies; 1) autonomous virtual character with natural behaviour, 2) text-to-speech synthesizer and voice recognition 3) gesture recognition. The integration of these technologies will enhance the user interface interaction and will improve the existing Digital Signage solutions offering a new way of marketing to engage the audience. The goal of this work is also to prove whether this new way of e-commerce may improve sales and customer fidelity.

Keywords: Multimodal Platform, User-Interface-Interaction, Digital Signage

1 Introduction

In the latest years, technology and especially new media has empowered marketing and commerce areas with new tools that go towards ubiquity and more and more faithful virtual representations of real products.

Nowadays, HTML5 and Web3D technologies are strongly pushing to web standardization of new media. Good and serious examples of efforts that are being done in this direction are the low level WebGL specification [1] and the high-level X3DOM architecture [2].

These new approaches could provide the basic platform for creating innovative marketing and e-commerce applications, which take advantage from potentiality of all technological channels and devices in a standardized way.

With this premise, the work presented in this paper consists on the development and integration of a web-based 3D engine and software modules that enable natural communication channels.

This technical work is built over three pillars:

- Coherent coexistence and communication among technologies coming from different disciplines and with different levels of maturity.

– Strong focus of usability, providing new interaction channels that make the human/computer communication more natural.
– Keep the message. That is, build technology that improves the way a message is transmitted to the user, not to condition the own message.

Therefore, these three pillars pretend to improve the channel related modules of the Shannon Weaver communication schema [3], as shown in Figure 1

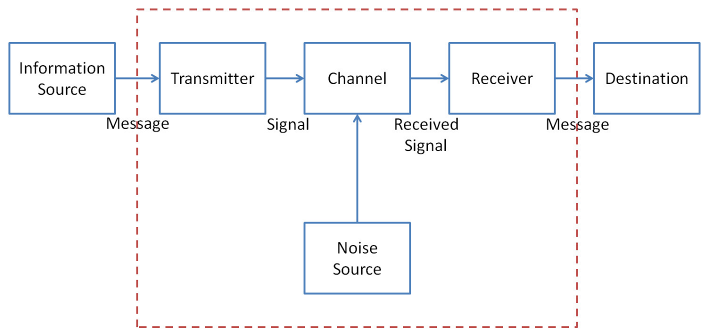


Fig. 1. Shannon Weaver communication schema on message transmission.

A realistic marketing use case has been built over this integration. The use case is designed as an interactive marketing platform to be shown in digital screens in public spaces. The platform allows the end-user to interact through natural channels, such as gestures and voice.

Therefore, the platform developed acts as a testbed to experiment interaction practices that maximizes the way a digital message is sent to the end-user. The use case is considered ideal when the end-user is a potential customer in this case and so, the importance of properly transmiting the digital message is even more critical.

Moreover, the standardized feature of the technologies developed keeps the coherence of the information when it is shown in any additional device, from PCs to smartphones. A modular design allows the content creator to abstract the digital message from the interaction channels, providing a platform that easily adapts to and takes advantage from the interaction capabilities of each device where it is running.

The paper shows the technical work carried out to get a stable version of the whole platform. Beyond the potential technical capabilities of these new technologies, the marketing use case is being validated in a controlled but real environment to check the usability in the marketing area during these months.

The paper is organized as follows; Section 2 analyzes the related work regarding digital signage and user categorization, Section 3 explains the architecture followed to accomplish the goals of this work, describing in detail each of the modules involved. The final section is about conclusions and future work.

2 Related Work

Research on Human Computer Interaction (HCI) goes back to the 1980s [4], however the growig affordability of the devices using these interfaces and the accessibility to software development kits [5] have led to the evolution of HCI into Natural User Interfaces (NUI), this new way of interaction operates through intuitive actions related to natural, everyday human behaviour such as; touch screen interaction, gesture recognition, speech recognition or brain machine interfaces. These new ways of interaction have gained broad interest within the HCI community [6] and various experiments have been done to prove its benefits.

The use of Microsoft Kinect sensors has been crucial in many of these experiments due to the availability of its open-source and multi-platfom libraries that reduce the cost of algorithm development. Keane, S. et al. [7] present a survey on the Kinect sensor. A gesture recognition module based on the motion sensor included in the Kinect is used in [8] to improve user experience in the management of an office environment.

NUI is also being introduced into digital signage systems. Satho, I. [9] presents a framework for building and operating context-aware multimedia content on digital signage systems in public or private spaces and to demonstrate the utility of the framework, he presents a user-assistant that enables shopping with digital signage.

Chen, Q. et al. [10] describe a vision-based gesture recognition approach to interact with digital signage. Bauer, C. et al. [11] also introduce a conceptual framework for interactive digital signage which allows the development of various business strategies.

Adapting content according to the audience is one of the objectives pursued by the companies that offer digital signage. There are several studies that personalize content according to the audience. For example, Müller et al., [12] present a system that automatically learns the audience's preferences for certain content in different contexts and presents content accordingly.

Ravnic, R. and Solina, F. [13] developed a camera enhanced digital signage display that acquires audience measurement metrics with computer vision algorithms. The system also determines demographic metrics of gender and age groups. It was tested in a clothing boutique where the results showed that the average attention time is significantly higher when displaying the dynamic content as compared to the static content.

The introduction of autonomous characters into user interface platforms is also a matter of study. In 2006, Gribaudo, C. and Manfredi, G. [14] patented a modular digital assistant that detects user emotion and modifies its behaviour accordingly.

3 System Overview

During this work a 3D avatar able to interact with the user through different channels has been implemented. It has been designed using a modular schema that includes three main components: a web component, the speech component, and the gesture component.

The web component is responsible for displaying the virtual character along with the content of the signage application. It supports any browser that implements WebGL technology.

The speech component allows the user to interact with the virtual character using voice commands. It integrates speech recognition and synthesis technologies.

The gesture component integrates computer vision technologies for face and hands tracking. It allows the user to interact directly with the content using hand gestures, at the same time allowing the system to estimate the emotional state of the user through his face.

All the modules are integrated in a HTML5 compliant application that is used as the frontend of the signage system. However, part of the core of the speech and gesture components are native applications and must be executed in a desktop environment.

Following sections describe each module in detail.

3.1 Virtual Character

A main aspect of this work is the introduction of an autonomous virtual character into digital signage systems to act as a natural interface between the user and the content offered by the device. The role of the avatar will be to ease the communication between the audience and the digital information provided. Thus, users will experience a more natural and intuitive interaction emulating the one between real people. Likewise, the user can customize this interaction by accessing information according to his interests or preferences.

To achieve this virtual character with natural behaviour an animation engine based on WebGL technology as the one presented in this work [15] has been developed. This animation engine allows realistic simulation of both the avatar's face and body expressions. The engine is capable of real-time rendering of the lips when the avatar is speaking and it also interpolates the facial expressions depending on the avatar's mood.

As for the body language the avatar performs gestures and movements as humans do when communicating with others. A thorough study regarding natural hand, arm and body gestures has been done and animations emulating these movements have been designed.

Figure 2 represents the introduction of a virtual character with natural behaviour into a WebGL compatible browser. Nowadays most commonly used browsers support this technology (Firefox, Chrome, Opera, Safari).

3.2 Speech Synthesis and Recognition

The platform includes technologies for both automatic speech recognition and speech synthesis.

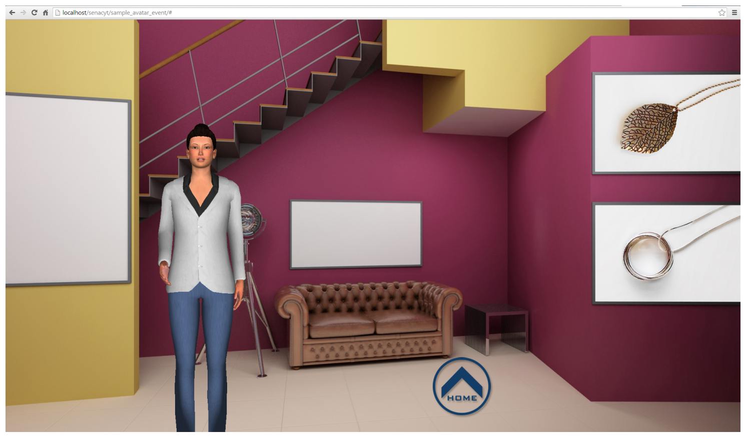


Fig. 2. Integration of the virtual character into a WebGL compatible browser.

Regarding speech recognition, the Google Speech Recognizer for Spanish was integrated adapting the publicly available java API to the needs of the platform. During the recognition process, the audio is collected from the microphone. It is then encoded to FLAC and passed via an HTTPS POST to the Google speech web-service, which responds with a JSON object with the transcription. The Google Speech Recognizer is speaker-independent and provides two language models to be used, based on (1) web searches for short phrases and (2) a generic language model for dictation. Considering the needs of the project, the generic language model was used to allow continuous speech recognition.

The integration with the web platform has been done using a regular text file. The recognition software writes into the file the transcription which is consumed by a script using long pooling techniques. This integration forces the component to be deployed on the same machine as the virtual character, but in the future it could be done using the new HTML5 standards for audio input.

Two possible solutions were included in the platform for speech synthesis in Spanish. Like for speech recognition, the Google Speech Synthesis was integrated for text-to-speech conversion. In this case, the text is sent to the servers of Google via an HTTP REQUEST and a speech file in MP3 format is returned through an HTTP RESPONSE. Since the Google Synthesizer is limited to a maximum of 100 characters, the API was modified to enable the platform to synthesize longer texts. For this purpose, the input text is previously splitted on the full stops. Each sentence is then synthesized and all the returned audios are concatenated in a unique WAV file at the end.

As an alternative to Google, the Microsoft Speech Synthesizer was integrated in the platform. This technology is provided through the Microsoft.Speech. Synthesis namespace, which contains classes that allow user to easily integrate functionalities for speech synthesis. In order to extend the voicebank of the platform, a module for voice transformation was also included. This module transforms the synthesized voices modifying some prosodic features like the fundamental frequency, the speech rhythm and the energy. As a result, this module is able to modify the source speakers speech to make it sound like that of a different speaker.

3.3 Gesture component

The platform implements a method for detecting and tracking the user's facial emotions and hand gestures. The system is composed by a Kinect device which captures video and a depth map, and it also includes a face detector for emotion recognition. The goal of the gesture component is to allow the interaction of the user with the avatar in both directions, resulting in a more natural experience. The avatar behaves according to the user's emotions and the user can perform gestures to communicate with the avatar.

The Kinect device captures video with an integrated camera and a depth map using infrared sensors. With the help of OpenNI and Nite APIs the system is able to detect and track human body parts and perform gesture recognition. Our system first detects the human body and then gets the head and the right and left hand positions. The 3D position is projected to 2D screen coordinates and the distance of the user with respect to the camera is obtained. This way the interaction is restricted to users that are facing the camera and close enough to it, avoiding interaction with people passing by.

In order to perform gesture recognition the user's hands are segmented from the rest of the body and tracked. The Nite API allows to track and detect the click gesture, the waving gesture and the rising hand gesture. The coordinates of the hand are also converted to screen coordinates so the user can use the hand as a mouse for selecting or clicking objects in a screen.

For the emotion detection process the face of the user is detected in combination with the head detection of the Kinect and a probabilistic face detector. First, the 2D position of the head is obtained from Kinect. If the distance to the camera is close enough the probabilistic detector is applied. Finally the face is detected and tracked and the system performs the emotion detection.

Our implementation of the emotion detection is based on the method proposed by [16]. When a face is detected in the screen a facial point mask is fitted to the face. This is achieved by first detecting facial features based on local image gradient analysis and then adjusting a deformable 3D face model to those features in the 2D plane. The mask represents the main facial features of a human face an it is able to track facial deformations computational efficiently and under challenging light conditions.

The emotion recognition method is based on the Facial Action Coding System developed in [17]. Every component of a facial movement is represented by an Action Unit (AU) and therefore every facial expresion can be decomposed into AUs. An AU is independent of any interpretation as they are the result of the contraction or relaxation of one or more muscles. In our program an AU is represented by the movement of a point of the facial mask. For example, the hapiness expression is detected if the threshold of the AUs "cheek raiser" and "lip corner puller" is exceeded. Using the facial point mask makes it trivial to measure AUs and detect if a facial emotion is being performed. A threshold is set to skip low intensity muscle actions. Although there are up to 100 AUs our system just measures a few AUs related to the seven universal emotions: fear, surprise, sadness, anger, happiness, disgust and contempt.

Finallly the system filters the detected emotions to reduce them to three emotions of interest for our application: the user can be interested, neutral or not interested. The avatar will behave differently depending on the emotions recognized on the user.

To avoid sending massive information to the avatar controller, the gestures and emotions are filtered over the frames to generate statistics that are sent every certain number of frames.

The integration has been done in the same way as the speech component. In this case the component is deployed with an executable that writes the gesture and face information in a text file. The web component reads the file using long pooling techniques.

4 Conclusions and Future Work

New multimedia sources are those that blend computer technology, with the audiovisual and telecommunications technology. They support a given language formed by image, sound, voice, written text, gestures and expressions and reach the user in a single marketer message.

Digitization is the universal tool that is profoundly transforming the international markets. This has eased the creation of new forms of marketing and industries dependent on these modes of information. The evolution of technology has changed the environment. The technological society of most developed countries live with new modes of experimental communication based on interactivity and the development of new forms of interdisciplinarity.

These digital technologies have completely changed the way information is transmitted. Due to its interactivity, the medium becomes the message itself. The multimedia models create a new, more powerful way to inform. New audiences are segmented and differentiated by gender, age and other components and the message focuses on this fact. These audiences are very selective with the message they receive because of their multiplicity.

New languages of human-machine communication are generated. Digital communication refers to the use of technology to achieve a certain purpose. The digital format of the medium indicates that the message content has something different and innovative. The format of the content is dependent on the medium, the distributor and the transmission system. The representation by digital screens in public spaces, allows private conversations and public environments simultaneously. It is a message created for the public, because the public participates by their expressions in the creation of the message. It is based on real-time interaction with an availability of 24 hours a day.

The transmitter, in this case the avatar, is in the same physical space as the receiver, so there is no distancing. A first approach takes place, it does not depend on the receiver who is surrounded by technology, but it depends on a transmitter / avatar searching for his receiver in their natural environment and a receiver that unintentionally becomes such.The figure of transmitter and receiver are constantly exchanged.

The message is not predetermined, it is dynamic, it is being created as the feedback happens. The receiver gives meaning to the message that has been sent massive / selectively. The audience tends to choose their messages, which improves the effectiveness of them. The range of possibilities the content displayed to the user is based on their characteristics and their willingness to receipt of the message, plus the response obtained along the communication between the avatar and the client.

Electronic technologies have a greater impact on the audience than more traditional media, since this support is also the means of reaching people and does not need more intermediaries. This way the message reaches out to the viewer, and the screen becomes the scenery for the reception.

The system has been tested in an academic atmosphere and it has proved to work recognizing and categorizing each user correctly. However, as future work we are planning to set the platform on a real scenario at a public space. This validation will serve the purpose of examining whether personalized marketing as the one proposed by our platform is better than traditional marketing systems.

To validate our system we will perform the following experiment; two groups of volunteers covering various age and gender ranges will be created. These volunteers will be invited to enter a mall in which the two trial systems have been set. Each group will try only one of the systems, once they have tried their corresponding system they will fill in a survey. In each corridor of the mall one of the marketing device systems will be set

1. did you stop in front of the device?
2. what struck your attention?
3. why did you leave?
4. did you find the experience amusing?
5. did you enjoy talking to the avatar?
6. did you find the communication with the avatar natural?
7. have you entered the store/s proposed by the system?
8. did you buy anything in any of the stores proposed by the system?
9. how would you improve the system?

The conclusions drawn from these surveys will direct further investigations in this field.

As mentioned in section 3.2 we also plan to standardize the voice capture from the microphone using WebRTC API [18], this communications standard developed by the W3C enables the embedding of audio and video in applications and websites. The WebRTC standard solves incompatibilities in real-time communications between browsers. This will also allow to integrate the Kinect device and video processing with HTML, which is currently handled by the Gesture Module plugin.

References

1. WebGL Specification, http://www.khronos.org/registry/webgl/specs/latest/1.0/ retrieved on March 2014.
2. X3DOM Specification, http://www.x3dom.org/x3dom/doc/spec/ retrieved on March 2014.
3. Weaver, W. Recent contributions to the mathematical theory of communication. In C.E. Shannon and W. Weaver (Eds.), The mathematical theory of communication, pages 128. 1949.
4. Myers, B. A. A brief history of human-computer interaction technology. interactions, pp. 44-54, 1998.
5. Goth, G. Brave nui world. Commun. ACM, pp. 14-16, 2011.
6. Seow, S. C., Wixon, D., Morrison, A., and Jacucci, G. Natural user interfaces: the prospect and challenge of touch and gestural computing. In CHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 4453-4456). ACM. 2010.
7. Keane, S., Hall, J., and Perry, P. Meet the Kinect: An Introduction to Programming Natural User Interfaces. 2011.
8. Re, G. L., Morana, M., and Ortolani, M. Improving user experience via motion sensors in an ambient intelligence scenario. 2013.
9. Satoh, I. A framework for context-aware digital signage. In Active Media Technology. Springer Berlin Heidelberg. pp. 251-262, 2011.
10. Chen, Q., Malric, F., Zhang, Y., Abid, M., Cordeiro, A., Petriu, E. M., and Georganas, N. D. Interacting with digital signage using hand gestures. In Image Analysis and Recognition. Springer Berlin Heidelberg. pp. 347-358. 2009.
11. Bauer, C., Dohmen, P., and Strausss, C. Interactive Digital Signage-An Innovative Service and Its Future Strategies. In Emerging Intelligent Data and Web Technologies (EIDWT), 2011 International Conference. IEEE. pp. 137-142. 2011.
12. Müller, Jörg, et al. "Reflectivesigns: Digital signs that adapt to audience attention." Pervasive computing. Springer Berlin Heidelberg, pp. 17-24. 2009
13. Ravnik, R., and Solina, F. "Audience measurement of digital signage: Quantitative study in real-world environment using computer vision." Interacting with Computers 25.3 pp. 218-228. 2013.
14. Gribaudo, Claudio; Manfredi, Giorgio. Virtual Assistant With Real-Time Emotions. U.S. Patent Application 11/617,150, 28 Dic. 2006.
15. Diez, Helen V., Sara Garca, Jairo R. Snchez, and Maria del Puy Carretero. "3D animated agent for tutoring based on WebGL." In Proceedings of the 18th International Conference on 3D Web Technology, pp. 129-134. ACM, 2013.
16. Unzueta, Luis, Waldir Pimenta, Jon Goenetxea, Lus Paulo Santos, and Fadi Dornaika. "Efficient generic face model fitting to images and videos." Image and Vision Computing 32, no. 5, pp. 321-334. 2014.
17. Ekman, P., Freisen, W.V. and Ancoli, S. "Facial signs of emotional experience." Journal of Personality and Social Psychology, 39(6), pp.1125-1134. 1980.
18. Web Real-Time Communications Working Group, et al. WebRTC 1.0: Real-time Communication Between Browsers. 2012. http://dev. w3. org/2011/webrtc/editor/webrtc. html, 2012.

Virtual Training of Fire Wardens through Immersive 3D Environments

- Authors: Helen V. Diez, Sara Garía, Andoni Mujika, Aitor Moreno and David Oyarzun
- Booktitle: Proceedings of the 21st International Conference on Web3D Technology
- Year: 2016
- Publisher: ACM

Virtual Training of Fire Wardens Through Immersive 3D Environments

Helen V. Diez *, Sara García †, Andoni Mujika ‡, Aitor Moreno §, David Oyarzun ¶
Vicomtech-IK4 Research Center, San Sebastián, Spain

Abstract

This paper presents an interactive application addressed to training experts on occupational hazard prevention and more precisely in fire safety in buildings. The platform allows the simulation of fires throughout a building. The fire warden trainee has to deal with that fire by a) finding the safest evacuation route or b) putting the fire out. The simulated fire will react to the environment, propagating in different ways depending on several factors; air currents in the building, material of the floor and walls, furniture, etc. The heat release of different materials has to be studied for this purpose. The fire will also react in real time to the response of the trainee depending on the method used to put it out. BIM standards will be applied to the logistic of the simulation. Thanks to this work methodology the application is provided with precise information regarding the building. The application is developed using a multiplatform game engine, so it can be used either as a web player serious game or in immersive environments such as the Oculus Rift HMD. In this publication we discuss the work methodology planed to fulfil the aforementioned issues.

Keywords: fire warden, training, evacuation plan

Concepts: •Human-centered computing → Virtual reality; Web-based interaction; •Applied computing → Interactive learning environments;

1 Introduction

The effectiveness of response during emergencies depends on the amount of planning, training, and drilling previously performed. Relatively speaking, small businesses may have more to lose than large companies when a disaster strikes. Because of high costs or lack of resources, many smaller companies have less rigorous business-continuity plans in place, and some have no formal processes at all. New technologies can help these smaller companies in their emergency plans as they are accessible and low cost.

Fire wardens have the role of coordinating the evacuation of people from the building and ensuring the safety of all occupants. The evacuation procedure is designed on the R.A.C.E. principle:

- R - Rescue people in immediate danger
- A - Alert others
- C - Contain the fire to the room or space of origin
- E - Extinguish if safe to do so or Evacuate the building

The application introduced in this paper aims to train fire wardens in emergency situations making use of new technologies. A web based serious game has been developed for such purpose. The game takes advantage of WebGL technology to visualize realistic 3D scenes on the browsers. Modern immersive technologies can also be used to train fire wardens, so the game is also accessible with VR devices such as Oculus Rift [Oculus 2015] or HTC Vive.

For this work, a simplified scenario has been defined. Building Information Model (BIM) methodology is interesting in such emergency situations as they offer complete information about the building: structure, plumbing, electricity, gas, exits, etc. BIM describes the process of designing a building collaboratively using one coherent system of computer models rather than as separate sets of drawings. BIM allows the entire supply chain of a building to better communicate with each other.

The building designed for this work consists of a warehouse unit on the ground floor and an office unit on the first floor. The ground floor is packed with wooden pallets and cardboard boxes; this material is highly flammable and dangerous if a fire were to take place. Two fire emergency situations have been defined. The fire warden trainee has to find the safest evacuation route taking into account the R.A.C.E. principle. An expert will evaluate the exercise in order to improve the preparation of the trainees.

This paper is organized as follows. Section 2 analyses the related work regarding fire fighting and BIM. Section 3 describes the work methodology followed to accomplish the goals of this work. Section 4 presents preliminary testing of the application with a simplified scenario. The final section is about conclusions and future work.

2 Related Work

Virtual environments have been widely used in training to quickly offer an impression of the layout of the building and reduce the hazardous effects of getting lost.[Bliss et al. 1997] [Waller et al. 1998].

A BIM based virtual environment (BIM-VE) was introduced by [Wang et al. 2014]. The BIM-VE aimed at improving building emergency management and demonstrated the capability (of the developed system) to train users, allowing them to quickly get familiar with the building and identify the right evacuation route.

[Li and He 2008] discussed the framework regarding 3D indoor navigation services, BIM models for building industry were directly imported into 3D GIS systems and they defined an ontology for 3D indoor navigation.

A hotel arson fire in Taiwan was simulated by [Shen et al. 2008]. The results demonstrated good prediction of fire development and smoke movement by comparing the combustion evidence of the scene and evacuees description. FDS [McGrattan et al. 2010] was used in this paper as it is a widely used simulator.

[Yuen et al. 2014] study demonstrated that FDS can be regarded as a reliable numerical tool for fire scene reconstruction. Their

*e-mail:hdiez@vicomtech.org
†e-mail:sgarcia@vicomtech.org
‡e-mail:amujika@vicomtech.org
§e-mail:amoreno@vicomtech.org
¶e-mail:doyarzun@vicomtech.org

Web3D '16,, July 22-24, 2016, Anaheim , CA, USA
ISBN: 978-1-4503-4428-9/16/07 ...$15.00
DOI: http://dx.doi.org/10.1145/2945292.2945296

simulation has been adopted as an innovative tool with which fire-fighters are trained to better understand the fire development in a compartment with specified amount of fuel loads. The study is being used at the E-Fire Investigation online training program (http://www.efireinvestigation.com.au/).

A hybrid method combining gaming technology, agent programming and fire science knowledge to design an evacuation training system for fire wardens was introduced by [Xi and Smith 2014].

[Cox and Terry 2008] created a BIM for emergency management for The Pentagon. The integrated system was based on building automation and control and fire protection systems to provide real time information about the state of the building. The resulting system gave the incident commander access to a large set of building data to respond to any threat with a complete picture of the situation.

3 Architecture

Figure 1 shows the architecture proposed for this application and the following subsections describe each module in detail.

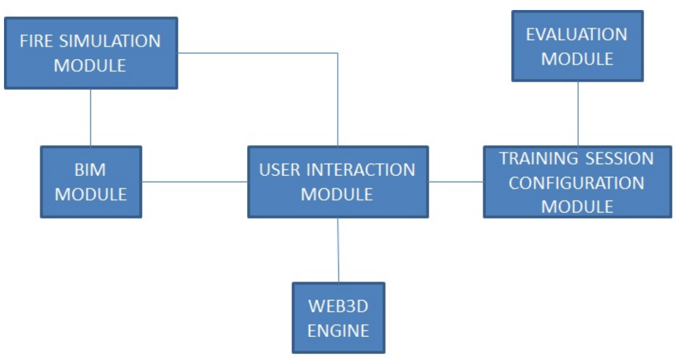


Figure 1: *Platform Architecture.*

3.1 Web3D Engine

The idea of this application is to train people on fire safety in an accessible and attractive way. This is why web 3D technology is suitable as it requires no installation or plugins, most commonly used web browsers nowadays support WebGL technology [WebGL (accessed April 2016)]. Furthermore, new immersive VR devices have appeared in the market, they have become very popular and an alternative to traditional learning techniques. Both platforms will be supported in this application.

On one hand, a lightweight multi-platform 3D engine is being developed (Figure 2). This engine is easily integrated into any solution requiring visualization capabilities and 3D animations. It supports desktop platforms (PC, Mac, Linux), mobile platforms (Android, iOS) and web browsers with WebGL support. The main features of this engine are:

- OpenGL 3, OpenGL ES 2.0, WebGL (three.js) backend
- 32 and 64 bit builds
- C++11 API
- Hierarchical scene graph
- Layer-based node management
- Materials: GLSL shaders
- Illumination: ambient, directional, point, spot lights
- Effects: render composition system, full-screen post-processing, render to texture
- various model format support: COLLADA, FBX, .obj, .ply, .stl, .3ds,
- Skinning-based animations
- Multilayer blending system

In addition, different tools have been developed including a web scene editor and conversion utilities for efficient model loading.

The final goal of this work is to introduce BIM models into the platform to take advantage of the information provided by them. Industry Foundation Classes (IFC) language is used for transferring information between BIM applications.

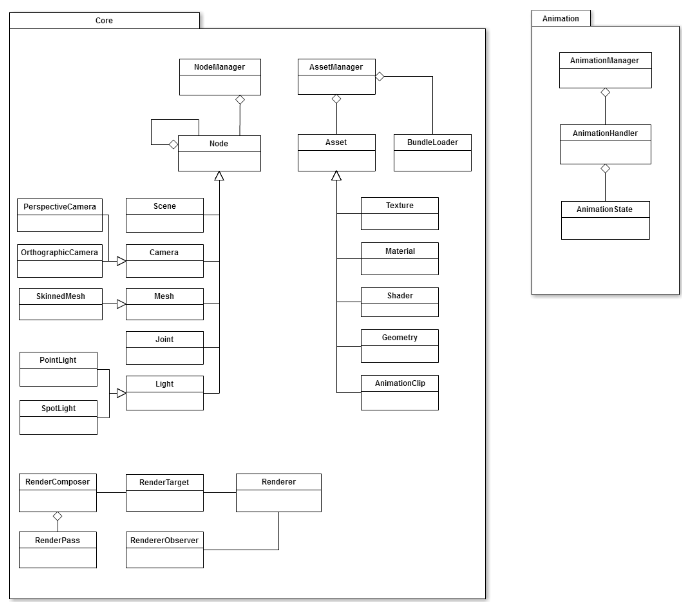


Figure 2: *Multi-platform 3D Engine Core Architecture.*

On the other hand, Unity3D engine has been used for the preliminary tests with the Oculus Rift VR device (Figure 3).

Figure 3: *Unity3D used to test the scene with Oculus Rift.*

3.2 User Interaction Module

This application is focused on training fire wardens. The wardens will have to face different challenges and will have to take decisions regarding the best way to solve them. As mentioned in the introduction, R.A.C.E. principle is applied. In some cases the best solution will be to extinguish the fire but in others the safest solution will be to find an evacuation route. These two situations are studied in this work.

3.2.1 Fire Extinguishing

As a general rule, fire fighting equipment should only be used to limit the spread of fire to enable safe evacuation. Before tackling any fire, it is vital that the alarm is raised and the fire brigade are called.

The most common form of fire fighting equipment are extinguishers. Fires are classified into 4 categories, depending on the fuels involved:

- Class A: Fires involving ordinary combustible solid materials (wood, cloth, rubber, paper and certain plastics or synthetic materials)
- Class B: Fires involving flammable liquids or liquefiable solids (gasoline, oil, paint, flammable gases or liquids and lubricants)
- Class C: Fires involving gases
- Class D: Fires involving burning metals

Different extinguishers are appropriate for different classes of fires [Tomorrow 2010]. However, type ABC multipurpose fire extinguishers are adequate for the purpose of this study.

In the application the trainee will have to find the fire extinguishers and put out the fire by using PASS formula for safe use. The course evaluator will check if the following steps have been completed correctly:

- Find the closest fire extinguisher
- Pull the pin
- Aim the extinguishers hose or nozzle at the bottom of the fire
- Squeeze the handle
- Sweep it slowly back and forth, covering the entire fire with the extinguishing substance

Leap Motion controller is used to capture and recognize the gestures of the trainee. [Weichert et al. 2013] studied the accuracy and robustness of this controller. The results showed a deviation between the desired 3D position and the average measured positions below 0.2 mm for static setups and of 1.2 mm for dynamic setups.

Leap Motion Controller allows the detection of the following gestures (Figure 4):

- Pulling the pin of the extinguisher: pinch and pull gestures have to be monitored
- Aiming the fire: an implementation based on [Dods 2013] will be developed
- Squeezing the handle: fist gesture will be monitored
- Sweeping the extinguisher: left and right swiping gesture has to be monitored

3.2.2 Building Evacuation

Fire extinguishers should only be used by trained personnel, so if this is not applicable the best procedure is to evacuate the building.

During an evacuation, fire wardens need to:

- Direct everyone to leave the building using all the appropriate routes and exits (avoiding inappropriate exits such as lifts)
- Check all accessible spaces in their area, including the bathroom, to make sure everyone has evacuated
- Close the doors to help to isolate the fire
- Guide everyone to the assembly area and assist in checking that everyone has arrived safely

Fire wardens in the game have to follow these safety rules to fulfil the goals of the exercise. Figure 5 shows Chrome browser in player mode in which the trainee has to exit the building following the signage.

Leap Motion Controller will also be used to detect if the trainee is closing doors and windows to contain the fire. Fist gesture is monitored for this purpose.

An evaluator will check the performance of the exercise and if will make the appropriate comments. If necessary, the trainee will have to repeat the exercise until it is proven that the goal of the lesson has been achieved.

3.3 BIM Module

A Building Information Model (BIM) is a semantic representation of a building. The geometric and semantic entities of the database describe elements such as walls, slabs, stairs, windows and doors; their materials and relationships between them. BIM models also provide highly accurate and detailed data about the current state of building elements. Their use in the architecture/engineering/construction (AEC) industry is rapidly increasing.

Industry Foundation Classes (IFC) system is a data representation standard and file format used to define architectural and construction-related CAD graphic data as 3D real-world objects. IFC was developed by buildingSMART and provides an interoperability solution between different software applications.

In [Jelenewicz et al. 2013] the Society of Fire Protection Engineers (SFPE) outlines how BIM technology is currently being used in the fire protection engineering profession. In this guide [Keung 2013] some fire protection BIM elements are described:

- System piping, droppers, fittings, valves and sprinkler heads, sprinkler inlets, sprinkler control valve set, subsidiary valves, flow switches.
- Pipe supports and brackets
- Fire alarm gongs and break glass unit
- Fire sprinkler pumps
- Sprinkler tanks
- Hydrants and hose reels (location of street fire hydrant determined by architects)
- Gas piping for suppression systems.
- Heat or smoke detectors, control panels, monitoring and control sensors, pump panels, check meter positions
- Fire extinguishers

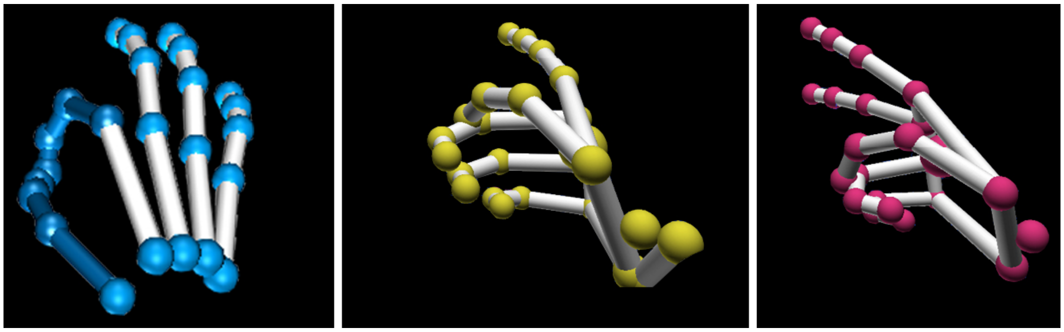


Figure 4: *Pinch, Fist and Swipe gestures controlled with Leap Motion.*

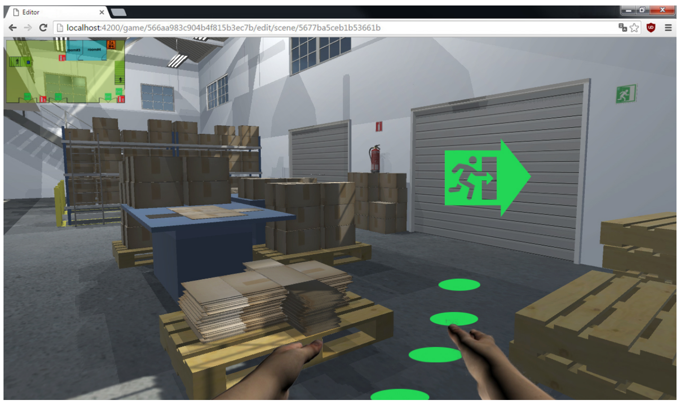


Figure 5: *Player mode web view.*

- Fire shutters and hoods above
- Smoke Curtains

The simplified test scenario used in this work includes some of these BIM elements.

3.3.1 Emergency Management

As proposed by [Cox and Terry 2008] an emergency management plan based on BIM is possible.

Figure 6 shows the emergency plan defined for the testing scenario used in this work. The plan includes the location of:

- Exits
- Fire Extinguishers
- Lift
- Staircases
- Room identifiers

The fire warden will have to be familiar with this plan before the virtual training session. The game will show this plan as a small image in the upper left-side of the screen.

3.3.2 Path planning

To evaluate the performance of the warden, it is important to know which route is the optimum one. As explained in section 4 the shortest path to the exit may not be the optimum path in fire safety as the warden might need to contain the fire or check if there are

Figure 6: *Emergency Floor Plan.*

people being left behind. This work plans to develop an algorithm to obtain a route that will combine the shortest solution with the proper solution regarding fire safety.

To decide which route the warden should follow, firstly, it is necessary to analyse the topology of the building. For that, the plan of each floor is used. Once doors and obstacles (such as columns) are identified, a graph that represents the connectivity of the building can be made, placing nodes in the doors that connect two rooms, in the beginning of the stair that goes to another floor and in the equidistant points among obstacles (like in a dual graph of a Voronoi Diagram). This way, connecting the nodes that do not have an obstacle in the middle, we obtain the a connected graph containing the paths that avoid obstacles and go to other rooms of the building.

Once the graph is constructed, the desired path (a sequence of nodes) has to be found. The list of nodes starts in the position of the warden, visits all needed places (windows that need to be closed, rooms that has te be checked, etc.) and finishes exiting the building. Computing this route is a classic problem: the Travelling Salesman Problem. Several algorithms can be found in the literature to solve this NP-hard problem ([Giardini and Kalmár-Nagy 2007], [Bui and Colpan 2005]).

However, the paths obtained from the graph are not natural paths, i.e., a person does not walk turning sharply in each node. That is why the paths are smoothed (and shortened). For example, if the path goes towards the next node of the sequence and at some point the subsequent node is visible from that point, the route changes the direction smoothly to go towards the node that is more advanced in the sequence (and makes the path shorter).

BIM information will be necessary in order to identify the obstacles and doors that will define the connectivity graph and to check what rooms, doors, windows, etc. the trainee should check before exiting the building.

3.4 Fire Simulation Module

Fire simulation is a complex task that uses the environment description and the physical or empirical behaviour of the fire spreading to calculate how the fire evolves through time.

There are a lot of ongoing research trying to find better and more complete algorithms at different time scale and geographic locations. Forest fires have been always a predominant target in a great number of previous works. We can mention FARSITE [Finney 1998], which uses Huygens principle of wave propagation. The continuous addition of complex information models (fuel model, changing wind patterns, humidity, terrain slopes...) have produced better results that can be used in short and long term predictions or in forensic analysis. The efficiency of such algorithms was not a primary concern, as normally supercomputers might be used to produce the results for a given time scale.

The introduction of buildings in the scenarios impacts the fire model and the algorithms that simulates the fire spread. Authors like Iwami [Iwami et al. 2004] and Ohgai [Ohgai et al. 2005] have introduced simplified building models in their scenario description and novel algorithms have been developed to deal with that possibility. Essentially, those works do not use the full interior of the buildings and they limit the fire spread to the building as a whole.

The fire spread simulation inside the buildings follows typical Computational Fluid Dynamics (CFD) calculations. The fire becomes volumetric, in contrast with the 2D vision in open fields like forests. The formulation behind these algorithms is complex as they are based in partial differential equations and diffusion systems. Consequently, they tend to obtain long computation times depending on the complexity of the building structure and its interior components (furniture, carpets, walls, windows...).

The fire simulation module presented in this work will use FDS [McGrattan et al. 2010] simulator to obtain off-line simulations of the fire spread inside the target building. The module uses this information to handle the visualization of a virtual fire in the scene at a given time. Thus, the module can be considered as an animation manager, providing the fire status at any requested simulation time.

The fire simulation module handles the interactivity in a simple but effective manner. When the user interacts with the virtual fire, some conditions are checked:

- that the chosen extinguisher is appropriate
- that its content is enough to put out the targeted fire
- that the extinguisher is correctly set-up (PASS)

If those conditions are met, the targeted fire spot is removed from the visualization, not instantly but with an animation in order to provide more realistic results. The perception for the users is that their actions have put out the fire.

3.5 Training Session Configuration Module

The platform presented in this work provides virtual training for fire wardens. The training course will be composed by a collection of exercises that will be configured, determined and validated by experienced tutors in the field. Each exercise targets specific concepts of the training program and progressively, the exercises become more complex and varied in relation with the presented fire emergency situations.

Prior to the virtual training course, a theoretical explanation of the purpose and objectives of the course will be presented. The theory will provide the essential knowledge on how to deal with emergency situations involving fire including the R.A.C.E. concepts. For each training session, a preliminary outlook of the main objectives will be introduced. The goals and the evaluation criteria will be clearly stated to the trainee.

3.6 Evaluation Module

The virtual training sessions will be supervised by an evaluator. The evaluation will be divided in objectives and subjective categories. The objective categories involve measurable actions of the trainee in the virtual environment:

- Simulation time till completion
- Interactions with the extinguishers
- Routing information including status of the doors, windows and other interactive element in the virtual scene

The subjective categories enclosed all the relevant information perceived by the evaluator that might not be translated to the virtual environment:

- The emotional situation of the trainee (nervous...)
- Doubts in taking decisions

The evaluation might be complemented with a questionnaire where the trainee's feedback could be collected. This feedback could be used to modify the evaluation. For example, the trainee might have experienced sickness in the VR setup and therefore, some situations in the virtual environment might be incorrectly handled. The feedback will be very useful to configure better exercises and to improve the platform itself.

4 Preliminary Tests

A two floor building has been designed for these tests. A warehouse unit takes most of the space on the ground floor and an office unit is disposed on the first floor. This distribution has been chosen as it shows a very common building distribution for this kind of work environments.

Furthermore, these environments where many wooden or plastic pallets are used are more likely to suffer severe fire spreads if a small fire takes place.

Table 1: *Features of the testing scenario.*

Ground Floor	First Floor
1 office unit	2 office units
1 bathroom	-
3 fire extinguishers	1 fire extinguishers
4 single-door windows	8 single-door windows
-	3 double-door windows
1 pedestrian door	1 pedestrian door
3 gateways	-

Pallets are an important material handling tool for warehousing. However, idle pallets present a significant fire hazard: they provide a source of dry fuel, their frayed edges are subject to easy ignition, and their open construction provides flue spaces through which fire can grow very hot and spread quickly. A pallet fire can severely test a sprinkler systems ability to contain the fire, and may result in severe damage to a buildings structure and its contents.

4.1 Testing scenario

The main features of both floors presented in the building plan in Figure 7 are shown in Table 1:

Both floors are connected by a lift and two staircases can also be found at the front and rear of the warehouse. Exit from the first floor is also possible by an external staircase.

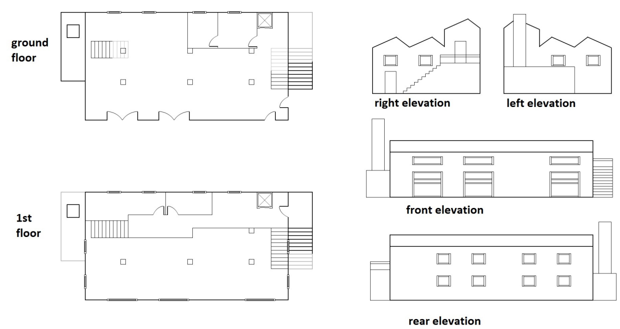


Figure 7: *Two floor building plan.*

Figure 8: *Testing scenario.*

The content of the warehouse also has to be taken into account as it may contain hazardous or very flammable substances. Figure 8 shows how the ground floor of the warehouse is packed with wooden pallets and cardboard boxes. The possible interaction of the fire with these objects has to be considered as it may influence the evacuation route.

According to Babrauskas [Babrauskas and Peacock 1992], the heat release rate (HRR) is the best predictor of fire hazard. The HRR it is expressed as the amount of heat (kW) released per time unit (s). It depends on various factors such as, the calorific value of the material being burned, shape and state of the material (big or small pieces, liquids, gases), the speed in which the material is burned and the air source available to feed the fire. The heat release rates of the objects in the testing scenario are needed as a prerequisite to estimate the growth of the fire. Data for this heat release rates is available in [Babrauskas and Grayson], [Bukowski et al. 1989] and [Kim and Lilley 2002].

- cardboard
- wooden pallets
- paper
- plastic
- glass

In these preliminary tests the fire warden is in charge of a single floor of the warehouse. The evaluator will value the trainees procedure. The features taken into account are: time needed to perform the exercise and accuracy regarding safety principles explained in Section 3.2.2. If the trainee chooses an incorrect route, the evaluator will explain again the fire safety rules to the trainee and the exercise will have to be repeated. The trainee has two attempts before the evaluator suggests the correct route.

At the time being, the wardens are in charge of evacuating the building following the safety principals. Oculus Rift glasses are used to navigate through the warehouse and Leap Motion Controller is used to test whether the trainee performs the fist gesture, this gesture is used to detect the opening/closing of doors and windows.

4.2 Scenario 1: Fire in the ground floor

In this first exercise a fire has started on the ground floor of the warehouse, the fire warden trainee must decide what evacuation plan to follow. The trainee has 30 seconds to complete the exercise.

The ground floor has 4 possible exits including gateways and pedestrian exits. As seen in Figure 9 in the first attempt, the trainee has chosen the orange route to evacuate the warehouse. This might be the shortest route but it is not the proper route.

The evaluator will point out that every room in the floor has to be checked to make sure there is no one left inside the building. Then the trainee will repeat the exercise. Now the trainee has followed the green route, which is correct, as rooms number 3 and 4 (office and bathroom) have been checked before exiting the building. The trainee has walked to those rooms and has performed the fist gesture in front of any closed door.

To simplify this exercise, the warden is only in charge of the ground floor, so there is no need to check the first floor, however in further exercises the warden will be in charge of the whole building.

The evaluator will take into account the time used by the trainee to complete the exercise and the number of attempts needed before selecting the correct route.

Figure 9: *Fire in ground floor: evacuation options.*

Figure 10: *Fire in first floor: evacuation options.*

4.3 Scenario 2: Fire in the first floor

In this second exercise, a fire has started on the first floor of the warehouse, the fire warden trainee must decide what evacuation route to follow. The trainee has 30 seconds to complete the exercise.

The first floor has one exit with access to the exterior staircase and is connected to the ground floor by two staircases and a lift. In normal circumstances the lift is not operable in case of a fire. However, if the warden were to help a person with mobility issues, the lift would become a possible option.

In the first attempt, the trainee chooses the red route to evacuate the building. This is obviously the shortest route but it is not accurate regarding R.A.C.E. formula. The evaluator explains that, if it is safe, the fire must be contained.

In the second attempt, the trainee chooses the orange route. Room number 2 is checked to make sure there is no one in and the door in room number 1 is closed (the trainee has performed the fist gesture) to contain the fire. These steps are correct. However, the trainee takes the stairs to the ground floor and exits the building. This choice is dangerous. The staircase chosen by the trainee is adjacent to the fire, so the fire can easily spread towards it.

In this case, the evaluator explains this issues to the trainee and finally the correct route is taken (green route). With this route rooms number 1 and 2 are checked and the trainee exits the building through the door in the first floor. This is the best route safety wise.

The performance of this exercise is shown in Figure 10.

Besides taking into account the objective performance of the trainee (time needed to evacuate the building, accuracy resolving the situation) the evaluator will examine subjective aspects (attitude, dubiousness, etc.) and will value the overall performance.

5 Conclusions and Future work

A fire extinguishing web based serious game has been presented in this article. The architecture of the application has been described including its main modules.

The preliminary testing of the application validates its usability. Although, a simplified scenario has been used for these tests, the results obtained prove that such an application could be of great interest for training fire wardens. The trainees benefit from the realistic experience offered by virtual reality and acquire concepts regarding fire safety in an interactive way. Moreover, the proposed platform offers a low-cost and safe solution, which is very convenient for the first stages in fire safety learning.

In the tests presented trainees have to find the best evacuation route and must act using R.A.C.E. formula. This execution is evaluated by an expert so the trainee can repeat the exercise with this feedback in order to improve.

Leap Motion Controller is used to detect the trainee's gestures. At the time being fist, swipe and pinch gestures are detected. Fist gesture detection is used to evaluate whether the trainee is containing the virtual fire by closing doors and windows. However, we plan to study whether other gestures might seem more natural.

At the moment fire is not interactive, meaning it only turns on or off depending on the users interaction. In the future we plan to make this fire simulation interactive so it will turn off gradually in a realistic way. We also plan to make it interact with the environment taking into account the heat release rates of the nearby objects. Fire will also have to react to air currents from windows or drafts.

As seen in Section 4.3 the shortest route is not always the best route in evacuation situations. In order to obtain the optimum evacuation route, we plan to work on an algorithm that will combine Voronoi diagram (to synthesize the topology of the building), a Travelling Salesman Problem solving algorithm (to get the ideal path in the graph) and the information from the BIM model (windows, doors and exits location) to obtain the shortest and safest route.

An IFC to .stl adapter is planned to be integrated in the web3D engine to ease the job of exchanging information.

WebVR is planned to be integrated into the 3D web engine proposed in this paper. This experimental JavaScript API provides access to Virtual Reality devices, such as the Oculus Rift.

The application has to be validated with real users to improve the training and evaluation modules. A local enterprise expert in occupational hazard prevention will collaborate in the validation of the platform.

References

BABRAUSKAS, V., AND GRAYSON, S. Heat release in fires, 1992.

BABRAUSKAS, V., AND PEACOCK, R. D. 1992. Heat release rate: the single most important variable in fire hazard. *Fire safety journal 18*, 3, 255–272.

BLISS, J. P., TIDWELL, P. D., AND GUEST, M. A. 1997. The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators and Virtual Environments 6*, 1, 73–86.

BUI, T. N., AND COLPAN, M. 2005. Solving geometric tsp with ants. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, ACM, 271–272.

BUKOWSKI, R., ET AL. 1989. The hazard-1 computer code package for fire hazard assessment. *NBS (NIST), Gaithersburg, MD*.

COX, B., AND TERRY, F. 2008. Creating a bim for emergency management. *Journal of Building Information Modeling (JBIM) National Institute of Building Sciences*.

DIJKSTRA, E. W. 1959. A note on two problems in connexion with graphs. *Numerische mathematik 1*, 1, 269–271.

DODS, V., 2013. Do you feel lucky?

FINNEY, M. A. 1998. Farsite: Fire area simulator-model development and evaluation. *Res. Pap. RMRS-RP-4. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station 1*, 47.

GIARDINI, G., AND KALMÁR-NAGY, T. 2007. Performance metrics and evaluation of a path planner based on genetic algorithms. In *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*, ACM, New York, NY, USA, PerMIS '07, 84–90.

HART, P. E., NILSSON, N. J., AND RAPHAEL, B. 1968. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on 4*, 2, 100–107.

ISIKDAG, U., ZLATANOVA, S., AND UNDERWOOD, J. 2013. A bim-oriented model for supporting indoor navigation requirements. *Computers, Environment and Urban Systems 41*, 112–123.

IWAMI, T., OHMIYA, Y., HAYASHI, Y., KAGIYA, K., TAKAHASHI, W., AND NARUSE, T. 2004. Simulation of city fire. *Fire Science and Technology 23*, 2, 132–140.

JELENEWICZ, C., ET AL. 2013. Building information modeling and fire protection. In *ASSE Professional Development Conference and Exposition*, American Society of Safety Engineers.

KEUNG, J., 2013. Bim essential guide for mep consultants.

KIM, H.-J., AND LILLEY, D. G. 2002. Heat release rates of burning items in fires. *Journal of propulsion and power 18*, 4, 866–870.

LI, Y., AND HE, Z. 2008. 3d indoor navigation: a framework of combining bim with 3d gis.

MCGRATTAN, K., HOSTIKKA, S., AND FLOYD, J. E. 2010. Fire dynamics simulator (version 5), users guide. *NIST special publication 1019*, 5, 1–186.

OCULUS, V. 2015. Oculus rift. *Available from WWW: ¡ http://www. oculusvr. com/rift*.

OHGAI, A., GOHNAI, Y., IKARUGA, S., MURAKAMI, M., AND WATANABE, K. 2005. Cellular automata modeling for fire spreading as a tool to aid community-based planning for disaster mitigation. In *Recent Advances in Design and Decision Support Systems in Architecture and Urban Planning*, J. P. Leeuwen and H. J. P. Timmermans, Eds. Springer Netherlands, 193–209.

SHAO, W., AND TERZOPOULOS, D. 2005. Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM, 19–28.

SHEN, T.-S., HUANG, Y.-H., AND CHIEN, S.-W. 2008. Using fire dynamic simulation (fds) to reconstruct an arson fire scene. *Building and environment 43*, 6, 1036–1045.

TOMORROW, G. 2010. Types of fire extinguishers.

WALLER, D., HUNT, E., AND KNAPP, D. 1998. The transfer of spatial knowledge in virtual environment training. *Presence 7*, 2, 129–143.

WANG, B., LI, H., REZGUI, Y., BRADLEY, A., AND ONG, H. N. 2014. Bim based virtual environment for fire emergency evacuation. *The Scientific World Journal 2014*.

WEBGL, (accessed April 2016). Webgl - 3d canvas graphics.

WEICHERT, F., BACHMANN, D., RUDAK, B., AND FISSELER, D. 2013. Analysis of the accuracy and robustness of the leap motion controller. *Sensors 13*, 5, 6380–6393.

XI, M., AND SMITH, S. P. 2014. Simulating cooperative fire evacuation training in a virtual environment using gaming technology. In *Virtual Reality (VR), 2014 iEEE*, IEEE, 139–140.

YUEN, A., YEOH, G., ALEXANDER, R., AND COOK, M. 2014. Fire scene reconstruction of a furnished compartment room in a house fire. *Case Studies in Fire Safety 1*, 29–35.

Augmented Reality System to Assist in Manufacturing Processes

- Authors: Ramón J. Ugarte, Nagore Barrena, Helen V. Diez, Hugo Álvarez and David Oyarzun
- Booktitle: Proceedings of the XXVI Spanish Computer Graphics Conference
- Year: 2016
- Publisher: Eurographics Association

CEIG - Spanish Computer Graphics Conference (2016)
Alejandro García-Alonso and Belen Masia (Editors)

Augmented Reality system to assist in manufacturing processes

R. J. Ugarte[1], N. Barrena[1], H. V. Diez[1], H. Alvarez[1] and D. Oyarzun[1]

[1]Interactive Computer Graphics Department, Vicomtech-IK4
20009 Donostia-San Sebastian, Spain

Abstract

This paper presents the initial results of a research project focused on bringing the Markerless Augmented Reality technology to the advance manufacturing sector. These initial results consist in a completely functional software system where a very simple use case is implemented to validate a set of base technologies. It has been implemented using a conventional camera without placing any markers to not modify the working environment and to not interfere with other industrial systems. Moreover, stability is a strong requirement to obtain a system that is actually useful in a manufacturing environment. Therefore, it is expected that an advanced version of this technology can be applied in an actual industrial use case in a short-to-medium term.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking

1. Introduction

In the very latest years, several international initiatives are promoting what is called the Fourth Industrial Revolution. Like the *Industrie 4.0* in Germany [KWH13], *Industrial Internet* in USA [EA13], *Industrial Value Chain Initiative* in Japan [bwo15], etc. These are roadmaps motivated by Governments, as country strategies, to make this concept real. This revolution is supposed to be driven by the application of different aspects and techniques over the actual factory, such as interoperable communication protocols, digitalization, advanced HCI techniques, etc. In this trend, it is considered that Visual Computing technologies should have a very important role.

Some authors carried out a recent research reflecting the usage of Visual Computing fields as Key Enabling Technologies -KET- for the Fourth Industrial Revolution [aotp15]. The work presented in this paper is a step forward, where a proof of concept for assisting manufacturing processes has been developed.

The proof of concept starts from the hypothesis that in the context of the new industrial trends, the advances on Human Machine Interfaces will play a key role, creating tools that improve the efficiency and competitiveness of workers using them.

The software system has been developed trying to fulfil these requirements:

- **To not use any marker.** The base technology is pure markerless. Therefore, flexibility is provided to use in any machine or process.
- **To place any kind of media content.** The tracking system should be robust enough to give support to any kind of media that is useful to assist in the manufacturing process.
- **To not interfere with other industrial software or hardware systems.** The final software and interface should be independent from any other industrial software, hardware or process. It will not be part of the critical steps in the production chain.
- **Importance of the stability.** The stability of the tracking will be crucial. It will be robust enough to provide support in processes that involve complex machines and granular operations.

Fulfilling these issues, the base system and the proof of concept to validate it have been developed as described in the rest of the paper.

The paper is structured as follows. The next section presents the state of the art on augmented reality technologies(from now on AR) specifically applied in the industrial sector. Section 3 details the technology involving our system and in Section 4 some experimental results are described. Finally, Section 5 presents the conclusions obtained and future work.

2. Related Work

Many authors have written about Markerless AR in the recent years. Good examples are provided by [TdMLA*07] or [CMPC06] where several markerless techniques are described in a general form, giving the initial clues for the developing of an AR markerless system that could be applied in many fields of knowledge.

Specifically, in the industry sector [NOCM12] makes a survey of AR systems and techniques that can be applied for several tasks such as robot path planning, AR collaborative design, plant layout or simulation. They also present the main challenges for the future AR systems on the industrial scope, focusing on the human interaction with the environment.

The present trend of manufacturing focuses on the modern ways of human interaction with the workstation in order to help on common tasks giving extra information about the workstation and the manufacture.

Given that, there are many existing solutions to reach that. For example, AR solutions for industrial purposes based on markers, such as the system proposed by [HWB08]. Here an AR system for mobile phones is presented, where the 3D virtual model is renderized in a position given by the detected marker.

Some authors have oriented their systems to several methods based on interaction in the manufacturing area such as [KVPK15], where the human-robot interaction is used to give the worker a real-time visualization of the robot's plans. In this case they place some artificial markers into the environment and extract features from it to recreate a 3D virtual representation, so that in the real-time stage, the markers are not necessary.

A similar strategy is used by [BPS05], where they place markers on the reconstruction stage and manually calibrate the system to place the CAD model on the environment correctly.

The main disadvantage of those methods is that you need to place the markers along the environment and in many cases it is not possible. Markerless AR techniques have evolved lately trying to overcome that situation.

For example, in [UWS09] a CAD-based recognition system is presented, where a 3D model can be directly tracked without the use of markers.

Moreover, [WLT*16] describes an AR markerless system, without the need of using markers on the reconstruction stage, that guides a worker through an assembly manual giving instructions with 3D virtual models placed directly on the workstation. The advantage with the previous system is that this solution can reconstruct itself the environment and does not need a CAD model to work. The main disadvantage is the need of an expensive RGB-D camera.

The reason that gives rise to this paper is the possibility to create a complete system using a conventional camera to create a 3D reconstruction of the environment and track it without placing any artificial markers. This paper presents a completely functional, simple to use and low-cost Markerless AR system for an interactive guiding of a worker on the assembly tasks.

3. System Overview

The main goal of this work is to create the base technology to develop tools that assist on manufacturing processes.

As mentioned previously, one key requirement is to not interfere in the production processes. Therefore, the system is conceptually defined as it is shown in figure 1, where the own production processes and machinery are separated from the assisted manufacturing processes.

In order to properly use AR techniques with industry or fabrication purposes it is necessary to set up correctly the virtual elements in the nearest environment, in this case the workstation. With this objective, the main goal of the system described in this section is to estimate accurately the camera pose in real-time. This pose must be calculated from the input frames that the camera is providing.

The aim of the design of the system presented in this work is to not pollute the environment with external marks. This way, the tracking system is based in the 3D representation of the nearest environment in order to estimate the camera pose, avoiding the use of marker based techniques.

In a synthesized way (figure 3), the system behaviour is as follows: it takes into account a set of images, which represent the nearest environment and it uses Structure from Motion (from now on SfM) techniques to obtain a 3D representation.

Therefore, the 3D representation of the environment is obtained, as well as the camera poses for all the images of the set used in the SfM process. Later, the camera pose will be obtained for each input frame and for this purpose it will be necessary to match the points of the 3D representation with their corresponding 2D points in the current images, as well as apply a template matching techniques to improve the pose estimation of the points.

In order to make the described process accurate, the system presented in this work follows two main stages:

- **Preprocess Stage**
  This stage is performed off-line and the main goal is to get the 3D representation of the environment. This is done using SfM techniques. A set of images of the nearest environment, denominated *keyFrames* are used as the input of this process. These *keyFrames* are used in the SfM process in order to obtain the 3D representation of that nearest environment, i.e the 3D point cloud (sample can be seen in figure 2). For this reason, the set of images must cover different perspectives of the environment (or object) where the AR application is going to work.
- **Real-time Stage**
  This stage is launched online and the *keyFrames* and the 3D point cloud obtained in the preprocess stage is used. For each input frame, the main goal is to calculate the camera pose. Two steps define this stage: detection and tracking. The first step is done matching the *keyPoints* (the most characteristic 2D points of the image) with their corresponding 3D points of the point cloud, obtaining an initially camera pose. In the second step template matching techniques are used in order to track the points of the point cloud through the previous input image to the current one, again obtaining the camera pose. When the tracking of the points is lost the system returns to the first step (figure 3).

The following sections are devoted to describe in detail the two stages of the described algorithm.

3.1. Preprocess Stage

The goal of the Preprocess Stage is to create a 3D representation of the environment from a given set of images. This representation needs to be as accurate as possible to avoid errors in the Real-time Stage. SfM techniques are used to achieve this purpose.

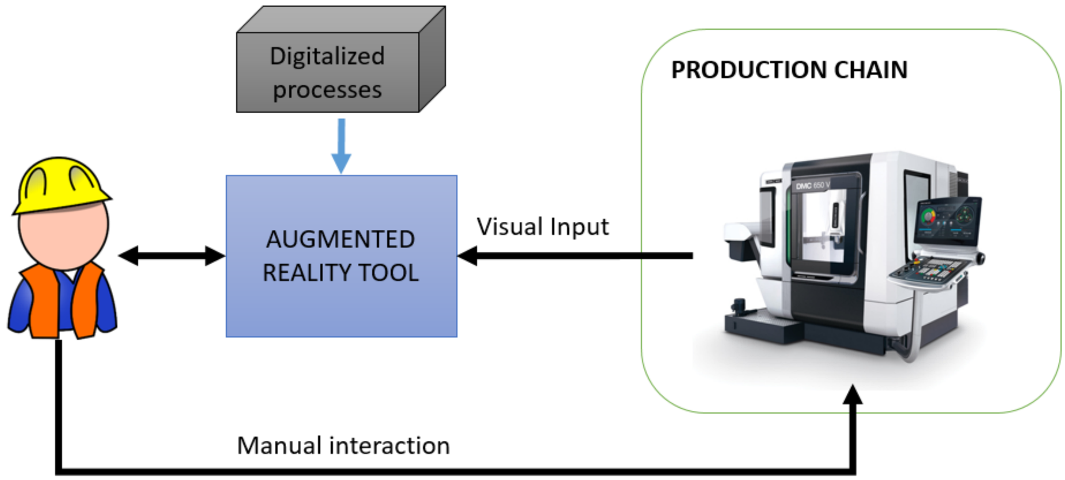


Figure 1: *Schematic of an industrial AR system*

These techniques use *keyFrames* as input data. For each *keyFrame*, *keyPoints* are extracted and theirs descriptors are calculated. In this way, FAST [TH98] feature extractor is used in order to extract the *keyPoints* of the *keyFrame*.

On the other way, the system has been built specifically for FREAK descriptors [AOV12] in order to work in real-time. This kind of descriptor is based on the human vision pattern. FREAK are binary descriptors calculated by the difference of intensity between the points of the pattern. Due to that it is very robust against illuminance and orientation changes but it is not scale invariant. To solve that issue, scale pyramidal reductions of the *keyFrames* are computed. For every scale level the *keypoints* are extracted. And for every *keypoint* its descriptor is calculated applying levels of Gaussian noise, taking the mean descriptor of the Gaussian levels as the descriptor of the *keypoint*.

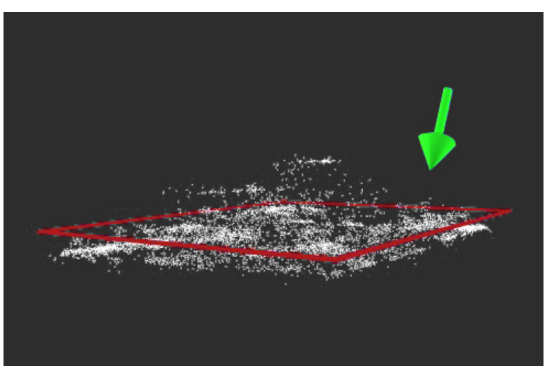


Figure 2: *Side view of the 3D point cloud*

Once the descriptors are calculated, in a sequential way, the algorithm tries to match those descriptors of the current *keyFrame* with the previous ones. In order to apply the SfM process as mentioned before, the same *keyPoint* had to be seen in several *keyFrames*. So, with the 2D positions of the same *keyPoint* in the different *keyFrames*, a triangulation method is applied in order to get their 3D coordinates. Then, a Bundle Adjustment [TMHF99] process is launched to get an estimation of the *keyPoints* 3D coordinates. Also a normal vector and a *reference keyFrame* has to be assigned for each *keyPoint*. In this case, the *reference keyFrame* for each *keyPoint* is the one where the *keyPoint* had been seen firstly and the normal vector associated to it faced the camera, i.e. is a vector from the *keyPoint* to the camera center.

The output of the Preprocess Stage is a file with the following data:

1. Point cloud (3D coordinates of the reconstructed points).
2. Surface normal estimation of each point of the point cloud.
3. *Reference keyFrame* for each point of the point cloud,
4. Camera pose of each *keyFrame* and,
5. Measurements of each point of the point cloud in each *keyFrame*.

It is important to remark that in order to simplify, the normal vector estimation for each *keyPoint* faces its *reference keyFrame* . Those normal vectors will be used later in the tracking stage.

Once the point cloud is calculated, all the *keyFrames* are processed to prepare them for the Real-time stage. For every *keyFrame* two actions are performed.

First, a 3D process in which a three level pyramidal reduction of the *keyFrame* is built. For each level, the *keyFrame* and its *keyPoints* from the 3D point cloud are resized to a half of the previous size.

Furthermore, each level is smoothed with a Gaussian kernel. The Gaussian smooth is applied to simulate intermediate scale levels

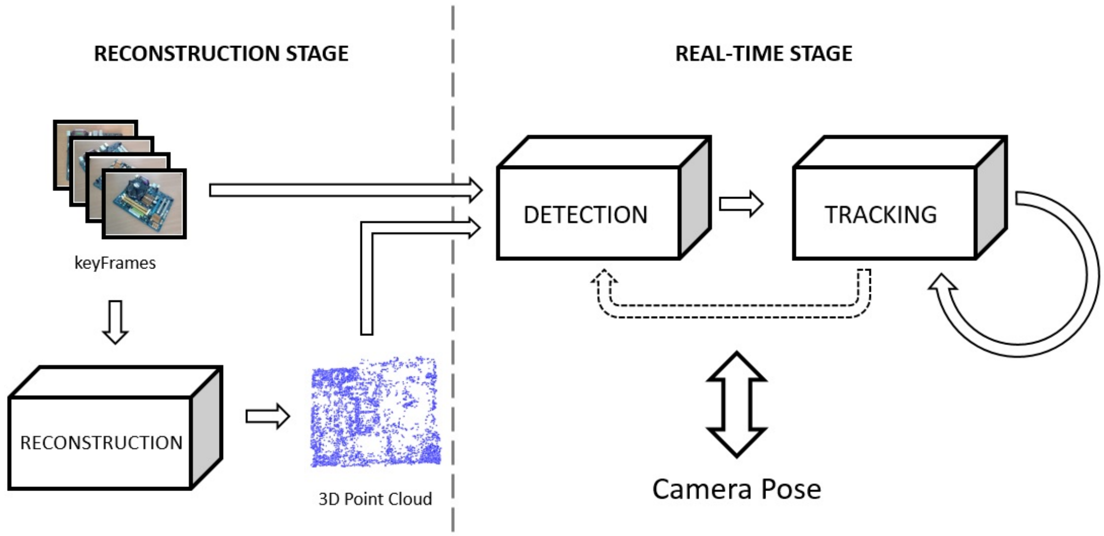


Figure 3: *Conceptual view of the system architecture*

and the effects of blur or noise during the live capture. The mean value obtained from the descriptors extracted in each smoothed image represents robustly the *keyPoint* and is the one that is saved as the representative descriptor of the *keyPoint*.

Second, a 2D process of the image where independent *keyPoints* of the *keyFrames*, not related with the 3D point cloud, are extracted. This *keyPoints* are used later for a refinement of the pose during the real-time stage and for the construction of a global database of all the *keyFrames*, using a fixed number of *keyPoints* for every *keyFrame*. This 2D process also applies Gaussian smooth to compensate blur, scale and noise during the capture and the mean value obtained by computing the descriptors of all the Gaussian levels is saved. These *keyPoints* are filtered by similarity of their descriptors and a score is given to each one of them. The ones with less score are discarded to reduce the number of *keyPoints* in the database.

3.2. Real-time stage

Once the pre-process is done, there is a middle stage needed to track correctly the 3D point cloud generated in the previous stage. The system can not associate automatically the 3D point cloud to the 3D virtual model, to be displayed and to match it with the 3D point cloud. So, an authoring tool is implemented to scale, rotate and translate the virtual models to match with the 3D point cloud. Once the transformation is complete, the real-time stage begins (figure 3).

The goal of this stage is to estimate accurately the camera pose for every input frame. This stage is divided in two sub-stages, the detection mode, where the system gives a first camera pose and the tracking mode, that estimates the rest of the camera poses until it fails and goes again into detection mode, repeating the process (figure 3).

The input needed in this stage are the *keyFrames* used in the preprocess stage and all the information calculated on the preprocess stage about the *keyFrames* and the point cloud. Once the database ready and the adjustment of the virtual model is done, the algorithm starts automatically on detection mode.

3.2.1. Detection mode

This stage uses the information extracted from the 2D process of the preprocess stage. With the input camera frame, the algorithm extracts the *keyPoints* and computes the descriptors. After that, it tries to match them with the global database of all the *keyFrames*.

Within this step, the algorithm obtains a histogram of the number of descriptors that are matched correctly for every *keyFrame*. This histogram shows the similarity of the input frame with the collection of the *keyFrames*. In this case, in order to gain robustness, the three *keyFrames* most similar to the input frame are taken. This helps the system to match correctly more *keyPoints* of the input frame with the *keyPoints* of the 3D point cloud.

With the three *keyFrames*, a function matches the descriptors stored with the input ones. The output of the function is the number of *keyPoints* correctly matched (from now on inliers) and a homography between the input frame and the *keyFrame*. This homography represents the difference of perspective between them.

If the number of inliers is higher than a fixed number, the system goes into the refinement step. Here, for the three selected *keyFrames*, the system takes initial homographies and warps the input frame with them. This process tries to adapt the view of the input frame to the *keyFrames* perspective.

On the new warped input frame, the system also warps the points of the 3D point cloud and recalculates their descriptors. Then, those are matched with the input descriptors. The inliers that match correctly are pushed into a vector of all the inliers of the three *keyFrames*.

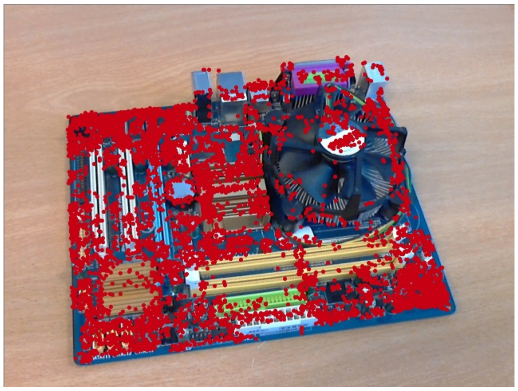


Figure 4: *Point cloud projected with the final camera pose*

Finally, the system will try to find the camera pose based on those inliers. For that, a *PnP* algorithm is applied, where the outliers are discarded using a *RANSAC* style procedure. If the final number of inliers is higher than a fixed minimum number, the final camera pose is taken as the correct one (figure 4), jumping to tracking mode.

3.2.2. Tracking Mode

The goal of this stage is to update the camera pose between the consecutive input frames. It is supposed not to be big shift differences between consecutive frames, so the 3D point cloud will be tracked following the points from one frame to the next.

To achieve that, based on the proximity of the camera pose between the previous frame and all the *keyFrames*, the algorithm uses the camera pose of the previous frame to find the current nearest *keyFrame*.

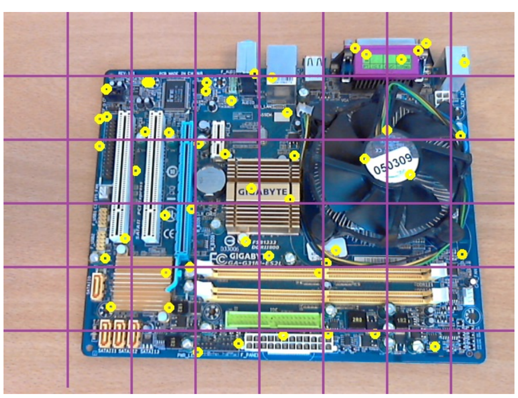


Figure 5: *Selected keyPoints taken from the 8x6 grid*

After that, the system selects a fixed number of *keyPoints* of the 3D point cloud that are visible in the selected *keyFrame*. With the aim of making the tracking properly the points are taken uniformly distributed from the 3D point cloud using a grid. This is done using a 2D grid with the projections of the 3D point cloud with the camera pose of the previous frame.

Figure 6: *Patch warped*

In this case we take 50 points using a 8x6 grid with 80x80 pixels square slots (figure 5). Jumping iteratively through each slot, the system takes the point with the best detectability and jumps to the next. If a slot does not have more points on it, the system jumps to the next one until it gets the 50 points or the maximum visible points.

Then, using template matching techniques, fixed size patches are taken around the selected projected points in the actual *keyFrame*(figure 5). More precisely, the patch size is 32x32 pixels being its center the projection of the 3D point. The corners of the patch are un-projected on the plane obtained with the normal vector associated to that 3D point and then projected to the plane of the current frame (figure 7), obtaining a warped square (figure 6). Then, the rest of the points of the patch are interpolated using the new coordinates of the corners.

Using a fixed searching area of 128x128 pixels, in our experiment, the patch and the searching area suffer a three level scale reduction in order to reduce computation complexity (figure 8).

In an iterative process, beginning with the smallest level and using a cross-correlation coefficient, the smallest reduction of the patch (8x8 pixels) is searched in the smallest level of the search area (32x32 pixels). Then, the position is propagated to the higher levels and finally to the original image, where the final position of the point is located.

Finally, the transferred points are used to feed the *PnP* algorithm that is executed in a *RANSAC* style. If the inliers are more than a fixed number, the resultant camera pose is taken as the new one. In other case, the system fails and the algorithm goes again to detection mode.

4. Experimental Results

This section evaluates the main characteristics of the system, such as, the robustness with regard of lighting or scale variations, partial occlusions, big image shift or the real-time requirement of the algorithm. All the code has been implemented in C++ and the PC used for the test is a Windows 10 with an Intel Core i5 3.3GHz 8GB.

The system has been tested with a CPU board as 3D real object.

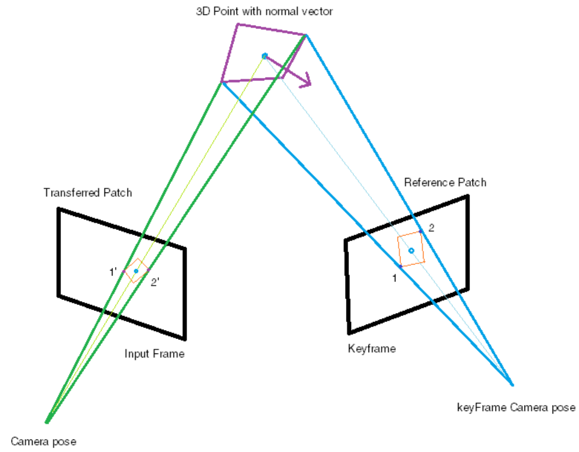


Figure 7: *Transferring patch process*

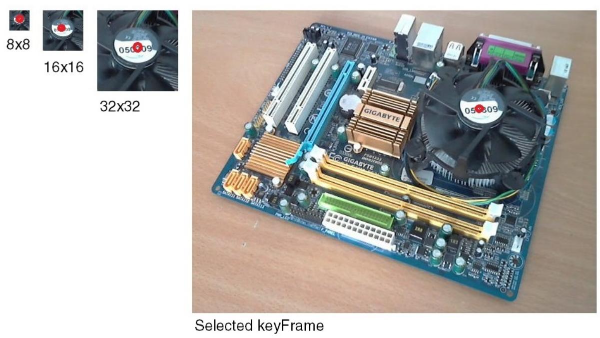


Figure 8: *Scale reductions of the patch*

The camera used for the test is a simple web camera with a VGA resolution (640x480 pixels). Twenty one photos of the board have been taken trying to cover the 360 degrees around the model and about 45 degrees from the aerial view.

In the real-time stage, the system takes a few minutes to compute the 3D point cloud and to process the *keyFrames* with that information. The 3D point cloud obtained once the images are passed through the reconstruction stage is a dense cloud with 6400 detectable 3D points.

In the real-time stage the detection mode, depending on the features of the input frame, runs approximately between 5 to 12 frames per second. This frame rate is a little below from real-time application but once the environment is detected, on the tracking mode, the system can run from 40 to 110 frames per second.

In the detection mode, to find the nearest *keyFrames* to the input frame, the global matcher takes 500 keypoints of each *keyFrame*, creating in this case a database of 4200 points. The bigger this database is, the slower the detection phase will run. After matching the input frame with the database, the experiments shows that no more than three *keyFrames* are needed to be extracted to perform the finest matching process. To reduce the compute time, this finest match is done parallelizing the three matching process. Once the first camera pose is obtained, the system jumps to tracking mode.

During the tracking mode, the system is continuously searching the *keyFrame* that is the closest to the input frame. It has been proved in the experiments, that in the process of transferring the warped patch, the complete tracking process is about two times faster if only the corners of the patch are transferred, while the rest of the points that belong to the patch are interpolated.

The system supports partial occlusions (figure 10), about 50% of occlusion, in both modes, detection and tracking. Also, in detection mode the system can make a correct match without the need of see all model. Furthermore the system can keep the tracking mode with heavy camera moves and can works in poor lighting conditions.

Finally, the result of the system is a stable tracking of the board where virtual objects can be placed (figure 9)

5. Conclusions and future work

This paper has described the initial results of a running research project whose objective is the development of an innovative technology that assists workers during manufacturing processes.

The partial results have consisted in the design and implementation of the base technology and a simple use case to validate it.

The system has been developed taking into account four critical requirements:

- To use pure markerless technology.
- To provide integration with any kind of media.
- To not interfere with industrial critical processes.
- To give enough stability.

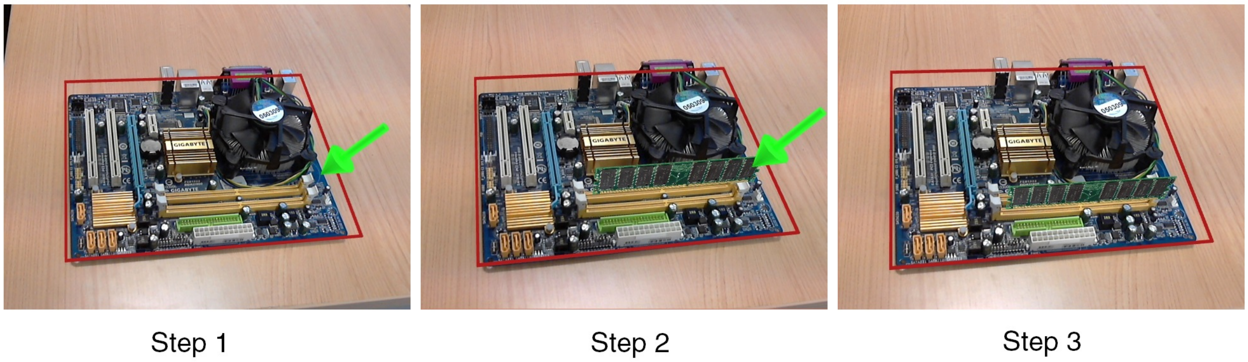


Figure 9: *Final application of the system; guiding on the assembly of a RAM memory piece by steps*

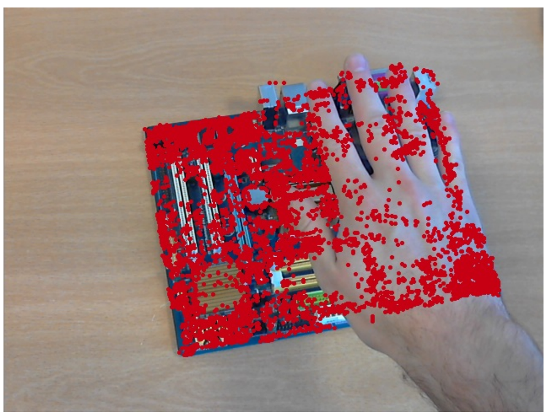


Figure 10: *Tracking mode with occlusion*

The results show the robustness of the system and they give sense to continuing the research towards a deployment in a real industrial use case.

Future work is focused on three main lines from the technological point of view:

- New tracking modules. A module that adds border tracking fed from machine CAD models is being implemented. This will strongly improve the tracking in cases where the CAD model exists.
- Optimization. The system is being optimized for its use in mobile devices, avoiding the need of having a PC during the assistance process.
- Web port. The tracker is being ported to web-based platforms.

In parallel, a real industrial use case is being developed and it will be validated in manufacturing companies with a closed relation to the author's research centre.

References

[aotp15] AUTHORS OF THIS PAPER S.: Visual computing as a key enabling technology for industrie 4.0 and industrial internet. *Computer Graphics and Applications, IEEE 35*, 2 (2015), 26–40. 1

[AOV12] ALAHI A., ORTIZ R., VANDERGHEYNST P.: Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012), Ieee, pp. 510–517. 3

[BPS05] BLESER G., PASTARMOV Y., STRICKER D.: Real-time 3d camera tracking for industrial augmented reality applications. 2

[bwo15] Industrial value chain initiative. "http://www.iv-i.org/", 2015. "[Online; accessed 20-February-2016]". 1

[CMPC06] COMPORT A. I., MARCHAND E., PRESSIGOUT M., CHAUMETTE F.: Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *Visualization and Computer Graphics, IEEE Transactions on 12*, 4 (2006), 615–628. 1

[EA13] EVANS P. C., ANNUNZIATA M.: Industrial internet: Pushing the boundaries of minds and machines. *General Electric Co. Imagination at work* (2013). 1

[HWB08] HAKKARAINEN M., WOODWARD C., BILLINGHURST M.: Augmented assembly using a mobile phone. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on* (2008), IEEE, pp. 167–168. 2

[KVPK15] KOŠNAR K., VICK A., PŘEUČIL L., KRÜGER J.: Markerless augmented reality for human robot interaction. In *Modelling and Simulation for Autonomous Systems*. Springer, 2015, pp. 185–195. 2

[KWH13] KAGERMANN H., WAHLSTER W., HELBIG J.: Recommendations for implementing the strategic initiative industrie 4.0. *Plattform Industrie 4.0* (2013). 1

[NOCM12] NEE A., ONG S., CHRYSSOLOURIS G., MOURTZIS D.: Augmented reality applications in design and manufacturing. *CIRP Annals-Manufacturing Technology 61*, 2 (2012), 657–679. 1

[TdMLA*07] TEICHRIEB V., DO MONTE LIMA J. P. S., APOLINÁRIO E. L., DE FARIAS T. S. M. C., BUENO M. A. S., KELNER J., SANTOS I. H.: A survey of online monocular markerless augmented reality. *International Journal of Modeling and Simulation for the Petroleum Industry 1*, 1 (2007). 1

[TH98] TRAJKOVIĆ M., HEDLEY M.: Fast corner detection. *Image and vision computing 16*, 2 (1998), 75–87. 3

[TMHF99] TRIGGS B., MCLAUCHLAN P. F., HARTLEY R. I., FITZGIBBON A. W.: Bundle adjustmentâĂŤa modern synthesis. In *Vision algorithms: theory and practice*. Springer, 1999, pp. 298–372. 3

[UWS09] ULRICH M., WIEDEMANN C., STEGER C.: Cad-based recognition of 3d objects in monocular images. In *ICRA* (2009), vol. 9, pp. 1191–1198. 2

[WLT*16] WU L.-C., LIN I., TSAI M.-H., ET AL.: Augmented reality instruction for object assembly based on markerless tracking. In *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (2016), ACM, pp. 95–102. 2

Interaction Challenges in the Development of Fire Warden VR Training System using a HMD

- Authors: Helen V. Diez, Aitor Moreno and Alejandro García-Alonso
- Booktitle: VISIGRAPP (2: HUCAPP)
- Year: 2017
- Publisher: Springer

Interaction Challenges in the Development of a Fire Warden VR Training System using a HMD

Helen V. Diez[1], Aitor Moreno[1] and Alejandro Garcia-Alonso[2]
[1]*Vicomtech-IK4, San Sebastian, Spain*
[2]*University of the Basque Country, San Sebastian, Spain*
{*hdiez, amoreno*}*@vicomtech.org, alex.galonso@ehu.es*

Keywords: Virtual Reality, Head-Mounted Display, Interaction Devices.

Abstract: This paper presents challenges encountered when interacting with 3D environments by means of Virtual Reality devices in a seated position. A use case has been devised and developed to support this work: a fire warden virtual training system. The users interact with the virtual environment with a combination of a Head-Mounted Display and other assisting devices to complement the virtual experience (keyboard, gamepad, leap motion). This work points out the benefits achieved using virtual reality in this environment and it exposes the limitations experimented by humans in such situations. With the knowledge gathered with these experiments this paper proposes interaction techniques to overcome them.

1 INTRODUCTION

In the last years, the new generation of Head-Mounted Displays (HMDs) have become popular devices to provide highly immersive experiences which are very difficult to achieve by other means. They are excellent tools to view virtual environments (VEs) in a realistic way but, by themselves, they do not offer interactive capabilities with the objects in the VE. Additional devices are attached to the HMD setup to overcome this situation.

When wearing a HMD, users cannot see the real world surrounding them, so it is difficult to locate and use the mouse, the keyboard, the gamepad or any other interaction device at their disposal. Additionally, even wireless devices, they are normally placed at physical location, breaking the virtual experience of the user. New interaction devices have been developed to solve this issue, such as, walking platforms, specific hand input devices, haptic gloves, etc. These solutions may be a good choice for gaming purposes, in which a higher level of activity is necessary or recommendable. However, depending on the application or final user these devices may not be suitable. For example, for educational purposes, with school students, it seems more likely for them to use their own computer screen and keyboard.

This work studies some interaction possibilities of HMDs in a natural seated position. To analyse the challenges of interacting with HMDs we consider three universal interaction tasks within 3D or VE applications (Hand, 1997), (Bowman et al., 2001), (Bowman et al., 2004) which Jankowski, J. and Hachet, M. (Jankowski and Hachet, 2013) defined as:

- Navigation: Related to the motor task of moving the viewpoint through an environment. If it includes a cognitive component, it is referred to as Wayfinding
- Selection and Manipulation: Related to the techniques of choosing and picking an object and specifying its position, orientation, and scale.
- System Control: Related to the communication between user and system which is not part of the VE.

For the use case presented in this work, we will study the methodology followed when planning the experiments and the decisions taken regarding these three tasks according to our work experience.

This paper is organized as follows. Section 2 analyses the related work regarding interaction in VEs and hand gesture recognition. Section 3 explains the interaction devices selected for the Setup of this work. Section 4 describes the fire warden use case developed and analyses each of the interaction tasks. The final section is about conclusions and future work.

Diez, H., Moreno, A. and Garcia-Alonso, A.
Interaction Challenges in the Development of a Fire Warden VR Training System using a HMD.
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) - Volume 2: HUCAPP*, pages 84-91
ISBN: 978-989-758-229-5

2 RELATED WORK

Jankowski, J. and Hachet, M. (Jankowski and Hachet, 2013) presented a state of the art of non-immersive interaction techniques for Navigation, Selection and Manipulation, and System Control. This work also summarizes the main 3D Web design guidelines.

Moya, S. et al. (Moya et al., 2014) analysed the influence of different factors in locomotion control in 3D VEs. They designed an automatic locomotion system and concluded that, for non-casual gamers, automatic locomotion is preferred, whereas gamers prefer to control locomotion themselves.

Regarding hand gesture recognition for HCI, (Sharma and Verma, 2015) proposed the tracking of six static hand gestures. They used images extracted from recorded video streams at different distances and positions. They are able to use their system to control applications such as power point, media player, windows picture manager, etc.

Manresa, C. et al. (Manresa et al., 2005) present a real-time algorithm to track and recognise hand gestures for interacting with the videogame. This algorithm is based on three main steps: hand segmentation, hand tracking and gesture recognition from hand features. The results of this work show that users can substitute traditional interaction metaphors with their low-cost interface.

Segura, I. et al. (Segura et al., 2007) developed a simulator of construction machinery for safety training. In this work the visualization was implemented as a chroma-key-based mixed reality system, combining a 3D VE, a real cabin interior, and some superimposed messages to the user. The overall impressions from the testers were positive.

Goenetxea, J. et al. (Goenetxea et al., 2010) presented an interactive and stereoscopic hybrid 3D viewer. In this work they used Nintendo Wii control and developed a dynamic gesture recognition procedure used for interacting with the weather animations.

Regarding fire simulation, it has also been a research topic in VR applications (Moreno et al., 2011), (Moreno et al., 2012). The simulation of real-time interactive fire is a challenging topic that has not been solved yet.

In this work we propose to use a HMD combined with usual interaction devices such as keyboards and gamepads. Leap Motion Controller is used for hand recognition. The following sections explain the setup we have developed for a fire warden virtual training use case.

3 USE CASE SETUP

In our use case, the tests are carried out autonomously by a fire warden trainee. Therefore, it is not a collaborative learning experience, as each user must be aware of their own actions and decisions.

The setup used in these experiments is composed of the following items:

- Chair. One goal of our work is to keep the setup as simple as possible. This is why we tried to limit the amount of extra devices needed. The users will interact in a seated position, needing no other devices.
- Computer. The recommended system requirements are: Windows 7 or newer, Compatible HDMI video output, Intel i5, NVIDIA GTX 770 or greater.
- Monitor. The final goal is for the user to visualize the 3D VE with the HMD, but it is necessary to have a monitor to make the first steps until the virtual experience is launched and to get some additional information after the training session.
- HMD. We have used Oculus Rift DK2. The Oculus CV version or any other HMD brand could be used, but the specifications of the computer might require modifications to match the HMD requirements.
- Keyboard. Anyone used to dealing with computers is familiarized to keyboards and there is no extra effort needed. Wireless or not is irrelevant, but it should be placed in a known and fixed place before the VR training session. Bare in mind that the user will not be able to see the real world with the HMD on.
- Gamepad. It is also a familiar device in video games and it is intuitive for navigation purposes. We have used the Logitech Rumblepad2.
- Leap Motion Controller. This device provides means for tracking the hand gestures performed by the user. It was attached to the Oculus Rift with the universal bundle. The latest Orion beta software has been used (Orion, 2016).

3.1 Interaction Devices

This section discusses the interaction devices selected for the use case presented in this work.

3.1.1 Keyboard

The keyboard is used to navigate through the VE using the up, down, left and right arrow keys. In traditional computers, they are a group of isolated keys, easily distinguishable from others. In more compact devices such as laptops, the arrow keys despite being closer to other keys are still easy to locate. The advantage with laptops is that the keyboard is attached to the screen so it cannot change position. Sometimes, WASD keys are also used for navigating. Normally, this keys are used with the left hand leaving the right hand free to interact with the mouse. This setup is typical used in gaming (see Figure 1).

Both arrow keys and WASD keys can be used with only three fingers of one hand, leaving the other hand free. In this use case, the free hand is used to perform the gestures defined to select and manipulate objects in the scene. This gestures are captured by the Leap Motion Controller.

3.1.2 Gamepad

Instead of using the keyboard to navigate, we have experimented with a gamepad. In this case both hands are needed to hold the gamepad and the user has to let go of one hand to perform the hand gestures for selecting and manipulating objects of the VE.

In most applications, the right stick is used for navigating (forward, backward, left and right) and the left stick for turning the camera view. This left stick is not necessary as the user can look around by moving the head with the 360 tracking system offered by the HMD. So, in the gamepad the user only has to use the right stick, none of the other buttons are needed. This makes it simple for the user, despite wearing the HMD.

3.1.3 Leap Motion

HMDs offer a really immersive experience. Users feel they are actually in the VE they are watching through the HMD device. This experience is so realistic and convincing that users try to interact with objects from the virtual world with their own hands. However, regular interaction devices as the ones described above do not allow this sort of interaction. This is why we have chosen Leap Motion Controller for our work. This device allows users to integrate their hands into the virtual world and unlike other devices, they require no practice or previous knowledge from the user.

The Leap Motion Controller is mounted on the front of the Oculus Rift glasses, as seen in Figure 2, and whenever the users lift their hands to grab an object from the virtual world, the virtual hands appear in the virtual world, performing exactly the same movements as the real hands.

This setup is more compact than other alternatives using external trackers for gesture recognition.

4 USE CASE: FIRE WARDEN TRAINING

This use case has been designed to train experts on occupational hazard prevention and more precisely in fire safety in buildings. The fire warden trainee has to perform the tasks described below:

- Locate the possible fires in the building.
- Extinguish these fires.
- Alert other people in the building.
- Evacuate the building.

These tasks imply different levels of interaction with the VE. In the next subsections we will explain the methodology followed to decide which kind of interaction devices to use, and we will explain the trial and error process carried out for each of the classical VR interaction categories.

4.1 Navigation

One of the tasks that the user has to perform is the location of the possible fires throughout the building, to discover these fires the user has to navigate through a two-floored warehouse. The floors are connected by two staircases which the user can use as many times as needed.

Another task that implies navigation is the evacuation of the building. For these tasks the user has to know the exits available in the building. Moreover, the users are asked to put the fire out, to do that the location of the extinguishers has to be found. The user has to select them and head towards the fire.

These tasks imply wayfinding, as the user has to have a cognitive overview of the building and has to be able to make decisions. To ease this job, the application displays some messages (images or text) on the screen of the HMD. These messages are explained in section4.4.

4.2 Selection

One of the goals of our fire safety training is to allow wardens to learn how to use fire extinguishers in a safe and efficient way. In real life a fire extinguisher must be manipulated in a precise way and fires must be put out following the PASS method (Safety and Administration, 2016):

- Pull the pin in the handle.
- Aim the nozzle at the base of the fire.
- Squeeze the lever slowly.
- Sweep from side to side.

Figure 1: Arrow Keys and WASD Keys in a Traditional Keyboard (left) and a Laptop (right).

Figure 2: Leap Motion Controller attached to Oculus Rift DK2.

Figure 3: Setup #1: Interaction with Keyboard and Leap Motion Controller.

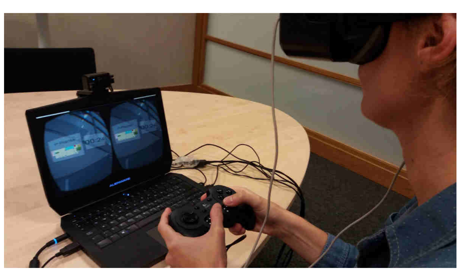


Figure 4: Setup #2: Interaction with Gamepad and Leap Motion Controller.

We have made several tests to see which hand gestures are most appropriate to simulate these actions in virtual life and we have used the Leap Motion Controller (LM) to track them.

In the first place, the warden has to grab a fire extinguisher. To allow this action in the fire safety training application, first the user has to navigate to one of the extinguishers, and get close to it to a certain distance. We have set 60 cm as the maximum distance to allow the user to grab the extinguisher. We have chosen that distance as it is the standard length of an adult's arm (McDowell et al., 2008),(NASA, 2016). Once the users are placed at reaching distance from the extinguisher they have to extend their arm towards it and grab it.

To represent this gesture, the first idea we thought of was a "palm closing" gesture. This gesture is easily trackable by the LM and it is easy to perform and understand by anyone: "I want to grab the extinguisher". However, we dismissed this gesture for two main reasons. First, the PASS method indicates the user has to pin in the handle. This action implies a more precise movement of the fingers and we wanted to force the user to do so. Second, as explained later, we selected the "palm closing" gesture for opening and closing doors or windows.

We want to enhance the interaction degree and the learning experience. For that we designed independent gestures for different actions. This way, the user can easily internalize the meaning and consequences of each virtual gesture in the real world. This is why we decided to define the "pinch" gesture for grabbing objects in the scene, in this case the fire extinguisher (see Figure 5).

So, summarizing the full action process, the user has to get close to the extinguisher, reach an arm towards it and pinch it by holding the index and thumb fingers together. The platform must detect a collision between the extinguisher and the virtual representation of the user's hand. If this happens, the extinguisher is moved towards the user and it is placed in the right side of the visual field, representing that the extinguisher has been grabbed and it is ready to be used. The actions that the user can do with the extinguisher are discussed in the next subsection 4.3.

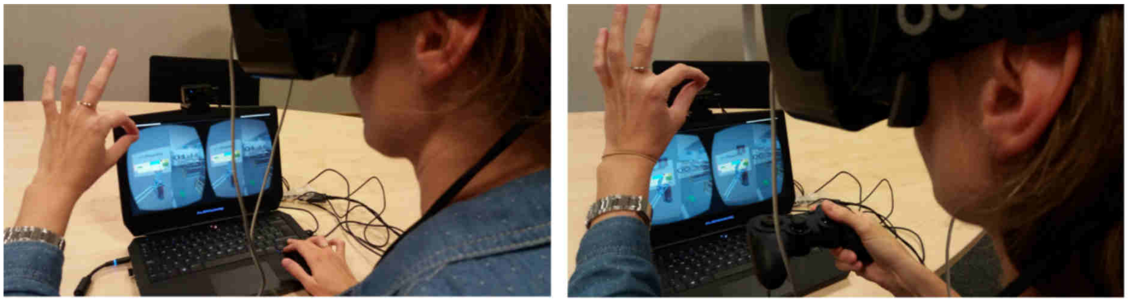


Figure 5: “Pinch” gesture with Setup #1 (left) and Setup #2 (right).

Another goal, is to teach the warden how to evacuate the building safely. In this work, to do this, the warden has to navigate to the closest exit and open the door. We have chosen the “palm closing” gesture for this action as it is the one we perform in real life to grab door handles. The user has to approach a door in the virtual scene and reach out the arm towards it. As with the extinguisher selection, the user has to be placed at least 60 cm. from the door.

At first, we thought on detecting the collision between the virtual hand and the door handle. We also thought it would be recommendable to track the hand turning down a certain angle as if turning a door handle to open it. But these actions resulted confusing and tedious. In the virtual scene, to detect a collision between the hand and the handle, the warden had to get very close to the door, which implied correcting their position and performing the “palm closing” gesture several times. Moreover, depending on the angle degree the hand was turned, this gesture could be confused with the “thumbs up” gesture. Finally, we decided to simplify this action by allowing the door to open by performing the “palm closing” gesture at the right distance from an exit (see Figure 6).

Additionally, we have the same action to represent a door by pulling it or by pushing it. Even if pulling or pushing would affect in the cognitive process of opening a door, we decided to omit this feature for the sake of simplicity. Furthermore, it is expected that the doors should be correctly designed, promoting “pushing” mode over “pulling”, as it may be required for safety regulations.

4.3 Manipulation

In this use case, the fire warden must learn to manipulate fire extinguishers. First, the user has to navigate with the fire extinguisher towards the fire and manipulate it following the PASS method. The correct distance to put out a fire safely is between three to one meters. So if the user is further than 3 meters or closer than 1 meter from the fire an alert text is displayed as a HUD. Once the user is at the right distance, the user has to aim at the fire, squeeze the nozzle and sweep from side to side.

In real life, we would need both hands to manipulate the extinguisher. However, in this virtual scenario the user also has to interact with the keyboard or gamepad. When using the keyboard, it is not convenient for the user to lift the fingers from the arrow keys and with the gamepad, the user needs one hand to keep holding it. To simulate the squeezing of the nozzle, we decided to track one hand of the user performing a “thumbs up” gesture (Figure 7), closing the hand with the thumb heading up. We chose this gesture because it is similar to the one users should perform in real life and to differentiate it from the “closing palm” gesture, in which the thumb is not visible.

Then, if the user is in possession of an extinguisher and performs the “thumbs up” gesture at the right distance from a fire, a graphical particle system representing the extinguishing jet will blast out from the user’s hand position. If the user stops performing the gesture, the particle jet stops. To increase the realism, the noise of a real extinguisher is played through the audio systems (integrated in the HMD or external).

While performing this gesture the user also has to move the hand from left to right to perform a sweeping gesture.

Our Unity-based 3D engine (Unity3D, 2016) detects whether a collision between the particle jet and the fire has taken place. This way we check if the user is aiming at the fire and sweeping the hand correctly. If there is a collision the fire is put out. We define each fire as a collection of particle systems, so they are put out gradually. If the user does not sweep from side to side, the fire will not put out correctly. Fire extinguishers contain around 10 seconds of extinguishing power. If the warden spends more than that time trying to extinguish the fire it will probably mean the sweeping is not being performed correctly. In this case an alert HUD appears to remind the warden the PASS method.

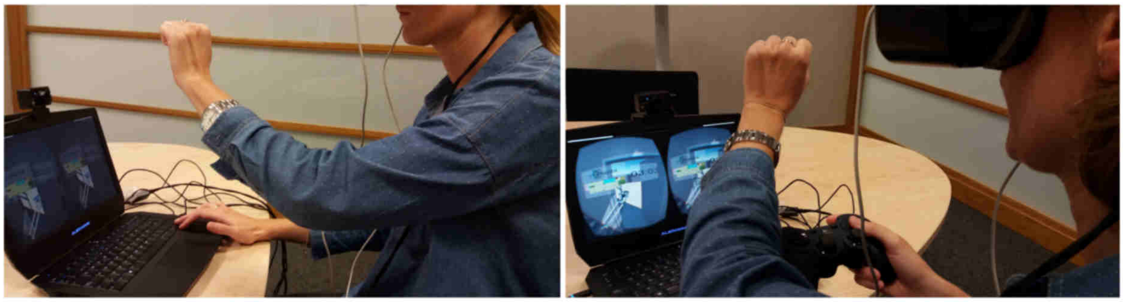


Figure 6: “Palm closing” gesture with Setup #1 (left) and Setup #2 (right).

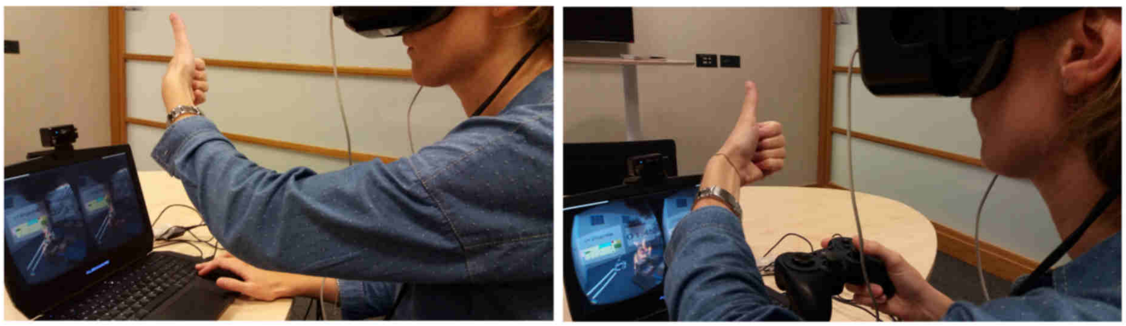


Figure 7: “Thumbs up” gesture with Setup #1 (left) and Setup #2 (right).

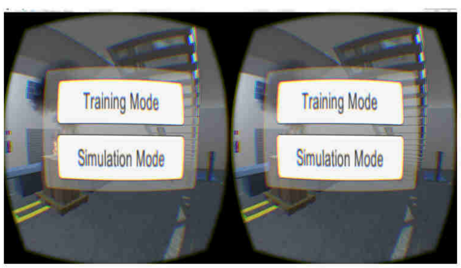


Figure 8: Mode Selection.

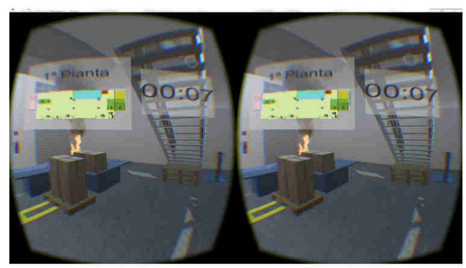


Figure 9: Upper Left: 2D mini-map. Upper Right: Timer.

4.4 System Control

This section explains the input messages we have defined to communicate the user and the system. These messages appear as HUDs (Heads-Up displays) on the screen.

Where to place each HUD is also important, as far as possible they should be integrated in the virtual scene (Yao et al., 2014).

The user can begin the simulation in training mode or in simulation mode (see Figure 8). The user selects between these options with the LM.

During the simulation the user can visualize a semi-transparent and non-intrusive 2D mini-map. This map has been designed as an egocentric map. The map shows the following signs to help the user:

- “You are here” marker.
- Exit location signs.
- Extinguisher location signs.

A timer is also visible at all times, so the user knows how long it is taking to perform the exercise. These HUDs are shown in Figure 9.

In the training mode, some green arrows are displayed on the 3D environment to show the user the path to the closest extinguisher, once the user has selected the extinguisher and put out the fire, again some arrows are displayed to show the path to the closest exit (see Figure 10). These signs are not shown in the simulation mode, as the user should have acquired knowledge about the environment in the training mode.

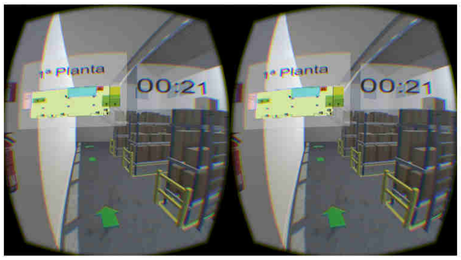


Figure 10: Training Mode. Green arrows placed on the floor guide the user to the exit.

5 CONCLUSIONS

In this work, a VR fire warden training system has been presented. Its main objective is to provide VR experiences to fire wardens in order to internalize and comprehend the concepts behind the standardized procedures that have to be followed when a fire emergency arises in a building. In our preliminary developments, a few actions have been tested as they target the main interactions within VR environments: Navigation for evacuation routes and fire finding; Selection and Manipulation to interact with the extinguishers and doors; and System Control interaction to display non intrusive 2D information as a HUD. All the actions have been developed as a combination of gestures recognized with the aid of Leap Motion Controller and keyboard/gamepad to navigate the 3D environment.

The seated position has been defined as a requirement from the very beginning as it is the most comfortable way of interacting with the VR environment. The stand-up position poses additional problems like i) problems with the cabling, ii) collisions with the real-world furniture and iii) anxiety regarding the cognitive disassociation of the virtual position and the real position, specially when stairs are included in the scenarios. The seated position is also recommendable if the interaction is going to last a long period of time.

More trials and validation of the VR setup is needed. However, we have found that the election of the keyboard or the gamepad is a matter of personal preferences. In any case, we have reports about the necessity to adapt the keys of the keyboard or buttons of the gamepad to each personal preference.

Some users found some limitations in the VR navigation. They report that turning commands conflict with looking to the sides in the VR environment. Nevertheless, it is a matter of time to get the users acquainted with the navigation system.

As future work, the following prototype will experiment alternatives to solve the problems reported by the users. A more extensive evaluation will be carried out.

The extension of the prototype to other users and purposes is also under consideration. The presented use case could be adapted easily to children in order to show them basic information about what they have to do when a fire emergency arises.

After the preliminary implementation, we are planning to introduce a different use case, oriented to the visualization of 3D models on the Web. This use case is very different from the one presented in this work and therefore, it will provide complimentary information about the VR techniques in a different VR environment.

REFERENCES

Bowman, D., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice, CourseSmart eTextbook*. Addison-Wesley.

Bowman, D. A., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. (2001). An introduction to 3-d user interface design. *Presence: Teleoperators and virtual environments*, 10(1):96–108.

Goenetxea, J., Moreno, A., Unzueta, L., Galds, A., and Segura, l. (2010). Interactive and stereoscopic hybrid 3d viewer of radar data with gesture recognition. In Graa Romay, M., Corchado, E., and Garcia Sebastian, M., editors, *Hybrid Artificial Intelligence Systems*, volume 6076 of *Lecture Notes in Computer Science*, pages 213–220. Springer Berlin / Heidelberg.

Hand, C. (1997). A survey of 3d interaction techniques. In *Computer graphics forum*, volume 16, pages 269–281. Wiley Online Library.

Jankowski, J. and Hachet, M. (2013). A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*.

Manresa, C., Varona, J., Mas, R., and Perales, F. (2005). Hand tracking and gesture recognition for human-computer interaction. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 5(3):96–104.

McDowell, M. A., Fryar, C. D., Ogden, C. L., and Flegal, K. M. (2008). Anthropometric reference data for children and adults: United states, 2003–2006. *National health statistics reports*, 10(1-45):5.

Moreno, A., Segura, Á., Korchi, A., Posada, J., and Otaegui, O. (2011). Interactive urban and forest fire simulation with extinguishment support. In *Advances in 3D Geo-Information Sciences*, pages 131–148. Springer Berlin Heidelberg.

Moreno, A., Segura, Á., Zlatanova, S., Posada, J., and García-Alonso, A. (2012). Benefit of the integration of semantic 3d models in a fire-fighting vr simulator. *Applied Geomatics*, 4(3):143–153.

Moya, S., Grau, S., and Tost, D. (2014). First-person locomotion in 3d virtual environments: a usability analysis. *J. UCS*, 20(7):1026–1045.

NASA (2016). Antrhropometry and biomechanics.

Orion (2016). Orion beta sw.

Safety, T. O. and Administration, H. (2016). Fire protection and prevention.

Segura, l., Moreno, A., Brunetti, G., and Henn, T. (2007). Interaction and ergonomics issues in the development of a mixed reality construction machinery simulator for safety training. In Dainoff, M., editor, *Ergonomics and Health Aspects of Work with Computers*, volume 4566 of *Lecture Notes in Computer Science*, pages 290–299. Springer Berlin / Heidelberg.

Sharma, R. P. and Verma, G. K. (2015). Human computer interaction using hand gesture. *Procedia Computer Science*, 54:721 – 727.

Unity3D (2016). Unity3d.

Yao, R., Heath, T., Davies, A., Forsyth, T., Mitchell, N., and Hoberman, P. (2014). Oculus vr best practices guide. *Oculus VR*.

Visual Computing Technologies to support the Operator 4.0

- Authors: Álvaro Segura, Helen V. Diez, Ander Arbelaiz, Jorge Posada and Alejandro García-Alonso
- Journal: Computers & Industrial Engineering
- Year: 2018
- Publisher: Elsevier

Visual Computing Technologies to support the Operator 4.0

Álvaro Segura[a], Helen V. Diez[a], Ander Arbelaiz[a], Jorge Posada[a], Alejandro García-Alonso[b]

[a] *Vicomtech, Mikeletegi 57,20009 San Sebastian,Spain*
[b] *Universidad del País Vasco UPV-EHU,Facultad de Informática,Paseo Manuel Lardizabal 1,20018 San Sebastian,Spain*

Abstract

Nowadays there is a clear trend for improving productivity and efficiency in the Industrial sector by integrating new advanced ICT technologies that are re-shaping the industrial production paradigms, as in the Industry 4.0 initiative. This new trend does not only affect production lines and machines but also operators. Markets demanding efficiency and flexibility would not be possible excluding the human-factor. Putting the operators in the centre of this new paradigm is mandatory for its success. The operators need to be empowered by giving them new tools and solutions for improving their decision-making processes. In this paper we show how Visual Computing technologies can play a key role in this empowering process, being therefore essential in the realization of the Operator 4.0 vision.

Keywords: Industry 4.0, Augmented Operator, Visual Computing, Digital Twin, Operator 4.0

1. Introduction

The Industry 4.0 initiative, described as the Fourth Industrial Revolution, started in Germany. It was defined by Kagermann et al. (2013) as the union of three key components in manufacturing environments: Internet of Things (IoT), Cyber-Physical Systems (CPS) and Smart Factories. Hermann et al. (2016) noted that while lot of academic and practical discussion on the topic of Industry 4.0 was made, the term was not concrete enough and different practitioners used slightly different concepts. They reviewed the literature to find the actual key components from the most frequent terms in related publications. Among the top terms found are *Cloud Computing*, *Big Data* and *Visual Computing*. Indeed, Posada et al. (2015) clearly defined *Visual Computing* as a key enabling technology for Industry 4.0.

Beyond the technological aspects, one key element in present and future manufacturing is the human operator. The Industry 4.0 concept does identify the human workforce as a relevant and active factor (Posada et al., 2015). Nevertheless, the integration of the human operator in the new *Smart Factories* poses ethical and societal challenges with scientific, technological and political implications (Manyika, 2016). Romero et al. (2016a) presented the concept of *Operator 4.0* to describe the worker in a Human Cyber-Physical System scenario, as an evolution of previous generations of operators. An Operator 4.0 is a *"smart and skilled operator who performs 'work aided' by machines if and as needed"*. In (Romero et al., 2016b) the authors go on to present different typologies of Operator 4.0 based on the technologies they work with. Following this typology classification this paper presents several practical developments we are carrying out where Visual Computing enables the *augmented operator*, the *virtual operator* and the *collaborative operator*.

The article is structured as follows. First the context and related works are reviewed in Section 2 followed by a description of our vision of Visual Computing concept and its relevance in enabling the Operator 4.0 in Section 3. Then five practical research and development use cases we have conducted or are working on are described and linked to our Operator 4.0 vision. Section 5 discusses the relevance of these works in the contribution of Visual Computing to operator enhancement. And finally conclusions are presented.

2. Related work

Thoben et al. (2017) presented an overview of Industry 4.0 regarding research issues and applications. The research issues in this work are structured in three main categories:

- Technological Issues, such as standards and interfaces, data analytics, data security, data quality, sensors and actuators (Zhang et al., 2017; Wuest et al., 2015).
- Methodological Issues, including reference models, visualization, service marketplaces or requirements engineering (Adolphs et al., 2015; Davis et al., 2012; Zhong et al., 2016).
- Business Case Issues, regarding privacy, investment or servitized business models. (Penzenstadler and Eckhardt, 2012; Boyer and Freyssenet, 1995).

The selected application scenarios show the wide scope of initiatives including new business models to offer the enhanced functionalities as services (Osterwalder and Pigneur, 2010), safe human-robot interaction (Khalid et al., 2017) or Cyber-Physical logistics Systems (Thoben et al., 2014). A central element of Industry 4.0 and its applications are Cyber-Physical Systems.

The term Cyber-Physical Systems (CPS) refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities (Baheti and Gill, 2011). They are systems of linked computational entities which are in close connection with the surrounding physical world and its on-going processes, providing and using, at the same time, data-accessing and data-processing services available on the Internet (Monostori et al., 2016). In the context of smart factories and Industry 4.0 CPS have many applications. Its current status and latest advancement in manufacturing have been analysed by Wang et al. (2015).

However, this work focuses on the operators of the future and the advances needed in human-machine interfaces (HMI) taking into account the evolution of smart factories in the Industry 4.0 era. Gorecky et al. (2014) point out a number of possible future HMI techniques and devices. Mobile devices such as smartphones, tablets and smart glasses will become essential. Mobile devices are Internet enabled and can both receive and process large amounts of information. Furthermore, mobile devices are equipped with high quality cameras and microphones, which again allow them to record and transmit information (Schmitt et al., 2013). Dialogue-driven voice controls and gesture recognition techniques will provide natural human-machine interaction enhancing the relationship between humans and Cyber-Physical Systems.

Romero et al. (2016a) define the concept of Human Cyber-Physical Systems (H-CPS) as systems engineered to improve human abilities and improve human physical sensing and cognitive capabilities by means of various technologies .

Posada et al. (2015) identified Visual Computing as a key enabling technology in Industry 4.0 and analysed the importance of its related technologies in modern digital manufacturing. Visual Computing can enhance human machine interaction in different levels of smart production (Diaz et al., 2015).

Visual Computing technologies such as Computer Vision have been traditionally applied to dimensional or surface defects identification inside automated manufacturing lines. Nowadays this technology is being extended to monitoring operators while operating in scenarios or environments such as collaborative robotics or *co-bots* (Mohammed et al., 2016). Most industrial robots today operate behind safety fences. With the introduction of co-bots these barriers have started to be removed. However, these types of robots are still not fully safe for the operator and additional actions need to be taken such as adding external or on-robot cameras to monitor operator movements and avoid risks. The availability of new sensors able to deliver high-speed full 3D information along with the development of new 3D real-time image processing are fostering the adoption of these types of solutions in the shop-floors.

Today's markets are demanding better production flexibility in order to support higher levels of product customization. These new trends (Martin, 2017) lead to the need of the integration of new flexible mechanisms to allow continuous training for workers. Fast changes in production lines for satisfying these demands are usually too fast for the operators to be fully efficient since the first production batch. In this regards, new tools need to be delivered to the operators in order to assist them during complex tasks such as new product assembly, new product quality assurance tests or new maintenance operations. Virtual reality is an appropriate technology to deploy continuous training procedures in the shop-floor, allowing the operators to learn beforehand unseen production situations. Also, virtual reality is a very suitable technology for training safety protocols or simulating dangerous operations (Matsas and Vosniakos, 2017).

At the core of both Industry 4.0 and Industrial Internet approaches there is the need to fully integrate and communicate every production entity all along the product-life-cycle, from product design to product operation. This vision can be supported by the so-called Industrial Internet of Things (IIoT) (Xu et al., 2014). In the IIoT scenario, where there is a close connection between physical world and digital world the paradigm of *Cyber-Physical Equivalence* or *Digital Twin* (NASA, 2012) emerges. This represents a seamless integration between both worlds meaning that the digital part can virtually replicate the behaviour of the physical counterpart, exchanging information indistinctly in both directions, from real in the field to digital and vice versa. This seamless integration allows the generation of new added value services such as remote monitoring, predictive maintenance (Lee et al., 2013), adaptive control strategies (Wang et al., 2016) or real-time data driven simulation (Xu et al., 2016).

3. Visual Computing for Operator 4.0

Visual imagery constitutes one of the most important sensory inputs for humans. This feature can be summarized in a very well-known sentence: *"An image is worth a thousand words"*. The term Visual Computing refers to a set of technologies that process or generate visual content or visual information. To cite some of the technologies that could fit under the Visual Computing term we can enumerate Virtual Reality, Augmented Reality, Computer Vision and Visual Analytics.

Fig. 1 illustrates the general idea of this paper: operators' tasks, in the inner ring are supported by Visual Computing technologies in the outer ring surrounding the worker, which is at the centre. The worker is thus empowered by Visual Computing to perform better or make decisions with stronger criteria. We have selected the following task categories as representative of the work carried out by operators in manufacturing environments.

1. Assembly
2. Maintenance
3. Quality control
4. Training
5. Inventory
6. Machine operation

From the set of Visual Computing technologies discussed by Posada et al. (2015), we have selected those we consider relevant and potentially beneficial to the operator. These are the technologies supporting the Operator 4.0:

Virtual Reality. Permits the vision and virtual use of elements out of reach, enabling safe use of hazardous equipment and enhanced learning of procedures.

Augmented Reality. Augments the workplace with relevant information not normally visible and useful for the work required at the moment.

Visual Analytics. Provides easy to understand depictions of large amounts of data and relations that are not immediately seen.

Collaborative Robotics Interaction. Co-bots enhance workers manual abilities enabling more precise or force-requiring operations to be performed.

HMI Interfaces. Act as the main interface between the worker and the supporting automation systems.

Media/Social Network. Industrial social networks allow communication between workers or between them and management increasing worker satisfaction and transfer of experience.

The above brief descriptions have in common a potential to enhance the workers' senses and abilities. They stand between the worker and the environment providing an interface to the manufacturing systems and bringing knowledge that would be hidden or hard to reach. With a proper application of Visual Computing, workers can understand the behaviour of a manufacturing system and act on it or carry out appropriate maintenance, they can perform assembly or maintenance procedures without remembering detailed instructions for each particular product, etc.

The Digital-Twin concept would not be possible without Visual Computing. A 2D/3D visual representation of the real counterpart is mandatory in order to support services such as simulation. Visual representation of simulation results can be combined with real-time production data into monitoring dashboards in HMIs as a powerful Visual Analytics tool for improving decision-making. As described

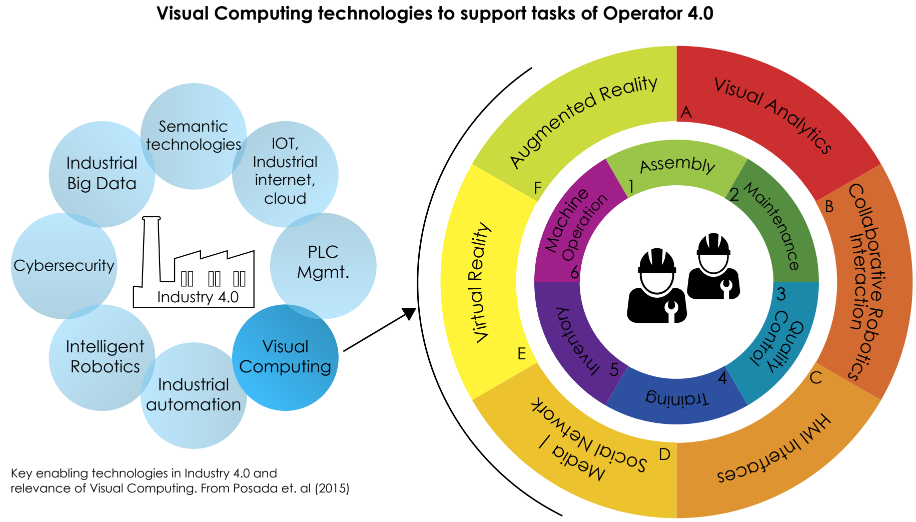


Figure 1: Left: General outline of Visual Computing technologies improving Operator 4.0 tasks. The application of Visual Computing advanced technologies in specific tasks empowers the operator to better address real industrial problems in Industry 4.0 scenarios.

by Xu et al. (2017) with the introduction of Industry 4.0 and the high connectivity between production entities, there is an unprecedented integration of sensing, computing and control.

In the next Section, we present five examples of research and development we have done in the past few years or we are currently developing that highlight the suitability of Visual Computing technologies enhancing industrial operators.

4. Use cases

The example use cases described below show how the Operator 4.0 is supported by Visual Analytics, HMIs, VR and AR in several different practical scenarios. As secondary support technologies, collaborative robots and social networks also play a role. Each particular scenario may involve more than one task category, and can benefit from more than one technology in combination.

4.1. Operator supported by Visual Analytics

This use case highlights how Visual Analytics integrated in a plant's HMIs can enhance the cognitive process of supervising a manufacturing line and applying corrective maintenance. In the context of Fig. 1, this scenario involves tasks dealing with machine operation and maintenance.

As introduced by Romero et al. (2016a), a new figure is now emerging named *augmented operator*. This type of operator will have access to the data coming from the machines and sensors in the shop-floor as well as information coming from other entities such as Manufacturing Execution Systems (MES) or even information such as lesson learned or best practices from other operators in the shop-floor or outside the shop-floor from professional social-networks. Nowadays there is a clear increase in data gathering and data flows through the manufacturing plants or factories. All this information needs to be delivered to the operator adapted, contextualized and transformed to make it easily understandable so that decisions can made quickly and reliably.

One clear way to transform or contextualize data is through visualization. We have implemented a novel Visual Analytics solution targeted at the application domain of Big Data analytics in manufacturing industries. We propose a new interactive visualization metaphor that enables operators to easily detect faulty production situations or define abnormal production conditions by evaluating the correlation of several production variables with respect to a key performance indicator (KPI) such as the Overall-Equipment-Efficiency (OEE), as shown in the screenshot of Fig. 2.

The interactive analytics tool provides information about the correlation of manufacturing varia-

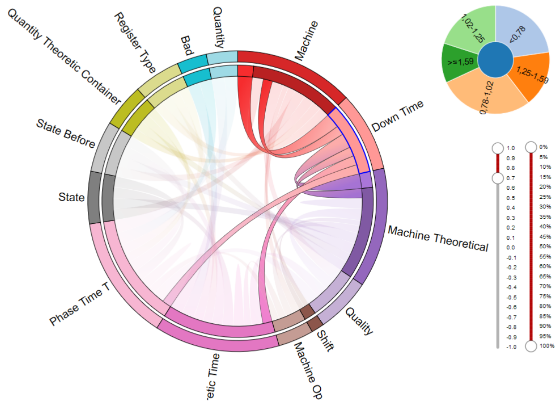


Figure 2: Visual analytics for OEE

bles. The user can i) select a span of OEE values by clicking in the upper-right gizmo, ii) select a range of values based on the importance with the first slider and iii) leave out a given percentage of values with the second slider.

The left part of Fig. 2 is updated dynamically and instantaneously whenever the user interacts with the graph. It shows a two-level hierarchical chord graph, being the outer ring the main categories and the inner ring the variables within that category. Clicking on any category or variable highlights the corresponding correlations with other variables in the graph. The colours, width and transparency are visual cues to the operator highlighting the importance of the correlation between the variables. Additionally, the user can click on any chord in order to retrieve additional textual information about the real data behind the link.

The analytics tool presents an innovative method that improves the cognitive process involved in understanding and getting comprehension of the manufacturing data. The tool enables the operator to analyse the data in a quick and interactive way, providing more insight from the data in less time and thus, improving decision-making efficiency.

The current version of the presented Visual Analytics tool is part of a collection of tools aimed at the data analysis in real-world factories.

4.2. Operator supported by Virtual Reality

This use case presents a Virtual Operator, following the definition given by Romero et al. (2016b), as the collaboration of an Operator with Virtual Reality (VR). The tasks relevant here are machine operation and training, while the enabling technologies are VR as HMI and collaborative robotics interaction.

There is a trend on Visual Computing related to virtual characters. These virtual characters are often used in training applications of various disciplines: medicine, safety, education, manufacturing, military (Kenny et al., 2007a; Diez et al., 2016; de Antonio et al., 2005; Kenny et al., 2007b; Wilson, 2008).

The introduction of virtual characters in VR applications helps to empathize with the situation and enhances the user experience. The point of view of the user regarding the virtual character can be in first person (Petersen et al., 2013) or in third person. The first person point of view offers a really immersive experience in which the user can interact directly with the environment, and the third person point of view allows the user to perceive a global perspective of the situation in which the virtual character can be representing the user himself or other workers, this could be applied in surveillance and supervision tasks.

In this work we have developed a scenario in which a collaborative robot picks up parts from the working environment. Then it places them in the platform of a 3D scanner or in a conveyor. From time to time, an operator replaces that part with another. In these cases, the robotic arm must detect the presence of the arm or hand of the operator and stop its movement to avoid the collision. Since trying out this situation with humans could be dangerous, we have developed a training program which allows the operator to work in a safe way.

In order to reduce risks during training, Digital Twins of the robot and the grasping device are used. Those Digital Twins are linked in real time with their real counterparts, providing real physics and synchronized behaviour to the Digital Twins.

As VR devices the Oculus Rift HMD and the Leap Motion sensor are used. Fig. 3 shows the interconnections between the real world, the virtual world and the virtual operator.

The developed application is based on *WebVR* and allows the real operator to visualize the virtual scene from two points of view as mentioned above (see Fig. 4).

In the first person point of view mode, the virtual operator can visualize the real movements of the robot as the robotic arm and its Digital Twin are connected. The operator views the scene in an

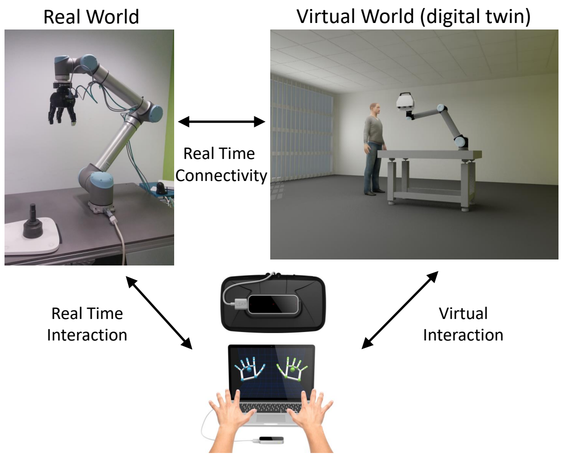
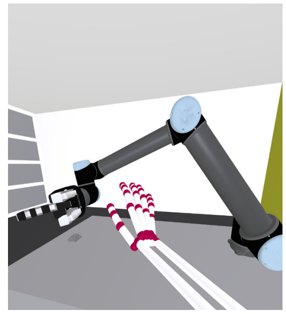
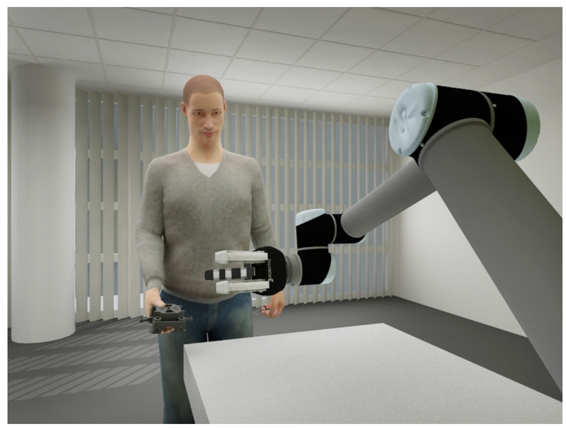


Figure 3: Setup for a VR training scenario involving a robot Digital Twin.



(a) First person mode virtual view.

(b) First person mode user view.



(c) Third person mode virtual character interacting with Digital Twin.

Figure 4: Virtual Operator using Oculus Rift and Leap Motion Controller.

immersive way through the HMD. The Leap Motion gesture sensor is attached to the HMD and allows real time tracking of the hands of the operator. The operator can introduce his hands in the virtual scene and, when the Digital Twin notices the presence of the operator, it automatically stops and therefore the real robot also stops.

In the third person point of view mode, the operator can view a Virtual Character interacting with the Digital Twin. This animated agent is endowed with natural behaviour and can perform realistic movements. The animated agent is also capable of reproducing facial expressions. This case is work in progress, but we believe this third person mode can be helpful for training and surveillance applications.

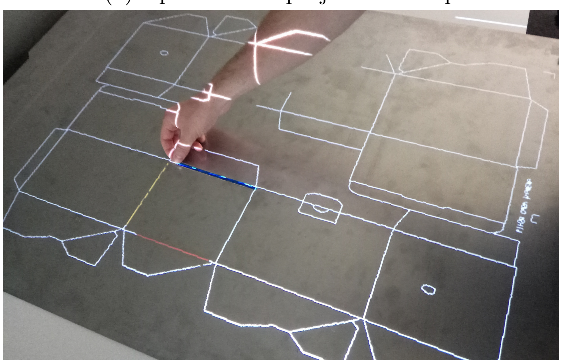
In both cases, the safety for the trainee is guaranteed since the interaction is purely virtual, but with real physical behaviour. The operator will learn very quickly and safely how the robotic arm behaves, reducing the stress produced in the initial interactions with the real robotic arm.

This system is currently demonstrated in a lab environment.

4.3. Operator supported by Augmented Reality

In this use case machine operation tasks are enhanced by AR technology in the form of projection mapping. In cooperation with a local packaging company, we developed a projection mapping system to assist workers when they have to tune the die cutter. These machines are usually misaligned after use, not exerting a uniform pressure in all their area of work, and therefore, the workers have to fit tapes of different thickness in certain locations to balance the pressure differences. Nowadays, the operator knows the state of each reference by accessing the store and consulting a large sheet of paper that contains the configuration of the tapes. Thus, instead of forcing the worker to waste time going to the store and looking for a sheet, and to interpret the information in the sheet to extrapolate it to the machine (increasing the mental load, and consequently, increasing the fatigue and the chance of making mistakes), the proposed system projects directly onto the machine the tapes to be placed and their positions, simplifying the task to the worker. We can think of the worker as an Augmented Operator because this projection mapping is a type of Augmented Reality. In Fig. 5 (a) we can see the die cutting machine and the projector at the top-right corner.

(a) Operator and projection set-up



(b) Projection close-up

Figure 5: Die cutter tuning using a projection mapping system that indicates which tapes to place (coded by a colour) and where.

Projection mapping can be defined as the act of projecting virtual content properly adjusted to the surfaces of the objects that are present in the real world using a video projector. Beyond traditional hand-held displays or HMDs, projection mapping detaches the display device from the user and integrates it into the environment, so it is also referred to as Spatial Augmented Reality in a more general sense (Bimber and Raskar, 2005). In an industrial environment, this kind of technology presents several advantages, such as the freedom of the worker's hands (they do not have to hold the screen) or the lower cognitive load of the worker, who does not have to look up and interpret information on a screen, but visualizes it directly on the target object. To ensure that projected content is correctly adapted to the target surfaces (*geometric registration*) a calibration step is necessary. In some scenarios a manual procedure is used, but this has the drawback that the projection will become misaligned if there is a minimal change in the scene (the projector or the target object move). Advanced projection mapping systems incorporate a camera that is able to record any change and automatically readjust the projection (Knibbe et al., 2015).

Our system incorporates a camera to automatically recognize the working area of the die cutter (the target object) and to adjust the projection to project the tapes into their correct position each time the reference is changed. The operator performs a camera-to-projector calibration in a preliminary off-line step with the simple push of a button in the HMI. Then, during the on-line step, the camera captures images of the scene and locates the target object by detecting special markers on it using Computer Vision techniques. Given the calibration and the location of the object, a projection model is applied to display the virtual content properly adjusted to the die. Fig. 5 (b) shows the projected die cutting pattern including the places where the operator has to put compensating tapes (in different colours).

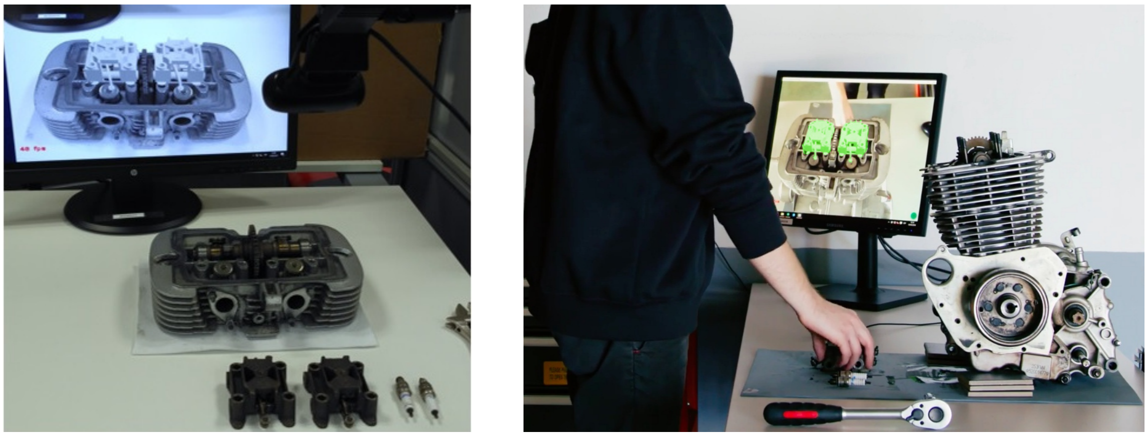
The system here described is successfully deployed in an actual factory.

4.4. Operator supported by Mixed Reality

In this use case showcases HMI interfaces based on AR applied to tasks involving assembly, maintenance or training. In this case, the type of AR used is more specifically Mixed Reality, which adds virtual 3D objects in real time on a real environment.

Manual assembly lines often contain multiple work stations in which one or more workers perform a set of tasks. These tasks may consist of several steps of different complexity and are not always repetitive, but may vary due to variations in the products manufactured. In this setting, an Augmented Operator can leverage AR technology to enhance his/her performance. Similarly, Mixed Reality using *head-mounted displays* has been deployed successfully in recent years in applications such as elevator maintenance, like Thyssenkrupp's operator assistance systems based on Microsoft's HoloLens (Achatz, 2017) and manual welding training, as reported by Okimoto et al. (2015) using the Soldamatic simulator. The latter example is highly immersive and can almost be considered VR.

The use case here presented has been developed for the automotive industry dealing with assembling or maintaining engines (See Fig. 6 (a)). The



(a) Preparation of augmented instructions　　(b) Augmented operator in the workshop

Figure 6: AR-assisted assembly of an engine

idea is to facilitate the training of newcomers or to assist operators through the assembly process, avoiding the need of physical manuals. Cognitive load is reduced by placing the elements of the assembly process directly in his perspective of the surrounding environment. This way the performance is not dependent on the operator remembering all operations required for a particular product.

The AR system has been specially designed using markerless technologies (Teichrieb et al., 2007) in order to avoid polluting the station with artificial markers and interfering with industrial critical processes. Any kind of media content can be displayed to the operator with stable and real time rendering. Also, the system has been conceived as an easy-to-use and inexpensive tool. Thus, the elements involved are just a web camera and a computer with a monitor. With those few elements the system can perform both main processes required for the application, described below.

The first process is performed off-line and its goal is to reconstruct an accurate 3D model of the surrounding environment where virtual elements will be superimposed (i.e. the object to be tracked). The inputs needed to perform the reconstruction are a set of images of the environment from different points of view and the intrinsic parameters of the camera used. From the set of images, the system extracts 2D features (Trajković and Hedley, 1998) and matches theirs descriptors (Alahi et al., 2012) between the different images to estimate the 3D position of those features (Dellaert et al., 2000). The output of this process is a 3D point cloud of the features.

The second process is performed on-line and tries to track the 3D object through the incoming images matching them with the information stored in the 3D point cloud. The result of this detection process is an initial pose of the camera with respect to the environment (i.e. the point of view of the operator). Once the first camera pose is obtained, less computationally expensive tracking techniques (Barrena et al., 2015) are used to follow movements of the operator through the environment until tracking is eventually lost. Then, the system has to find a new camera pose, repeating the pipeline in a cyclic way. Having an estimated camera pose with the current assembly operation is shown as animated virtual objects over the live real image on the monitor (as seen in green in Fig. 6 (b)). The operator has to repeat on the physical product the behaviour of the animated objects.

This application is currently available for demonstration in a lab environment.

4.5. Augmented assistance of disabled operators

This last use case describes ongoing work where a broader set of technologies are being developed and combined into a platform aimed at enabling human-automation symbiosis. The Operator 4.0 type here will not be one, but a combination of an augmented operator, a virtual operator, a collaborative operator and a social operator. This project aims at optimizing workplaces and human-automation load balance through the application of key enabling technologies to support these operators. Our work focuses on empowering operators with disabilities in assembly tasks. The main enabling technology in this case is AR, but Collaborative Robots and Industrial Social Networks will also play a role in increasing worker performance and satisfaction.

The specific use case for development and demonstration centres on an electrical cabinet assembly line where intellectually disabled operators have to wire the different components. One key aspect in this scenario is the preparation of digital material describing the wiring process. Currently, experienced engineers prepare paper-based simplified documentation from the original schematics.

The system in development is a distributed software framework that includes authoring components and visualization and interaction components supporting several input and output devices. The authoring components allow importing existing assembly instructions and interactively preparing an augmented interactive presentation that will guide operators step by step. This preparation involves defining a sequence of instructions based on the imported data, specifying where visual elements will be displayed and writing rules that will govern the process.

Fig. 7 shows the current experimental set-up that includes two projectors, two cameras and a tablet. One visible light camera captures a view of the workspace and tracks the position and orientation of the cabinet while an additional depth camera assists in tracking and user gesture recognition. Having two projectors permits projecting images on surfaces that would be hard to reach without excessive occlusions by a single projector such as vertical and horizontal surfaces.

Wiring a cabinet involves inserting wires in small holes and tightening screws in the right order. Each wire is labelled and corresponds to a pair specific terminals, as dictated by the documentation. Unambiguously highlighting the correct holes to connect each wire in a cabinet that is not in a fixed position would require a tracking system with a precision not currently achievable. Therefore, the assembly instructions do not try to project on precise positions. Instead, the system projects virtual 3D representations of the cabinet components on the workspace and shows an animated screwdriver on the terminals to connect in the current step. Simultaneously, a video depicting the procedure for the current step is played.

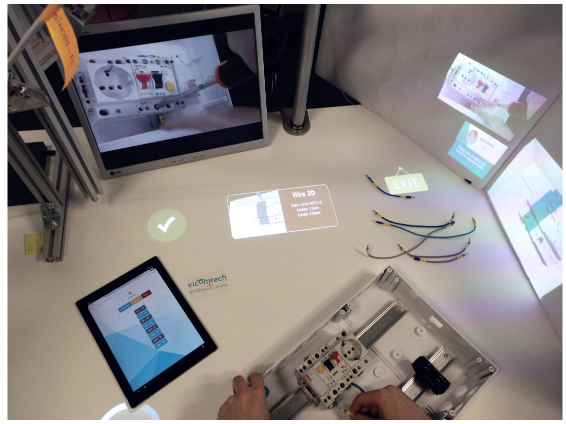


Figure 7: Assembly operator assistance set-up for electric cabinet wiring.

At the moment, initial in-lab tests are being conducted. This will result in refinements to what and how is visually presented to enhance assembly performance. Later, tests with actual workers with different disabilities will be carried out and the AR framework will be integrated with other systems including collaborative robots for further assistance to physically disabled workers.

5. Discussion

The previous Section has shown how the ideas introduced in Section 3 have been put into practice in practical scenarios. Each scenario shows how a combination of Visual Computing technologies enhance industrial operators in particular types of tasks. They establish links between the inner and outer rings of Fig. 1 that support the Operator 4.0. Table 1 summarizes the Visual Computing technologies used in each use case, and the current status of these developments, whether they are in-lab developments or already deployed in industrial environments.

Visual Computing includes technologies such as Virtual Reality, Augmented Reality, Computer Vision and Visual Analytics. Not all are relevant in empowering operators. Even though Computer Vision contributes indirectly as part of AR, pure computer vision systems are mainly important in manufacturing automation and quality assurance. The selection of technologies made in this paper has shown to be relevant in enhancing the capabilities of human operators. It is important to note that the technologies shown do not act in isolation:

Table 1: Summary of technologies used and current state of the use cases presented

Use case	Technologies	Status
4.1	Visual Analytics, HMI Interfaces	Deployed
4.2	Virtual Reality, Collaborative Robots	Lab
4.3	Augmented Reality, HMI Interfaces	Deployed
4.4	Augmented (Mixed) Reality, HMI Interfaces	Lab
4.5	Augmented Reality, Industrial Social Networks, Collaborative Robots	Lab

they work in connection with other digital and physical parts of the plant environment.

In Section 4.1 we have seen how Visual Analytics can give an operator insight into the complexities of performance tuning a manufacturing line at a quick glance.

In Section 4.2 immersive Virtual Reality is demonstrated as an alternative way of enhancing perception of the operator and operating machines. Deploying this kind of solution in a manufacturing environment can raise safety concerns due to the user being largely isolated from its near surroundings. Care must be taken in this case to make sure operation is safe. However, this set-up allows a safe tele-operation where the operator can stand in a safe room while remotely monitoring and operating potentially hazardous equipment. The case of immersive Mixed Reality based on head-mounted displays (as in Achatz (2017) and Okimoto et al. (2015)) may be less problematic due to the partial view of the real environment but also requires care if used in industrial scenarios.

Three of the use cases described leverage AR but employ different technical approaches. We can see how AR can enhance the performance of the operator in two different ways. First, the knowledge required by the operator, e.g. the know-how about the procedures to assemble different products, is lowered. The system stores that information and displays it when needed and directly in the field of view of the worker. Two approaches have been shown: Mixed Reality in Section 4.4 and projection-based AR in Section 4.5). These systems allow new inexperienced workers to quickly start performing tasks. In the case of intellectually disabled workers the cognitive support provided by AR is of particular importance.

The system described in Section 4.3 demonstrates additional benefits. In this application the calibration of the relation between projector and projection plane is crucial because the position of projected markings requires high precision. Such precision is not generally required in Augmented Reality systems displaying information, hints or notifications. This system has been deployed in an actual manufacturing plant and is saving time, physical storage space and costs.

Similarly to the Industry 4.0 where CPS connect physical and digital worlds, the figure of the Operator 4.0 emerges and can succeed when the above technologies are combined in suitable ways and connect the operator to the Cyber-Physical environment. We argue that Visual Computing provides ways to enable this connection.

6. Conclusions

In this article, we have presented a conceptual framework, supported by specific practical examples, showing the application of different Visual Computing technologies in industrial operator tasks. The application of these technologies can empower operators in the context of Industry 4.0 scenarios. Visual Analytics, Augmented Reality, Virtual Reality, HMI interfaces, Media/Social Networks and Collaborative Robotics Interaction are some of the most relevant technologies for the new Operator 4.0, since they not only improve productivity and efficiency, but are essential to tackle the social, inclusion and interaction aspects that are central to these new socio-technical systems. The application of Visual Computing technologies can contribute decisively to the enhancement of operator ability to perform traditional tasks, and to the definition of new tasks and scenarios.

The selected cases that we present are good real-world examples of how a “Visual Computing enhanced” Operator 4.0 can have a central role in future industrial production scenarios. The Overall-Equipment-Efficiency case shows how a proper visual analytics interactive platform can help the Operator (as well as other roles, such as managers) to better understand and easily detect faulty production situations. The Robotic Digital-Twin for Virtual Operators shows how an operator can be trained in collaborative robotics interaction to be prepared (technically and psychologically) for a novel interaction with machines and robots. The projection-based AR system supports operators in a critical maintenance application for machines in the paper industry. In another example, we show how a Mixed Reality fast pipeline system can be prepared with a set of images to build AR support for assembly operations based on robust markerless tracking. Finally, the system being developed in the European R&D Project MANUWORK will help industries employing operators with disabilities by complementing their intellectual or physical limitations. Hence, we are directly addressing the need of adaptation and balancing of new production forms with the enhancement of the skills and abilities of the operators. Future work includes the measurement and analysis the actual impact of these different examples in the factory, including both productivity/efficiency and social/psychological aspects.

Acknowledgements

This work has been partially funded by the research projects LANA and LANA II of the Basque public agency SPRI. The OEE Visual Analytics example is part of the R&D project Thinking Factory, partially funded by the Basque Government's HAZITEK programme. The "Basque Industry 4.0" programme also funds part of the research for the projection-based AR system. The MANUWORK EU Project is partially funded by the European Commission (Grant Agreement no. 723711).

References

Achatz, R. (2017). Digital Transformation at thyssenkrupp: Challenges, Strategies and Examples", booktitle="Advanced Information Systems Engineering. Springer International Publishing, Cham, pp. 3–12.

Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., Heidel, R., Hoffmeister, M., Huhle, H., Kärcher, B., et al. (2015). Reference Architecture Model Industrie 4.0 (RAMI 4.0). VDI/VDE Society Measurement and Automatic Control (GMA).

Alahi, A., Ortiz, R., Vandergheynst, P. (2012). Freak: Fast retina keypoint. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. Ieee, pp. 510–517.

Baheti, R., Gill, H. (2011). Cyber-physical systems. The impact of control technology 12, 161–166.

Barrena, N., Sánchez, J. R., García-Alonso, A. (2015). Wapt-surface normal estimation for improved template matching in visual tracking. In: VISAPP (3). pp. 496–503.

Bimber, O., Raskar, R. (2005). Spatial Augmented Reality: Merging Real and Virtual Worlds. A. K. Peters, Ltd., Natick, MA, USA.

Boyer, R., Freyssenet, M. (1995). The emergence of new industrial models. Actes du GERPISA 15, 75–144.

Davis, J., Edgar, T., Porter, J., Bernaden, J., Sarli, M. (2012). Smart manufacturing, manufacturing intelligence and demand-dynamic performance. Computers & Chemical Engineering 47, 145–156.

de Antonio, A., Ramírez, J., Imbert, R., Méndez, G., Aguilar, R. A. (2005). A software architecture for intelligent virtual environments applied to education. Revista Facultad de Ingeniería-Universidad de Tarapacá 13 (1), 47–55.

Dellaert, F., Seitz, S. M., Thorpe, C. E., Thrun, S. (2000). Structure from motion without correspondence. In: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Vol. 2. IEEE, pp. 557–564.

Diaz, J., Posada, J., Barandiaran, I., Toro, C. (2015). Recommendations for sustainability in production from a machine-tool manufacturer.

Diez, H. V., García, S., Mujika, A., Moreno, A., Oyarzun, D. (2016). Virtual training of fire wardens through immersive 3d environments. In: Proceedings of the 21st International Conference on Web3D Technology. ACM, pp. 43–50.

Gorecky, D., Schmitt, M., Loskyll, M., Zühlke, D. (2014). Human-machine-interaction in the Industry 4.0 era. In: Industrial Informatics (INDIN), 2014 12th IEEE International Conference on. IEEE, pp. 289–294.

Hermann, M., Pentek, T., Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In: 49th Hawaii International Conference on System Sciences (HICSS). pp. 3928–3937.

Kagermann, H., Wahlster, W., Helbig, J. (2013). Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 – Securing the Future of German Manufacturing Industry. Final report of the industrie 4.0 working group, acatech – National Academy of Science and Engineering, München.

Kenny, P., Parsons, T. D., Gratch, J., Leuski, A., Rizzo, A. A. (2007a). Virtual Patients for Clinical Therapist Skills Training. In: Intelligent Virtual Agents. Springer Berlin Heidelberg, pp. 197–210.

Kenny, P. G., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S. C., Piepol, D. (Nov. 2007b). Building Interactive Virtual Humans for Training Environments. In: Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).

Khalid, A., Kirisci, P., Ghrairi, Z., Pannek, J., Thoben, K.-D. (2017). Safety requirements in collaborative human–robot cyber-physical system. In: Dynamics in Logistics. Springer, pp. 41–51.

Knibbe, J., Benko, H., Wilson, A. D. (2015). Juggling the effects of latency: Software approaches to minimizing latency in dynamic projector-camera systems. In: Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. UIST '15 Adjunct. ACM, New York, NY, USA, pp. 93–94.
URL `http://doi.acm.org/10.1145/2815585.2815735`

Lee, J., Lapira, E., Bagheri, B., Kao, H.-a. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters 1 (1), 38–41.

Manyika, J. (2016). Technology, jobs, and the future of work.

URL http://www.mckinsey.com/global-themes/employment-and-growth/technology-jobs-and-the-future-of-work
Martin, P. (2017). Empowering the Workforce to Control the Future of Automation. HMI & Operator Interfaces in Factory Automation.
URL http://www.automation.com/portals/factory-discrete-automation/hmi-operator-interfaces/empowering-the-workforce-to-control-the-future-of-automation
Matsas, E., Vosniakos, G.-C. (2017). Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. International Journal on Interactive Design and Manufacturing (IJIDeM) 11 (2), 139–153.
Mohammed, A., Schmidt, B., Wang, L. (2016). Active collision avoidance for human–robot collaboration driven by vision sensors. International Journal of Computer Integrated Manufacturing, 1–11.
Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K. (2016). Cyber-physical systems in manufacturing. CIRP Annals-Manufacturing Technology 65 (2), 621–641.
NASA (2012). The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles.
Okimoto, M. L. L., Okimoto, P. C., Goldbach, C. E. (2015). User Experience in Augmented Reality Applied to the Welding Education. Procedia Manufacturing 3, 6223 – 6227, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.
Osterwalder, A., Pigneur, Y. (2010). Business model generation: a handbook for visionaries, game changers, and challengers. John Wiley & Sons.
Penzenstadler, B., Eckhardt, J. (2012). A requirements engineering content model for cyber-physical systems. In: Requirements Engineering for Systems, Services and Systems-of-Systems (RES4), 2012 IEEE Second Workshop on. IEEE, pp. 20–29.
Petersen, N., Pagani, A., Stricker, D. (2013). Real-time modeling and tracking manual workflows from first-person vision. In: Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on. IEEE, pp. 117–124.
Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., de Amicis, R., Pinto, E. B., Eisert, P., Dollner, J., Vallarino, I. (2015). Visual Computing as a Key Enabling Technology for Industrie 4.0 and Industrial Internet. IEEE Computer Graphics and Applications 35 (2), 26–40.
Romero, D., Bernus, P., Noran, O., Stahre, J., Fast-Berglund, Å. (2016a). The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation towards Human-Automation Symbiosis Work Systems. In: APMS (Advances in Production Management Systems).
Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., Gorecky, D. (2016b). Towards an Operator 4.0 Typology: a human-centric perspective on the fourth industrial revolution technologies. In: International Conference on Computers & Industrial Engineering CIE 46.
Schmitt, M., Meixner, G., Gorecky, D., Seissler, M., Loskyll, M. (2013). Mobile interaction technologies in the factory of the future. IFAC Proceedings Volumes 46 (15), 536–542.
Teichrieb, V., do Monte Lima, J. P. S., Apolinário, E. L., de Farias, T. S. M. C., Bueno, M. A. S., Kelner, J., Santos, I. H. (2007). A survey of online monocular markerless augmented reality. International Journal of Modeling and Simulation for the Petroleum Industry 1 (1).
Thoben, K.-D., Pöppelbuß, J., Wellsandt, S., Teucke, M., Werthmann, D. (2014). Considerations on a lifecycle model for cyber-physical system platforms. In: IFIP International Conference on Advances in Production Management Systems. Springer Berlin Heidelberg, pp. 85–92.
Thoben, K.-D., Wiesner, S. A., Wuest, T. (2017). Industrie 4.0 and Smart Manufacturing - A Review of Research Issues and Application Examples. International Journal of Automation Technology 11 (1).
Trajković, M., Hedley, M. (1998). Fast corner detection. Image and vision computing 16 (2), 75–87.
Wang, L., Törngren, M., Onori, M. (2015). Current status and advancement of cyber-physical systems in manufacturing. Journal of Manufacturing Systems 37 (Part 2), 517–527.
Wang, S., Wan, J., Zhang, D., Li, D., Zhang, C. (2016). Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination. Computer Networks 101, 158–168.
Wilson, C. (2008). Avatars, virtual reality technology, and the US military: Emerging policy issues. Library of Congress Washington DC Congressional Research Service.
Wuest, T., Wellsandt, S., Thoben, K.-D. (2015). Information quality in plm: A production process perspective. In: IFIP International Conference on Product Lifecycle Management. Springer, pp. 826–834.
Xu, J., Huang, E., Hsieh, L., Lee, L. H., Jia, Q.-S., Chen, C.-H. (2016). Simulation optimization in the era of Industrial 4.0 and the Industrial Internet. Journal of Simulation 10 (4), 310–320.
URL http://dx.doi.org/10.1057/s41273-016-0037-6
Xu, L. D., He, W., Li, S. (Nov 2014). Internet of Things in Industries: A Survey. IEEE Transactions on Industrial Informatics 10 (4), 2233–2243.
Xu, P., Mei, H., Ren, L., Chen, W. (Jan 2017). ViDX: Visual Diagnostics of Assembly Line Performance in Smart Factories. IEEE Transactions on Visualization and Computer Graphics 23 (1), 291–300.
Zhang, Y., Ren, S., Liu, Y., Si, S. (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of Cleaner Production 142, 626–641.
Zhong, R. Y., Lan, S., Xu, C., Dai, Q., Huang, G. Q. (2016). Visualization of rfid-enabled shopfloor logistics big data in cloud manufacturing. The International Journal of Advanced Manufacturing Technology 84 (1-4), 5–16.

3D model management for e-commerce

- Authors: Helen V. Diez, Álvaro Segura, Alejandro García-Alonso and David Oyarzun
- Journal: Multimedia Tools and Applications
- Year: 2017
- Publisher: Springer

Multimed Tools Appl (2017) 76:21011–21031
DOI 10.1007/s11042-016-4047-1

3D model management for e-commerce

Helen V. Diez[1] · **Álvaro Segura**[1] · **Alejandro García-Alonso**[1] · **David Oyarzun**[1]

Received: 27 April 2016 / Revised: 22 September 2016 / Accepted: 5 October 2016 /
Published online: 19 October 2016

Abstract This paper contributes to the efficient visualization and management of 3D content for e-commerce purposes. The main objective of this research is to improve the multimedia management of complex 3D models, such as CAD or BIM models, by simply dragging a CAD/BIM file into a web application. Our developments and tests show that it is possible to convert these models into web compatible formats. The platform we present performs this task requiring no extra intervention from the user. This process makes sharing 3D content on the web immediate and simple, offering users an easy way to create rich accessible multiplatform catalogues. Furthermore, the platform enables users to view and interact with the uploaded models on any WebGL compatible browser favouring collaborative environments. Despite not being the main objective of this work, an interface with search engines has also been designed and tested. It shows that users can easily search for 3D products in a catalogue. The platform stores metadata of the models and uses it to narrow the search queries. Therefore, more precise results are obtained.

Keywords Cad · BIM · 3D model conversion · WebGL technology

Electronic supplementary material The online version of this article (doi:10.1007/s11042-016-4047-1) contains supplementary material, which is available to authorized users.

✉ Helen V. Diez
hdiez@vicomtech.org

Álvaro Segura
asegura@vicomtech.org

Alejandro García-Alonso
alex.galonso@ehu.es

David Oyarzun
doyarzun@vicomtech.org

[1] Vicomtech-IK4 Research Center, Paseo Mikeletegui 57, 20009 San Sebastian, Spain

1 Introduction

People involved in e-commerce do not need to have technical background regarding CAD or BIM data. However, they do need to create and manage online catalogues to offer their products and allow potential customers to view and interact with them.

Embedding CAD/BIM data into web-based repositories has a number of challenges in contrast to common element types such as text and images.

- CAD/BIM file formats are not included in web standards and browsers do not directly support them.
- There is no standard 3D model file format that can be embedded in web pages, unlike image or video formats.
- The WebGL 3D rendering API can render polygonal models (triangle meshes), but CAD/BIM models are usually not composed of polygons but are defined in terms of higher complexity parametric shapes.
- When parametric CAD models are transformed into polygonal meshes, the result often contains a large number of polygons and the file is too large for an efficient use by a web client.

The main challenges for using such content on the web are thus to transform CAD/BIM files into formats suitable for web-based handling and rendering, and to present this content to client viewers in an efficient and user-friendly repository. That is, from the point of view of the CAD/BIM designer it has to be easy to add content to the repository. And from the point of view of the visitor, searching, accessing and viewing content has to be easy and quick. The main bottleneck for the last challenge is transfer time. This can be reduced with mesh simplification algorithms, but at the expense of visual quality. The system will have to find a compromise between shape quality and file size (i.e. transfer time).

In order to deal with the above challenges, the integration of CAD/BIM data involves solving several non-trivial issues:

- **Polygonal conversion:** Convert non-polygonal models into a polygonal model. For example, curved surfaces into triangle meshes.
- **Conversion to web compatible file formats:** Convert different native CAD and BIM file formats into a format that can be interpreted and rendered in a web page script.
- **Polygonal mesh simplification and Semantics:**

- Reduce the number of polygons that describe an object maintaining its appearance and making it suitable for the Web: fast transmission rate and efficient rendering.
- Keep as much semantic information as possible.

These issues make it difficult to create and update web catalogues. Sales agents and designers, from now on users, have to deal with technical issues that exceed their knowledge field.

The web platform proposed in this paper allows users – with no particular knowledge – to view, manage and interact with complex 3D models. It allows users to upload CAD or BIM models into the web. So, 3D model catalogues for e-commerce can be managed in an easy and intuitive way. It also provides a collaborative mechanism to share models.

The e-commerce oriented platform proposed in this paper solves the issues presented above in an automatic way. It offers a device-independent, interoperable, plug-in free solution. The platform serves the purpose of an online catalogue; users can easily upload their 3D models by simply dragging their native 3D formats into the online platform. The system automatically converts these file formats into a web compatible file format and stores them in a database. WebGL technology has been chosen to render 3D models in browsers.

Even more, this web based platform makes sharing 3D models immediate and creates a collaborative work environment in which anyone with an internet connection enabled device (PC, smartphone, tablet) can interact with the models from the online catalogue.

To overcome these limitations, the platform presented in this paper performs a sequence of operations on the original models to make them adequate to the Web. These operations include tessellation, polygon simplification and scene graph reduction.

To validate the platform, a particular use case has been designed. Several tools have been developed to improve the users experience with the platform; colour tool, measure tool, cut tool and snapshot tool. A search engine has also been included in the application since it is a common and necessary functionality in any online catalogue.

It should be kept in mind that the goal of our work is to prove that the "user-friendly CAD/BIM to Multimedia web based publishing" challenge can be achieved and to propose and validate a possible architecture. The same concept applies to the searching utilities. So, although our work makes use of different geometric and topological tools and algorithms, our contribution in not focused on these problems.

Next Section deals with background regarding the issues stated in the introduction. The proposed platform is described in detail in Section 3. Section 4 presents some results of this work using the proposed platform and finally, Section 5 analyses the conclusions drawn from the results and introduces future work.

2 Background

Sections 2.1, 2.2, and 2.3 address the issues presented in the introduction and Section 2.4 presents some existing platforms.

2.1 Polygonal conversion

BIM and CAD models are normally designed as parametric models. Parametric models offer very precise and realistic models. However, interactive visualization requires converting them into a polygonal representation by approximating these surfaces into many polygons. This process is also known as tessellation.

There are essentially two methods [11, 12] to convert a parametric surface into its polygonal representation.

2.2 Web compatible file formats

These are some of the most widely used CAD exchanging formats:

- AutoCAD DXF [10] (Drawing eXchange Format) is a CAD data file format developed by Autodesk for enabling data interoperability between AutoCAD and other programs.

- IGES [36] (Initial Graphics Exchange Specification) defines a vendor neutral data format that allows digital exchange of information among CAD systems (version 6.0 1998–01-05).
- STEP (STandard for the Exchange of Product) was developed as the successor of IGES; it is an ISO 10303–21 [17] (last reviewed in 2012) standard for the computer-interpretable representation and exchange of product manufacturing information.
- VRML (Virtual Reality Modeling Language) [6] is a text file format where vertices and edges for a 3D polygon can be specified along with the surface colour, transparency, etc.
- In 2001, X3D [38] arrived as an XML encoding of VRML. X3D provides both the XML-encoding and the Scene Authoring Interface (SAI) to ena-ble web applications to incorporate real-time 3D, presentations and con-trols into non-3D content.
- COLLADA (COLLAborative Design Activity) [9] defines an open standard XML schema for exchanging digital assets among various graphics software applications. Unlike X3D, COLLADA does not define the semantic in the 3D scenes, it is an intermediate format whose primary goal is to represent rich data in multiple forms, to enable the transformation of assets as they journey from content tools that use higher level description paradigms to applications that require platform-specific optimized descriptions [1].

Regarding BIM, Industry Foundation Classes (IFC) is a commonly used collabo-ration file format. It is an object-based file format with a data model developed by buildingSMART [5] to facilitate interoperability in the architecture, engineering and construction (AEC) industry.

IFC files are SPF (STEP Physical File) or XML representations defined against a schema. The schema gives meaning, names and relations on top of the knowledge contained in the IFC file. IFC4 [18] was released in March 2013.

None of the above formats are directly embeddable in web pages. The challenge is to select one or more native CAD/BIM input formats and to provide a tool that converts them into a format that can be interpreted and rendered in a web page.

2.3 Polygonal mesh simplification and semantics

Algorithms used to simplify meshes are based on the so called simplification operators; the most popular ones are vertex decimation, edge collapsing and vertex clustering.

- Vertex decimation operator was first proposed by Schroeder [31] Vertex decimation operates on a single vertex by deleting that vertex and re-triangulating the resulting hole.
- After vertex decimation, edge collapsing became the most common mesh simplification algorithm, such that today nearly all iterative algorithms use some sort of edge collapsing [15].
- In vertex clustering the bounding box of the mesh is divided into a grid, and all of the vertices in a given cell are replaced with a single representative vertex. Faces that become degenerate are removed from the resulting simplified mesh. Lindstrom [19] showed how Quadric Error Metric (QEM) simplification algorithm can be used to generate higher quality results.

Software tools can also be used to simplify the polygonal complexity of 3D models. Some of these tools work as plug-ins, as stand-alone applications or as programming libraries:

- Polygon Cruncher [28]
- Simplygon [2]
- MeshLab [25]

Considerable research has been conducted on polygonal mesh simplification [8, 22, 23]. Shamir A. [32] presents a state of the art on mesh segmentation techniques, and concludes that the key factor for choosing both the algorithm and the criteria for mesh segmentation is the application in mind. Thakur A. et al. [35] present a list of CAD model simplification techniques relevant for physics-based simulation problems and characterize them based on their attributes. They state that there are many open research issues such as the lack of formal analysis of computational complexity or the lack of application-specific error measures.

Posada J. et al. [29] present an ontology based compression system that uses STEP compliant standards for the compression and design review visualization of large CAD data sets. The introduction of semantics in CAD models has also been analyzed to achieve interoperable systems and exchangeable data [39].

2.4 Related work: platforms

As for visualization technologies are concerned, WebVR is an experimental JavaScript API that provides access to virtual reality devices. This technology has been used to visualize CAD models [37, 40] and also to manage big city information, proving the usability of such technology for 3D city visualization [20].

As applications are concerned, CyberCAD [34] in 2003 aimed to establish a virtual synchronous collaborative design environment to overcome geographical constraints, shorten product development time and cost through the Internet. This project developed a proprietary framework for networked CAD.

Han et al. [14] built a pilot real-time 3D system to promote the Internet-based collaborative engineering design (modelling) using STEP standard to store database. The system provides a web-based search tool with the concept of metadata for navigating the product data.

Lu, Zhihan et al. [21] proposed Open3D platform to enable the collaborative curation of large-scale city models. This platform allows simultaneous city modelling allowing multiple users to work on different aspects of the same 3D model.

GrabCAD [13] was founded in 2009 as a marketplace to connect engineers with CAD-related jobs. In 2011 it evolved into a community for engineers to share CAD models. It is short of social network where engineers can create a personal profile, store different projects, view, upload and download models from other members, communicate with each other and collaborate in other user's projects. In 2013, GrabCAD released Workbench, a cloud-based product data management (PDM) solution. This platform offers visualization of CAD models using original format viewers. This workbench offers visualization of CAD models using original format viewers. One of its features is section cutting. They resolve this issue by applying shaders to the visualization of the 3D model. However, we perform an accurate geometric cut of the model applying Binary Space Partitioning algorithm [16, 26].

BIMserver [4] is an open and stable software core to easily build reliable BIM software tools. BIMserver uses open standards, it is built as a plug-in framework and it offers an administrator configuration panel and SDKs. It features a number of modules such as visualizations, clash control and flexible queries. The platform has been installed in several companies for test purposes and feedback has been provided by developers and mailing list subscribers. BIMserver is therefore a very specialized solution for AEC industry. On the other hand, our platform is oriented to multimedia service.

The systems currently available do not provide a solution to the objectives described in the introduction of this paper. There was a need to find solutions that allowed a user-friendly connection between CAD systems and online marketing applications.

In the long run, our work does not intend to compete with commercial platforms. Our research work shows that there are new challenges that should be addressed by them. Meanwhile our research tools can provide a useful service. We hope that our work may push the industry towards these new challenges. We hope user-friendly CAD/BIM to Multimedia web based publishing and searching tools might become common tools in the near future.

3 The proposed platform

In this Section, we present a platform design that provides a structured solution to the problems related to e-commerce catalogues.

The platform is designed following a client-server model. To ensure interoperability between modules, the platform is based on Web Services technology. Specifically, SOAP messaging protocol is used.

The above mentioned functionalities will be composed by three software modules within a common platform: Authoring Tool, Web Visualization Tool and Content Management. Section 3.1 explains the geometric issues taken into account to perform the Authoring Tool, Section 3.2 describes the Web Engine developed for the platform, Section 3.3 deals with semantic issues and finally, Section 3.4 briefly introduces the Content Manager module (Fig. 1).

3.1 Geometric issues

As mentioned above, CAD/BIM models contain complex modelling descriptions and structures. These models are defined by primitives, surfaces, curves, etc. In real time, 3D rendering models must be represented as flat polygons, normally triangles, so curves, spheres, cylinders, etc. must be approximated to polygonal meshes. These polygons are normally subdivided into triangle sets so that the graphics hardware can render them onto the screen.

Real time graphics rendering engines are libraries used to ease application development. They include functionalities such as loading models, scene and camera management, animations, visual effects, etc. These libraries support polygonal formats which are oriented to interactive visualization. Therefore, CAD formats are not usually supported natively by the graphical engines.

Hence, the proposed platform must take a CAD/BIM format as an input and convert it into a web 3D compatible file format. Figure 2 illustrates the whole automatic process followed since the reception of the original CAD file until the web visualization of the 3D model. Decisions taken to accomplish this process are explained bellow.

Many CAD design programs own proprietary formats which cannot be interpreted by other applications. However, as mentioned in Section 2.3, these programs also support 3D standard file formats. Table 1 lists the exporting file formats supported in the most common CAD applications.

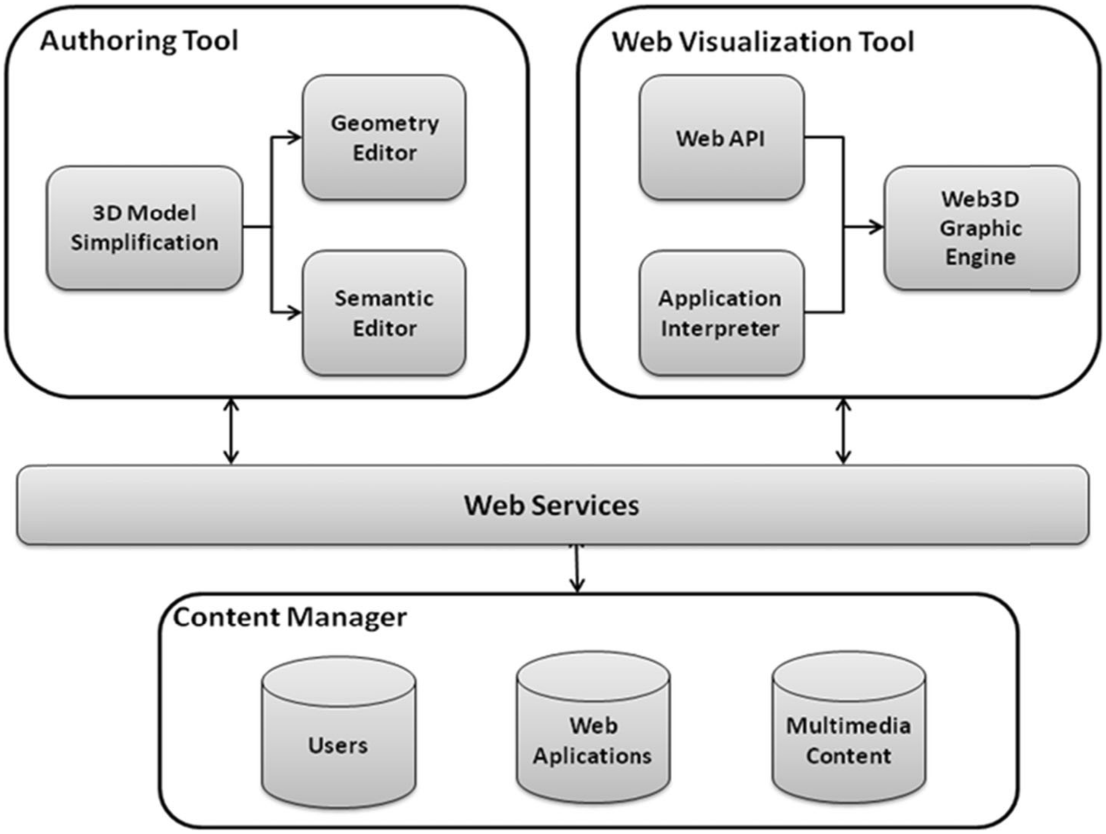


Fig. 1 3D Model Management Platform Architecture

STL, U3D, VRML, OBJ and COLLADA are polygonal file formats for visualization purposes, not CAD specific. But they are supported by some CAD applications. As shown in Table 1, the file formats supported by all applications are IGES and STEP. This fact simplifies the conversion process. Hereafter for the rest of the paper, any CAD o BIM file will be treated as a STEP file, since there is a valid exporting option to obtain a STEP file.

However, CAD programs export IGES models not only as polygonal meshes: they also use primitives and complex curves to define the 3D model. In order to use IGES as input to the platform, a tessellation process must be performed.

Graphics processing libraries have been used to make these conversions. Open CASCADE [27] and OpenSceneGraph [24] have been selected. Open CASCADE Technology is a software development platform applied in development of specialized CAD/CAM/CAE applications. On the other hand, OpenSceneGraph is an open source high performance 3D graphics toolkit, used by application developers in fields such as visual simulation, games, virtual reality, scientific visualization and modeling.

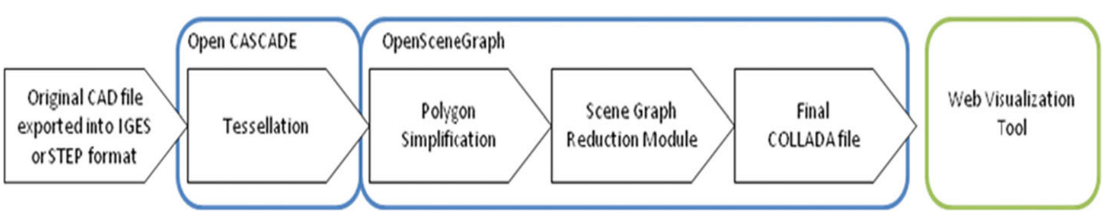


Fig. 2 Original CAD model conversion into Web compatible format

Table 1 Exportable formats in a selection of CAD software

		CAD application				
		SolidEdge	SolidWorks	CATIA	Unigraphics	Pro-Engineering
File Format	Parasolid	×	×			×
	JT	×			×	×
	ACIS	×	×			×
	CATIA	×		×		×
	IGES	×	×	×	×	×
	STEP	×	×	×	×	×
	ProE-assem		×			
	STL	×	×	×		
	U3D	×	×			×
	VRML		×	×	×	
	OBJ					×
	COLLADA					

As shown in Table 2 STL format is the only exchangeable format between these two libraries.

Finally, as mentioned in the introduction, WebGL technology has been chosen to render 3D models into web browsers. A higher level library must be chosen to ease the programming job. Table 3 shows the formats supported by some of these higher level JavaScript libraries.

Taking the afore-mentioned information into account, the process developed to convert original CAD/BIM models into web 3D compatible models is explained in the following paragraphs.

First, the original CAD/BIM models are exported into IGES or STEP files, since these two file formats are supported by most commonly used software. Then, OpenCASCADE is used to tessellate or triangulate these files. CAD/BIM complex primitives and surfaces are

Table 2 Import/Export file formats in open CASCADE and OpenSceneGraph

		Graphics Library / Engine	
		OpenCASCADE	OpenSceneGraph
File Format	Parasolid	×	
	JT		
	ACIS	×	
	CATIA		
	IGES	×	
	STEP	×	
	ProE-assem		
	STL*	×	×
	U3D*		
	VRML*	×	×
	OBJ*		×
	COLLADA*		×

Table 3 Formats supported by JavaScript libraries

		High level Web Library			
		X3DOM	O3D	GLGE	Three.js
File Format	X3D	×			
	COLLADA		×	×	×
	OBJ			×	
	JSON		×		×
	UTF8				×

approximated into triangles obtaining a polygonal mesh. This process is controlled by modifying the tessellation process. Increasing the level of quality parameter, the obtained result will be more similar to the original 3D model but at the cost of having more triangles, and hence, an oversized file. Lower quality value generates smaller and easier to handle files. To finish this part of the conversion process, the tessellated results are exported to STL format, which as seen in Table 2 is an exchangeable file format between Open CASCADE and OpenSceneGraph.

Next, OpenSceneGraph is used to import this STL file, optimize it and convert it into a COLLADA file.

The optimization process joins equal or very close vertices. This process reduces the number of vertices obtaining a simplified mesh. Deleting too many triangles may result in a deformed model with little resemblance with the original. To prevent this undesired effect, an approximation error parameter is set. This parameter will control the balance between the simplification obtained and the shape accuracy.

This optimization process requires a high amount of memory. In order to address this issue, a batching utility has been developed. When the system has to deal with a largest STL file, it is automatically divided into smaller text files; each of which contains a section of the original geometry. Each sub-file is optimized following the process previously explained. After every sub-file has been optimized they are converted into COLLADA file format and merged into one single file containing the optimized model.

This COLLADA file is sent to the web client where it must be loaded and parsed by JavaScript graphic software. ASCII file formats are not particularly efficient to store huge data sets as those required to describe 3D models (space coordinates, vertices, normals, texture coordinates, colours, etc.). These ASCII files may also contain duplicated or too precise information regarding geometries or vertices. To solve these problems and speed up the interpretation of text files a Scene Graph Reduction Module (SGRM) has been developed.

This SGRM module performs, among others, the following optimizations to de COLLADA file:

- *Fixed point precision*. The precision of vertex coordinates, texture coordinates and normal vector coordinates can be controlled.
- *Coordinate or normal vector simplification*. The SGRM searches for duplicated groups of coordinates or normal vectors; it leaves only one group and references the rest.
- *Geometry simplification*. The SGRM searches for duplicated geometries; it leaves only one and references the rest.

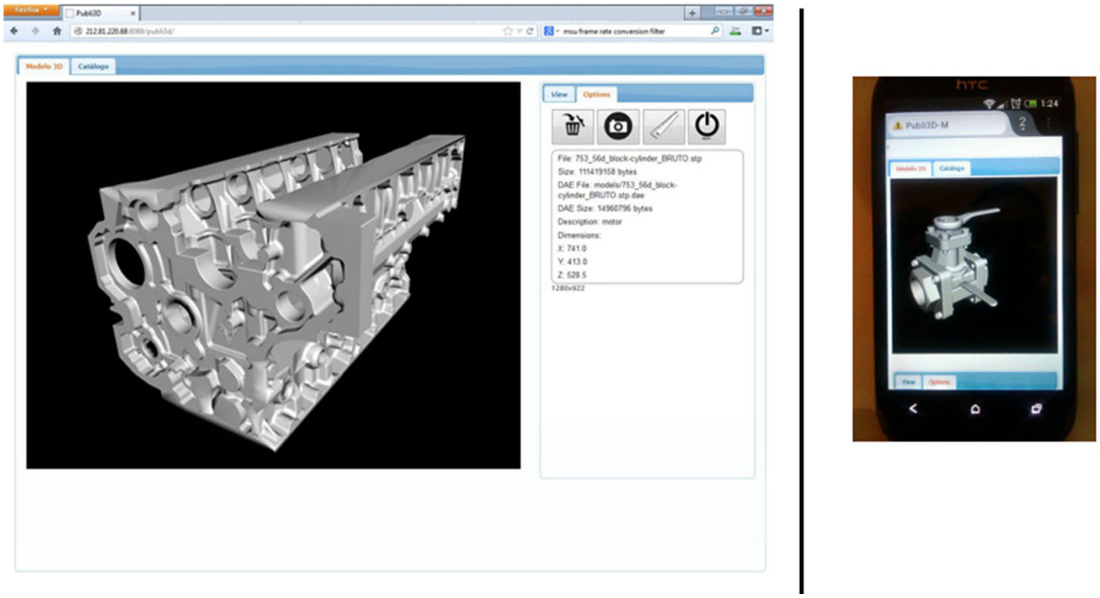


Fig. 3 Web Visualization Tool **a** Firefox web browser and **b** Smartphone

- *Multi-material simplification*. Some objects with multi-materials are not well interpreted by browsers, so the SGRM may simplify them.

Finally, this pre-processed COLLADA file is sent to the web application for rendering.

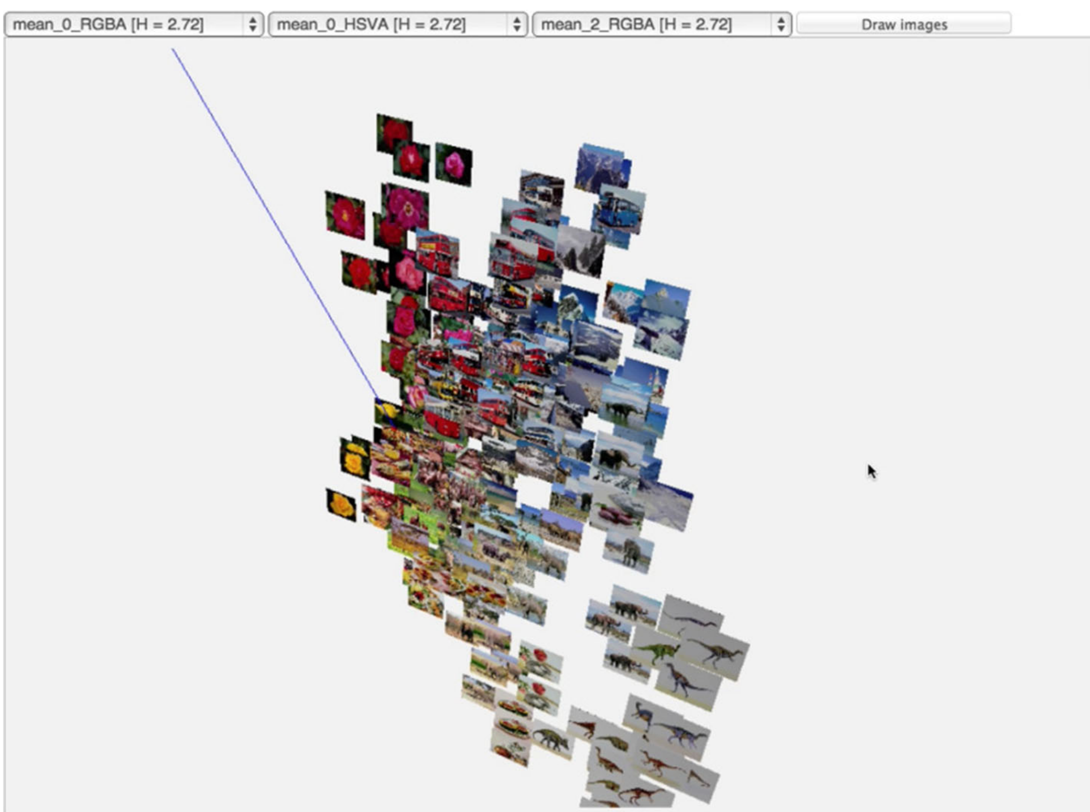


Fig. 4 Image cloud distributed in a 3D space. Images are grouped based on 3 descriptors regarding their colour similarity

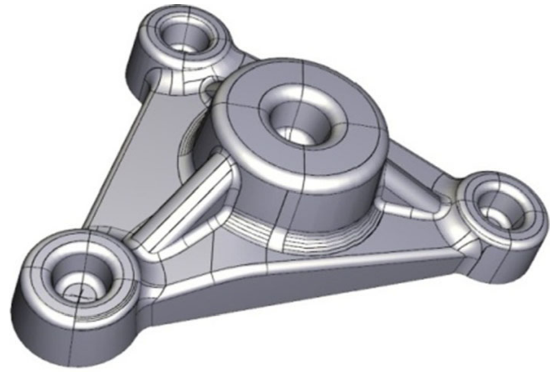


Fig. 5 Original STEP model (1.2 MB, 2145 vertices)

3.2 Web3D engine

The engine developed in this work uses the very well-known Three.js library. This engine is used to visualize the 3D model on a web browser (Fig. 3). The engine is offers the following functionalities:

- 3D Model loading. Three.js includes a function to load COLLADA models. Nevertheless, this function has been slightly modified to prevent errors reading the textures of materials.
- Real time rendering. Three.js library provides backends for Canvas, SVG, CSS and WebGL. This platform uses the WebGL backend to render the models. This allows creating complex 3D scenes in a much simpler way.
- Interactivity. Several functionalities have been designed so that users can interact with the 3D models in various ways.

- Camera controls: In order to explore the CAD/BIM model, users can interact with the model; zoom in, zoom out, view it from different perspectives, etc.
- Cut tool: With this tool the user can perform several cuts to the 3D models. When this functionality is activated, a cutting plane is visualized over the model. The user can interact with this plane with three degrees of freedom to perform and visualize different cuts on the model.
- Measure tool: Picking is used to select different faces or points of the model. This information is then used as a measuring tool, calculating the distance or the angle between the selected faces.

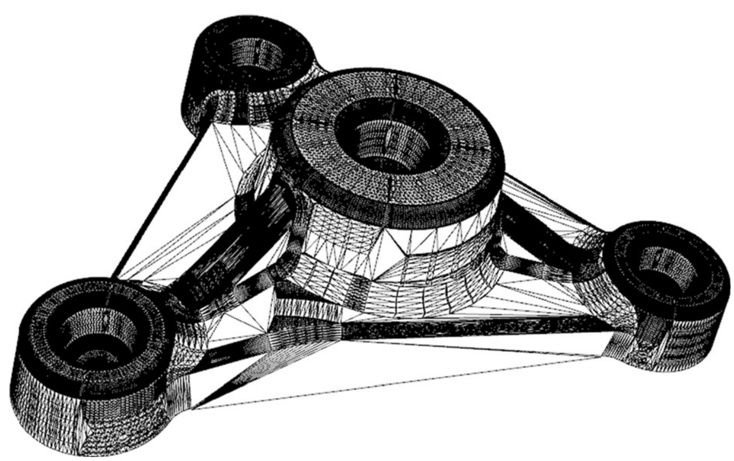


Fig. 6 Tessellated model; high quality selected (154,000 polygons)

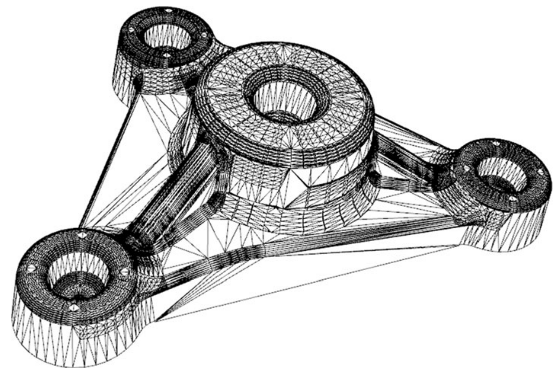


Fig. 7 Tessellated model; medium quality selected (42,069 polygons)

- Snapshot tool: The user can choose to store a thumbnail of the CAD model. This tool allows taking a snapshot of the model in the desired position and associating this image with the metadata of the model.

3.3 Semantic issues

The following Section addresses the problem of storing and searching for the previously loaded 3D models.

CAD models contain metadata which typically provide semantic information including a brief description of the data set, the area covered by the data set, the data structure and file format, the coordinate system or projection of the data, the time when the data was collected and method of collection or the quality or accuracy of the data.

On the other hand, in Computer Vision, image descriptors describe certain elementary characteristics of the images such as shape, colour, texture or size. The advantage of these descriptors is that they can be mathematically compared: similar images will have similar descriptors. This quality is widely used by search engines for 2D image classification.

In our case, 3D projections are stored with the CAD/BIM converted model as metadata. These projections can be generated automatically. Users can easily take snapshots of specific views and store them. These images can be used by classical image classification systems. In this way, content based retrieval processes can be integrated into our platform [30] (Fig. 4).

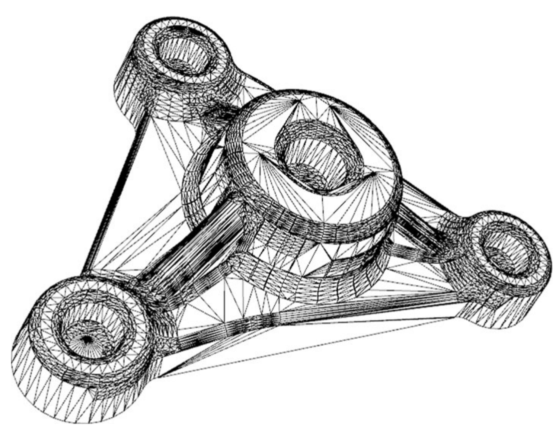


Fig. 8 Simplified polygonal model (10,340 polygons)

Fig. 9 Tessellated model (579 MB, 2.2 million polygons)

When the user takes a snapshot of the loaded 3D model, this image is stored in Cloudinary. It is a proprietary cloud platform to manage large numbers of images. Each image file is associated with a unique identifier and URL. Cloudinary extracts the main metadata from the images (codec, size, etc.). Clearly, this data offered by Cloudinary is not enough to search images by similarity. The Semantic Editor Module proposed in this work adds content descriptors: colour and texture histograms and number of detected faces and their orientation.

In essence, the descriptor of each image is a vector of numbers that represents all those features. The enriched metadata is stored in a MongoDB database in the cloud. When a user initiates an image search in our platform a query with the searching criteria is sent. The platform then retrieves all the images from the database with similar features to the one in the query.

A way to evaluate the level of similarity between the retrieved images is to calculate the Euclidean distance between the descriptors in the query and those from the result. As mentioned before, these descriptors are represented as a vector of numbers so their Euclidean distance can be easily obtained.

The Semantic Editor Module in this work calculates the distance between the image used in the query and those in the database and retrieves the 20 more similar images. This collection is done applying the methodology proposed by Silva, J. L. [33] which is based on a probabilistic k-nearest neighbour supervised classification algorithm.

3.4 Content management

The platform presented in this work serves as an online marketing catalogue. The contents of this catalogue are stored in a MongoDB database. MongoDB [7] is a cross-platform document-oriented database system. Classified as a NoSQL database, MongoDB eschews the traditional

Fig. 10 Simplified model (852,393 polygons)

table-based relational database structure in favour of JSON-like documents with dynamic schemas, making the integration of data in certain type of applications easier and faster. These databases hold a set of *collections* and these collections hold a set of *documents*. A document is a set of key-value pairs. Documents have dynamic schema which means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a document collection may hold different type of data.

The designed database has a collection named *catalogue* in which all the information regarding the 3D models is stored. When a new model is uploaded into the platform a new document is added to the catalogue collection. Initially, these documents are described only by two fields; name and description. These fields are edited by the user when uploading a new model. When the model has been processed as explained in Section 3.1, extra information regarding various aspects (optimized size, path, snapshot image, etc.) is added to this document.

4 Results

This Section presents the platform proposed in this paper using two examples.

4.1 Simple STEP model conversion

The input for this first example is a simple STEP file. As seen in Fig. 5 this model has not been designed with a very high detail. The number of vertices used to define its surface is very low: 2145 vertices. However, it includes surfaces which must be processed for interactive visualization. The first automatic step tessellates this file, approximating its surfaces to triangulated meshes. In Fig. 6 a high quality tessellation is shown with 154,000 polygons. This number can be reduced selecting a lower quality level. With medium quality for this model achieved 35,700 polygons (see Fig. 7).

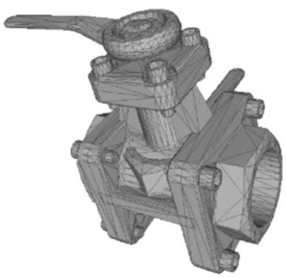
It may happen that a huge polygonal STEP file does not contain the original surfaces description. In these cases, the scene graph reduction module is applied to the tessellated model. Triangles and vertices are deleted maintaining a bounded approximation error.

Table 4 Conversion Times for model #1.

	Size (KB)	Time (s)
Initial	87	Transfer: 0.0.10
Tessellation	128	0.158
Polygon Simplification	30	0.143
Optimization (SGR)	24	0.199

Final number of Polygons: 1,374

Table 5 Conversion Times for model #2.



	Size (KB)	Time (s)
Initial	1,120	0.035
Tessellation	20,101	2.015
Polygon Simplification	1,623	3.543
Optimization (SGR)	1,325	0.566

Final number of Polygons: 77,355

Using the previous model (35,700 polygons), an automatically simplification achieves a model with 10,340 polygons (see Fig. 8).

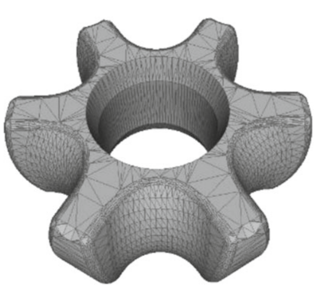
4.2 Complex STEP model conversion

The following example presents a much more complex situation. The initial input model in this example is a 165 MB STEP file. Its tessellated result, shown in Fig. 9 , is composed by 2.2 million polygons and its STL ASCII file size is 579 MB. The model is too heavy and complex to be visualized even by a native desktop application. Following the algorithm described in Section 3, the model is automatically split in various sub-models and treated separately. In this case, the model was divided into 15 sub-models. Each of these pieces was processed and finally these pieces were merged to obtain the resulting optimized file (see Fig. 10).

4.3 Conversion time rates

This section shows the time needed to convert several CAD models of diverse size and complexity into COLLADA web compatible format. Each example includes the size of the models (KB) after each conversion/optimization process and the final number of polygons (units).

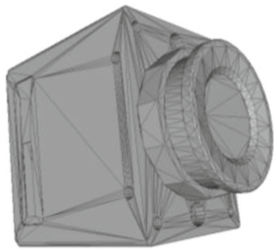
Table 6 Conversion Times for model #3**.**



	Size (KB)	Time (s)
Initial	2,245	0.069
Tessellation	5,350	1.559
Polygon Simplification	557	2.512
Optimization (SGR)	459	0.084

Final number of Polygons: 18,831

Table 7 Conversion Times for model #4.



	Size (KB)	Time (s)
Initial	22,296	0.282
Tessellation	34,192	15.940
Polygon Simplification	4,463	5.016
Optimization (SGR)	3,748	0.789

Final number of Polygons: 229,887

A medium quality level of conversion has been applied to these models. As defined in Fig. 2 the conversion process includes:

- Tessellation
- Polygon simplification
- Optimization or Scene Graph Reduction (SGR)

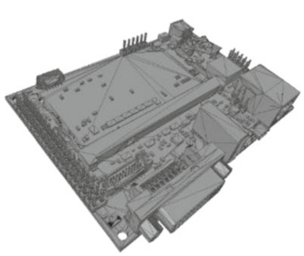
As shown in the following examples, tessellation and polygon simplification are the most time consuming steps of the process (see Tables 4, 5, 6, 7, 8, 9 and 10).

Next we analyse the most attention-getting data of these examples.

The size of the models after the tessellation process is bigger than the initial size. Original CAD/BIM files are defined as parametric models; these curved surfaces must be converted into polygons. In order to obtain a precise approximation of the model, a vast number of polygons is needed. However, this model is used as an intermediate process of the conversion. It is not transferred at any time.

In most cases, the size of the final multimedia file is smaller than the original one. However, in Table 5 the final size of the converted model is slightly bigger than the initial size. This happens with models designed in high detail. In order to preserve the appearance of the original model, a bigger amount of polygons is needed and the final size might be slightly

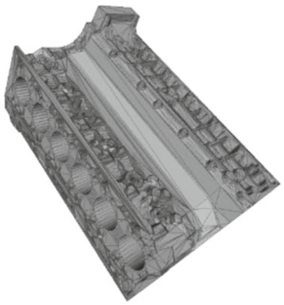
Table 8 Conversion Times for model #5.



	Size (KB)	Time (s)
Initial	62,006	0.767
Tessellation	70,253	107.329
Polygon Simplification	3,538	67.093
Optimization (SGR)	2,888	0.445

Final number of Polygons: 193,323

Table 9 Conversion Times for model #6.



	Size (KB)	Time (s)
Initial	108,808	1.196
Tessellation	530,791	149.263
Polygon Simplification	13,025	573.354
Optimization (SGR)	10,827	0.609

Final number of Polygons: 486,579

increased. However, the size of the converted models remains manageable for multimedia transmission purposes.

5 Conclusions and future work

This paper presents a platform for CAD and BIM model management for online marketing. It eases the development of e-commerce systems that contain catalogues with 3D content. As stressed in the introduction, the goal of this work is to offer a user-friendly CAD/BIM to multimedia web based publishing solution and to solve some of its challenges.

The platform has been tested with diverse CAD and BIM models from different locations, devices, operating systems and browsers. The platform has been tested both through wired internet connection and through 3G mobile connection.

The import process considers automatic primitive tessellation, mesh reduction, huge model splitting, scene graph reduction and Web3D compatible model generation (COLLADA).

Model uploading rates vary depending on the size of the model: light models (90 KB) take only a few milliseconds to upload (example in Table 4), however larger files (160 MB) can take up to 20 min (example in Table 10). Then, converted models are smoothly managed and rendered in the client side.

Table 10 Conversion Times for model #7.

	Size (KB)	Time (s)
Initial	165,613	1.749
Tessellation	881,030	356.989
Polygon Simplification	19,934	1002.751
Optimization (SGR)	16,199	1.653

Final number of Polygons: 852,393

To test the platform various models have been dragged and dropped onto the web application. Loading rates on the client vary between a few seconds and a few minutes depending on the size of the file. Large files can be compressed to reduce the loading times [3].

Tests have been made on Mozilla Firefox version 45 and Google Chrome version 50 browsers. Additionally, the platform has also been tested, except the model loading functionality, on smartphones with Android operating system and Firefox Mobile or Chrome Mobile browsers.

A solution for 3D model management has been considered based on virtual 2D images of the 3D models, evaluating image descriptors.

As for future work there are several points to work on, for instance, a level-of-detail based simplification software could be used: a few models with different quality rates could be generated. Later on, depending on the client specifications the most appropriate ones would be sent for local visualization. Regarding model materials there is also further research to be done; intelligent scene graph reduction could be implemented by merging geometries with the same material. The platform is ready to provide model management based on an appropriate image descriptor analysis. Better descriptor and image search tools are being researched to integrate them in the platform.

References

1. Arnaud, R., and Parisi T. (2007). "Developing web applications with collada and x3d." White paper.
2. Automagic 3D Optimization. Simplygon. http://www.simplygon.com/Accessed 30 Mar. 2015.
3. Bayardo, R. J., Gruhl, D., Josifovski, V., and Myllymaki, J. (2004). “An evaluation of binary xml encoding optimizations for fast stream based xml processing”. In Proceedings of the 13th international conference on World Wide Web, pp. 345–354. ACM.
4. Beetz, J., van Berlo, L., de Laat, R., and van den Helm, P. (2010). BIMserver. org–An open source IFC model server. Proceedings of the CIP W78 conference.
5. Building SMART, U. K. (2010) Investing in BIM competence Building SMART: a guide to collaborative working for project owners and building professionals. Endeavour, UK
6. Carey, Rikk, and Gavin Bell. (1997)."The annotated VRML 97 reference."
7. Chodorow K (2013) MongoDB: the definitive guide. O’Reilly Media, Inc.
8. Cignoni P, Montani C, Scopigno R (1998) A Comparison of Mesh Simplifica-tion Algorithms. Comput Graph 22(1):37–54
9. COLLADA 3D Asset Exchange Schema. https://www.khronos.org/collada/. Accessed 16 April 2015.
10. DXF Reference (2007): 1–296
11. Frey PJ, George PL (2000) “Mesh Generation. In: Application to Finite Element”. Hermes Science Publishing, Paris, Oxford
12. George PL, Borouchaki H (1997) Triangulation de Delaunay et maillage: applications aux éléments finis. Hermes, Paris
13. GrabCAD Helps Mechanical Engineers Build Great Products Faster. (2015). CAD Collaboration Solution That Accelerates Product Development.

14. Han S, Choi Y, Yoo S, Park N (2002) Collaborative engineering design based on an intelligent STEP database. Concurr Eng: Res Appl 10(3):239–250
15. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. (1993). “Mesh optimization”. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques. pp. 19–26. ACM.
16. Hughes, J. F., Van Dam, A., Foley, J. D., and Feiner, S. K. (2014). “Computer graphics: principles and practice”. Pearson Education.
17. ISO, ISO. "10303–21: 2002." Industrial Automation Systems and Integration–Product Data Representation and Exchange–Part 21: 72.
18. Liebich, T. (2013). “IFC4 – the new building SMART Standard”. http://www.buildingsmart-tech.org/specifications/ifc-releases/ifc4-release/buildingSMART_IFC4_Whatisnew.pdf. Official Release Date of building SMART’s IFC4–12″. Date of ISO 16739–21.
19. Lindstrom, P. (2000). Out-of-core simplification of large polygonal models. Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co.
20. Lu Z, Li X, Zhang B, Wang W, Zhu Y, Hu J, Feng S (2016a) Managing big city information based on WebVRGIS. IEEE Access, vol 4:407–415
21. Lu, Z., Guerrero, P., Mitra, N. J., and Steed, A. (2016b). Open3D: crowd-sourced distributed curation of city models. In Proceedings of the 21st International Conference on Web3D Technology. pp. 87–94. ACM.
22. Luebke D. (2001). “A developer’s survey of polygonal simplification”. In IEEE Virtual Reality Conference Tutorial Vol 7.
23. Luebke D (2003) Level of Detail for 3D Graphics. Morgan Kaufmann
24. Martz, P. (2007). "OpenSceneGraph quick start guide." PMARTZ Computer Graphics Systems.
25. MeshLab. MeshLab, the Mesh Processing System for 3D Scanning and Printing. http://meshlab.sourceforge.net/ Accessed 30 Mar. 2015.
26. Naylor BF (2005) Binary space partitioning trees. Handbook of Data Structures and Applications pp:20–21
27. Open CASCADE Technology, 3D Modeling & Numerical Simulation.. http://www.opencascade.org/ Accessed 30 Mar. 2015.
28. Polygon Cruncher for 3ds Max / 3ds Max Design, Maya, Lightwave, 3D Browser and as a Stand Alone Software. http://www.mootools.com/plugins/us/polygoncruncher/index.asp Accessed 30 Mar. 2015.
29. Posada J, Toro C, Wundrak S, Stork A (2006) Using ontologies and STEP standards for the semantic simplification of CAD models in different engineering domains. Appl Ontol J 1(3–4):263–279
30. Quartulli M, Olaizola IG (2013) A review of EO image information mining. ISPRS J Photogramm Remote Sens 75:11–28
31. Schroeder, W. J., Zarge, J. A., and Lorensen, W. E. (1992) Deci-mation of triangle meshes. Siggraph Computer Graphics. Vol. 26. No. 2. ACM.
32. Shamir A (2008) A survey on mesh segmentation techniques. Computer Graphics Forum 27(6):1539–1556
33. Silva JL, Bengoa NA, Quartulli M, Olaizola IG, Zulueta E (2015) Web-Based Supervised Thematic Mapping. IEEE J Sel Top Appl Earth Obs Remote Sens 8(5):2165–2176
34. Tay, F.E.H., and Roy, A. (2003). CyberCAD: a collaborative approach in 3D–CAD technology in a multimedia-supported environment. Computers in Industry, 52 vol. 2 pp. 127–145
35. Thakur A, Banerjee G, Gupta SK (2009) A survey of cad model simplification techniques for physics-based simulation applications. Comput Aided Des 41(2):65–80
36. US Product Data Association. (1996)."Initial Graphics Exchange Specification. IGES 5.3." ANSI.
37. Wei-dong WU (2010) Research and Application of Engineering Drafting Teaching Mode based on Technology of Three Degree CAD [J]. Sci Technol Inf 23:479
38. X3D V3.3 Abstract Specification. http://www.web3d.org/content/x3d-v33-abstract-specification. (2013). Accessed 20 April 2015.
39. Yang QZ, Zhang Y (2006) Semantic interoperability in building design: Methods and tools. Comput Aided Des 38(10):1099–1112
40. Zhang, M., Lv, Z., Zhang, X., Chen, G., and Zhang, K. (2009) Research and Application of the 3D Virtual Community Based on WEBVR and RIA. Computer and Information Science, vol. 2, no 1, p. 84.

Helen V. Diez received a M.S. degree in Telecommunication Engineering from the University of Navarra, Donostia-San Sebastian, Spain, in 2009. In 2013 she received a Master in Advanced Computer Systems from UPV-EHU. In 2010 she joined Vicomtech-IK4 as a Researcher, working on interactive 3D visualization, simulation, computer vision and Web3D. She is currently a researcher at the Interactive Computer Graphics department of Vicomtech-IK4. Her current research interests include virtual reality, human-computer interaction, 3D visualization and BIM.

Álvaro Segura received a M.S. degree in Electrical Engineering from the University of Navarra, Donostia-San Sebastian, Spain, in 1999. From 1998 to 2001 he worked as a Junior Researcher for San Sebastian's Technology Park, on interactive Web technologies, including 3D graphics, and then on intelligent mobile robotics, including visual and audio perception and navigation. In 2002 he joined Vicomtech-IK4 as a Researcher, working on interactive 3D visualization, simulation, computer vision, GIS, Web3D and computational geometry. He is currently a Principal Researcher at the Industry and Advanced Manufacturing department of Vicomtech-IK4. His current research interests include virtual reality, 3D visualization, simulation and interaction.

Alejandro García-Alonso Mechanical Engineer (U. Navarre, 1980), Ph. D. (U. Navarre, 1990), research assistant (CEIT, 1983), research leader (CEIT, 1992), associate professor of computer graphics (Euskal HerrikoUnibertsitatea, 1995), vice Dean Computer Science Faculty (2008). He has collaborated in seven books, two video, seventeen papers in journals, six invited conferences, more than fithy conference communications, supervisor in ten Ph. D., collaborates with many research institutions (CEIT, Tecnalia, Ikerlan, VICOMTech). He has imparted invited lectures in ten universities (USA, Canada, Europe), worked or directed many I+D projects for the industry: Mechanical Dynamics Inc. -MDI, Michigan- (i.e. "ADAMS/Animation"), Aerospatiale, Alenia, CASA, (i.e. "Distributed Simulation for Visualization of Space Activities in Near-Real-Time", 1995), European Space Agency ("Investigations in the Production of Realistic 3D Scenes. Survey of Methods and Algorithms", 1986). He has collaborated with more than thirty companies. His research interests include computer graphics, virtual reality, geometric problems (collisions), and haptics. He is a member of the IEEE, ACM, and Eurographics.

Dr. David Oyarzun is the director of the Interactive Computer Graphics Department at Vicomtech-IK4. He studied Computer Engineering in the University of the Basque Country. He received his Ph.D. in the field of 3D Virtual Environments standardization from the University of the Basque Country. From 2001 to 2003 he gained experience in topics such as biometric access, control systems for networks or Artificial Intelligence. From 2003 onwards he is working at Vicomtech-IK4. He is also member of the Board of Directors of Web3D Consortium. He has dozens of publications in topics related with Virtual and Augmented Reality in international conferences and journals, and he is also regular member of several IPCs. Moreover, he is external advisor of ANEP, the Spanish Agency that evaluates some of the main national R&D calls for projects.