Politecnico di Bari

Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Operator 4.0: Industrial Augmented Reality, Interfaces and Ergonomics

This is a PhD Thesis

Original Citation:
Operator 4.0: Industrial Augmented Reality, Interfaces and Ergonomics / Manghisi, Vito Modesto. - ELETTRONICO. - (2019). [10.60576/poliba/iris/manghisi-vito-modesto_phd2019]

Availability:
This version is available at http://hdl.handle.net/11589/161097 since: 2019-01-18

Published version
DOI:10.60576/poliba/iris/manghisi-vito-modesto_phd2019
Politecnico di Bari

Terms of use:
Altro tipo di accesso

(Article begins on next page)

19 April 2024

Politecnico di Bari

Department of Mechanics, Mathematics and Management
MECHANICAL AND MANAGEMENT ENGINEERING
Ph.D. Program
SSD: ING-IND/15 - DESIGN METHODS FOR INDUSTRIAL ENGINEERING

Final Dissertation

OPERATOR 4.0: INDUSTRIAL AUGMENTED REALITY, INTERFACES, AND ERGONOMICS

by
Manghisi Vito Modesto

Referees:
Prof. Francesco Ferrise
Prof. Maurizio Muzzupappa

Supervisors:
Prof. Antonio E. Uva
Prof. Michele Fiorentino
Prof. Vitoantonio Bevilacqua

Coordinator of Ph.D Program:
Prof. Giuseppe P. Demelio

Course n°31, 01/11/2015-31/10/2018

To my beloved Wife

Index

Index **1**
Introduction **5**
Chapter 1: IAR for the Augmented Operator **11**
1.1. IAR supporting the Operator 4.0 ..... **11**
1.2. The IAR System components **14**
1.2.1. Tracking 18
1.2.2. Authoring 21
Chapter 2. Design and test of a projective AR workbench for manual working stations 25
2.1. Introduction **25**
2.2. Related works **27**
2.2.1. Applications of SAR in the industry 27
2.2.2. User evaluation of AR in the industry 28
2.3. System requirements **30**
2.3.1. The choice of the SAR technology 30
2.3.2. Requirements 31
2.4. The Prototype **32**
2.4.1. Technical features 32
2.4.2. Contents 32
2.4.3. System Architecture 33
2.4.4. The Physical setup 35
2.4.5. The User Interface 36
2.5. Preliminary trial and prototype improvement **37**
2.6. Evaluation of the SAR based MWS effectiveness **38**
2.6.1. Research questions 38
2.6.2. Material and methods 39
2.6.3. Participants 40
2.6.4. Procedure 40
2.6.5. Measures 43
2.7. Analyses and results **44**
2.7.1. Completion time 44

2.7.2. Error rate 45
2.7.3. Users' acceptance 46
2.8. Discussion and conclusions 47
Chapter 3. Text legibility study for monocular Optical See-Through Displays for IAR 51
3.1. Introduction 51
3.2. Related works 53
3.3. Our approach 56
3.4. Design of the experiment 58
3.4.1. Participants 59
3.4.2. Apparatus and materials 59
3.4.3. Text settings 61
3.4.4. Procedure 63
3.5. Results 64
3.6. Discussion and Conclusions 66
Chapter 4. An IAR Framework for P&ID enhanced comprehension. 69
4.1. Introduction 70
4.2. Materials and methods 72
4.3. Preliminary user test results 75
4.4. Conclusions 77
Chapter 5. The Operator 4.0 and the Human Machine Interactions 78
5.1. Introduction 80
5.2. Research aim 82
5.3. The gesture vocabulary design 82
5.3.1. Interface requirements definition 83
5.3.2. Gestures elicitation procedure 83
5.3.3. Vocabularies definition 88
5.4. The NUI implementation 88
5.4.1. User-leader definition 89
5.4.2. User's state definition 90
5.5. User's actions triggering 93
5.6. NUI testing and evaluation 95
5.6.1. Experimental procedure 95
5.6.2. Metrics: Administered Questionnaires 96
5.6.3. Results 97
5.7. Discussion and Conclusions 100
Chapter 6. A general framework for mid-air gestures-interfaces design 102

6.1. Introduction 102
6.2. Related works 103
6.3. The proposed framework 105
6.4. The case study: Interface Requirement Definition 107
6.5. Gestures Elicitation 108
6.5.1. Collection of gesture proposals 108
6.5.1. Gestures labelling 109
6.5.2. Gestures clustering 112
6.5.3. Agreement analysis 114
6.6. Vocabularies definition 115
6.6.1. Composing vocabularies 116
6.6.2. Ranking vocabularies 118
6.6.3. Validation procedure 119
6.7. Discussion and Conclusions 123
Chapter 7. A Postural Risk Assessment Tool supporting the Healthy Operator 126
7.1. Introduction 126
7.2. Related works 127
7.3. Materials and Methods 131
7.3.1. K2RULA software 131
7.3.2. The RULA method 132
7.3.3. Data retrieval 133
7.3.4. Functionalities 136
7.4. Experiment 1: validation with an optical motion capture system 137
7.4.1. Equipment 137
7.4.2. Procedure 138
7.4.3. Data analysis 140
7.5. Experiment 2: validation with RULA expert and comparison with the Jack TAT 141
7.5.1. Equipment 142
7.5.2. Procedure 142
7.5.3. Data analysis 142
7.6. Results 142
7.6.1. Experiment 1 142
7.6.2. Experiment 2 143
7.7. Discussion and conclusions 144
7.7.1. Main contributions 144

7.7.2. Limitations of the study 145
7.7.3. Conclusions and research developments 146
Conclusion and future works 148
Acknowledgements 152
Appendix A 153
Appendix B 158
Appendix C 160
Bibliography 162

Introduction

The German program Industry 4.0 and the corresponding international initiatives (such as Smart Manufacturing in the USA, Smart Factory in South Korea, and Industria 4.0 in Italy) will continue to transform the industrial workforce and their work environment through 2025 (Lorenz, Ruessmann, Strack, Lueth, & Bolle, 2015). This will significantly involve the nature of work in the industry as Industry 4.0 will transform design, manufacture, operation, and service of products and production systems. At the same time, the demography is changing, especially in Europe and Japan, which brings forth additional challenges for manufacturing companies.

Therefore, 'smart factories' that play the role of socio-technical systems, will need to form and adapt a social perspective to be proficient in assisting ageing, disabled and apprentice operators. By using advanced digital and industrial enabling technologies smart factories will help people to remain in, return to or incorporate into the modern manufacturing workforce. Meanwhile, considering the developments from a technical perspective, new connectivity and interaction technologies among parts (cf. smart products), machines (cf. smart machines) and humans (cf. smart operators) will make production systems more lean, agile, traceable, and adaptable (De Weck, Ross, & Rhodes, 2012; Romero, Stahre, et al., 2016).

According to Lorenz et al. (2015), Industry 4.0 (I4.0) enables new types of interactions between operators and machines. These interactions will transform the industrial workforce and the nature of work. An important part of this transformation is the emphasis on human-centricity of the factories of the future (Ridgway et al., 2013), allowing for a paradigm shift from independent automated and human activities towards a human-automation symbiosis (or 'human cyber-physical systems') characterized by the cooperation of machines with humans in work systems and designed not to replace the skills and abilities of humans, but rather to co-exist with and assist humans in being more efficient and effective (Tzafestas, 2006).

In parallel with the evolution of the industry, the history of the interaction of operators with various industrial and digital production technologies can be summarized as a generational evolution (Romero, Bernus, Noran, Stahre, & Fast-Berglund, 2016). Thus, Operator 1.0 generation is defined as humans conducting 'manual and dexterous work' with some support from mechanical tools and manually operated machine tools. Operator 2.0 generation represents a human entity who performs 'assisted work' with the support of computer tools, ranging from

CAx tools to NC operating systems (e.g. CNC machine tools), as well as enterprise information systems. The Operator 3.0 generation embodies a human entity involved in 'cooperative work' with robots and other machines and computer tools, also known as - human-robot collaboration. Finally, the Operator 4.0 generation represents the 'operator of the future', a smart and skilled operator who performs work 'aided' by machines if and as needed (Figure 1).

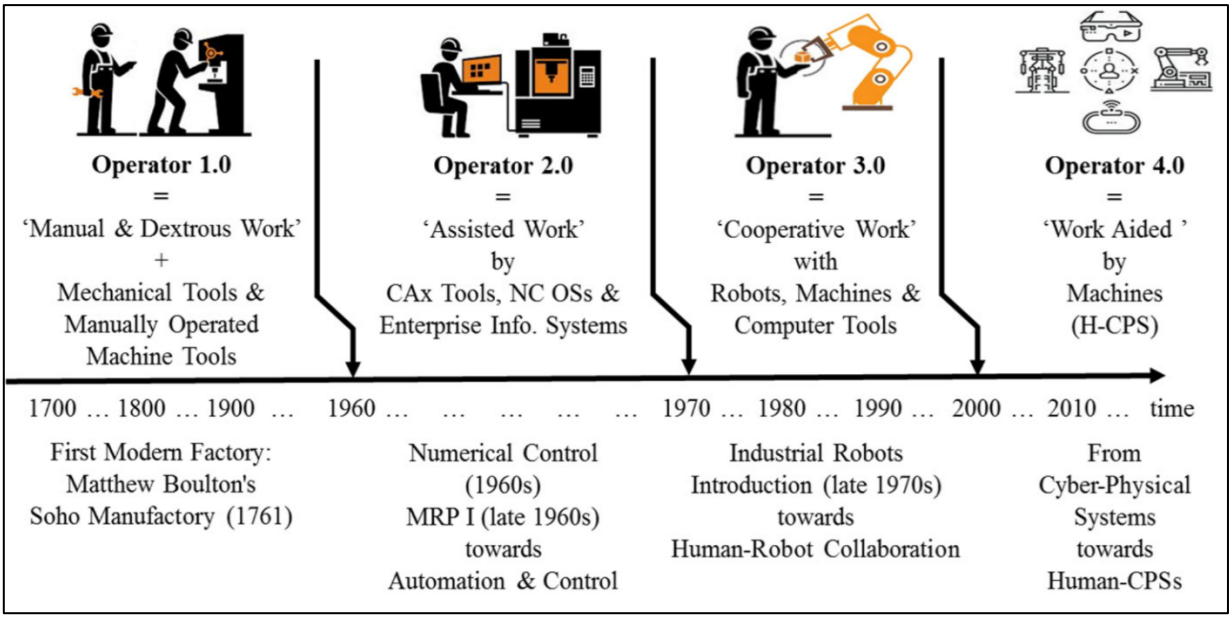


Figure 1 -The Operator Generations (R)Evolution. Source: (Romero, Bernus, et al., 2016).

This work aims at applying the enabling technologies of Industry 4.0 in order to design and develop, methods and applications supporting the figure of the Operator 4.0.

Romero, Stahre, et al. (2016) defines a multifaceted operator where I4.0 enabling technologies cooperate to compose these facets (Figure 2).

Those typologies include:

- **The Super - Strength Operator**, where a human-robotic exoskeleton powered by a system of motors, pneumatics, levers or hydraulics works cooperatively with the operator to allow for limb movement, increased strength and endurance;
- **The Augmented Operator**, where Industrial Augmented Reality (IAR) enriches the real-world factory environment of the smart operator with digital information and media (sound, video, graphics, GPS data, and so on.) that is overlaid in real-time in his/her field of view;

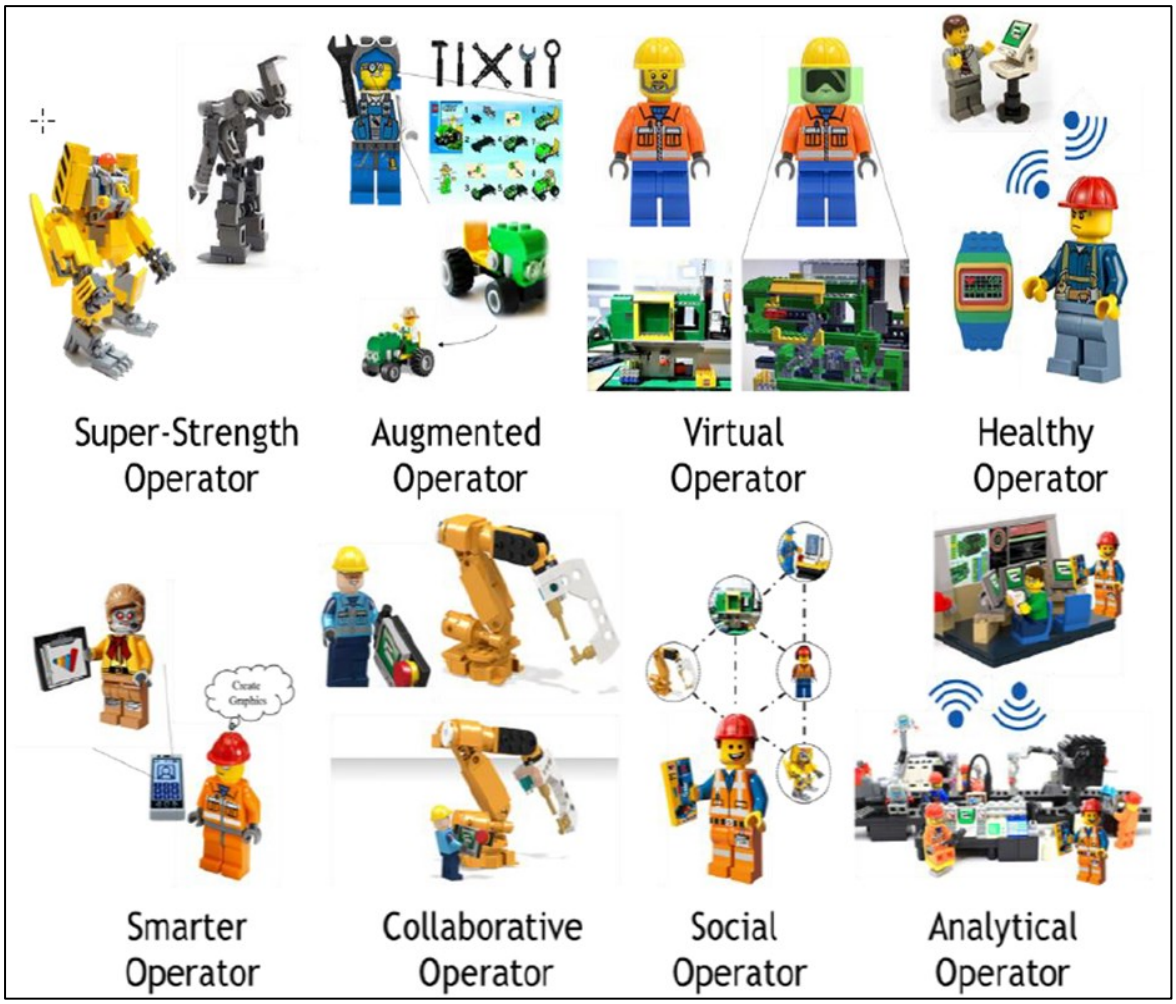


Figure 2 - The Operator 4.0 typologies, Source: (Romero, Stahre, et al., 2016).

- **The Virtual Operator**, where Virtual Reality (VR) digitally replicates a design, assembly or manufacturing environment and allows the smart operator to interact with any presence within (e.g. a blueprint, a hand-tool, a product, a machine tool, a robot, a production line, a factory), with reduced risk and real-time feedback;
- **The Healthy Operator**, where smart solutions (i.e. wearable trackers for health-related metrics) including data analytics capabilities together with advanced Human-Machine and Human-Automation Interfacing/Interaction are used to monitoring bio-data (i.e. physiological data, postures, and workload). Thus, driving positive change in terms of improved productivity, wellbeing, and proactive safety measures at smart workplaces;
- **The Smarter Operator**, where an Intelligent Personal Assistant, that is a software agent or artificial intelligence helps a smart operator in interfacing with machines, computers, databases and other information systems as well as managing time commitments and performing tasks or services in a human-like interaction;

- **The Collaborative Operator**, where industrial robots (cf. humanoid or robotic arm) capable of performing a variety of repetitive and non-ergonomic tasks and that have been specially designed to work in direct cooperation with the smart operator by means of safety (e.g. force sensing and collision) and intuitive interaction technologies, including easy shop-floor programming, will allow co-working spaces and interaction with their human counterparts without the need for traditional safety barriers;
- **The Social Operator**, where social networking between smart operators, enabled by real-time mobile communication capabilities, can empower the workforce to contribute their expertise across the production line and to the shop-floor, can accelerate ideas generation for product and processes innovation and can facilitate problems-solving;
- **The Analytica Operator**, where Big Data analytics may help smart operators (e.g. production managers) to achieve better forecasts, understand the smart factory performance (shop-floor control), fuel continuous improvement (Six Sigma), provide greater visibility of KPIs (data visualization and interactive dashboard) and real-time alerts based on predictive analytics (fault detection and quality improvement) in order to leverage real-time information for driving the right response to prevent mistakes, quickly identify problems and call for the right decisions to improve operational efficiency.

These types of Operators may exist on the shop-floor as either single- or hybrid- types. The augmentations can be combined and, it is also very likely that the future Operator 4.0 may only be augmented in one specific area whereas the other aspects are neglected (Romero, Stahre, et al., 2016).

The Operator 4.0 represents a futuristic vision of how the Industry 4.0 technologies can assist operators to become 'smarter operators' in the factory workplaces. Despite the recent availability of efficient and mature technologies, their application to augment the operator in the shop floor still remains confined to few prototypes. The feasible use of such applications in the real world requires prototype validation, and moreover, a design paradigm based on a user-centric approach.

In this work we describe the methods and applications designed and developed to contribute to the Augmented the Virtual and the Healthy Operator facets. This thesis can be divided into two macro-areas. The first one, composed of chapters from 1 to 4, describes the researches carried out in the field of Industrial Augmented Reality (IAR) supporting the figure of the Augmented Operator.

In Chapter 1, we introduce the researches carried out in the IAR field. We describe the Augmented Reality (AR) technology and its application in the field of the Industrial Augmented

Reality (IAR). In chapter 2, we present a Spatial Augmented Reality (SAR) workbench prototype designed in the early stage of this research and we describe the experiments carried out to validate its efficiency as support to the Operator 4.0.

In Chapter 3, we describe the experiments carried out to optimize legibility of text shown in AR interfaces for optical see-through displays. In this research, we propose novel indices extracted from the background images, displayed on an LCD screen, and we compare them with those proposed in the literature by designing a specific user test.

In Chapter 4, we present an AR framework for handheld devices that enhance users in the comprehension of plant information traditionally conveyed through printed Piping and Instrumentation Diagrams (P&ID).

The second section of this thesis, from chapter 5 to chapter 7, describes the researches carried out in the fields of HMI and Ergonomics supporting the role of the Virtual and Healthy Operator.

In Chapter 5 we describe the research carried out in the field of HMI related to the use of Natural User Interfaces in Virtual Reality. We designed and developed a gesture interface for navigation of virtual tours made-up of spherical images. We compared the developed interface with a classical mouse-controlled one to evaluate the effectiveness of such an interface in terms of user acceptance and user engagement.

In Chapter 6, we describe a general framework to design a mid-air gesture vocabulary for the navigation of technical instructions in digital manuals for maintenance operations. A validation procedure is also proposed and utilized to compare gesture vocabularies in terms of fatigue and cognitive load.

In Chapter 7, we treat the facet of the Healthy Operator. We describe the design and development of a semi-automatic software tool able at monitoring the operator ergonomics in the shop-floor by assessing Rapid Upper Limb Assessment (RULA) metrics. We start by introducing the issues related to the monitoring of operator's ergonomics in the shop-floor and we analyze the works previously carried out in this field. Then we describe the design and development of our software prototype based on a low- cost sensor, the Microsoft Kinect v2 depth-camera. Subsequently, we validate our tool with two experiments. In the first one, we compared the K2RULA grand-scores with those obtained with a reference optical motion capture system. In the second experiment, we evaluate the agreement of the grand-scores returned by the proposed application with those obtained by a RULA expert rater. We also briefly introduce the development of a software

prototype for the assessment of Range of Movement (ROM) in the recovery process of injured operators.

Finally, we draw our conclusions regarding the work carried out and try to map out a path for the future development of our researches in these fields.

Chapter 1: IAR for the Augmented Operator

1.1. IAR supporting the Operator 4.0

Industry 4.0 applies the newest technologies related to ubiquitous sensing, Internet of Things (IoT), robotics, Cyber-Physical Systems (CPS), 3D printing or Big Data to improve the efficiency of the many processes that occur in the factory shop floor. In this new environment the operator is required to accomplish tasks whose complexity increases daily and demands the capacity to adapt quickly to the requirements of an often more flexible production process. This paradigm shift involves non neglectable changes of the way that operators perform their daily tasks. The new technologies can aid operators facing these new challenges by equipping with devices that act as an interface for human-machine communication and collaboration, and even as a Decision Support System (DSS) that would help to optimize their actions.

Augmented Reality (AR), and specifically Industrial Augmented reality (IAR), is one of the I4.0 enabling technologies that can support operators by providing powerful tools aiding the operators, helping them in assembly tasks, context aware assistance, data visualization and interaction (acting as a Human-Machine Interface (HMI)), indoor localization, maintenance applications, quality control or material management (Fraga-Lamas, Fernández-Caramés, Blanco-Novoa, & Vilar-Montesinos, 2018).

Azuma et al. (2001) defined the AR Systems to have the following properties: i) to combine real and virtual objects in a real environment; ii) to run interactively and in real time; iii) to geometrically align virtual objects and real ones in the real world.

Fite-Georgel (2011) defined IAR as the application of AR in order to support an industrial process.

By exploiting IAR technology, enterprises may achieve significant advantages:

- supporting the smart operator in real-time during manual operations by becoming a digital assistance system for reducing human errors;
- reducing the dependence on printed work instructions, computer screens and operator memory, which need to be interpreted first by a skilled worker;

- reducing defects, rework and redundant inspection by offering intuitive information and combining operator intelligence and flexibility with error-proofing systems to increase the efficiency of manual work steps, while improving the quality of work (Gorecky et al., 2013);
- incorporating a new human-machine interface to manufacturing IT applications and assets, displaying real-time feedback about smart manufacturing processes and machines to the smart operator in order to improve decision-making (Gorecky, Schmitt, Loskyll, & Zühlke, 2014).

One of the main advantages of AR is that it can help workers to accomplish several tasks, making it possible the shift from mass production to mass customization.

AR technologies for supporting operators have been an academic research topic for around 50 years now. Thanks to the major progress in the last decade the AR technology is getting closer to being implemented in industry and shows many industrial tasks and sectors where it can bring value (Figure 3).

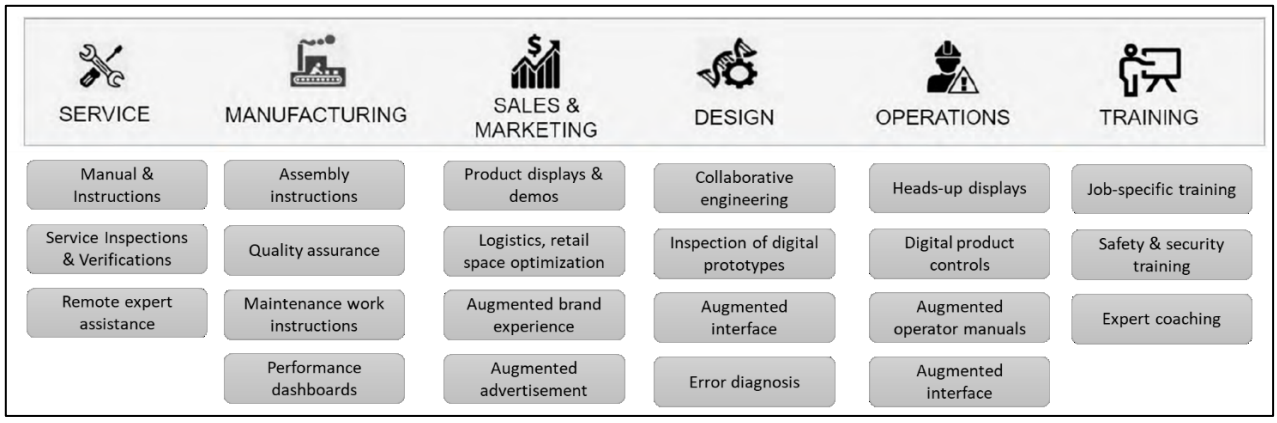


Figure 3 - Value of IAR across Industry 4.0. Source: (Fraga-Lamas et al., 2018).

Among these sectors we can underline those able to support the role of the Augmented Operator. One of the IAR most common applications for the Augmented Operator is the assistance to workers in assembly/maintenance/repair/control tasks through instructions with textual, visual, or auditory information (Aleksy, Vartiainen, Domova, & Naedele, 2014). By rendering this information ubiquitously, the worker perceives the instructions with less effort, thus reducing the operator's workload by avoiding the change from a real context to a virtual one where the relevant data is accessed.

Another relevant field of application of IAR is remote assistance. It enables monitoring, operating, and repairing machines remotely-located with the minimum amount of people on-site

(Bordegoni et al., 2014). Furthermore, IAR can help by easing remote collaboration between workers (Smparounis et al., 2008) or for collaborative visualization in engineering processes during stages related to design or manufacturing (Schneider, Rambach, & Stricker, 2017).

IAR can also assist workers in the decision making process, combining the physical experience together with the display of information retrieved from databases (Moloney, 2006). Furthermore, IAR can provide quick access to documentation like manuals, drawings or 3D models (M. Fiorentino, Uva, Gattullo, Debernardis, & Monno, 2014; Henderson & Feiner, 2011).

IAR may demonstrate useful during the training process, by giving step-by-step instructions to develop specific tasks (Tallig, Zender, & Runge, 2017). This helps especially when training workers to operate machinery like the one used for assembly in sequence, what reduces the time and effort dedicated to checking manuals (Horejš*ı*, 2015) and can reduce the training time for new employees.

IAR may also help to hand down the practical knowledge acquired by the most skilled workers. Proper IAR solutions can include such a knowledge in applications to train new hiring (Besbes, Collette, Tamaazousti, Bourgeois, & Gay-Bellile, 2012; Boulanger, 2004).

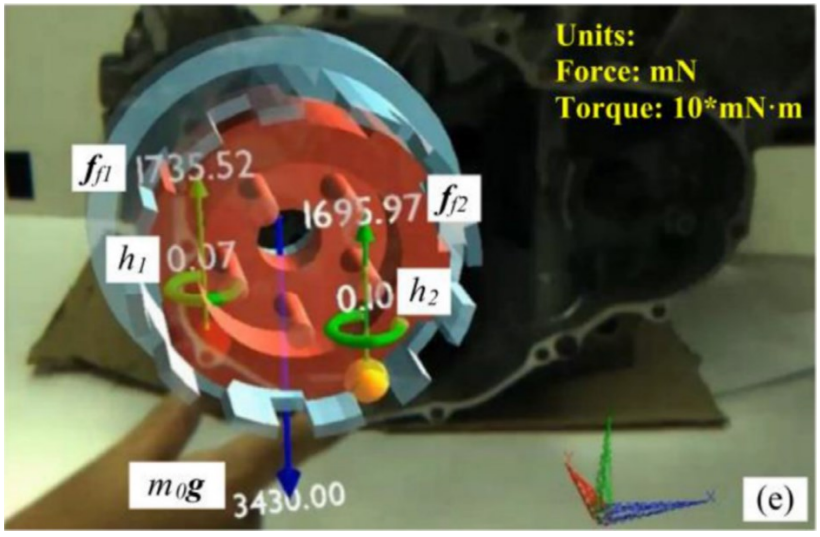


Figure 4 - Assembly planning through AR. The virtual component is overlaid on the real component. Forces are shown as arrows. Their magnitude is reported numerically. Source: (X Wang, Ong, & Nee, 2016).

By placing a virtual object (3D model) anywhere and observing at full scale whether it fits or not in a specific scenario, IAR can also support engineers while creating and evaluating designs and products (Nee, Ong, Chryssolouris, & Mourtzis, 2012) or during the assembly planning (X Wang et al., 2016) as shown in Figure 4. Furthermore, IAR enables providing on-site CAD model corrections, thus improving accuracy, alignment and other details of the model (Wuest, Engekle, Wientapper, Schmitt, & Keil, 2016).

By delivering the right information at the right time IAR can help manufacturing to avoid mistakes and increase productivity (Loch, Quint, & Brishtel, 2016; Paelke, 2014). Furthermore, in dangerous manufacturing tasks, IAR solutions could be used as monitoring and diagnosis tools.

IAR can support operators also by enhancing the efficiency of the picking process in warehouses by providing an indoor guidance system (Subakti & Jiang, 2016).

Despite the several studies aimed at introducing AR in the industrial environment, there are still some technical issues that prevent AR from being suitable for industrial applications (Palmarini, Erkoyuncu, Roy, & Torabmostaedi, 2018). In the following paragraph we will briefly describe the main components of an IAR system and will analyze the drawbacks related to its actual application in the industrial environment.

1.2. The IAR System components

An IAR system provides a view of a real-world physical environment that is combined with virtual elements. The overall system is composed of four main components:

- an element to capture a view of the real world (i.e., a Charge-Coupled Device (CCD), stereo, or depth-sensing camera);
- a visualization device aimed at returning the augmented view of the real world;
- a processing unit that outputs the virtual information to be added to the real world;
- a set of activating elements (e.g., images, GPS positions, QR markers, or sensor values from accelerometers, gyroscopes, compasses, altimeters or thermal sensors) that trigger the display of virtual information.

The hardware used as a visualization device can be categorized as follows:

- Hand-Held Displays (HHD). A screen is embedded in the device that can be handled by the user's hands (e.g., tablets, smartphones);
- Spatial Displays. They use digital projectors to show graphic information about physical objects. These displays ease collaborative tasks among users, since they are not associated with a single user;
- Head-Mounted Displays (HMD). They are the displays included in devices like smart glasses and smart helmets, which allow users to see the entire environment that surrounds them;
- Desktop PC or standalone video displays.

It is worth to define a topology for the display technologies that can be divided into two types: video-mixed (or video see-through) and optical see-through displays (e.g., projection-based systems).

In a video-mixed display, the virtual and real information, are digitally merged and represented on a display (Figure 5). This technology presents, among others, the disadvantage of providing a limited field of view. Moreover, because of the processing time needed to merge the acquired view of the real world and the augmented content, it introduces an unwanted latency (i.e. a time gap between what is happening on the real world and what is perceived by the eye) that induces unpleasant side effect on the user (for instance dizziness, nausea, and vomiting), especially for stereoscopic head-worn displays.

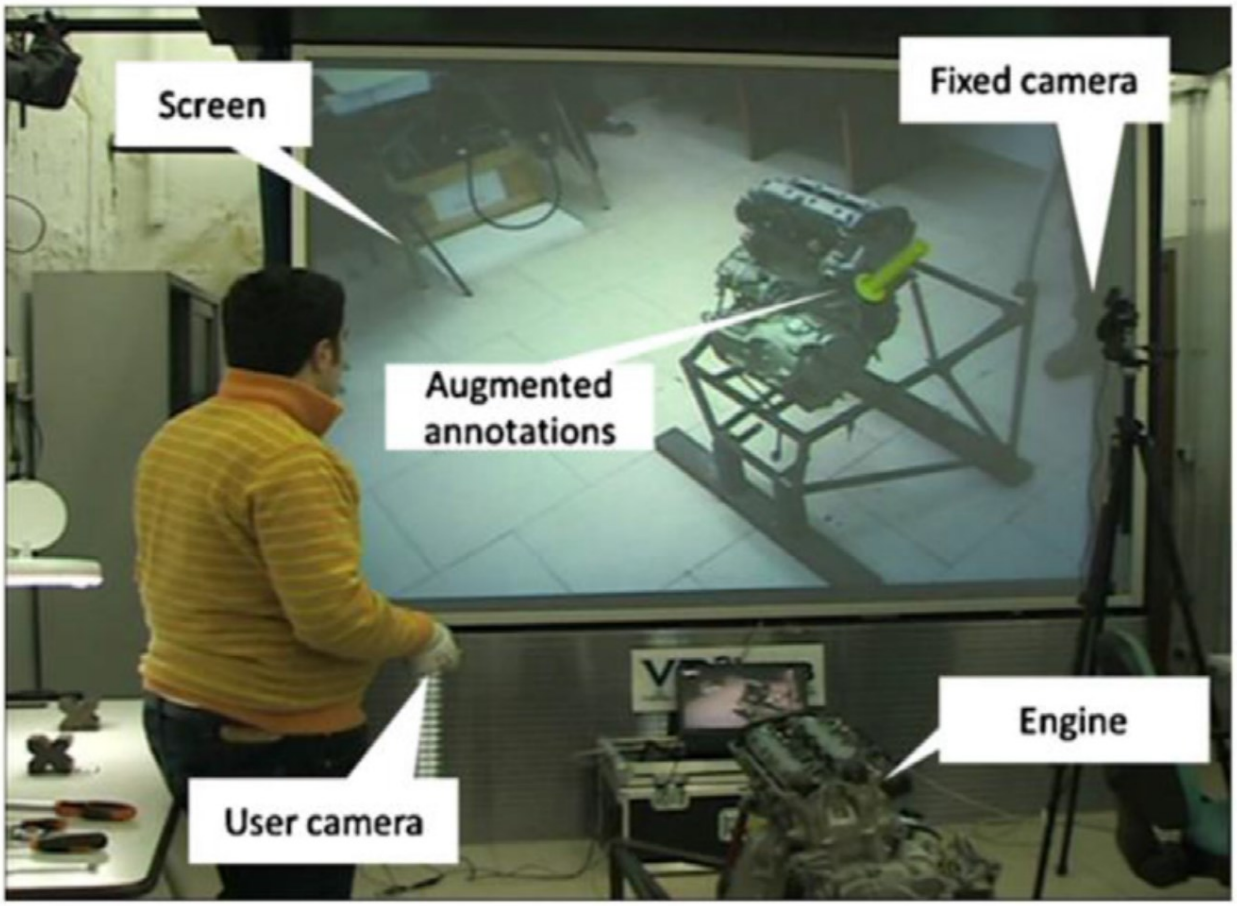


Figure 5 - Interactive AR instructions on a large screen using video-mixed technology. A motorbike engine on the bottom is captured by the user camera end projected on the screen. Fixed cameras enhance the tracking. Source: (M. Fiorentino et al., 2014).

Differently, optical see-through displays superimpose the augmented information on the user's field of view by means of an optical projection system. The technology is based on semi-transparent mirrors which allow the operator to “see-through” and, at the same time, are able to reflect computer generated images into the user’s eyes. They show a limited latency with respect

to the video-mixed technology but suffer from ergonomics issues related both to the different position of the focal plane of the augmented contents with respect to the real world and the weight of the devices that have to be head-worn. Furthermore, the use of such technology may suffer from the limited legibility of the augmented contents when superimposed on a background strongly illuminated or with texturization.

The internal logic of an AR system can be described as a pipeline composed of the functional modules shown in Figure 6.

The first component is the device camera that captures a frame of the real world. This frame is then processed by the AR software composed both by the Digital Image Processing and the Tracking modules. These modules estimate the camera position regarding a reference object (e.g., an AR marker) and track its position in the real world. Such an estimation can also make use of internal sensors, which also help to track the reference object. The accurate camera positioning is crucial while displaying AR content, since it also involves proper rotation and scaling according to the scenario.

The next element in the pipeline is the Interaction Handling module. It allows interactions with the displayed image and contents. It is crucial, while designing this module, considering the human-centric paradigm that characterizes the Industry 4.0 program in such a way to optimize the usability of the interactions and consequently the user's acceptance of the AR system.

The Information Management module is devoted to obtaining the information required to augment the real world. It acts as an IPA and can retrieve the needed information locally or remotely.

The last components of the pipeline are the rendering module and the display. They are devoted to render the image and the augmented content from the appropriate perspective and to present them to the user via the visualization device, respectively.

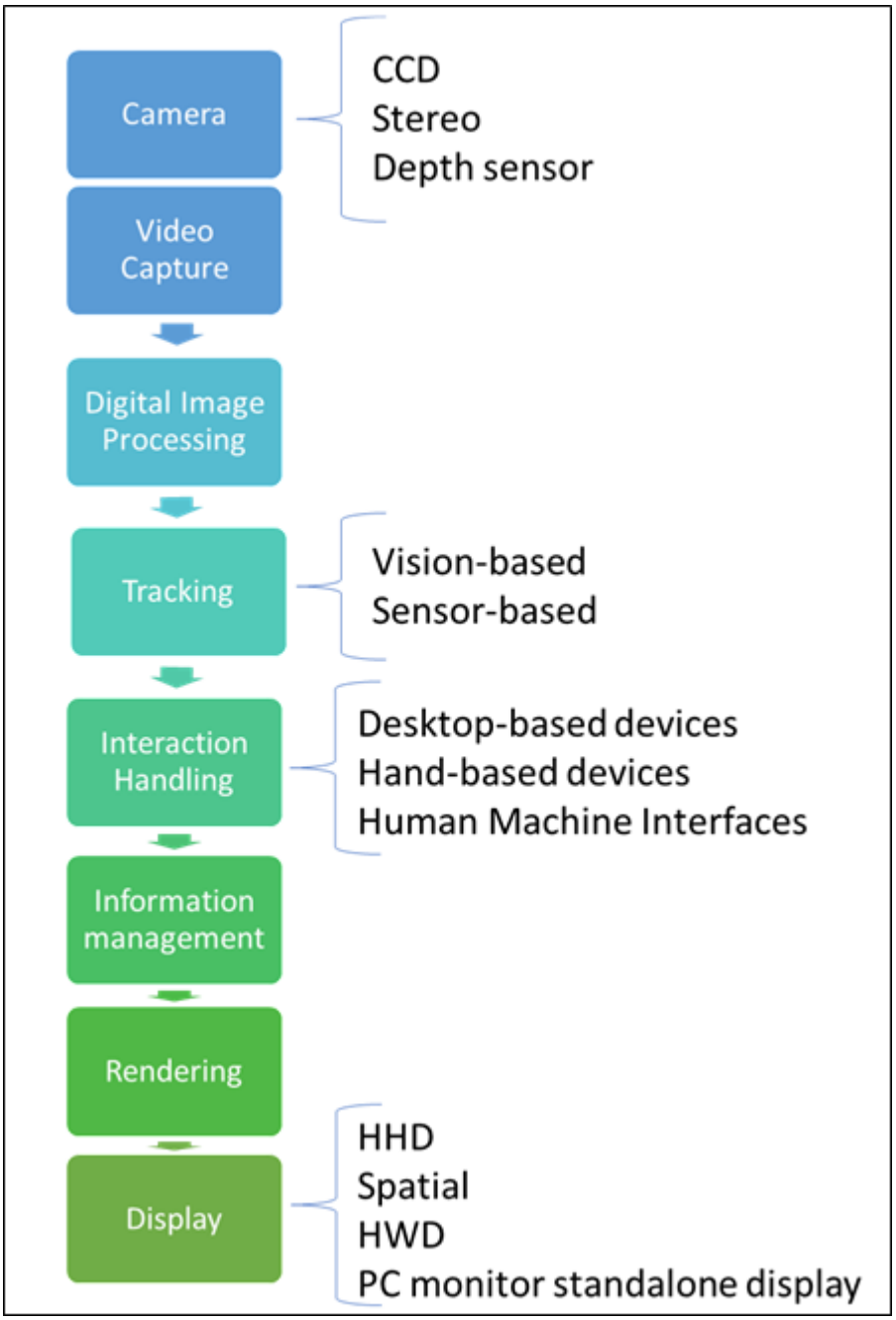


Figure 6 - A simplified AR pipeline.

The major technological challenges in this simplified pipeline are the internal registration of the objects displayed by the system, the tracking of visual elements and features, and the development by means of the authoring module of the information management system (Benko, Ofek, Zheng, & Wilson, 2015). Another technological challenge consists in the rendering (i.e., photometric registration, comprehensive visualization techniques, or view-management techniques) and the real-time data processing. Indeed, their performance is key when superimposing graphic elements in the environment in a rapid, consistent and realistic way.

Although not present in the simplified AR pipeline, there is another fundamental component involved in the development of an AR system, the authoring. By authoring is meant the process of creating digital contents for augmenting the reality (Ramirez, Mendivil, Flores, & Gonzalez, 2013).

1.2.1. Tracking

As already stated, the tracking process plays a crucial role in the development of an AR system. According to Ong, Yuan, & Nee (2008), an accurate tracking, which locates the users and their movements in reference to their surroundings, is a crucial requirement for an AR application. Siltanen (2012) defines it as the "heart" of these systems: it calculates the relative pose of the camera in real time. The pose consists of the position and orientation (6 DOF) of an object. Tracking relates to the ability of a system to anchor virtual content in the real world such that it appears to be a part of the physical environment (Billinghurst, Clark, Lee, & others, 2015).

Tracking techniques can be visual-based and sensor-based (Figure 7), additionally, it is possible a hybrid approach by using both the techniques at the same time.

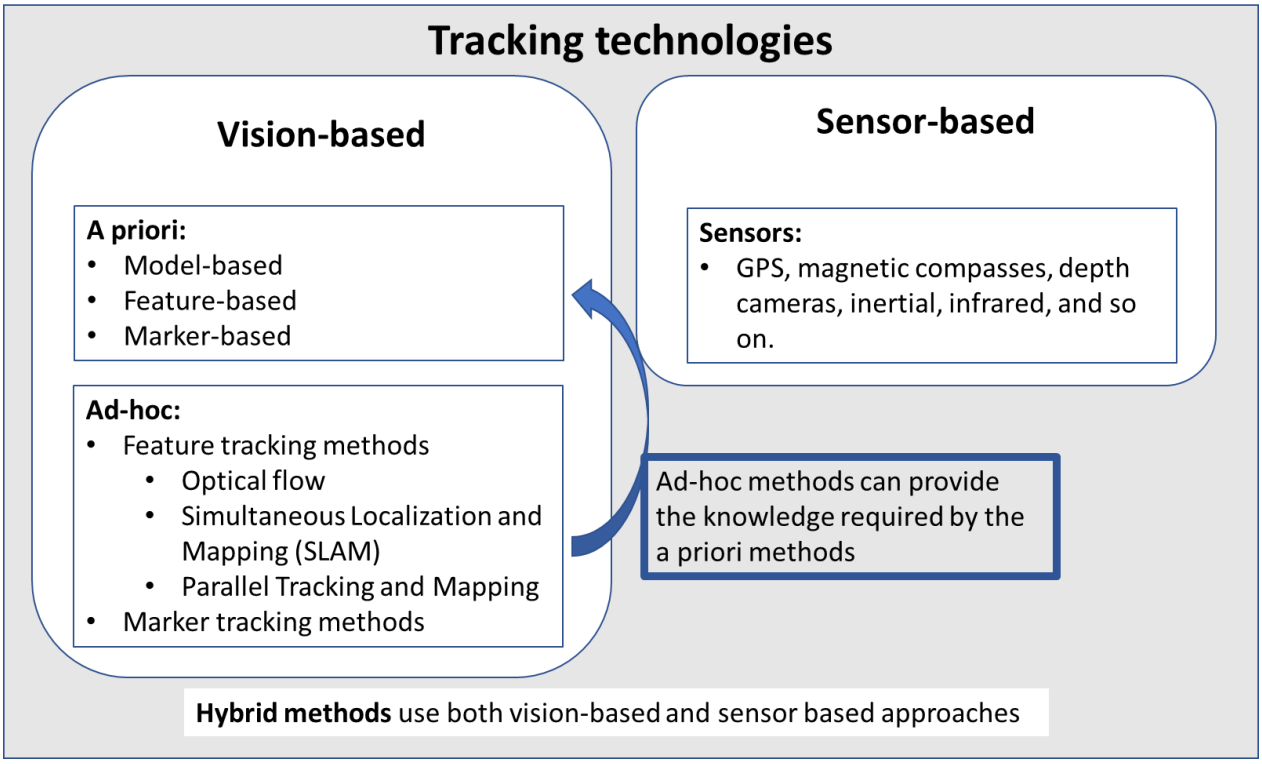


Figure 7 - Scheme of the tracking approaches. Source: (Palmarini et al., 2018).

Visual-based tracking techniques can be further divided into two categories:

- a priori methods;

- and ad-hoc methods.

The first category implies that the AR system has an “a priori” knowledge about the object that will be tracked. On the bases of this knowledge they can be divided into: model-based, feature-based and marker-based. It means that the information available a-priori are respectively: a model, a feature-map and a marker. The information needed in an “a priori method” can be created utilizing an “ad-hoc” visual tracking method hence providing the initialization of the a-priori visual tracking method (Siltanen, 2012).

The first step before the tracking process consists in the recognition that aims at estimating the camera pose without relying on any previous information provided by the camera. Recognition represents the initialization phase of the AR system and happens whenever there is a tracking failure. The tracking process aims at tracking the camera pose based on the previous frame provided by the camera.

In its systematic literature review on AR industrial maintenance applications, Palmarini et al. (2018) analyzed 30 articles and found that 90% made use of “a priori ”vision-based tracking techniques (Figure 8).

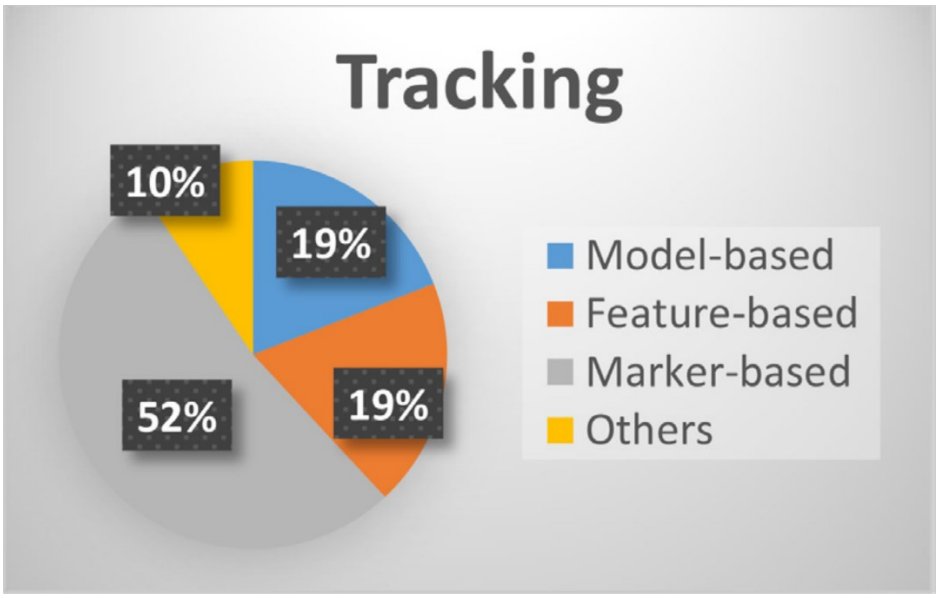


Figure 8 - Tracking techniques mentioned throughout the Systematic Literature Review provided by Palmarini et al. (2018).

Thanks to the diffusion RGB cameras across the different hardware utilized for AR, the vision-based methods are generally preferred with respect the other ones. The a-priori tracking is considered more robust and reliable than the model-based and a-priori feature-based ones since it is independent of environmental conditions (lighting, materials, etc.). The limitation resides

on the availability of the CAD models. Anyway, CAD based tracking may solve issues such as partial occlusions and rapid motion (Platonov, Heibel, Meier, & Grollmann, 2006).

The marker-based approach consists of placing physical markers onto the scene that has to be augmented. The configuration of markers requires a proper design. These markers, their position and orientation to be maintained on the real object, are registered a-priori on the AR system. Hence, by recognizing the marker the system is able to recognize the object. This method that is robust and accurate in controlled conditions, might not be so in an industrial environment. Indeed, its limitations rely on the visibility of the marker which might not always be in the frame of the camera. In an industrial environment, the markers require to be maintained (clean and not damaged) in order to perform properly. Furthermore, there are a lot of objects which could occlude the vision of the marker (people, tools, machinery, etc.) According to Liverani, Amati, & Caligiana (2006), this would cause the tracking failure of the AR system. For these reasons, the marker-based approach must be applied with caution in the industrial environments.

An alternative to artificial markers is provided by natural markers consisting of fiducial images which already exist in the environment and that do not need to be placed in the facility. The natural markers tracking-technique utilizes the Natural Feature Tracking (NFT) to extract characteristic points from images (Okuma, Kurata, & Sakaue, 2004). Such points are then used to train the AR system to detect them in real time. NFT techniques have the disadvantage of being more computationally intensive and slower than other alternatives. Furthermore, these techniques are less effective at long distances, but they provide a seamless integration of the augmented information with the real world, since there is no need for placing an artificial marker visible in the scene (Z. Chen & Li, 2010). To speed up the NFT technique it is possible to add small artificial markers in the scene. These markers, called fiducial markers, help to accelerate the initial recognition, improve the performance of the system, and reduce the algorithm computational requirements.

An alternative to the vision-based tracking consists of the sensor-based techniques. These techniques rely on sensors such as magnetic, acoustic, inertial, optical and/or mechanical ones together with the Global Positioning System (GPS). However, the GPS localization is unfeasible in indoor places such as the industry shop floor, furthermore, the precision and the update rate are not sufficient for accurate tracking.

Magnetic trackers exploit the properties of magnetic fields in order to calculate the pose of a receiver with respect to a transmitter, which is used as the real-world anchor (Billinghurst et al., 2015). These sensors provide high update rates, are invariant to occlusion and optical

disturbances, and the receivers are small and lightweight. However, the strength of magnetic fields falls of with the cube of distance and resolution falls off with the fourth power of distance. Furthermore, magnetic trackers are prone to measurement jitter, and are sensitive to magnetic materials and electromagnetic fields in the environment. Consequently, the factory shop floor does not represent their best application scenario.

Inertial tracking uses Inertial Measurement Unit (IMU) sensors such as accelerometers, gyroscopes and magnetometers to determine the relative orientation and velocity of a tracked object. Inertial tracking allows the measurement of three rotational degrees of freedom (orientation) relative to gravity, and the change in position of the tracker.

These sensors have "no range limitations, no line-of-sight requirements, and no risk of interference from any magnetic, acoustic, optical or Radio Frequency interference sources. They can be sampled as fast as desired, and provide relatively high-bandwidth motion measurement with negligible latency." (Foxlin, Altshuler, Naimark, & Harrington, 2004).

Inertial sensors have the limitation of being very susceptible to drift over time for both position and orientation. This is especially problematic for position measurements, as position must be derived from measurements of velocity.

Another available tracking technology is that of 3D structure tracking. It relies on sensors capable of detecting 3D structure information from the environment. These sensors commonly utilize technologies such as structured light (for instance the Microsoft Kinect V1) or time-of-flight (for instance the Microsoft Kinect V2) to obtain information about the three-dimensional positions of points in the scene.

Hybrid tracking systems fuse data from multiple sensors to add additional degrees of freedom, enhance the accuracy of the individual sensors, or overcome weaknesses of certain tracking methods (Billinghurst et al., 2015).

1.2.2. **Authoring**

According to the classification presented by Palmarini et al. (2018), the authoring approaches for the development of an AR system can be broadly categorized into four categories, based on the procedure and methods utilized by the authors to create the augmented contents:

- Manual;
- By annotations;
- By “boxes”;

- Automated.

In the manual authoring process the contents are manually generated. It includes not only the creation of the 3D/2D dynamic/static models, but also their implementation in the AR system (location, orientation, etc.). This authoring approach is demanding in terms of time and required skills such as programming, modelling and animations (Ramirez et al., 2013).

The annotations, the "boxes" and the automation approaches have been developed to overcome the authoring problems related to the manual method.

The annotation approach consists of the capability of adding virtual annotations to a real environment. According to Klinker et al. (2001) annotations are a set of primitive tasks (for instance a plant maintenance set of primitive tasks are: highlight, label, display information (text), clear information, edit information, set compass, hide/show.) The annotation approach fails with 3D dynamic and static contents that cannot be generated through annotations. These limitations can be overcome by manually attaching 3d annotations to an image and by using SLAM techniques for the correct registration into the environment as proposed by (Alvarez, Aguinaga, & Borro, 2011), or by using an ad-hoc web-based annotating system for attaching notes to 3D models as proposed by Jung, Gross, & Do (2002) and Nee et al. (2012).

The "boxes" method, by applying task by task procedures, aims at reducing the computer programming knowledge required in the development of an AR system. In this approach the operations involved in the operational tasks correspond to the boxes. This concept, utilized in the field of maintenance by Havard, Baudry, Louis, & Mazari (2015), models maintenance operations for AR as:

- Entity: the smallest part of the system to maintain (e.g. nuts, plates);
- External Entity: the smallest part external to the system to be maintained (e.g. tools);
- Actions: the activities to be performed (e.g. push, pull);
- Maintenance: a series of actions;
- Operation: list of maintenance operations.

Each different maintenance task or different operation can be obtained by switching the boxes or changing their order. Previously, our research group (M. Fiorentino et al., 2014) applied the same approach, even though utilizing a different nomenclature. The proposed authoring tool consisted of a set of actions that could be recalled to the AR application though an excel table. This authoring solution does not require any programming skill. An attempt to improve this approach has been made by Zhu, Ong, & Nee (2014) that proposed "A context-aware augmented

reality system to assist the maintenance operators" by allowing technicians and operators to access the authoring log (Figure 9), and modifying the contents provided by the AR developer in each box.

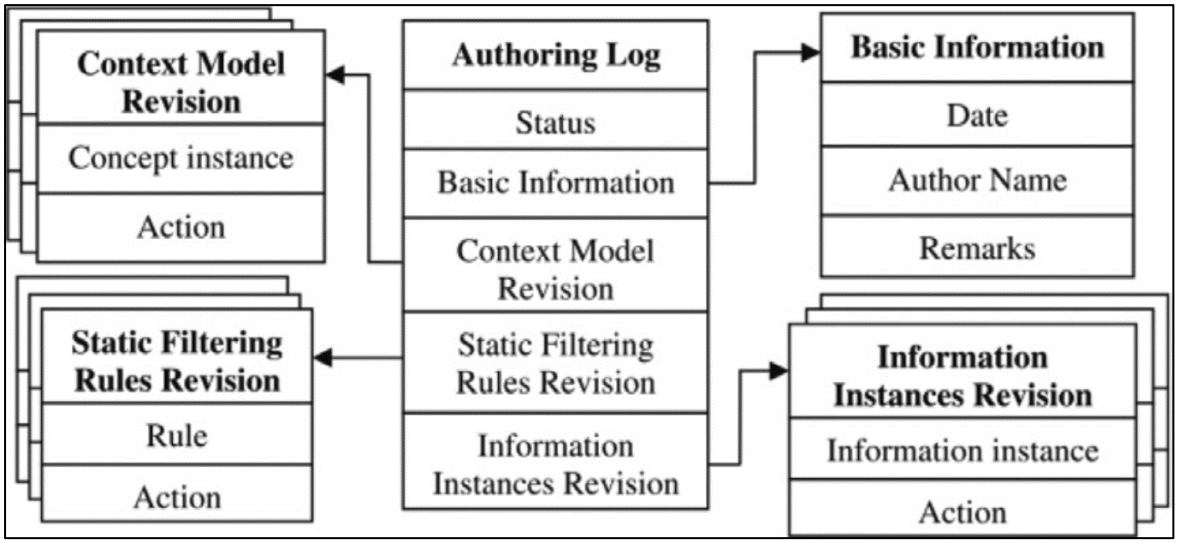


Figure 9 - Technicians authoring log proposed by (Zhu et al., 2014). *In the center, the Authoring log. On the sides, the menu connected with different modules of the authoring log. This is visualized by the technician through a device and the interaction is made by buttons.*

This AR system is mainly based on textual information, but the designed interface also allows to insert media files, modify visual properties and apply rendering rules. In any case, the smallest entities or nodes of an authoring solution by boxes have to be available or manually created. The reconfiguration of a procedure is limited to the boxes available in the system.

The automated authoring solution has been applied only to assembly and disassembly procedures (Palmarini et al., 2018). These procedures are created automatically based on the CAD models and dis/assembly planning theory. Starting from all the possible configurations of the CAD model, (Alvarez et al., 2011) has been able to automatically extract the disassembly procedure by merging the information of the disassembly-planning module and the CAD model constrains.

As noticeable after reviewing the state of the art, AR applications may provide a useful tool for the Augmented Operator in the smart factory. Anyway, developing a successful IAR application requires keeping in mind some fundamental aspects (Fraga-Lamas et al., 2018):

- The use cases and applications selected must provide added-value services;
- Functional discontinuities or gaps in the operating modes that can affect the functionality should be avoided;
- Reduce cognitive discontinuities or differences between old and new work practices. Learning new procedures may hinder the adoption of the technology;

- Reduce physical side-effects caused by the devices on users in the short and long term (e.g., headaches, nausea, or loss of visual acuity);
- Avoid unpredicted effects of the devices on users unfamiliar with the technology, like distractions, surprises or shocks;
- Take into consideration the user perception regarding ergonomic and aesthetic issues;
- Make user interaction as natural and user-friendly as possible, avoiding lapses or inconsistencies.

Additionally, there are still some problems that prevent a large spread of AR applications across the industrial scenario. Among these, we must remind those related to the tracking systems, those related to the application of a user-centric design-approach, and those related to the legibility of the augmented contents.

In the next chapter we will describe the design and development of a prototype solution to serve the Augmented Operator in assembly/maintenance tasks, and the experiments we carried out to verify the effectiveness of such tool in enhancing the operator's performance.

Chapter 2. Design and test of a projective AR workbench for manual working stations

During our doctoral program we cooperated at the design and prototype of a projective AR workbench for an effective use of the AR in industrial applications for Manual Working Stations. The proposed solution consists of an aluminum structure that holds a projector and a camera that is intended to be mounted on manual working stations. The camera, using a tracking algorithm, computes in real time the position and orientation of the object while the projector displays the information always in the desired position. We also designed the data structure of a database for the managing of AR instructions, that can be accessed interactively from our application. Furthermore, we carried out user studies to assess both the effectiveness of conveying technical instructions with this SAR prototype as compared to paper manual and users' acceptance. In this chapter[1] we will describe the design of such a prototype and the user test carried out for validating its effectiveness.

2.1. Introduction

In his survey of IAR applications Fite-Georgel (2011) observed that AR technology could support the operator throughout all the phases of the product life cycle:

- Product design and factory planning;
- Manufacturing;
- Commissioning;
- Inspection and maintenance;

[1] The results of the studies described in this chapter have been published in the following articles:
Uva, A. E., Fiorentino, M., Gattullo, M., Colaprico, M., Ruvo, M. F. de, Marino, F., Trotta, G. F., et al. (2016). Design of a projective AR workbench for manual working stations. International Conference on Augmented Reality, Virtual Reality and Computer Graphics (pp. 358–367). Springer;
Uva, A. E., Gattullo, M., Manghisi, V. M., Spagnulo, D., Cascella, G. L., & Fiorentino, M. (2018). Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. The International Journal of Advanced Manufacturing Technology, 94(1-4), 509–521.

- Redesign and decommissioning.

Looking at how AR has already been used for each one of them, those that are less dependent from the product itself, are manufacturing, commissioning, inspection and maintenance. We can observe that in those three phases, except for certain product categories (e.g. cars, planes, plants), the product is handled into a Manual Working Station (MWS). During the execution of the common operations that are accomplished into an MWS (for instance: assembly, welding (especially spot welding), packing, testing, repairing, inspecting), the operator has to follow some strict procedures and s/he is supported by information that can be provided in an AR mode, with all the benefits, which are widely discussed in the literature (Azuma et al., 2001; Billinghurst et al., 2015; Dünser, Grasset, & Billinghurst, 2008; Fraga-Lamas et al., 2018; Navab, 2004; Ohshima, Kuroki, Yamamoto, & Tamura, 2003; Regenbrecht, Baratoff, & Wilke, 2005; Romero, Stahre, et al., 2016; Thomas, 2009; Webel, Becker, Stricker, & Wuest, 2007). The work described in this chapter aims at confirming and making substantial the opportunity to use AR in the industrial sector. The motivation for this optimism arises from the literature and was strengthened by the previous work of our research group that found how AR instructions reduced significantly participants' overall execution time and error rate in manual assembly tasks.

The scenario of the MWS allowed us to develop a prototype that can easily meet the requirements requested by Navab (2004) for developing IAR applications, i.e.:

- reliability: high accuracy, fallback solutions;
- user-friendliness: AR system safe and easy to set up, learn, use, and customize;
- scalability: setup easy to reproduce and distribute in large numbers.

The MWS scenario is effortlessly reproducible in a laboratory, so we could look for all the possible bottlenecks and try to solve them in order to have a reliable application.

Furthermore, we could test the application with focused user tests, and thanks to their feedbacks, it was possible to create a user-friendly solution.

Finally, about the scalability, we aimed at finding the hardware and software solutions that were the less dependent as possible to the product and operations to be accomplished in the MWS.

We analyzed the available AR display technologies and, basing on their pros and cons, on the previous researches carried out by our group, and on the requests deriving from interested enterprises, we decided to use Spatial Augmented reality (SAR). SAR is based on digital projectors that superimpose the augmented contents (text, symbols, indicators, etc.) directly on

the real environment (Bimber & Raskar, 2006). However, being SAR, a relatively new technology it still requires some feasibility studies and optimization processes before its feasible application in the industrial environment. One of the most important issues is related to the correct visualization of technical information. In particular, we evaluated the possibility to project text directly on workbench surfaces (without the need to calibrate the scene), comparing users' performance with that deriving from the use of a normal LCD monitor. This because, in a real working environment, the operator stands in front of the workbench and is currently assisted by instructions on monitors usually placed on their workbenches or on tool carts.

The main contributions of this research are both the description of the SAR-aided MWS design and prototype and its evaluation to verify if the proposed solutions were effective to aid users in the visualization of technical instructions. In fact, it is known for a long time that AR can help for the support of technical operators. However, this was plenty demonstrated for traditional AR, but scarcely for SAR.

The distinction is not negligible because the visualization capabilities of SAR are poorer than those of traditional AR. For example, it is not possible to display mid-air virtual objects, but only 2D objects on a physical surface. Furthermore, surface-based distortions remain an issue, and the user could misunderstand the projected instructions.

2.2. Related works

The related works have been split into two sections: the first one relates to the use of SAR in IAR domain; the second one regards user evaluation of AR systems in industrial applications.

2.2.1. Applications of SAR in the industry

In the recent years, we are assisting to an increase in the number of works where SAR is used to perform assembly and maintenance tasks.

The traditional use of SAR in the industry is for the spot localization in spot-welding operations. Doshi, Smith, Thomas, & Bouras (2017) improved the precision and accuracy of manual spot-weld placements with the aid of projected visual cues. The prototype system, deployed at a General Motors (GM) plant, allowed a reduction of 52% of the standard deviation of manual spot-weld placement when using augmented reality visual cues. This reduction demonstrates the benefit of having an AR projection system on the production line as opposed to not having one. The visual cues helped to identify spot-weld locations quicker and exhibited the potential to be used as a training tool for new operators or new vehicle types. The SAR projection system can

be positioned at any manual production station for spot welding, welding, adhesive application, and inspection.

Schwerdtfeger, Pustka, Hofhauser, & Klinker (2008) explored the use of laser projectors as an alternative to head-worn displays for industrial augmented reality applications. They used two scenarios to gather practical experiences with AR laser projectors: quality assurance and maintenance. They concluded that laser-based augmentations have much potential and they said they were in the process of installing, evaluating, and further developing the system in the context of real industrial applications.

Sand, Büttner, Paelke, & Röcker (2016) presented a prototype that has many similarities with our concept, such as the frame holding the projector and the camera. "smARt.assembly" is a projection-based augmented reality (AR) assembly assistance system for industrial applications. This system projects digital guidance information such as picking information and assembly data into the physical workspace of a user. Compared to their previous prototype with smart glasses, the projection-based system seemed to be much more robust on changing light conditions. Furthermore, users working with the projection-based instructions were able to assemble products faster and with a lower error rate, compared to the system based on smart glasses.

In summary, the use cases of SAR found in the literature are the following:

- Spot-welding;
- Quality assurance;
- Maintenance;
- Designing;
- Assembly.

2.2.2. User evaluation of AR in the industry

The integration of augmented reality with assembly and maintenance operations is a topic that has been widely addressed in the last years. Xin Wang, Ong, & Nee (2016) presented an extensive review of researches on AR in assembly tasks reported between 1990 and 2015. They divided the literature related to AR assembly research into three categories: AR assembly guidance; AR assembly training; and AR assembly simulation, design, and planning. At the end of this review, they remarked the importance of the familiarization for both users and developers with the new proposed AR assembly solution.

A previous study of our research group (M. Fiorentino et al., 2014) performed a user evaluation test using video-based AR with a large screen display near the operator workbench, and a combination of multiple fixed and mobile cameras. Participants performed similar operations in two modalities: paper manuals and AR instructions. In the AR mode, visual labels, 3D virtual models, and 3D animations supported tool selection, bolts removal, part (dis)assembly. Statistical analyses showed that AR instructions displayed on the large screen significantly reduced participants' overall performance-time and error rate. Furthermore, user acceptance of the AR system was higher than the paper manual.

Arthur Tang, Owen, Biocca, & Mou (2003)compared assembling operations of toy blocks with four different instructional modes: i) printed manual, ii) 2D static images on a laptop with 15" display, iii) static images on a see-through HMD, and iv) a spatially aligned 3D model on the see-through stereo HMD using magnetic tracker (full AR). They measured, time of completion, the number of errors, and mental workload. As far as regards the time of completion, their test verified a relevant improvement passing from the paper manuals to the computer-supported ones. However, only with full AR, there was a significant reduction even of error rates and mental workload with respect to the printed manual. On the contrary, full AR did not show a significant time advantage compared to other computer-assisted approaches.

Tegeltija et al. (2016) implemented AR in the disassembly of a heating circulation pump. The experiment was carried out on three different types of the pump and with the disassembly procedures done in four variant modes. These modes were the combination of using AR and hard copy documentation and performing the disassembly sequence with the worker who has and has not previously performed pump disassembly. Observing the results of the experiment, the authors concluded that the implementation of AR in disassembly systems could lead to a series of time reductions in the disassembly process. They did not make analyses about the accuracy of the procedures.

Elia, Gnoni, & Lanzilotto (2016) proposed a multi-criteria model to support quantitatively production managers and researchers in evaluating the performance of different AR devices based on both technological and organizational performance required by the specific manufacturing process. They developed an expert-based tool to support, in a more efficient way, the design of AR applications in specific manufacturing issues. They validated the model with a test case of a complex maintenance task carried out in a harsh workplace. Even if their test case aimed only at demonstrating the actual applicability of the proposed multi-criteria model, their results are interesting for our work. In fact, they found that projectors are the best AR alternative

in a set also containing HWDs, haptic and handheld devices. Projectors were also the most "agile" devices, whereas their reliability was lower respect to handheld and haptic devices.

In the next section we list the system requirements and in section 3 we describe the proposed system and the prototype. Section 4 presents the preliminary trial results and the subsequent improvements.

2.3. System requirements

2.3.1. The choice of the SAR technology

As noticed by (Palmarini et al., 2018) most of the AR solutions in literature employ Head Mounted Displays (HMDs or Head Worn Displays HWDs), which have several drawbacks in terms of tracking, ergonomics, cost, limited field of view, low resolution, encumbrance and weight, and limited/fixed focal depth (Van Krevelen & Poelman, 2010). Hence, we discarded this option for three main motivations:

- in the application scenario (i.e. maintenance, inspecting, repairing, testing) HWDs prevent the operator from having a real perception of the objects that s/he is handling, hence involving potential dangers;
- it is likely that, when accomplishing her/his working task, the operator could occlude the field of view of the camera devoted to the tracking system;
- at the time of designing the prototype, HWDs needed a cable connection to the computer where the application is running. This physical link, along with wearing issues may cause ergonomic problems.

An alternative approach to HWDs is to use handheld devices like smartphones and tablets. Although this kind of AR is very easy to implement due to the availability of low cost and powerful devices, in practice, it has various limitations. One of the most important is that operators should employ one or even two hands for the visualization, thus limiting their ability to operate.

The use in an AR application for maintenance tasks of large screen display technology significantly improves the operators' performance with respect to a traditional approach based on paper manuals (M. Fiorentino et al., 2014).

However, the use of large screens is not an easily scalable solution in various industrial scenarios. Indeed, not everywhere there is enough space for a large screen. Furthermore, such displays suffer from a high angular offset, i.e., the offset between the real world being observed and the display through which the user sees it. A high angular offset may lead to alignment problems: users may not readily understand the relationship between what is seen directly in the real world and what is shown on the screen when comparing both, which may require difficult mental rotation and scaling (Kruijff, Swan, & Feiner, 2010).

A display solution that does not suffer from the problems above (ergonomics, not free hands, angular offset) is represented by spatial augmented reality (SAR) using digital projectors. Thus, it could be an optimal solution for the visualization of both instructions and technical information directly on the industrial workbench.

The main disadvantages of SAR, as pointed out by Kruijff et al. are the following (Kruijff et al., 2010):

a) surface-based distortions;
b) brightness, contrast, and visibility of projection;
c) lower color fidelity;
d) higher latency;
e) limited FOV of the projector;
f) registration.

Surface-based distortions may cause problems with legibility of projected text. A previous study of our research group evaluated if it were possible to effectively read the instructions directly projected on the workbench. It revealed that projecting text directly on workbench surfaces guarantees the same legibility performance of a traditional LCD monitor (Di Donato, Fiorentino, Uva, Gattullo, & Monno, 2015).

Basing on these observations we chose the SAR technology and designed the SAR-aided MWS trying to solve all the other potential issues.

2.3.2. **Requirements**

Taking into account the issues related to the sensitivity to lighting conditions of vision-based tracking methods we added an external lighting system to control the environmental lighting.

As regards contents, previous researches evidenced that the use of animated 3D virtual objects is not essential. The time required to create an animation is often more than the benefits that the

animation gives, especially for experienced workers. This encouraged us to use projective AR and the consequent use of 2D graphic signs to indicate objects or operations to accomplish in the real scene.

2.4. The Prototype

2.4.1. Technical features

The main technical features that we defined for the prototype are:

- A light frame to be clasped on a normal workbench of a Manual Working Station; the dimensions of the reference workbench are 1200 mm (L) x 1000 mm (D). The frame should be designed to hold multiple cameras and projectors at an adjustable distance;
- A turntable where the product to be assembled/maintained should be fixed. In many cases, this additional frame is already present to facilitate the operators in the maintenance tasks. We decided to use it in our system also to locate the tracking markers. This is because marker position on the product may vary during the working steps, (e.g. assembly);
- The Vision-based tracking-system is based on fiducial markers (artificial markers) that are glued on the turntable. A multi-marker technique is used with the markers placed so that they are almost always visible from the camera at least one of them;
- An additional lighting system must be provided, to have a uniform lighting on the markers;
- The virtual contents that the projector should display are 2D graphic signs such as circles, arrows, squares, crosshairs, and so on;
- The user interface is projected directly on the workbench and contains a menu with the tree of the operations performed/to perform and the possible subtasks of the current operation;
- The contents navigation could be done with virtual buttons or other Natural User Interfaces like voice recognition and gesture recognition.

2.4.2. Contents

The application framework developed for this prototype is based on the general structure of an assembly/maintenance manual. Previously, our research group, by reorganizing the information about the maintenance steps in a tree-like structure with different levels of detail experiences (M. Fiorentino et al., 2014), allowed the operators to gain in efficiency as compared to paper manual,

skipping well known details while accessing specifics only if needed (learning, troubleshooting, and so on.). Therefore, we assumed that the following four commands may be sufficient to effectively navigate the manual:

1. Next. A maintenance task, at the current level of detail, is clear or completed. The user wants to access the information about the next task at the same level of detail;
2. Previous. The user wants to access the information about the previous task at the same level of detail;
3. Go down (to a lower level). A more detailed information is required, therefore the current task is expanded in a more detailed sequence of sub-tasks (e.g., unknown task or troubleshooting);
4. Go up (to an upper level). The user needs fewer details, therefore s\he navigates through a sequence of less detailed tasks.

Going up to the first level brings the user to the root node of the manual. Considering this structure for the manual, the user of the application can go back and forth from a step to the previous/following, up a level to go the main menu, down a level to access further details for that step.

2.4.3. System Architecture

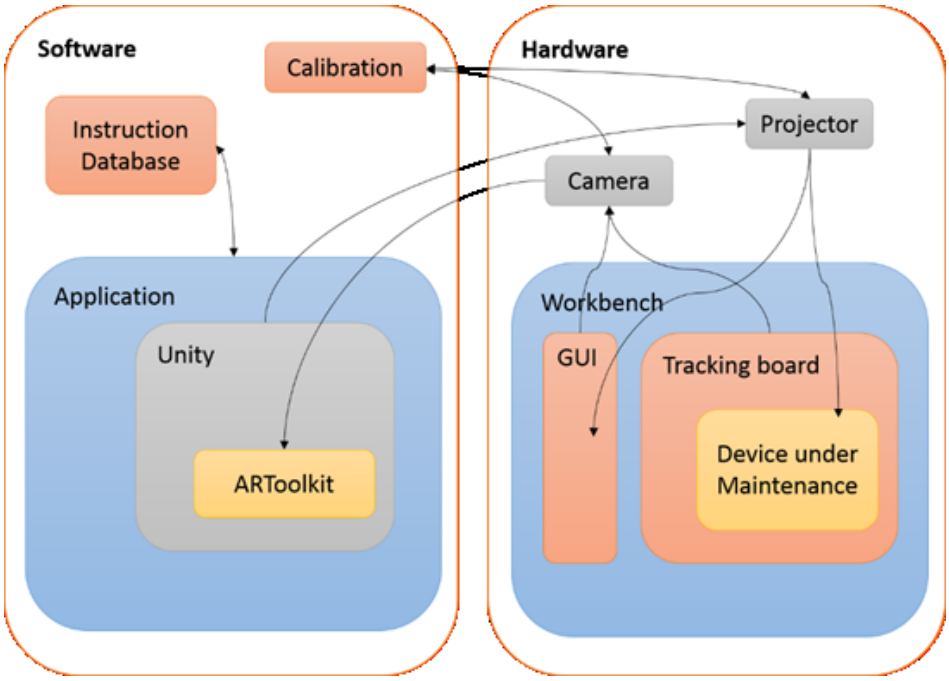


Figure 10 - The System architecture.

The system architecture is schematized in

Figure 10. The hardware side of the system architecture is composed by a camera for tracking and a projector for visualizing the augmented contents on the workbench. The device under maintenance is located on a tracking board is supported by four wheels allowing for easy translation and spinning on the work-bench. Four artificial markers are stuck on the corners of the boar to allow a feasible tracking system robust with respect occlusions. Furthermore, we took into consideration structural issues and the geometrical location of components for visibility and occlusions when designing the physical structure with the CATIA CAD software (Figure 11). The beamer projects also a GUI for managing the procedure and the info. The software side, running on one PC, is composed by an application based on Unity 3D and the tracking module based on ARToolkit. In our framework, each working step is a Unity scene file that includes all the contents needed for the display of the information, namely CAD models of the virtual objects, their placement on the scene (position/orientation), their possible animation, a text list of the needed tools, and text instructions. Each scene component is conveniently stored in a SQL database managed by a MySQL RDBMS, linked using an MYSQL connector for Unity 3D. Each interaction between the operator and the system correspond to a SQL query. The result of a query is a collection of data used to instantiate at run-time a scene component.

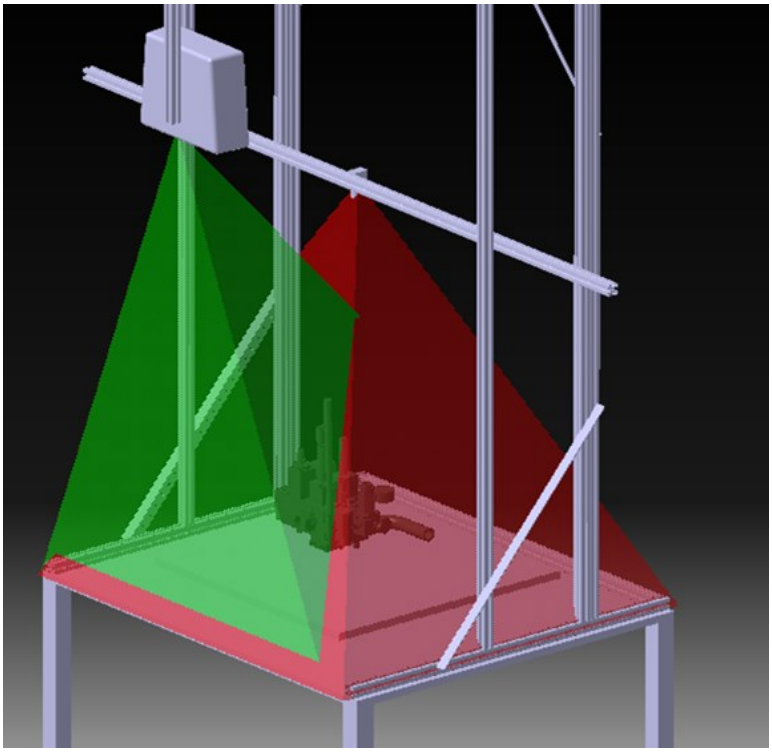


Figure 11 – The CAD design of the SAR-aided MWS prototype.

2.4.4. **The Physical setup**

As workbench, we used a table adjustable in height (the height was set at 750 mm from the floor), 1200 mm large and 1000 mm deep. The components are installed on a physical structure realized with Bosch Rexroth aluminum profiles 30x30 providing a very good physical stability of the structure (Figure 12).

As tracking board, we used a square table (side 650 mm) where we put on the product to be maintained; it can freely move on the table due to four wheels mounted beneath the base. At the corners of the table we fixed four 140 mm artificial markers used for the tracking. In this way, the markers move jointly to the product, but they are not fixed on it, so this solution is scalable to different products to be maintained.

Figure 12 - The MWS prototype.

We used a Benq W1080ST+ projector mounted at a distance of about 1300 mm from the table and from the side of the user. In this way the projector illuminates mainly the area in front of the user. The distance from the table is set in order to illuminate all the width of the table, whereas part of the height is not illuminated. The resolution used is 1600 x 1200. The projector was fixed at the end of an aluminum bar whose height can be changed, in order to have a scalable solution, i.e. we can use the same frame even if we change the projector. For tracking purpose, we used an Imaging Source DFK 23U445 (1/3" CCD sensor, resolution 1280 x 960, USB 3.0) with a 4 mm optic. The camera was mounted on a horizontal bar at a height of 1000 mm from the table to frame the entire table.

We also add two lamps (30W energy-saving, 5200K), mounted inside an aluminum reflector (Ø 26 cm) with a diffuser to have a softer light and very uniform illumination, at the height of 1150 mm.

2.4.5. The User Interface

In the first prototype the User Interface (UI) is based on a Graphical User Interface (GUI) devoted to providing the augmented contents and on an interaction system based on marker occlusions.

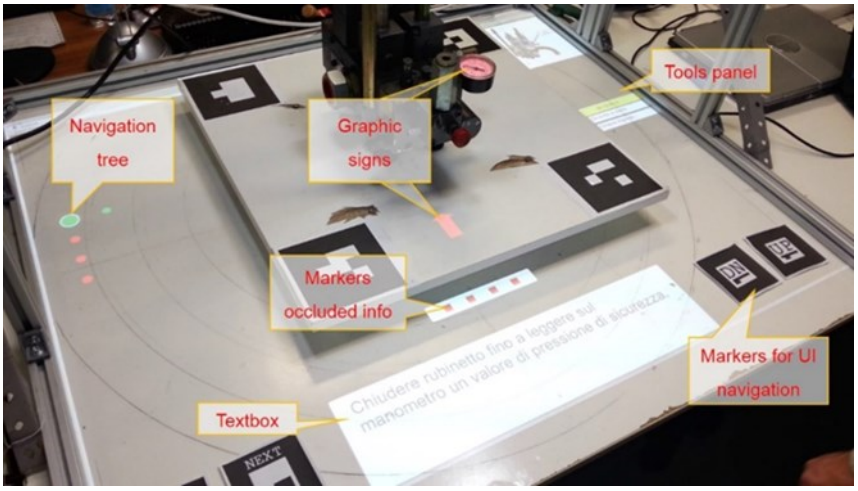


Figure 13 - The MWS prototype GUI.

The GUI (Figure 13) includes the following elements:

- Graphic signs (arrows, circles, etc.) that are projected to point at the parts of the product where to operate. Their placement is computed in the authoring phase by overlapping the virtual object on the CAD model of the object to be maintained. The CAD model is also used in the Unity scene as occlusion model to have a more realistic rendering of the scene;
- A textbox at the bottom of the table with instructions for the operator;

- A navigation tree, projected on the left side of the workbench surface, as the bullets of a multilevel bulleted list. The bullet corresponding to the visualized step has a different color (acting as a bookmark) to understand which step of the maintenance procedure the operator has reached and if there are any sublevels for that step;
- A panel with the list of needed tools; this panel may also provide additional information or images;
- Indicators of the marker occluded for the navigation of the user interface.

As to the interaction system, we used a technique based on marker occlusion. We used four different markers for the four directions of the application: next, previous, up a level, down a level. When the operator occludes with her/his hand the marker, s/he activates the corresponding trigger and the corresponding scene is loaded. We inserted also a visual feedback on the GUI to check if a marker has been occluded. It consists of four squares; a square is red if the related marker is not occluded, green otherwise.

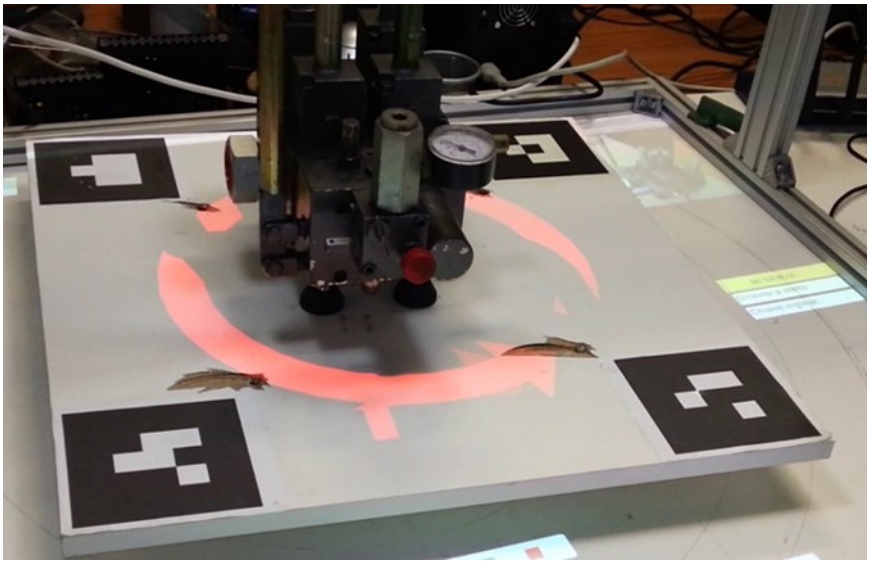


Figure 14 - The red circular arrow suggests the operator rotating counterclockwise the rotating base.

Just as an example, on each new step, the system suggests the operator moving the rotating base in order to adjust the point of view to optimize the projected information. A virtual red circular arrow indicates the direction of rotation, and it disappears when the correct position is reached (Figure 14).

2.5. Preliminary trial and prototype improvement

After prototyping the SAR-aided MWS, we carried out a preliminary user test. As a test case, we implemented the maintenance procedure of an oleo-dynamic valve (model GMV 3010 ¾").

The maintenance procedure was quite simple with four macro-steps, with several sub-levels. Nevertheless, the test case was able to test the functionality of the complete system. In this case, we measured a working volume of about 600 mm (L) x 500 mm (D) x 600 mm (H). Common assembly/maintenance instructions usually contain information about the localization of the sub-product involved in the operation (e.g., a screw), the tools needed, and the operation itself to accomplish. Our system proved to be very effective to locate specific points or objects in our product, for example heads of screws, holes, sensors, buttons, etc. However, these points must lie on a surface, which is directly illuminated by the projector.

As to the interface navigation, in this initial prototype we were not able to create a reliable solution. This type of navigation of the interface it is not reliable because an occlusion occurs whenever something (often user's arms) stands between the marker and the camera, leading to frequent false positives (unwanted triggers).

We then modified the prototype interaction interface by substituting the marker occlusion system with a simple joystick that accomplishes the content navigation Figure 15.

2.6. Evaluation of the SAR based MWS effectiveness

2.6.1. Research questions

The goal of our research was to support the Augmented Operator with an efficient technical documentation based on SAR. It should contain the same information contained in a paper manual, for example, text instructions and pictures, directly projected on the workbench. The added value of SAR is the possibility to locate with graphic signs on the real product, the parts involved in the task. Then, it is important to evaluate if this added value produces improvements in operator performance and if operators well accept this visualization modality.

We can say that the point of user evaluation of AR in the industrial domain is well addressed in the literature. Most of the works show that with AR, it is possible to reduce times and error rates in task execution. Furthermore, related works show that the use of SAR in industrial scenarios is continuously growing, mainly due to brighter, lighter, and cheaper projectors. However, there are no works about user evaluation of a SAR system for technical instructions, versus a traditional mode of presenting instructions, based on printed manuals. It is not possible to transfer the results of user evaluations on traditional AR to SAR since most of the works in section 2.2.2 make use of 3D CAD models that are not displayable in a SAR interface.

Then, also considering the increasing relevance of SAR technology in the industrial contest, we felt crucial to make such an evaluation. Therefore, we formulated these research questions:

1) Are SAR-conveyed technical instructions more effective than paper ones?

2) Is SAR well accepted by users as technical guidance compared with a paper manual?

2.6.2. **Material and methods**

To compare the two instruction modes (hereafter, we call them paper and SAR), we conducted a mixed design experiment evaluating user completion time and error rate, and users' ratings. In both approaches, a within-subjects experiment was carried out to test our hypotheses.

As regards the first research question, we formulated the following hypotheses:

> *(H1) SAR system will significantly reduce the amount of time to complete a set of maintenance tasks, compared to paper instructions;*
> *(H2) SAR projection system will significantly reduce errors of a set of maintenance tasks, compared to paper instructions.*

As regards the second research question, we formulated the hypothesis:

> *(H3) SAR projection system will significantly improve user acceptance.*

As a system we used the SAR-aided MWS prototype with the joystick-based interaction interface.

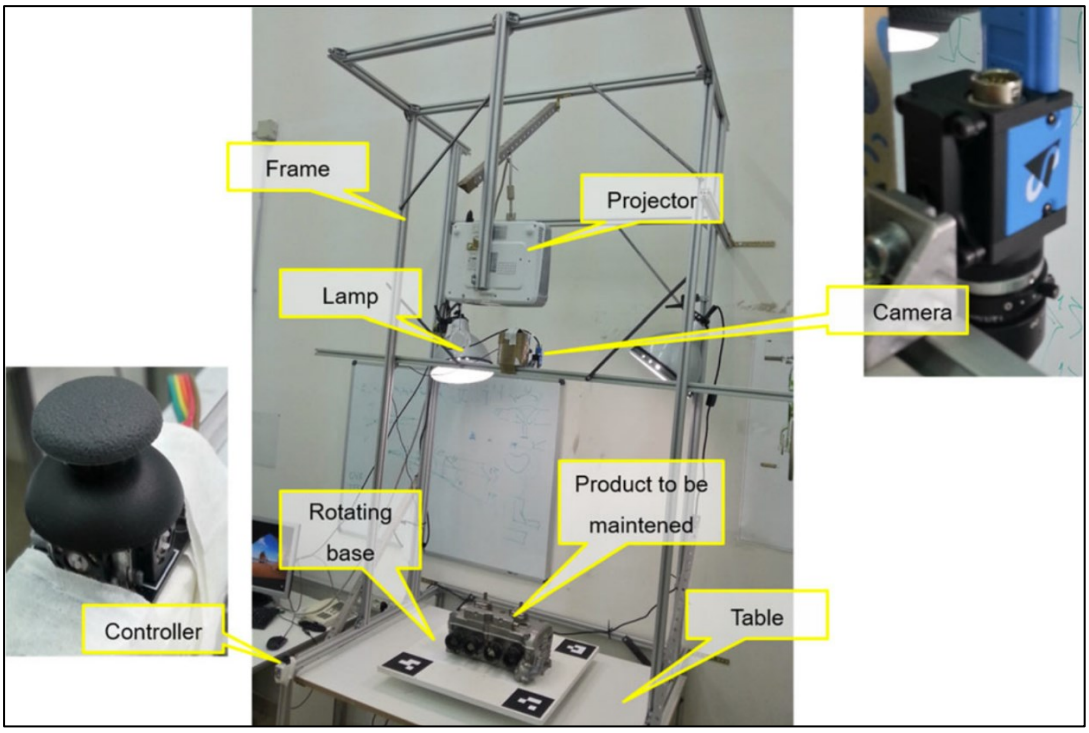


Figure 15 - The prototype of the SAR based MWS used in the experiment (on the left the controller substituting the marker-occlusion based control system of the first prototype).

2.6.3. **Participants**

A total of 16 voluntary participants (all males) were recruited among engineering students at Politecnico di Bari, using convenience sampling. The average age was 24.6 (min 22, max 32, SD = 2.17). We interviewed the subjects about their personal experience in maintenance operations: 14 participants had previous experience; 2 had no experience, half of the participants had previous experience with maintenance manual instructions, 14 participants were novices to AR and computer-based maintenance guidance systems.

The participants in this study were all engineering students and therefore, highly interested and favorable to modern technologies. Furthermore, only half of them had previous experience with maintenance instructions. Considering this background of our participants, we designed an experiment with very basic tasks so that the experience and the motivation of the participants would not affect the results.

2.6.4. **Procedure**

Participants were assigned to both treatments, paper and SAR one, using a Latin square design. Then, we had eight participants performing first the paper mode and then the SAR mode and the other eight vice versa. They were told to complete each task as quickly and as accurately as possible. Each participant could familiarize with the system for 10 min before the test. This

training phase helped the participants to get accustomed with the user interface—the controls to navigate the tree of instructions and the visual and textual information given by the system. At the end of the training phase, an experimenter checked that the participant was able to easily use the system.

We video recorded each user test. Completion time was measured by carrying out an offline analysis on these videos. We measured both the completion time of each task and of the overall procedure. A human assistant supervised each test and checked for errors in real time, using a notepad to register types and causes of mistakes. The assistant observed the user from a distance, without hindering her/his movements. After completing the test, users were asked to respond to a questionnaire to get subjective evaluations about the treatments.

We chose as a case study, part of the maintenance procedure of a Honda CBR 600 motorbike engine. We arranged a task sequence based on the maintenance procedure for the inspection of camshafts. The operations used in the test include common and crucial maintenance tasks: part localization, components identification, sequences identification.

The sequence of the tasks is the following:

> Task 1. Cylinder head-cover removal: pull out six bolts in no particular order and remove the cover;
> Task 2. Cam chain guide removal: pull out two bolts in no particular order and remove the guide;
> Task 3. Removal of the exhaust and intake camshafts and their holders: pull out eighteen bolts in a specific order and remove the camshafts in a given order;
> Task 4. Assembly of the exhaust and intake camshafts in a specific order and a proper position;
> Task 5. Assembly of the exhaust and intake camshafts holders: assemble the camshafts and insert eighteen bolts in a specific order;
> Task 6. Cam chain guide assembly: insert the guide and two bolts in no particular order;
> Task 7. Cylinder head-cover assembly: insert the cover and first two bolts, then the other four.

We designed an experiment that could be the least affected by users' skills and dexterity in workshop operations. For this reason, we simplified the un/screw operations by replacing the original bolts with plain unthreaded ones that only needed to be pulled out/inserted from/into

their slots. We added an instruction about the correct position of the lobes of two cams for the proper installation of the camshafts (Task 4).

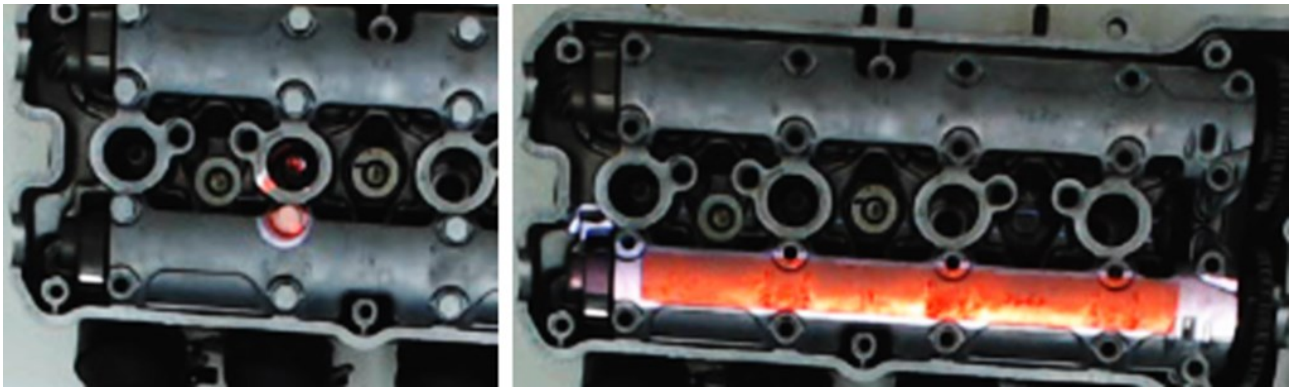


Figure 16 - Colored graphic signs used in the SAR mode to indicate a bolt (left) and a camshaft holder (right) to be removed

For this reason, we separated this task from that of the assembly of the holders. We prepared two similar sequential procedures for paper and SAR instructions, available in Appendix 1. The instructions about the identification of a component (e.g. a bolt), in the SAR mode, are given by red "coloring" the interested area, as can be seen in Figure 16.

We considered tasks with different levels of difficulty. We used the definition of task difficulty made by Kantowitz (1987): "task difficulty increases with the capacity investment needed to obtain a specified level of performance." In our test, the level of performance desired is that of a skilled operator doing the task. Task difficulty is an intrinsic property of a task, and it must not be confused with task complexity, which is related to instruction mode. The complexity, according to Kantowitz (1987), refers to "the (hypothetical) mental system architecture used to perform a task." As additional internal stages or processes are required, task complexity increases.

In our experiment, we evaluated, in a range from 1 to 3, difficulty and complexity indices associated with our seven tasks with a think-aloud procedure in a team of maintenance experts.

For example, we can observe the difficulty and the complexity for Task 1 and 3:

Table I - An example of task difficulty and complexity.

	Difficulty	**Complexity**	
		Paper	SAR
Task 1	*	*	*
Task 3	***	***	*

Task 3 requires the removal of eighteen bolts in a given order. It is more difficult than Task 1, which requires the removal of six bolts in no particular order. As regards complexity, Task 3 is more complex to accomplish with the paper manual than with the SAR because of the number of mental processes required to execute the task increases. In the paper, the user has to read the text instruction, to make a mental match of the camshaft holder and its printed image, and to remember the progressive number. With SAR, the user has only to read the instruction (less text) and to localize the red circle projected on the engine. On the contrary, Task 1, presents the same low complexity in paper and SAR instruction mode.

2.6.5. **Measures**

The experiment's independent and dependent variables, collected for the subsequent statistical analysis, are shown inTable II. We used the following error rate definition:

$$ER\% = \frac{n.errors}{n.targets} \cdot 100$$

The *n.errors* is the sum of all the participants' errors observed for each task. The *n.targets* is the maximum number of errors that a user could make for each task, multiplied by the number of the participants that performed the experiment. For example, for Task 5, the possible errors are:

Wrong assembly of the exhaust camshaft;
Wrong assembly of the intake camshaft;
Wrong assembly of a screw (for all the eighteen screws).

Then, the maximum number of errors that each user could make for Task 5 is 20 (1+1+18); multiplying this number by the number of participants (16), we have that *n.targets*=320 for Task 5.

Table II - Independent and dependent variables of the experiment.

INDEPENDENT VARIABLES		
Participant users	16	16 males, mean age 24.6
Instruction Modes	2	Paper, SAR
Tasks	7	Dis/assembly tasks, as described in the previous section
Total trials	**224**	16x2x7
DEPENDENT VARIABLES		
Completion time	time (in seconds) spent on the execution of each task and the overall procedure	
Error	wrong task completion	
Users rating	votes assigned by users in a 5 points Likert scale, for each instruction mode as regards ease of use, intuitiveness and global satisfaction	

2.7. Analyses and results

We used two separate statistical models to analyze the effects of the instruction mode and the task type on time and error rate.

2.7.1. Completion time

Completion time was analyzed using the ANOVA. Instruction mode and task-type were modeled as fixed effects. Participant ID was modeled as a random effect. In Task 6 one user had a very bad performance due to an error in the accomplishment of the previous task. This bad performance resulted in a strong outlier (i.e. distance from the third interquartile greater than three times the interquartile distance). Thus, we decided to neglect the completion times of this user for the statistical comparison of this task and the overall procedure.

To compute the ANOVA, we initially had to verify normality and homoscedasticity. If both were verified, we could proceed with a classic homoscedastic parametric ANOVA. Otherwise, we had to perform other tests.

For the normality test, since the size of the group we wanted to analyze was limited (N=16), we chose to use the Shapiro-Wilk test. All samples positively passed normality and homoscedasticity tests.

We conducted a planned pair-wise comparison of completion time between instruction modes for each task and the overall procedure (Table III).

Table III - Statistical analysis for the completion time

Task	One-way ANOVA		Mean completion time	
	F	p	Paper Mode	SAR Mode
Task 1	0.024	0.879	36.9 s	37.4 s
Task 2	1.723	0.199	23.3 s	20.0 s
Task 3	51.247	<0.001	**190.4 s**	**128.6 s**
Task 4	1.987	0.169	95.1 s	82.4 s
Task 5	14.207	0.001	**201.2 s**	**157.7 s**
Task 6	4.179	0.050	24.3 s	19.3 s
Task 7	3.662	0.065	48.4 s	41.3 s
Overall	22.973	<0.001	**610.9 s**	**486.7 s**

The results allow us to confirm the null hypothesis H1. Comparing the means, for these operations, we found a completion time improvement with the SAR mode over the Paper mode.

For the overall procedure, we found a significant [$F(1,30)=22.973$, $p<0.001$] reduction of 20.3% of completion time.

As to the sub-steps, we found a significant [$F(1,30)=51.247$, $p<0.001$)] reduction of 32.5% of time completion for the Task 3, and a significant [$F(1,30)=14.207$, $p=0.001$)] reduction of 21.6% for the Task 5.

2.7.2. **Error rate**

To make statistical analyses about the influence of instruction mode on the accuracy of the operations, we grouped the error data of all the users involved, for each task type. Then, for each task, we computed the error rate as defined in Section 2.6.5. We used the method of "nx2 contingency tables" to make statistical inference.

Comparing the error rates of the experiment tasks (Table IV), we observed an error rate reduction with the SAR mode over the Paper one. This result allows us to confirm the null hypothesis H2. Instruction mode has a significant effect on the overall procedure [$\chi2(1)=33.518$, $p<0.001$]. As to the specific tasks, we found a significant difference for Task 3 [$\chi2(1)=21.646$, $p<0.001$], for Task 4 [$\chi2(1)=8.085$, $p=0.004$], for Task 5 [$\chi2(1)=5.530$, $p=0.019$], and for Task 7 [$\chi2(1)=3.865$, $p=0.049$]. For all the other tasks, we did not find a significant difference between error rates for the two instruction modes.

Table IV - Statistical analysis of error rate

Task	Chi-squared test		Error rate	
	χ^2	p	Paper Mode	SAR Mode
Task 1	0.000	1.000	0%	0%
Task 2	0.000	1.000	6.25%	6.25%
Task 3	21.646	<0.001	**5.97%**	**0%**
Task 4	8.085	0.004	**20.00%**	**2.08%**
Task 5	5.530	0.019	**3.75%**	**0.94%**
Task 6	0.368	0.544	6.25%	12.50%
Task 7	3.865	0.049	**43.75%**	**12.50%**
Overall	33.518	<0.001	**7.03%**	**1.17%**

Table V reports a comparison of the results for the single tasks with their difficulty index and their complexity, as defined in Sections 2.6.4 and 2.6.5.

Table V - Comparison of task difficulty, complexity and performance improvement

Task	Difficulty	Complexity		Significantly different?	
		Paper	SAR	Completion time	Error rates
Task 1	*	*	*	NO	NO
Task 2	*	**	*	NO	NO
Task 3	***	***	*	YES	YES
Task 4	***	***	**	NO	YES
Task 5	***	***	*	YES	YES
Task 6	*	**	*	NO	NO
Task 7	**	***	*	NO[1]	YES

[1] towards significance (p=0.065)

2.7.3. **Users' acceptance**

The post-experiment questionnaire featured five-point Likert scale questions (1 = most negative; 5 = most positive) to evaluate ease of use, satisfaction level, and intuitiveness for each Instruction Mode. The summary results from these ratings are reported in Figure 17. Medians of the ratings for SAR mode were always higher (5) than Paper mode (3 for ease of use and intuitiveness, 2.5 for satisfaction) for any of the three characteristics.

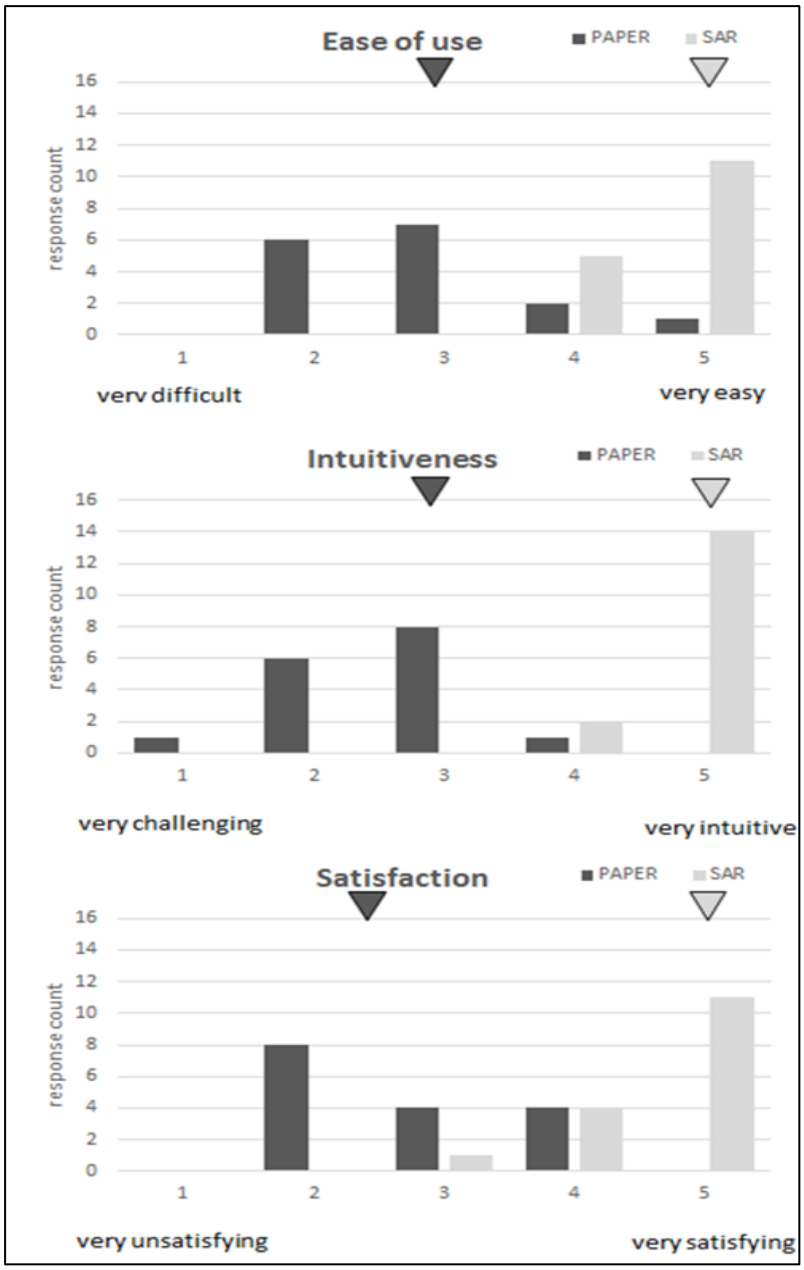


Figure 17- User ratings about ease of use (top), intuitiveness (middle), satisfaction (bottom). Median values for each condition are shown as triangles.

2.8. Discussion and conclusions

The results showed that the SAR presentation mode is significantly better than paper one in completion times (reduction of 20.3% in overall completion time) and error rates (83.3% fewer errors with the SAR mode) confirming H1 and H2. Subjective evaluation supported H3 (at least 2 points higher for SAR mode), confirming good user acceptance of SAR technology for conveying technical instructions.

Tasks with low difficulty (tasks 1, 2 and 6) show no significant difference ─ both for completion times and for error rates.

Tasks with high difficulty (tasks 3, 4, 5 and 7) show a significant difference in error rates. Tasks with high difficulty show a significant difference for completion times only if there is a substantial reduction in complexity due to SAR.

Tasks 3, 5 and 7 require the understanding of a correct sequence of the bolts. Indeed, using the paper manual, either the mental workload associated with these instructions is high if users tried to memorize the sequence of the bolts, or the cognitive distance is high if users switched from information space (the manual) to physical space (the engine) for every bolt or both. The SAR highly reduces the complexity of these tasks because the technical information is conveyed illuminating the bolt to be pulled out/inserted. With the SAR, we have then two great benefits: 1) the reduction of the mental workload associated with a task because only the task-relevant information is displayed; 2) the reduction of the cognitive distance because the information space and the physical space coincide. This result is consistent with what suggested and proved by Kim & Dey (2009).

In task 4 the reduction in complexity is lower than in the other tasks. Task 4 shows contrasting results: there is a significant and relevant decrease in the error rate but the difference in the completion times is not significant. Task 4 is also a difficult task, but the textual information is the same for paper and SAR, and therefore the time needed to read and understand the instruction is similar. However, SAR provides explicit component identification (green illumination of the portion of the engine to control), thus reducing error rates.

From these results, we can conclude that the use of SAR brings more advantages for difficult tasks than for simple ones. The main advantage of SAR is related more to the reduction of error rates than to completion times. These results were never found before in the literature for SAR, thus supporting the motivation for our research.

For traditional AR, in which 3D CAD models are used for displaying information, we found works where AR reduces both times and error rates and works showing only error rate reduction. Our result agrees with what found by Tang et al. (A. Tang, Owen, Biocca, & Mou, 2003): in their experiments, AR does not appear to have an advantage in time of completion comparing with other traditional media; nevertheless, they observed a strong reduction of error rates. The tasks in our and their experiment are very similar, based on the identification of a specific part and its positioning in a specific position. However, we achieved this result without using the CAD model of the assembled parts (screws), but just that of the engine used as occlusion model and as a projection mapping reference geometry, with a strong effort reduction in the authoring

phase. Indeed, we experienced authoring times much shorter in modeling and positioning of the colored graphic signs than in modeling and animating virtual screws.

Thus, we can conclude that SAR, similarly to traditional AR, always ensures low error rates and consequently guarantees operator accuracy. This result is important because it is in line with the objectives of Industry 4.0 and supports the role of AR as a key technology for a smart factory. Our industrial partners always considered overriding the reduction of the errors made by the operators and optional the reduction of times.

The result of this work for maintenance dis/assembly operations should be read together with what found in the literature for other manufacturing processes, such as data visualization of CNC machines (Olwal, Gustafsson, & Lindfors, 2008), spot-welding operations (Doshi, Smith, Thomas, & Bouras, 2016), human-robot collaboration (Andersen, Madsen, Moeslund, & Amor, 2016), and other manufacturing processes. Therefore, we can feel that SAR technology can be integrated shortly into industrial processes and product lifecycle. In particular, the proposed SAR-based MWS prototype could be, with further improvements, a good candidate for a future introduction of SAR in a smart factory as a support tool for the Augmented Operator. However, the advantages of SAR found with our validation are just the starting point to be sure of the path taken. Other problems remain open such as the authoring of contents, the tracking, the usability, and the interface design. In the authors' view, when all these issues will be addressed, industries will accept the introduction of SAR in their practices.

The SAR solution proposed and tested can be used for both training new maintenance operators and assisting operators in ordinary maintenance. In fact, we designed a multilevel tree-like structure of the procedure-instructions. It allows regular technicians a fast and effective browsing of instructions, by skipping well-known details and by accessing specific information only if necessary. While for new operators, there is the possibility to provide further detailed information to learn new procedures. In our work, the information detail is high because both text, images and graphic signs were used and users were asked to browse all the procedure steps.

Even if the case study used here is very basic, there are some tasks, as Task 3 and Task 5, where the mental workload is high because a dis/assembly order should be memorized. In these cases, SAR is helpful also for experienced maintenance operators, who usually can perform the procedures even without the support of instructions because they learned them over time. Furthermore, in the age of Industry 4.0 manufacturing is mainly based on mass customization. Thus, even experienced operators could work on many various products or versions of the same product, and they could not memorize all the relative procedures.

However, in this study, we did not consider as a variable the experience of operators, as well as the learning effect due to the repetition of a procedure at different times, or the possible fatigue of operators using SAR.

Chapter 3. Text legibility study for monocular Optical See-Through Displays for IAR

HMDs represent the most used visualization technology in IAR (Palmarini et al., 2018). Among the others, text legibility is one of the problems that can potentially undermine the effectiveness of AR applications based on Optical See-Through Displays (OST). The recently introduced devices, such as the Samsung BT Moverio and the Microsoft Hololens, present improved visualization performance, anyway text legibility in augmented reality with OST displays can be challenging due to the interaction with the texture on the background. Literature presents several approaches to predict legibility of text superimposed over a specific image, but their validation with an AR display and with images taken from the industrial domain is scarce. This chapter[2] describes the study carried out, during our doctoral program, aimed at identifying novel indices able to predict text legibility over textured backgrounds.

3.1. Introduction

The role played by AR as one of key the enabling technologies of Industry 4.0 and its proved effectiveness as support to the operator, is pushing big companies and research facilities to invest in AR hardware technology (such as display development and input devices) and AR software development (such as authoring, tracking, and applications). Unfortunately, on the other hand, few studies have addressed AR usability and ergonomics, especially if applied in real industrial scenarios. However, basic human-related issues, as text legibility, can spoil the AR experience and its effectiveness on demanding and critical applications (such as industrial maintenance and medical applications).

[2] The results of the studies described in this chapter have been published in the following article:
Manghisi, V. M., Gattullo, M., Fiorentino, M., Uva, A. E., Marino, F., Bevilacqua, V., & Monno, G. (2017). Predicting text legibility over textured digital backgrounds for a monocular optical see-through display. Presence: Teleoperators and Virtual Environments, 26(1), 1–15.

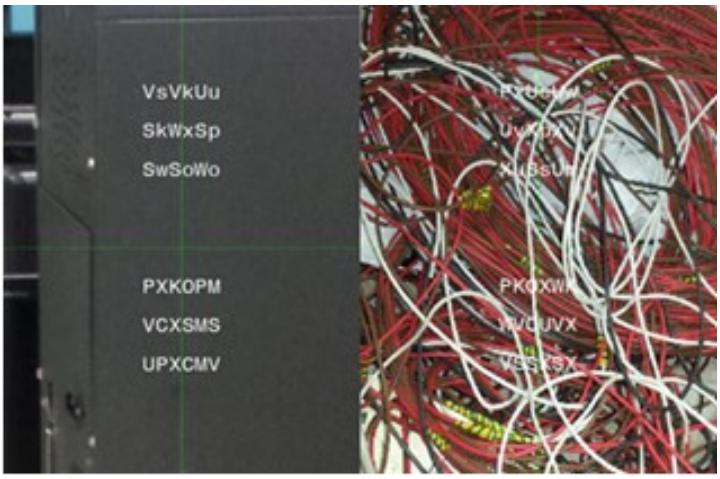


Figure 18 - The influence of the background in augmented reality: good legibility (left) and bad legibility (right).

As well argued by Albarelli et al., the key feature for a professional usage of a new class of AR devices is that the user should be able to access data without losing the focus on what he/she is doing. In AR technical documentation, the focus of the user must switch (even subconsciously) between the real world (product to be inspected) and the overlaid data. They called this behavior with the term competitive see-through, in contrast to the cooperative see-through found in standard augmented reality, where the virtual information is designed to blend with the real scene in a seamless way (Albarelli, Celentano, Cosmo, & Marchi, 2015). Due to the user focus switching and to the continuously changing background, the legibility of even simple text in AR is challenging (

Figure 18). Gabbard, Swan, Hix, Kim, & Fitch (2007) considered the text as one of the most fundamental graphical elements in any user interface, and therefore, they focused their study on text legibility in AR. AR text discrimination is likely to be affected by luminance contrast and texture of the stimulus and background (Chubb, Sperling, & Solomon, 1989; Hill & Scharff, 1999). Previous studies addressed the influence of text style (M. Fiorentino, Debernardis, Uva, & Monno, 2013) and background lighting on text displayed on video see-through (VST) versus OST head worn displays (HWDs) (Gattullo, Uva, Fiorentino, & Monno, 2015). With common illuminance levels (400 lx, suggested by IESNA standards for common visual tasks (Rea, 2000)), legibility with OST displays is better than that with VST displays. However, OST displays demonstrated to be very sensitive to high levels of illuminance (4000 lx, recommended for demanding visual tasks), and in such conditions, high contrast text styles like outline or billboard are ineffective (Debernardis, Fiorentino, Gattullo, Monno, & Uva, 2014). Nevertheless, VST displays, because of the lack of a real-world vision, proved that they are not well accepted in

industrial environments (Schwald & De Laval, 2003a). For this reason, we limited our study to the legibility of text on OST HWDs and background texturization only.

The influence of background texturization on legibility is complex to measure in a quantitative way. In fact, when the background is not or slightly texturized, there is an almost uniform contrast between all the letters of the text and the background, and the global background illuminance only influences legibility. When the background is heavily texturized, contrast can change locally at the character level, and it can change rapidly according to the user point of view. Literature presents several approaches to predict text legibility, but they miss an experimental validation with AR setups and in an industrial context. Therefore, the main contribution of the research described in this chapter is to analyze the existing approaches, to propose novel indices to predict legibility, and to validate their effectiveness. We do not want to make a display comparison nor to find a way to optimize legibility with OST displays. Our aim is to find an objective method to classify backgrounds according to their texturization. From the results of this work, it should be possible to make a ranking of backgrounds using the proposed method, hence supporting AR developers.

In the following section, we present the related works on the topic. Section 3.3 presents our approach followed in section 3.4, by the design of the experiment. In section 3.5, we present the results of the test and a discussion about the better indices to use as legibility predictor.

3.2. Related works

The problem of text legibility over textured backgrounds, both in video footage and AR domain, has been widely addressed in the literature. The proposed solutions are:

- Use of a solid background behind the text, to create a more uniform contrast between all the letters and the real background (Albarelli et al., 2015; Jankowski, Samp, Irzynska, Jozwowicz, & Decker, 2010).
- Use of text drawing styles to enhance contrast, like anti-interference, shadow, outline, billboard (Gabbard et al., 2007; Jankowski et al., 2010).
- Move the text in the most readable position on the background (Albarelli et al., 2015; Leykin & Tuceryan, 2004; Orlosky, Kiyokawa, & Takemura, 2014; Tanaka, Kishino, Miyamae, Terada, & Nishio, 2008; Thanedar & Höllerer, 2004).

For AR applications, a solid background behind text occludes the real world. Albarelli et al. tested the possibility to occlude just one single eye and let the brain merge the displayed with the real scene. However, results showed that users preferred other display solutions with a

transparent background. Moving text in different positions of the GUI is not always possible in industrial application. In fact, workers are used to handling manuals with a fixed layout where each GUI area is assigned a specific function (such as tool list, task sequence, and warnings). Use of text drawing styles like outline or shadow can be a more viable solution for the industry. For instance, Gattullo et al. (2015) proposed the outline to improve legibility with high background illuminance levels. The effects of texturization on legibility are harder to evaluate, and there is not a continuum as in contrast. For this reason, different solutions based on the different texturization of the background should be found. The use of different outline widths could be an example. However, before testing possible solutions, there is a research question to answer: find a method to classify backgrounds based on the legibility of text over them. To answer this question, we evaluated some indices used in the literature to predict legibility.

The work of Tanaka et al. (2008) inspired our search. They proposed a method to determine the best region for displaying information in an OST HWD. They acquired the background image with a digital camera fixed in front of the user with a very wide optic and evaluated an index (Tanaka index, hereafter named "T"). The average of RGB components, the variance of saturation "S" in HSV color space, and luminance "Y" in YCbCr color space were used in the computation of T. We appreciated the proposed approach, but we found some limitations as shown in Figure 19: although left and right backgrounds have the same T index, legibility of text is clearly different.

Figure 19 - Left and right backgrounds share the same Tanaka (T) index but have very different legibility.

We found two possible issues in the T index:

a. the use of colors
b. the proposed region of analysis.

As regards to the first issue, we considered that the mean color of the image is more representative of the luminance contrast between text and background than of the actual

texturization. As to the second issue, to read and understand a portion of text appropriately, all characters in the text must be readable. Therefore, if the background has a non-uniform texturization, a local analysis at word level versus a global averaging one is more desirable.

Haralick, Shanmugam, & others (1973) proposed a set of general texture classification indices for analyzing satellite and medical images. They observed that texture and tone (related to Y in YCbCr color space) are related. When a small area of an image had little variation of discrete gray tone, the dominant property of that area was tone. When a small area had a wide variation of discrete gray tone, the dominant property of that area was texture. The size of the small area patch, the relative sizes of the discrete features, and the number of distinguishable discrete features were crucial to this distinction. The textural features extraction was based on the assumption that the texture information on an image is specified by a set of gray-tone spatial-dependence matrices. The features are all functions of distance and angle. From these angular nearest-neighbor gray-tone spatial-dependence matrices, the authors derived fourteen textural features. As far as the author knows, Haralick features have never been applied to predict legibility.

Other authors addressed the problem of texture classification, using methods based on spatial frequency analysis. According to Aach, Kaup, & Mester (1995), information about the distribution of texture power in different spectral bands can be obtained by applying a set of bandpass filters tuned to different orientations and spatial frequencies to texture signal, and by computing the amplitude envelopes of the filter outputs. Gaussian-shaped Gabor filters are often used for this purpose as they minimize the uncertainty relation, thus optimizing the tradeoff between resolutions in the spatial and spectral domains.

Petkov & Westenberg (2003) studied the effect of bandlimited noise on the perception of contours and, in particular, the effect of such noise on the perception of the text. They conducted psychophysical experiments to conclude that the frequency range of most efficient suppression was related to the letter stroke (or weight) and not to the letter size. Solomon et al. showed that the luminance contrast threshold above which the letters become readable depends on the intermediate frequencies (Solomon, Pelli, & others, 1994).

Leykin and Tuceryan presented an approach to automatically determine if overlaid AR text will be readable or unreadable, given dynamic and widely varying textured-background conditions (Leykin & Tuceryan, 2004). To determine the legibility of overlaid text, they employed a real-time classifier that used contrast features for the text, font features such as size and weight, and Gabor filter-based texture features, at various frequencies and orientations, of the background

image. They conducted a series of experiments in which participants categorized overlaid text as “readable” vs. “unreadable,” and used their experimental results to train the classification system. They found that the dominating frequency range that affects the readability ranged from 1.5 to 3.0 cycles per letter. Their result was in accordance with Petkov and Solomon (Petkov & Westenberg, 2003; Solomon et al., 1994). Dominant orientation was $\pi/4$, and the least correlated orientation was $\pi/2$.

Scharff et al. compared different discriminability measures of image backgrounds: text contrast, background RMS contrast, global masking index, text energy in spatial frequency bands of lines and letters, spatial-frequency-selective masking index (Scharff, Ahumada Jr, & Hill, 1999). They found out that each of the approaches to predicting readability led to similar correlations with reading speed. Thus, if someone wanted to determine the cost-benefit of texture and contrast choices, they would recommend using the first three simpler indices. The image measure regressions indicated that text contrast alone is a poor predictor of reading speed and that RMS contrast energy in the background better predicts reading speed. Furthermore, when they determined the contrast energy in spatial frequency bands roughly corresponding to letters based on (Solomon et al., 1994), words and lines, they found that the different textures did not predominantly contain energy in any of these bands. Therefore, although the correlations with reading speed were slightly higher for the spatial frequencies corresponding to letters and lines, they could not make firm conclusions about the relative contributions of the different bands.

In conclusion, there is not a proven and a commonly accepted index that can be used to predict legibility of AR text on a textured background. The proposed approaches are very different: some of them are simple and fast to compute (Scharff et al., 1999; Tanaka et al., 2008), others are complex and need high computational resources (Haralick et al., 1973; Petkov & Westenberg, 2003) not always available in mobile AR systems.

3.3. Our approach

We planned to evaluate a set of 37 parameters (see Table VI) to predict AR legibility. Seven of them (CRMS, T, TSY, TS, GM, GE, and GEI) were proposed in the literature. We tested the use of the other 30 in the field of text legibility.

Table VI- The list of the AR legibility indices under study, in gray the novel ones

INDEX NAME	NUMBER	DESCRIPTION
CRMS	1	Background root mean square contrast
Tanaka (T)	1	Tanaka index
Tanaka without RGB components (TSY)	1	Summation of the variance of luminance and the variance of saturation components
Tanaka index variance of Y in YCbCr (TY)	1	Variance of luminance component
Tanaka index var S in HSV (TS)	1	Variance of saturation component
Haralick image features (from FM1 to FM14)	14	Mean of the Haralick features over the four directions
Haralick image features (from FR1 to FR14)	14	Range of the Haralick features over the four directions
Gabor filters feature (GM)	1	Mean of the filtered images over the four directions
Gabor filters feature (GSD)	1	Mean of the standard deviation of the filtered images over the four directions
Gabor energy (GE)	1	Mean of the Gabor energy over the four directions.
Gabor energy with anisotropic inhibition (GEI)	1	Mean of the Gabor energy with anisotropic inhibition over the four directions.

We calculated the T index value as described in (Tanaka et al., 2008) and then we analyzed its components as indices on their own, by considering the variance of luminance, the variance of saturation, and the summation of these two components. In addition, we proposed the variance of Y in YCbCr color space (named TY).

We thought to evaluate the effectiveness of Haralick features as legibility predictors. We computed the gray level co-occurrence matrix with a distance of 1 pixel and with angular directions of 0, $\pi/4$, $\pi/2$, and $3\pi/4$ between the gray levels. We set the gray levels to 8 after heuristic testing. We calculated the 14 sets of 4 measures as described in (Haralick et al., 1973), then we proceeded by computing the mean and the range of each of these measures, averaged over the four directions, obtaining a total set of 28 features, named from FM1 to FM14 (mean) and from FR1 to FR14 (range).

We also considered the use of Gabor-based features with four orientations (0, $\pi/4$, $\pi/2$, and $3\pi/4$), and with a spatial frequency of 4.35 cpd (cycles per degree). We chose this value heuristically in our exploratory tests. We calculated the first Gabor-based feature (GM) by computing the mean of the values in the four directions as proposed in (Leykin & Tuceryan, 2004). As a novel

index, we also introduce their standard deviation (named GSD). The other two Gabor-based features, already presented in literature, were obtained using the Gabor energy filters (GE) and the anisotropic inhibition (GEI) as described in (Petkov & Westenberg, 2003) by using a bank of 4 filters with four orientations (0, π/4, π/2, and 3π/4), and with a spatial frequency of 4.35 cpd.

We applied the formula presented in (Scharff & Ahumada, 2002) to compute the background RMS contrast (CRMS).

We computed our indices in the region of interest (ROI) directly underneath the text, instead of considering macro regions or the user full visual angle. The background ROI used for computation was obtained from the text block projected area enlarged of one-character height in all direction (Figure 20).

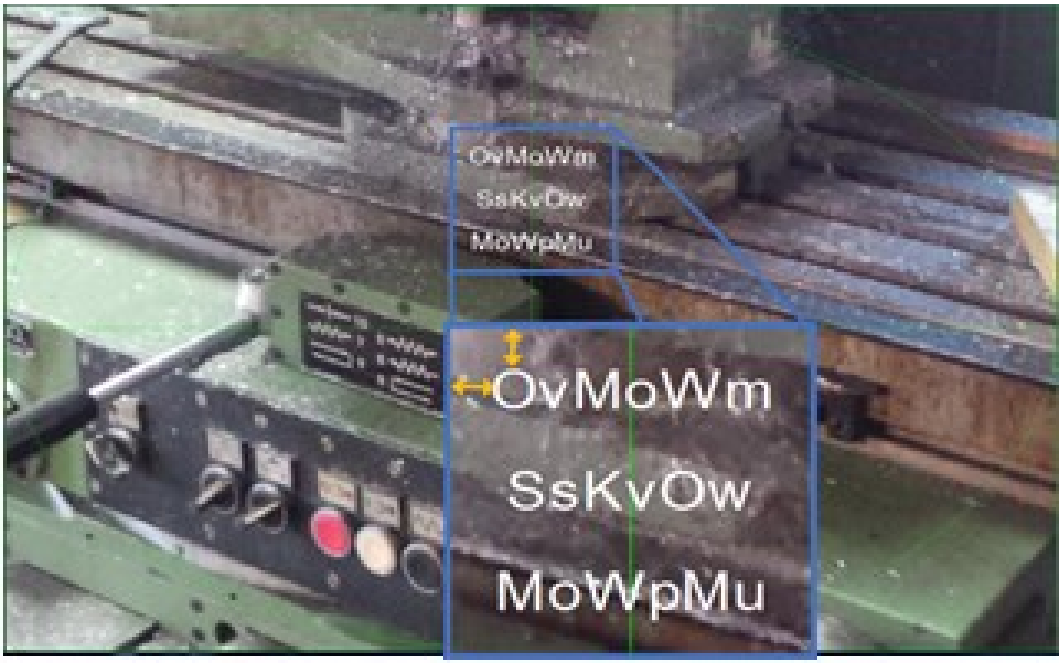


Figure 20 - An example of the ROI, one font height larger than the text block area.

This extended ROI allows to compute the effects of the background on the borders of the text block and to tolerate minimal misalignments of the user point of view. We implemented the digital image processing algorithms in a Matlab® script.

3.4. Design of the experiment

To validate the legibility indices, we evaluated their correlation to the real user ratings acquired in a dedicated experiment.

3.4.1. **Participants**

We recruited nineteen unpaid participants for the study (seventeen males and two females, mean age 26, standard deviation of 9.24 in the range from 21 to 47). They were undergraduate and graduate students in technical subjects in mechanical engineering. Eleven participants had natural normal visual acuity, whereas the others used their glasses to correct it. Seventeen users had right eye dominance. No participant was color-blind, and all had normal stereoacuity.

3.4.2. **Apparatus and materials**

We selected 13 images of industrial backgrounds (Figure 21), with a resolution of 1600 x 1000 pixel. Background images and test software are openly available on our development website (Michele Fiorentino, 2016).

Figure 21 - The 13 background images used in the experiment.

In watching real world scenes, there are differences in depth that will cause the background to appear blurred and dynamic range will be considerably different. The dynamic range in photography represents the limits of luminance range that a digital camera can capture. Human sight has a very high dynamic range. Modern cameras increase the dynamic range of a picture, using High Dynamic Range (HDR) techniques in postprocessing: they combine by software multiple exposures of the same picture. We tried to minimize the effects of dynamic range in our design of the experiment, using High Dynamic Range in picture acquisition. Additionally, to minimize the effects of defocus blur, we focused on the zone of the image of the ROI where the text is displayed. We set on the ROI the focal plane of our OST device. We computed the indices

considering only the pixels in that ROI: it is just a small area of the overall real background, and then we expect that the effects above on legibility should be little.

We used an optical see-through HWD (Figure 22): the Liteye LE 750A (Liteye Inc., n.d.), OLED display, (resolution: 800 x 600 x 60Hz, contrast ≥ 100:1, transmission 70\30, max luminance 300 cd/m2, 28° diagonal field of view, monocular).

Figure 22 - The setup of the test: the user, the optical see-through head-worn display, and the monitor.

We set, for each user, the diopter adjustment to focus the target plane of the LCD Screen. The HWD was mounted on a fixed frame at the height of 130 centimeters from the floor so that each user could perform the test with her/his dominant eye. To make the test execution as comfortable as possible, each participant sat on a stool with adjustable height.

Each background image was displayed in true size on a color calibrated Dell UltraSharp 2408WFP 24 inches LCD monitor with fixed settings. Gabbard, Swan, & Hix (2006) did a similar experiment to evaluate the effects of text drawing styles, background textures and natural lighting on user performance in outdoor AR. They used large posters as backgrounds, but for our experiment, we preferred to display the background images on a color-calibrated screen, and to avoid the color errors due to the print process. In fact, we computed the indices in Table VI with image processing algorithms on the original acquired digital image.

For our experiment, the use of real backgrounds complicated a lot the execution of the experiment, forcing us to reduce the number of backgrounds analyzable. However, we tested if this design choice affected the results of the study. Then, we did another experiment to compare legibility with a real industrial background and the corresponding image displayed on the monitor. We chose three backgrounds with diverse levels of texturization and adjusted the monitor luminance to ensure a similar luminance (~60 cd/m2) of the real scene. We gathered legibility ratings with both digital and real backgrounds from ten users. They were seven males

and three females, mean age 32 with a standard deviation of 6.97 in a range from 25 to 46. Four participants had normal visual acuity, whereas the others used their glasses to correct it. We compared the legibility ratings of digital and real setup for the three backgrounds with a Kruskal-Wallis test. The results, reported in Figure 23, show that there is not a significant difference between legibility on the real and digital background for all the three backgrounds.

This result confirms that displaying background on an LCD monitor does not affect legibility. Thus, we can ignore the influence of focal depth and dynamic range to a large degree.

Gabbard et al. (2006) found that there was no main effect of viewing distance from the background, then we located the monitor on a desk at 1.3 m from the user's eye. It was the right distance to match the OST field of view with the monitor size. Two crosshairs ensured the alignment between the images on the monitor and the OST HWD; one was visualized on the OST device and another one on the monitor.

The experiment was held in a laboratory with a controlled illuminance level of 400 lx, (as suggested by IESNA standards for common industrial tasks). We also measured the luminance level of the monitor itself, with a Konica Minolta CS-200. We pointed the color meter on the area where the text is displayed with an angular opening of 1 degree. In this way, we covered an area greater than text height. The higher luminance was measured for background 7 and was about 55 cd/m2.

Figure 23 - Setup of the test with the monitor placed in front of the real backgrounds, with their digital image displayed on it. The results of the Kruskal-Wallis test showed no significant difference between the two conditions.

3.4.3. **Text settings**

In the legibility test we used white text, sans serif non-proportional typeface (Monospace Typewriter), without billboard background or border. Font height was 16 points (16 pixels on the OST device), corresponding to a visual angle of 29 arcminutes. The visual angle was larger than the 20 arcminutes required to ensure the recognition of a symbol chain according to ISO

9241-8. The letter stroke was five arcminutes as in the work by (Petkov & Westenberg, 2003) and overlaid five pixels on the background image.

We used just one text color and one text height because this is a starting work to explore the possibility to use some novel indices to classify backgrounds according to their texturization. In a follow-up study, we would validate those indices with other text settings. We chose a white color to have the highest luminance for the text.

In this experiment, we wanted to isolate, as much as possible, the influence of background texturization on legibility, from the influence of luminance contrast. For this reason, we set the luminance of the monitor to ensure a minimum contrast ratio of 4.5:1, as recommended by W3C (W3C, 2012). We measured the maximum and minimum luminance transmitted by the OST display for all the backgrounds, with a Konica Minolta CS-200 (Figure 24): maximum luminance (L_{max}) is achieved displaying a white image on the OST display (all pixels turned on), whereas minimum luminance (L_{min}) is achieved displaying a black image on the OST display (all pixels turned off). Then we computed Contrast Ratio ($CR = L_{max}/L_{min}$) and Contrast Sensitivity (reciprocal of Michelson Contrast, $C = (L_{max}-L_{min})/(L_{max}+L_{min})$) for all the backgrounds. The minimum contrast occurs when a white image is displayed on the LCD monitor. In this case, we have CR = 4.90:1 higher than the minimum 4.5:1 recommended by W3C. Contrast ratios for all the used backgrounds, reported in Figure 25, were higher than this value.

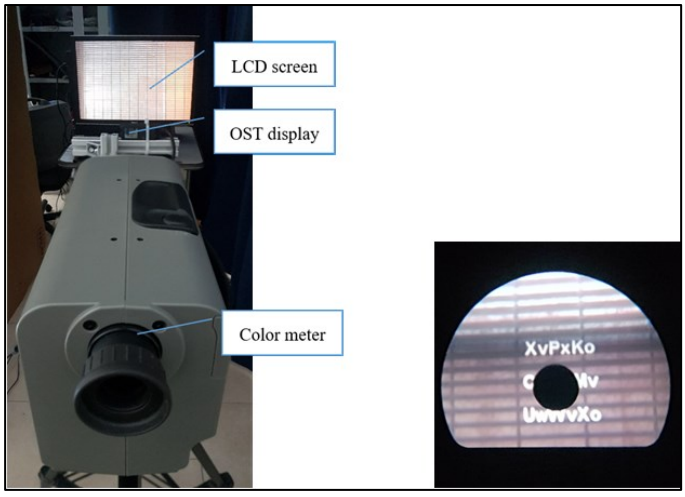


Figure 24 - Luminance contrast measurements: setup of the system (left) and the view from the color meter; the black point is the measurement area of the color meter.

As to Contrast Sensitivity, we reported all the computed values on a Contrast Sensitivity Function plot and compared our values with the legibility threshold curves found by (Legge, Rubin, & Luebker, 1987): they were all grouped in a small region well below the legibility threshold. This result supports our hypothesis that, with this kind of design, differences in text

legibility among backgrounds are not due to contrast, but to texturization of the background, that is what we want to measure with this experiment.

3.4.4. **Procedure**

Each participant looked through the OST HWD displaying a text block, while a background image was displayed on the LCD monitor (Figure 22). Three random generated strings with alternating uppercase and lowercase letters composed the text block.

BACKGROUND IMAGE	CONTRAST RATIO	CONTRAST SENSITIVITY
1	35.3	1.06
2	24.6	1.08
3	10.6	1.21
4	8.3	1.28
5	21.3	1.10
6	18.2	1.12
7	7.9	1.29
8	24.0	1.09
9	12.9	1.17
10	33.1	1.06
11	16.0	1.13
12	12.3	1.18
13	15.2	1.14

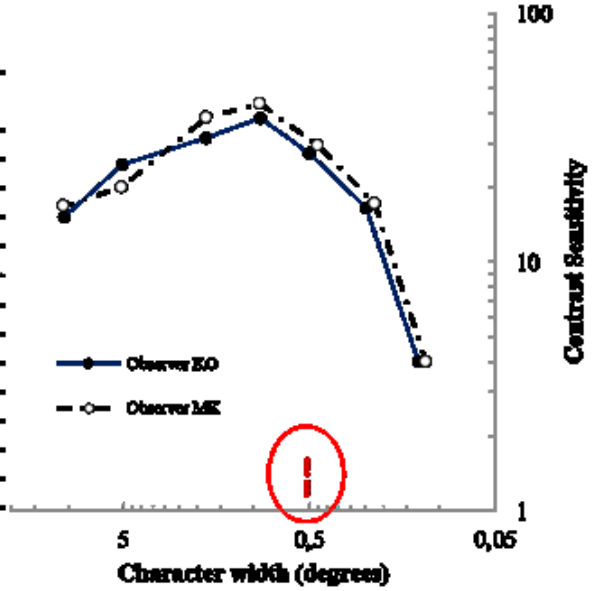


Figure 25 - Contrast measurements for the background images used in the experiment (table on the left); the plot of Contrast Sensitivity values (red points in the image on the right) reveals that with all the backgrounds, we are well below the legibility threshold in the readable zone, defined by the curves found in (Legge et al., 1987)

After a trial period of 3 minutes to get used to the system, each user had to browse the 13 backgrounds and to identify which of them gave the worst and the better legibility. After this phase, each user was asked to rate the legibility on each background image giving an ordinal score ranging from one (for the worst legibility) to five (for the best one). Participants had no time limits, and they were free to change their ratings during the test. The variables of our experimental test procedure are listed in Table VII.

Table VII - Independent and dependent variables for the legibility test experiment

INDEPENDENT VARIABLES		
Participants	19	17 males, 2 females
Backgrounds	13	Various industrial backgrounds
DEPENDENT VARIABLES		
User Legibility ratings	13	Median value of legibility ratings assigned by users for each background image

3.5. Results

Figure 26 reports the median values of the user ratings.

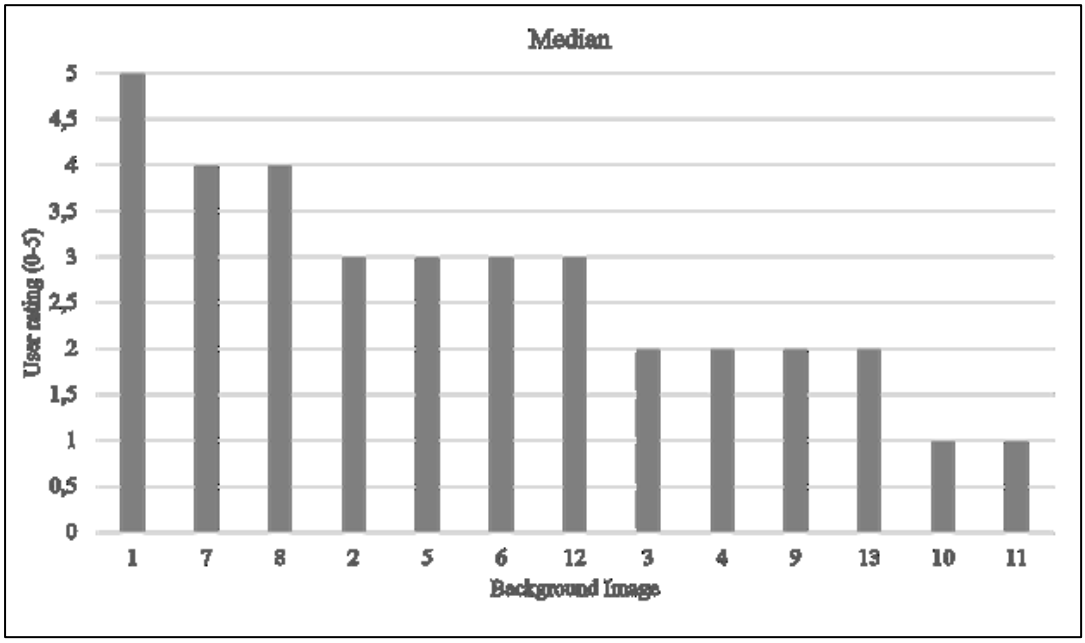


Figure 26 - Median values of user ratings for each background image in the legibility test.

The next step was to evaluate the correlation between the user responses and each of indices described in Table VI. For each background image, we cropped the ROI corresponding to the extended text block area (200 x 200 pixels). We computed all indices only in the background image ROI and not considering the overlaid text.

Legibility ratings are in the ordinal format, therefore we used the Spearman approach (Hauke & Kossowski, 2011) as a correlation metric. Spearman's rank correlation coefficient ρ is a nonparametric (distribution-free) rank statistic proposed as a measure of the strength of the association between two variables. It does not require the assumption that the relationship between the variables is linear and it does not require variables to be measured on interval scales. The null hypothesis of the Spearman test is that there is no monotonic correlation in the population. To reject this hypothesis with a significance greater than 95% for 13 samples, the Spearman correlation coefficient absolute value should be greater than 0.56 (Zar, 1972). Table VIII presents the obtained Spearman correlation coefficient absolute value for each index, where non-significant correlation values are marked in gray.

The statistical analysis showed that most of the indices (26 out of 37) have a significant (i.e., $\rho > 0.56$) correlation with user legibility ratings. This result proves that it is possible to predict legibility, just analyzing the background.

Table VIII: The Spearman correlation coefficients for each of the indices, in gray the non-significant ones (i.e. $\rho <$ 0.56).

INDEX	SPEARMAN ABSOLUTE CORRELATION COEFFICIENT ρ
CRMS	0.700
T	0.737
TSY	0.723
TS	0.638
TY	0.867
FM1	0.816
FM2	0.839
FM3	0.198
FM4	0.867
FM5	0.816
FM6	0.184
FM7	0.829
FM8	0.731
FM9	0.850
FM10	0.755
FM11	0.755
FM12	0.102
FM13	0.368
FM14	0.051

INDEX	SPEARMAN ABSOLUTE CORRELATION COEFFICIENT ρ
FR1	0.448
FR2	0.862
FR3	0.181
FR4	0.607
FR5	0.850
FR6	0.947
FR7	0.850
FR8	0.312
FR9	0.816
FR10	0.755
FR11	0.738
FR12	0.099
FR13	0.607
FR14	0.023
GM	0.434
GSD	0.915
GE	0.754
GEI	0.788

On the other side, it shows that several indices may be good predictors and therefore a deeper analysis is needed. Figure 27 depicts the 13 indices with strong (i.e., $\rho > 0.80$) correlation values. The Haralick feature FR6 showed the best correlation ($\rho = 0.95$). As to Gabor features, GSD performed better ($\rho = 0.92$) than the other indices based on Gabor filters did. Finally, the Tanaka derived TY ($\rho = 0.87$), i.e. the variance of Y in YCbCr color space, showed a higher correlation than the original Tanaka index ($\rho = 0.74$).

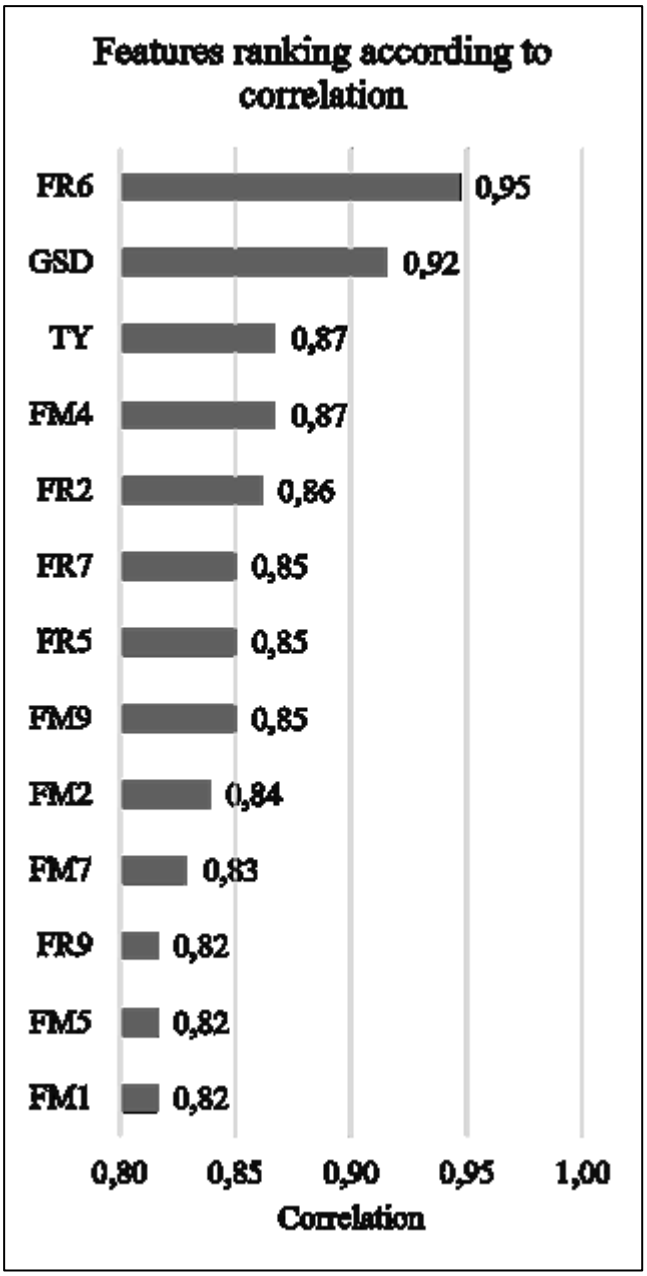


Figure 27 - Bar chart of the indices with strong correlation values ($\rho > 0,8$).

3.6. Discussion and Conclusions

We designed the experiment, trying to isolate as much as possible the effect of texturization on legibility from that of contrast. We measured contrast ratios for all the backgrounds and found that there is no correlation (Spearman correlation coefficient $\rho = 0.18$) between Contrast Ratio and user legibility ratings. Thus, the differences in legibility measured in our experiment are due to background texturization, as desired.

As expected, the background with the best legibility was the image 1 ("software house") that presented a very low texturization (Figure 21). On the opposite side of the ranking, we found image 10 ("automotive workshop") and 11 ("electric repairs") that presented a marked

texturization. These results are not generalizable due to the limitations of our experimental setup (e.g. we did not use real backgrounds, we employed just one device and just white as text color). However, they are in accordance with literature and our hypotheses. This research wants to be a starting point for further exploration with other devices and text styles.

In this experiment, we did not use a setup with real backgrounds because, in this research step, we wanted to test many indices on as many backgrounds as possible. Indeed, there are little differences in depth between the real background and the focal plane of the OST device. We displayed a picture of the background on an LCD screen during the test. In this case, the defocus blur and the dynamic range are different from what perceived by the human eye on real backgrounds. However, we minimized the effects of this mismatch, as previously explained. In the literature there are no studies about differences in legibility among real backgrounds, images on monitor and poster board, then it could be an interesting follow-up study. For each background image, we computed some indices with image processing techniques: we examined 37 indices of legibility, and most of them have a high correlation with user legibility ratings.

According to Haralick et al. (1973), the entropy features (FM9 and FR9) should take higher values for more complex images. Then, we expected a high correlation for FM9 and FR9. Our results confirmed it (ρ = 0.85 for FM9 and ρ = 0.82 for FR9), but we found that other features had higher correlation values. In particular, the Haralick feature FR6 showed the best correlation (ρ = 0.95) in our test. It is also interesting to notice that most of the Haralick features (18 out of 28) had a significant correlation in legibility and 15 of them correlated equal to or more than the AR dedicated T index. If also confirmed for other experimental setups, this is a non-trivial result considering the general origin of the descriptor.

We wanted to test the GSD index, based on Gabor filters because we hypothesized that a feature based on standard deviation of the values was more representative of the texturization influence than a feature based on the mean. As the second outcome, our test confirmed that our novel proposed index GSD could be a good legibility predictor (ρ = 0.92).

The Tanaka derived TY, i.e. the variance of Y in YCbCr color space, showed a very good correlation (ρ = 0.87) and surprisingly higher than the original Tanaka index (ρ = 0.74). As argued in the introduction, the limited contribution of colors to legibility could be the main reason. Considering that the goal of Tanaka et al. was to reduce the computational effort, we consider that the simplification in the implementation of TY may be a significant improvement.

As a final remark, the CRMS index was significantly correlated (ρ = 0.70) in accordance with the literature (Scharff et al., 1999). However, in our tests, it correlates less than other indices. It is important to notice that, differently from their paper, we did not consider the text in the

computation because it was extremely difficult to measure the luminance of the text as seen by the user through the OST HWD.

These results were obtained measuring the correlation between indices computed with image processing and user ratings about legibility with a Liteye LE 750A OST display. This measure of legibility was used as the reference value. We tried to compare this result to what found in the literature for other OST devices. Livingston, Gabbard, Swan, Sibley, & Barrow (2013) say that the optics play a critical role if the experiment also tests contrast and that the visual acuity of the user in perceiving virtual objects (text) is affected by the angular resolution of the display elements. Then, we supposed that with other OST devices with a similar angular resolution to ours (like the Nomad used in (Livingston, 2006)) and high contrast values, like those reported in our experiment (Figure 25), we would obtain similar legibility measures. However, simply measuring the angular resolution of a head-worn AR display is not sufficient to characterize the visual acuity or the broader experience a user will have when wearing the display. While it would theoretically place a limit on the performance a user could achieve, the human visual system is quite exceptional at filling in and interpolating information (Livingston et al., 2013).

Even considering the limitations of this work, the provided results are useful to define an objective method for background classification.

The main goal of this work was to find an index, extracted from a digital background image, to predict text legibility in augmented reality with optical see-through displays in order to enhance the performance of IAR applications for the Augmented Operator. We reviewed literature for a well-accepted solution to the problem, but we found still scarce knowledge on this topic and the lack of specific user tests.

This exploratory study provided interesting results; however, it also demonstrated how current approaches might be improved to provide simple and well-accepted guidelines for the upcoming generation of AR application developers.

Chapter 4. An IAR Framework for P&ID enhanced comprehension.

As already stated, Hand Held Devices in general do not represent a feasible solution inside the industrial environment, anyway there are some tasks in which their use could turn useful to the Augmented Operator. During the doctoral program, the research carried out in this field aimed at exploiting such a visualization technology to enhance users in the comprehension of plant information that is traditionally conveyed through printed Piping and Instrumentation Diagrams (P&ID).

In our experience in the industrial field, we noticed that printed P&ID have the disadvantages to convey limited information by means of graphical signs. Such information is static (its update requires a new drawing by a field expert), is easily understandable only by a limited number of expert operators and is limited to the topology of the machinery. In this application scenario, information retrieval by means of the P&IDs it is a cumbersome task. For instance, if the operator needs the machinery model, the location and the maintenance record, s/he must, in sequence: identify the machinery on the P&ID, read its unambiguous plant-id, identify the corresponding machinery model and its location inside the plant, and, finally, retrieve from the machineries-documentation archive the maintenance records.

When consulting a P&ID an operator does not have to handle other tools, thus s/he can easily use other devices such as a tablet or a smartphone, hence, we believe that this is an application scenario in which HHDs can be a feasible solution. In this chapter[3] we will describe the research carried out to design an IAR framework for handheld devices that enhances users in the comprehension of plant information.

[3] The results of the studies described in this chapter are under publication in the following article:
A. Boccaccio, G. L. Cascella, M. Fiorentino, M. Gattullo, V. M. Manghisi, G. Monno and A. E. Uva, "Exploiting Augmented Reality to display technical information on Industry 4.0 P&ID" Advances on Mechanics, Design Engineering and Manufacturing II, Lecture Notes in Mechanical Engineering series by Springer, JCM Cartagena, June 2018.

4.1. Introduction

One of the biggest impacts of the fourth industrial revolution on industrial companies is the shift from mass production to mass customization of their products. This process will involve a revision of the production chain management models as well as the use of innovative technologies. The new production lines then will be suitable for rapid change in their configuration to satisfy customer requirements. In these smart factories, plants will be even more complex, and their configuration will change over the time (e.g. in case of maintenance, plant upgrade, and so on). It is important to provide operators, working on the plant, with all the updated information about it. For example, designers that are planning a new production need to know the layout and the interconnections between the components of the plant, maintenance operators need information about the history of maintenance of a machine, new operators need to understand how the plant is made, and so on. Currently, this information is stored in the P&ID (Piping and Instrumentation Diagram or Process and Instrumentation Diagram) and in the documentation stored in the factory archives. P&ID is a drawing showing the interconnections between the equipment of a process, the system piping and the instrumentation used to control the process itself.

According to Weber (2016), P&ID are widely used in the planning and maintenance processes in the industry. Common tools for the creation of these graphical plans for hydraulic systems are (amongst others) Autodesk AutoCAD, Microsoft Visio, and Lucidchart. However, the representation through the P&ID of a plant is not the best visualization method, especially for complex plants. Indeed, a deep knowledge of the plant is necessary to quickly understand the function of each machine. The P&ID does not contain additional information regarding machinery, such as the description of the machine's functionality or the maintenance history. It also requires constant updating because of system modifications.

Many companies use P&ID in paper form, for which the recognition of the various components and their functions is often tied to the know-how of the technicians working in the company. To improve the comprehensibility of P&ID, other works have already been presented in the literature. Many specialists have tried to develop systems that automatically transform the P&ID from a paper to a digital form, including the automatic recognition of the component. Arroyo et al. presented a method based on optical recognition and semantic analysis, which is capable of automatically converting legacy engineering documents, specifically P&ID, into object-oriented plant descriptions and ultimately into qualitative plant simulation models (Arroyo, Hoernicke, Rodrıguez, & Fay, 2016). Tan et al. proposed a novel framework for automated recognition of

components in a P&ID of raster form, based on image processing techniques to make a mathematical representation of the scanned image(Tan, Chen, Pan, & Tan, 2016). They further extended this method to acquire also the connectivity among the components (Tan et al., 2016)

With this tool, technicians can easily understand the components and connections in the plants even if they do not know the coding of the symbols used in the P&ID. However, this tool does not help operators in the support of decisions, since it does not provide further information, for example for the planning of maintenance procedures.

In this work, we propose to use Augmented Reality (AR) to help operators in the correct understanding of a plant and to retrieve useful information about the plant (e.g. machines layout, history of maintenance, and so on). AR has been successfully used to support operators in all the phases of the product lifecycle. Nee and Ong provided a review of some studies of AR applications in manufacturing operations, such as product design, robotics, facilities layout planning, maintenance, CNC machining simulation and assembly planning (Nee & Ong, 2013a)(Nee & Ong, 2013b).

Hou et al. developed an AR/VR training system and carried out an experimental study that demonstrated how it could enhance the quality of training with respect to traditional training paradigms, based on P&ID They observed that it could save the manpower, decrease the travel distance during the work, reduce the learning time, and improve the learning performance (Hou et al., 2017). Andaluz et al. developed a VR application to create virtual environments to allow undergraduate students to start familiarizing with physical connections, instrumentation, and equipment as they would be in real process (Andaluz, Castillo-Carrión, Miranda, & Alulema, 2017). The virtual environment was created by converting a P&ID into 3D CAD models using AutoCAD Plant 3D. These training systems provided a high level of details for operators that should make step-by-step procedures on the plant. The main drawbacks of this application are the effort required in the authoring process and the difficulty of the update.

Li et al. made a review of VR/AR prototypes and the related training and evaluation paradigms within the research and construction industry in the past two decades. They found that AR has been effectively used for on-the-job training of operators in the conversion of safety information directly from paper-based plans to actual work. It is important to note that the authors underline the concern about what kind of educational methods, theories, and tools could be smoothly embedded into VR/AR systems to improve the performance of training and education (X. Li, Yi, Chi, Wang, & Chan, 2018).

This concern may be solved with the development and testing of training platforms in different domains, like the one described in this work. We developed a dedicated Augmented Reality framework for an efficient visualization of technical plant information, through a handheld device, augmenting the printed P&ID without modifications to the drawing. The proposed framework helps unexperienced operators to understand complex plants in less time and to retrieve technical information in a more efficient and engaging way with respect to the traditional manner based on the reading from paper documents. Furthermore, technical information is always up to date thanks to the connection to a database where all the plant's data are stored.

The main features of the application that we developed are the followings:

- Display of virtual hotspots in correspondence of plant elements on the P&ID; the hotspots could be of assorted colors to indicate different elements: e.g. pumps, conveyors, filters, and so on;
- Filtering of the hotspots displayed at the same time, grouped either by category (e.g. all the pumps, all the conveyors, and so on) or by subsections of the plant;
- Display of technical information of a selected component of the plant; this could be either plate data (e.g. model number, supplier, efficiency, and so on) or history data (such as maintenance and modifications);
- Display of a 3D representation of a selected component; this could be done through a 3D CAD model of the component, if available, a 3D reconstruction of a scan, or finally a 360° picture of the selected section.

4.2. Materials and methods

The framework was designed using Unity 3D and Vuforia for the AR behavior. As a tracking method, we tested both the Vumark tracking by Vuforia and the image-based tracking, using the P&ID image as trackable. In order to have a precise tracking, the Vumark technique required the camera to frame the Vumark. Thus, it was hard to visualize virtual hotspots far from the Vumark on the drawing. To overcome this problem, we used the extended tracking, but there was a drift, so the location of the virtual object was not always correct. Adding Vumark on existing drawings would not be accepted by companies because it removes space for the drawing unless it was not placed in the title block. However, in this case, it was hard to track the rest of the drawing that in most cases are printed on large sheets. For these reasons we decided to use the image-based tracking using the digital version of the drawing as trackable (Figure 28); an important remark is that all the lines in the drawing should be black in order to achieve the highest tracking quality. Black lines on white sheets are mostly used in technical drawings, according to the drawing standards (UNI EN ISO 128-20:2002), however for P&ID other line colors are often used because many lines may overlap and to distinguish the fluids flowing.

We exported a datasheet of the (X, Y) coordinates of the plant components from “AutoCAD plant 3D” (“AutoCAD Plant 3D,” n.d.) and we used them for the positioning of the virtual hotspots in the Unity scene.

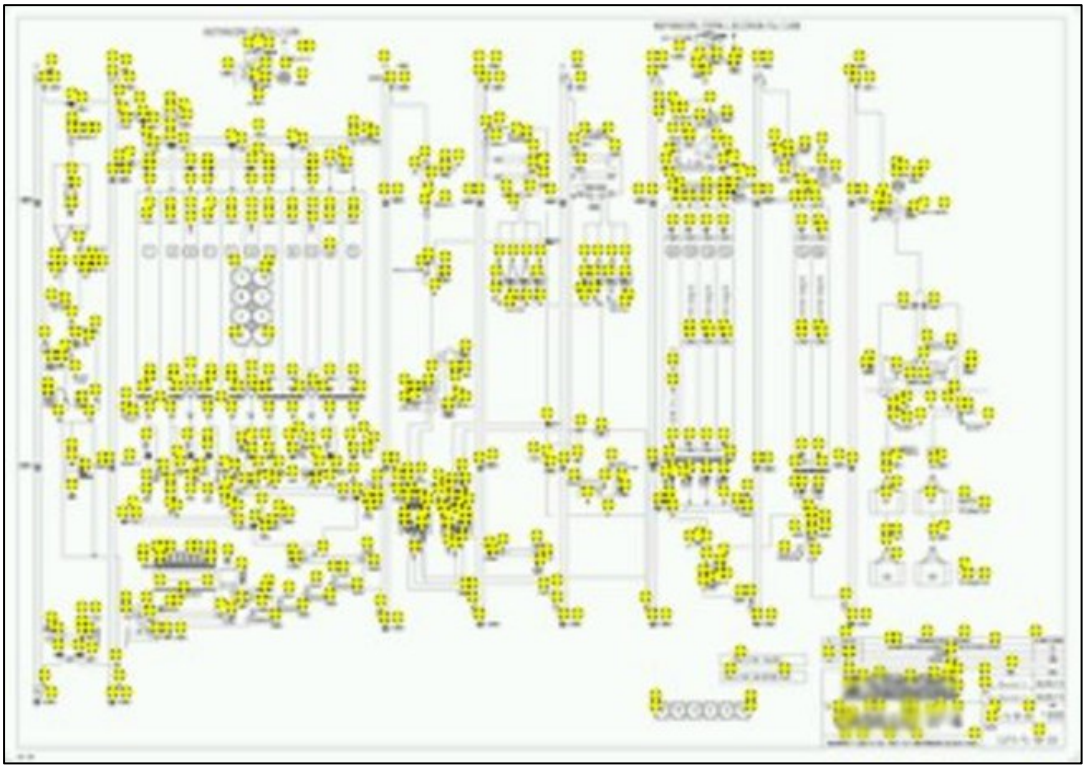


Figure 28 - Feature points used by Vuforia for the tracking

We associated a precise color to the hotspots according to their category, as resulting from AutoCAD classification. Then, we made a script to filter the visualization of the components displayed at the same time (

Figure 29). In this way, users would have fewer troubles in the identification of components in the P&ID.

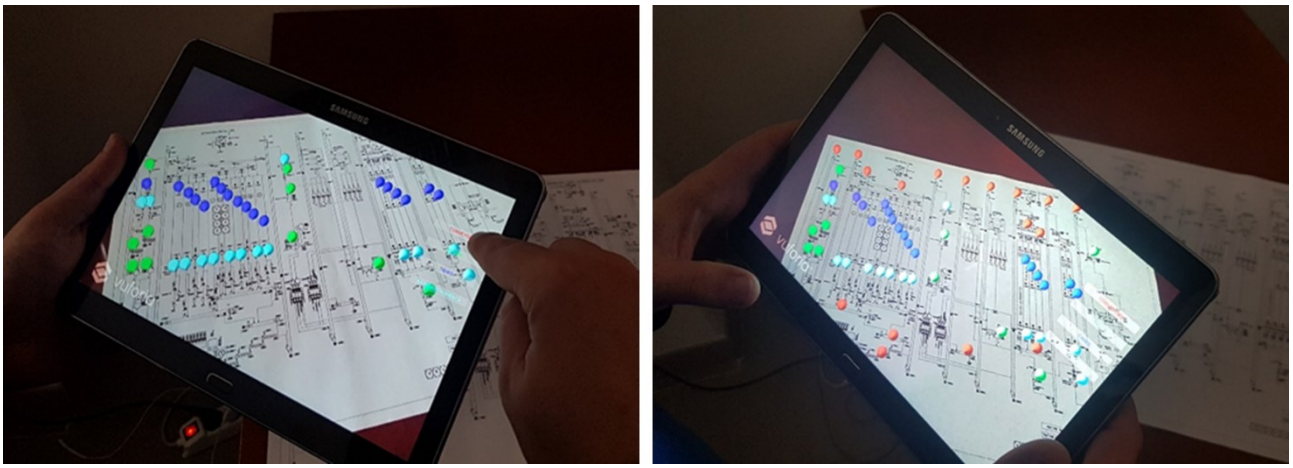


Figure 29 - Example of usage of the application: users can filter the visualization of the hotspots through a menu displayed on the GUI.

When users tap on the virtual hotspot on the device, that hotspot gets bigger, whereas the others get smaller and become not selectable, and the name of the plant component is displayed. A menu appears on the screen with three selectable buttons. We tested both 2D buttons, i.e. with a fixed position on the GUI, and 3D buttons (pie menu) that are registered on the trackable as a generic virtual element (Figure 30).

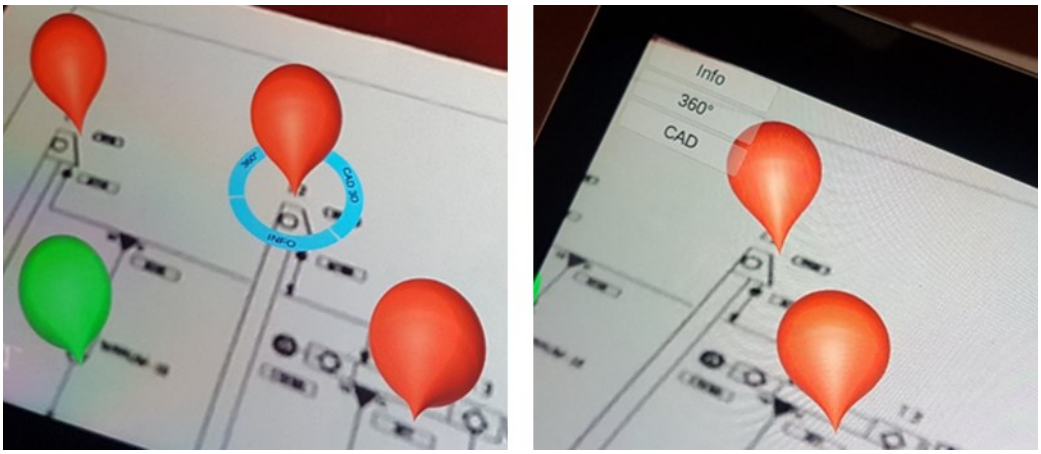


Figure 30 - Comparison of the hotspot menu layout: 3D on the left and 2D on the right.

In the final application, we decided to use the latter because they cause less occlusion of the real world and their usage is more intuitive since they appear just close to the selected hotspot. In fact, when the user clicks on the hotspot the buttons appear (Figure 31) while clicking a second time they disappear.

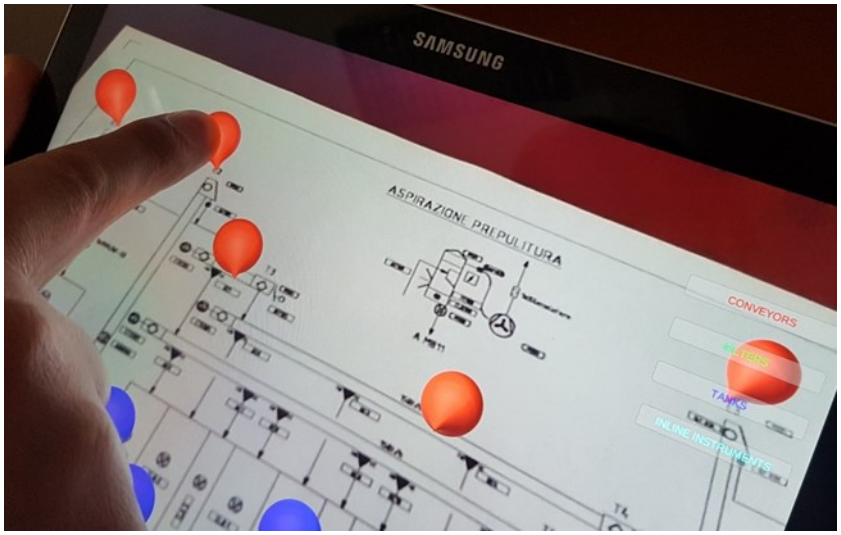


Figure 31 - User tapping on the hotspot to visualize the menu to access technical information.

A first button opens a technical chart of the component with all the information retrieved from a database (Figure 32a). We used an SQLite database automatically generated from AutoCAD Plant 3D. The information can be added either in AutoCAD Plant 3D or in the database since they are synchronized.

A second button opens a navigable 3D CAD model of the selected component to learning how it is made (Figure 32b). The CAD model could be either a model generated from scratch or a mesh reconstruction of a point cloud deriving from the scanning of the real machine.

A third button opens a 360-degree image of the component and its surroundings in the real plant (Figure 32c). In this way, it is possible to learn how the component is connected to the rest of the plant and its real location within the plant.

Figure 32 - Visualization of additional technical information through a pie menu: a technical chart with information retrieved from a database (a), a navigable 3D CAD model (b), and a 360-degree image of a plant section (c)

4.3. Preliminary user test results

As a case study we tested the application developed in this work for the following scenario: unexperienced operators are asked to retrieve information on some plant components displayed on the P&ID. We simulated that they were in the plant, i.e. they could not access to personal computers, but just to paper manuals or handheld devices (smartphone, tablet). The P&ID was that of the cleaning section of a milling plant.

The AR application is that described in section 4.2, and the displayed hotspots are on conveyors, tanks, filters, and inline instruments. From the menu on the GUI users could choose to display the hotspots of the desired classes, e.g. only conveyors. The pie menu allowed the operators to display useful information from a database (e.g. the floor where the selected component is located), a CAD model of the component, and a 360-degree image of the component within the real plant. The application was tested on the tablet SAMSUNG Galaxy Note 10.1, and the smartphone OnePlus 3.

We made a study with undergraduate students to test the usability of the application and gather useful feedbacks to improve the application design. The task consisted of finding information on five different components indicated on the P&ID.

The main results of the usability study are:

- users found the pie menu not so comfortable, especially when the button to be pressed is behind the hotspot;
- users suggested to leave the user the possibility to freely rotate the pie menu around the hotspot;
- users preferred the smartphone from an ergonomic point of view because it is more comfortable to handle;
- users preferred to read information on the tablet because the screen is larger.

The case study developed is just an example of how this application could be used for the training of new operators. However, the application can be used also for other industrial activities. The other potential beneficiaries of this application are:

- Maintainers: they can know the position and all the needed information of the components to be repaired. In this way, maintenance operations would be more accurate and require less time to be accomplished.
- Designers: they can easily make changes to the layout of the plant through the correct understanding of the single components and their connections.
- Security technicians: they can be aided in the updating of the plant security devices through the knowledge of the number, location, and features of the machines.
- Supervisors: they can control remotely the state and the history of all the machines on the plant.
- Visitors: they can be trained on the functioning of the plant through a direct association between the functional information of the drawing and the real form of the components in the plant.

Future works will involve the optimization of the Graphical User Interface through user studies and the development of the framework on other AR devices.

4.4. Conclusions

In this work, we present an Augmented Reality framework for the displaying of technical information of plant components on the P&ID of the plant. Nakai et al. observed the importance of sharing plant information is for quick decisions-making and prevention of miscommunications (Nakai, Kajihara, Nishimoto, & Suzuki, 2017). However, they proposed as a communication mean text messages and P&ID, whereas we propose to enhance the P&ID with digital and updatable information displayed in Augmented Reality with a consequent reduced cognitive load for final users, as shown by the literature in the field (M. Fiorentino et al., 2014; Henderson & Feiner, 2011).

The main development issues addressed in this work, supported by a usability study, are:

- Tracking of the printed P&ID: we compared marker based-tracking with image-based tracking, choosing the second one for this application. Contrary to marker-based tracking, dozens of feature points may be found from a single image, which means that the picture can be partly covered up, and the tracking will still work. With the increased computing speeds of the processors in the modern mobile devices, the tracking speed is no longer an issue with image-based tracking, as it was in the past.
- Choice of the device for the visualization in AR: we designed an application for handheld devices since they are readier to use in an industrial context. Users did not show particular preferences for the smartphone or the tablet. In future developments, we will explore the possibility to implement this framework on other AR devices (e.g. Head-Worn displays, projection tables) with a specific study on user interaction.
- Development of a dedicated user interface: we derived the shape of the hotspot from that of web mapping services like Google Maps. We also tested two types of menu buttons, 2D and 3D, to access technical information about a specific component. We decided to use 3D buttons in the form of a pie menu since they cause less occlusion and are more engaging for the final user. However, users did not feel comfortable the pie menu and they suggested to leave the user the possibility to freely rotate it.

Chapter 5. The Operator 4.0 and the Human Machine Interactions

It is a commonly accepted assumption that the role played by the operator in the smart factory of the future will be accompanied by changing tasks and demands. As the most flexible entity in cyber-physical production systems, operators will be faced with a large variety of jobs ranging from specification and monitoring to verification of production strategies. The new I4.0 technologies are crucial to support the capability of the operators to realize their full potential and adopt the role of strategic decision-makers and flexible problem solvers (Gorecky et al., 2014).

Hence, operators should be integrated into the cyber-physical structure in such a way that their individual skills and talents can be fully realized. A cyber-physical structure describes the relationship between humans and a Cyber-Physical System. The interplay between human and CPS occurs either by direct manipulation, or by means of a user interface. Such an interface can be created with the help of VR and AR. VR allows the user to simulate and interactively explore the behavior of a CPS-based production system. AR, by augmenting the real environment with specific contents, provides a tool for effectively visualizing the information stream generated by the CPS.

Given the complexity of the production process, the operator needs to focus her/his attention on the real environment by minimizing as more as possible the mental effort applied to browse information and interact with the CPS. Traditionally, industrial user-interfaces are based on unimodal interactions with mechanical inputs to the system (by keyboard, mouse, or touch screen) and a visual reply (displayed on a screen). The auditive channel plays a marginal role and is used for instance to alert the user with warning signals.

In the smart-factory working-scenario, users' interactions must be as intuitive as possible and must take into account manufacturing-specific requirements (especially concerning robustness and security). Such requirements call for suitable human-technology-solutions, which will provide transparency for humans concerning the networked and distributed manufacturing systems (Gorecky et al., 2014).

In the factory of the future new user interfaces will flank and sometimes replace the classical ones. These interactions belonging to the Natural User Interfaces (NUI), can be broadly categorized into three main typologies based on the interaction medium:

- Touch-based;
- Voice-based;
- Gestures-based.

Touch-based interactions will exploit new technologies – such as the Dispersive Signal Technology – allowing the use of touch screens in raw, industrial environments. Some hardware manufacturers already provide hardware solutions for mobile applications in manufacturing and logistics areas, with specific attributes such as dust and water splash protection. The drawback of such interfaces relies on the use of the hands to interact with.

Voice-based interactions have the main advantage of enabling the control of relevant applications by speech input. This turns particularly useful in the case where the users' visual attention or their haptic capabilities are fully occupied by the task to be accomplished. The main drawback of such interfaces is related to their use in noisy environments such as the factory shop-floor.

Gesture-based interfaces do not suffer from this problem. They allow controlling devices with natural gestures similarly to speech recognition, furthermore these interfaces, if designed according to a human-centered approach, are particularly intuitive and immediate.

The gesture recognition process is based on the human body tracking techniques that can be image-based or device-based. The device-based gesture-recognition techniques use wearable acceleration or position sensors to record and recognize user's movements. Image-based methods use techniques for object recognition and image processing. These methods may use highly accurate, but expensive techniques (such as OptiTrack® and VICON®), and low-cost, but flexible interaction means (such as Microsoft Kinect®).

Body tracking not only allows for direct user interaction, but, such technology, embedded into the corresponding assistance system, can be used to track, check, and classify user actions into the correct use context.

NUIs could provide effective user-interactions, especially in the factory shop-floor where devices such as mice and keyboards are hardly employable. Anyway, there is a need for ergonomically designed user interfaces, which would enhance user productivity, user acceptance, and user satisfaction, in one world *usability*.

In this chapter[4] we will describe the work carried out to develop a gesture-based interface for the navigation of virtual environments, and the experiments conducted to access its acceptance among users. Although not specifically designed for the industrial environment, the developed interface shows how it could effectively replace the classical mouse based one. Furthermore, this research paved the way for the work presented in chapter 6 that describes a general framework for mid-air gesture interface design.

5.1. Introduction

The Cultural heritage of a country represents a priceless patrimony both for its inhabitants, to understand and explain the origin of customs and traditions, and for the touristic industry. The use of new technologies, as well as novel interaction paradigms, were recognized to be one of the key-points to approach the mass audience and hence to adequately promote and present cultural heritage (Carrozzino & Bergamasco, 2010). Among the technologies recently utilized to this purpose, Virtual Reality (VR) plays a role of predominant relevance (Carrozzino & Bergamasco, 2010; Mortara et al., 2014; Pullambaku & Tsering, 2016). By allowing the user to interact with the virtual world in an immersive environment, VR represents one of the most appealing and effective ways to improve the users' engagement and attract their attention on specific cultural heritage subjects.

VR allows replacing the real visit with the exploration of reconstructed historical places and virtual museums (Styliani, Fotis, Kostas, & Petros, 2009), by means of virtual tours (i.e., immersive 360-degree panoramas (S. E. Chen, 1995)) and the manipulation of digital artifacts. Consequently, VR can provide both the preservation of these sites and the access of general public, and add all the potentialities related to the multiple virtual experiences, in terms of interaction and immersion (Guerra, Pinto, & Beato, 2015).

Natural User Interfaces (NUI) that are at the basis of VR-based exhibitions, - such as virtual tours or virtual museums (Barbieri, Bruno, & Muzzupappa, 2017) -, are the object of the study of a large number of researchers throughout the world.

[4] The results of the studies described in this chapter have been published in the following articles:
Manghisi, V. M., Fiorentino, M., Gattullo, M., Boccaccio, A., Bevilacqua, V., Cascella, G. L., Dassisti, M., et al. (2017). Experiencing the Sights, Smells, Sounds, and Climate of Southern Italy in VR. IEEE computer graphics and applications, (6), 19–25;
Manghisi, V. M., Uva, A. E., Fiorentino, M., Gattullo, M., Boccaccio, A., & Monno, G. (2018). Enhancing user engagement through the user centric design of a mid-air gesture-based interface for the navigation of virtual-tours in cultural heritage expositions. Journal of Cultural Heritage, 32, 186–197.

Three principal approaches are followed in the design of interfaces: (i) user-centric, (ii) individual or customized and (iii) centrist or authoritarian. The first approach utilizes an elicitation procedure where the user is asked to propose possible gestures to execute specific commands/referents. Interface inputs are then designed on the basis of the gesture proposals. In the individual approach each individual defines his/her own vocabulary (Kahol, Tripathi, & Panchanathan, 2006). The centrist analytical approaches designs interfaces by basing upon technological constraints and/or the experience of experts in the specific application field. However, the user-centric approach represents the standard methodology in NUIs design (Nielsen, Störring, Moeslund, & Granum, 2004) (Morris, Wobbrock, & Wilson, 2010) as it allows pursuing two important objectives: (i) lowering the cognitive load and (ii) improving the user experience.

In Touristic/cultural events, VR technologies can be conveniently utilized to 'immerse' the user in virtual tours reproducing the typical environment of a country thus letting him feel as physically present in 'that' place that can be thousands and thousands of miles away. During the doctoral program, we also worked at the design and development of the Multisensory Apulia Touristic experience (MATE). The MATE consists of a Smart-Multisense Ubiquitous System where a 20-foot container is utilized to host visitors giving them visual, climatic and olfactory stimuli (Figure 33).

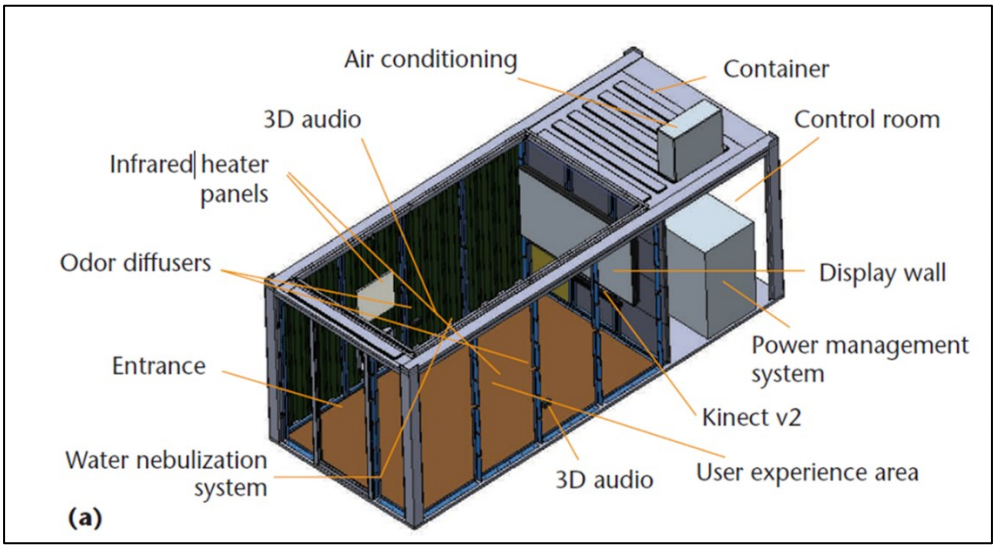


Figure 33 – The configuration of the Multisensory Apulia Touristic Experience (MATE) mobile container.

In this chaprer, we will describe how we further developed this system by designing and implementing a gesture-based interface for the navigation of virtual tours on a display wall (the display was hypothesized to be fixed on one of four walls of the container) and investigating the capability of such an interface to enhance the engagement/enjoyment of user's experience.

A number of studies utilized gesture-based interactions with display walls (Pullambaku & Tsering, 2016; Ren, Li, O'Neill, & Willis, 2013; Vatavu, 2012), but their application scenario is quite different from the navigation of virtual tours. Indeed, scenes in virtual tours almost consist of spherical panoramas that users can explore by using a zooming function, changing gaze direction or selecting active items. Other studies (Koehl et al., 2013; Kwiatek & Woolner, 2009), also reported in the literature, utilize VR technologies to carry out virtual tours but do not implement gesture-based interfaces.

5.2. Research aim

From the review of the state of the art, it appears that no specific studies are reported in the literature focused on the design, the implementation and the validation of gesture-based interfaces for the navigation of virtual tours in cultural heritage exhibitions. In detail, no studies are available that adopt the user-centric gesture-elicitation procedure for the definition of the vocabulary of gestures capable of guaranteeing the users' acceptance and consequently, the users' engagement/enjoyment. Our hypothesis is that including an "immersive" gesture-based interface improves the user's experience thus giving her/him the sensation of "exploring" in a seamless manner the virtual world.

Therefore, the aims of this work are:

1. To describe the application of such an elicitation procedure to the design process of a mid-air gesture-based NUI;
2. To test the developed NUI and evaluate its usability and hence its effectiveness in improving the engagement and the enjoyment of users with respect to the classical mouse-controlled interface.

5.3. The gesture vocabulary design

The gesture vocabulary, i.e., the set of gestures utilized by the user to interact with the interface, has to be straightforward and intuitive in such a way that the resulting NUI satisfies the general requirements of the low cognitive load and the low physical fatigue (Blake, 2012). The design process of the vocabulary was structured into three principal phases:

- The *Interface requirements definition* phase. In this phase, the set of commands that are the most suited to the specific scenario/application are defined. Three aspects must be considered in this phase: (i) the specific tasks that have to be performed, (ii) the

environment/context where tasks must be accomplished, and (iii) the set of commands required to perform the tasks.

- The *Gesture elicitation* phase. Spontaneous gestures the users would intuitively use to trigger the interface commands are proposed and collected. The agreement analysis then follows.

- The *Vocabulary definition* phase. The gesture proposals, collected in the previous phase, are ranked according to their guessability. The Vocabulary of gestures will include, for each command, the most intuitive gesture among those proposed for that command.

Hereafter, the following definitions will be utilized:

- Referents: interface commands, in this study five referents were hypothesized to be necessary to conduct and control a virtual tour (see next Section);

- Gesture proposals: gestures proposed by the participants during the elicitation phase as interacting metaphors to "execute" a given referent;

- Vocabulary: a combination of different gestures devoted, each, to "execute" a given referent.

5.3.1. Interface requirements definition

The following five interface commands were hypothesized to be strictly necessary to properly conduct and control a virtual tour on a wall display:

- move the pointer on the screen;
- zoom-in;
- zoom-out;
- change gaze direction (i.e., the solid angle of the spherical pano visualized on the display);
- select items.

5.3.2. Gestures elicitation procedure

The context/environment where the virtual tour was hypothesized to take place is a container with an average space of 8 m^2 available for users (see Figure 33). Containers are "portable" installations that can be easily transported to the location where cultural/touristic events are organized and properly equipped with all the devices required for a virtual tour. Following the hypothesis that the container can host up to 4 visitors simultaneously, the proposed virtual

exhibition system was designed to switch the control among users according to a defined policy. Users were hypothesized to be not familiar with gesture-based interfaces which were, therefore, designed as intuitive as possible.

For the gesture elicitation process, a population of 29 participants (average age 21.3 years, SD=3.11) was recruited, including 18 males and 11 females, all right-handed. All of them were students in Mechanical and Computer Engineering: five participants declared to use an Xbox Kinect for recreational purposes regularly; nine utilized the device a few times; fifteen never used it.

Before performing the elicitation, a preliminary investigation was carried out to evaluate the users' preferences on the two control modes that are available in the software suite *krpano* (Reinfeld, 2016) utilized to implement the virtual tour. In detail, this suite includes two main components: (i) a set of tools to create and edit virtual-tours, and (ii) a viewer, embeddable into HTML pages, which allows the navigation of virtual tours using a web browser. In the default configuration, the tour is only sensitive to system events, such as those triggered by the keyboard and/or the mouse. Two control modes are available to change the gaze direction. The first one, which we will call as "drag&drop" mode, permits the user to grab the scene by keeping the left mouse-button pressed, and to change the gaze direction by moving it together with the mouse pointer. In the second control mode, which we will call as "move-to" mode, while the left mouse-button is kept pressed and the mouse is moved, a vector is defined, the direction and the amplitude of which allow controlling the changes of gaze direction and the speed of its movement, respectively. After explaining the two control modes to all the participants, they were asked to test both of them using the standard mouse-controlled interface. The test had a duration of about 4 minutes. At the end, participants were asked to express their preference. 27 participants (i.e., 93.1 % of the recruited population) preferred the drag&drop mode which led us to elicit gestures just for this control mode.

The elicitation procedure adopted a user-centric conscious bottom-up approach (Nielsen et al., 2004) where referents are first explained to the participants, then the gesture proposals for each referent are collected. The Wizard of Oz study-setup (Dahlbäck, Jönsson, & Ahrenberg, 1993) was followed to carry out the elicitation procedure.

The experimenter asked each participant to think of the possible mid-air hand-gesture she/he would use to trigger each of the five referents. Then, staying in front of the display wall (a professional UHD 85 inches Samsung QM85D monitor), - at the distance of 2 meters- showing a scene of the virtual tour, each participant was asked to execute the gesture previously thought

for each referent. According to the Wizard of Oz study-setup, a hidden experimenter triggered the corresponding commands via the mouse thus giving the participant the impression of activating her(him)self the command via the executed gestures. All the gestures executed during this phase were recorded and analyzed. To this purpose, a desktop PC with a CPU Intel Core i7 6700 (3.4 GHz) 16GB RAM and a GeForce GTX 970 graphic adapter were utilized.

Figure 34 shows and briefly describes all the gestures proposed for the execution of the five referents while a more detailed description of them is given in Appendix B.

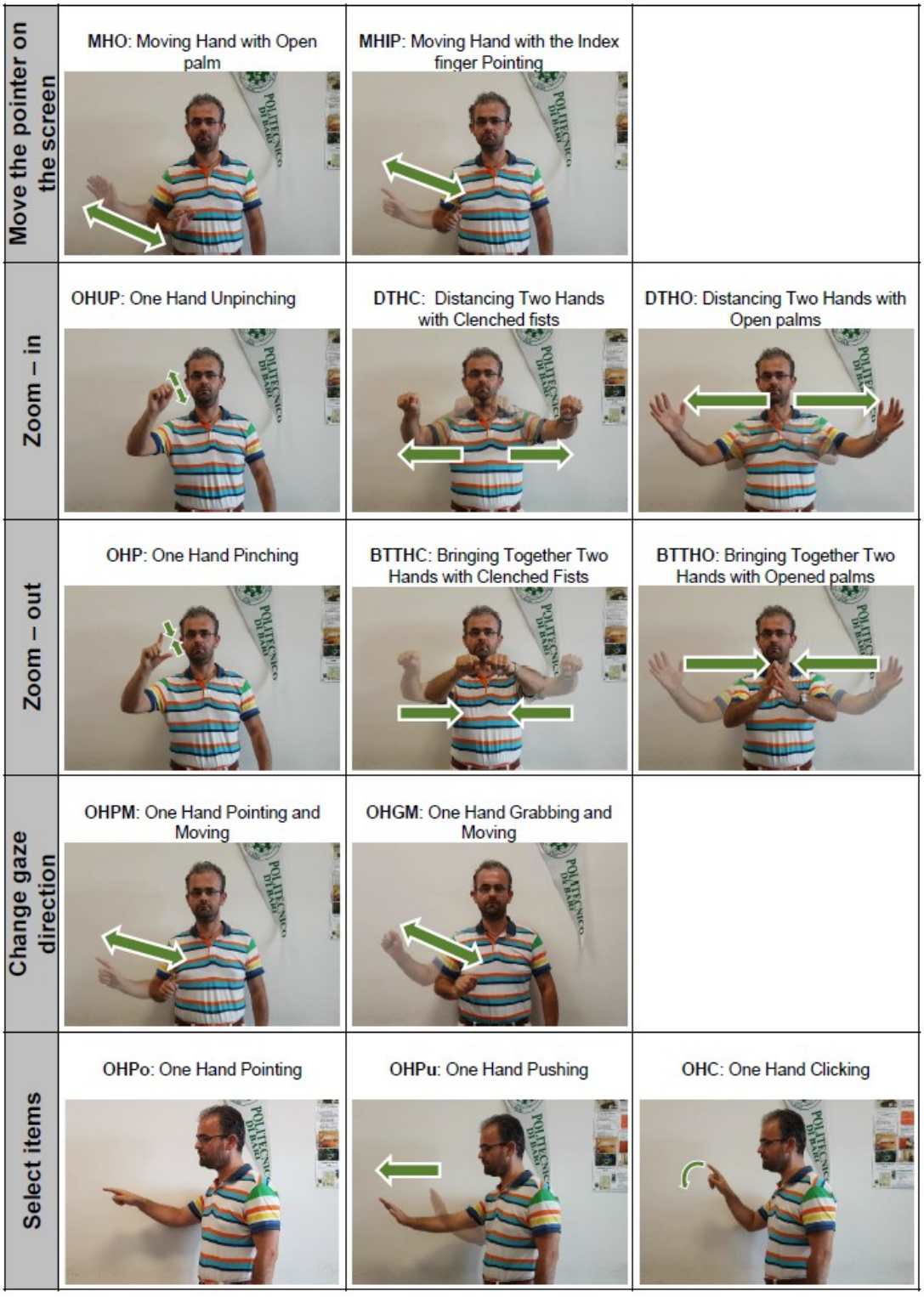


Figure 34 - Gesture proposals collected for each of the five hypothesized referents. The acronyms utilized to identify each gesture proposal are reported on the top of each picture.

Acronyms were utilized to make more compact the notation. Figure 35 shows the distribution of the gesture proposals, i.e., the number of times a given gesture was utilized to execute a specific referent.

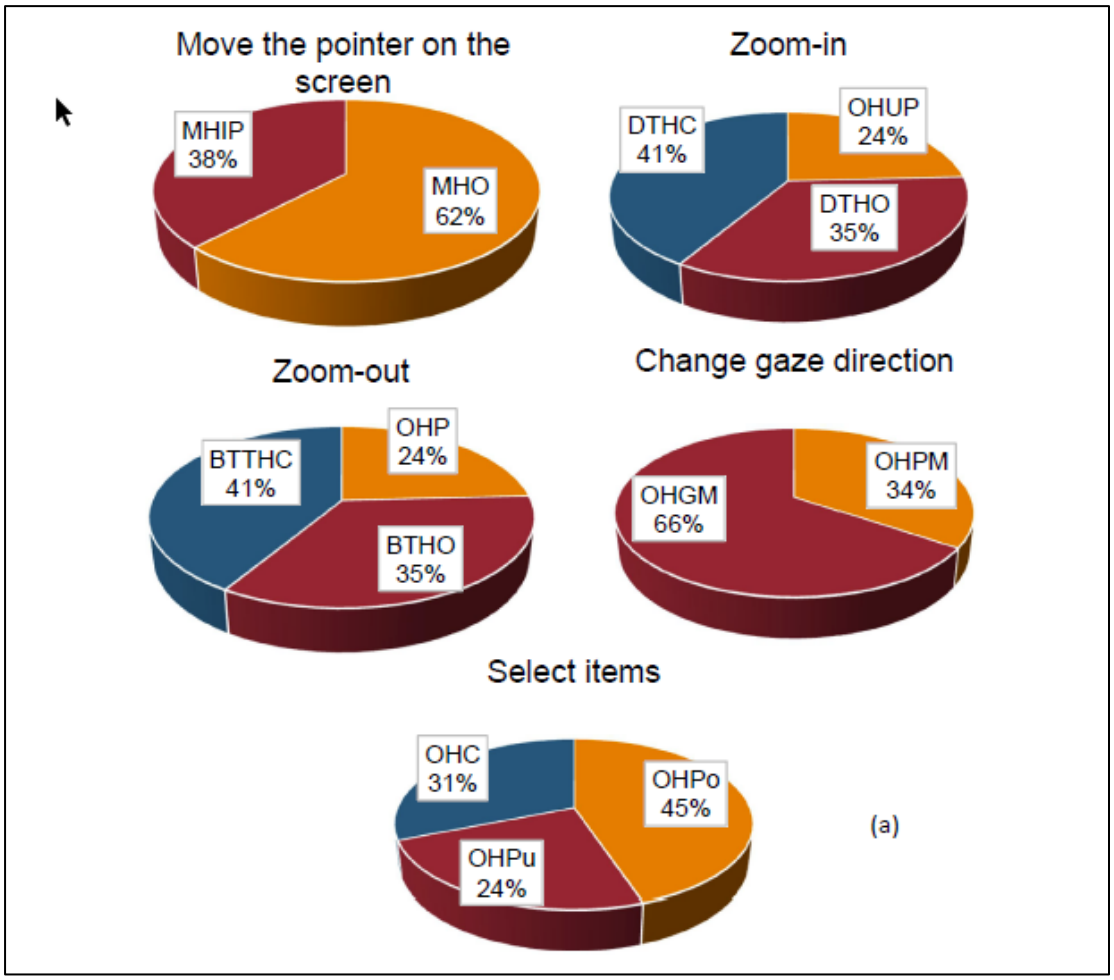


Figure 35 - Percentage of gesture proposals collected for each referent. The acronyms utilized are specified in Figure 34.

For each referent, the Agreement Rate (AR) was finally computed which is a coefficient that measures the agreement that exists between the gestures proposed for each referent (Vatavu & Wobbrock, 2015). AR ranges in the interval [0, 1]; for a given referent r_k, the value $AR(r_k) = 0$ means that all the proposals collected for the referent r_k are different from each other, on the contrary, the value 1, indicates that all the proposals (gathered for r_k) are the same. Generally speaking, a low value of AR implies a high cognitive load and hence the necessity of re-designing the set of referents. The values of AR computed in this study were large enough, which led us to conclude that the five hypothesized referents do not require a high cognitive load. Further details on the computation of the Agreement Rate values are given in Appendix C.

It is worthy to note that 3 out of the 13 proposed gestures (Figure 34) are almost the same: One Hand Pointing (OHPo), One Hand Pointing and Moving (OHPM) and Moving Hand with Index finger Pointing (MHIP).

5.3.3. Vocabularies definition

To select the best gesture for each referent, the *gesture guessability G* (Piumsomboon, Clark, Billinghurst, & Cockburn, 2013; Wobbrock, Aung, Rothrock, & Myers, 2005; Wobbrock, Morris, & Wilson, 2009) was computed, which is a factor that 'measures' the intuitiveness of each gesture with respect to the corresponding referent. The *gesture guessability G* is given by:

$$G_i^k = \frac{P_i^k}{P_{TOT}^k} \; ; \; G \in]0, 1]$$

where P_i^k is the number of times the ith gesture was proposed for the kth referent and $P^k{}_{TOT}$ is the total number of the proposals collected for the kth referent.

Ranking each gesture according to the guessability (Figure 36) led us to define the best vocabulary of gestures utilized to carry out the virtual tour.

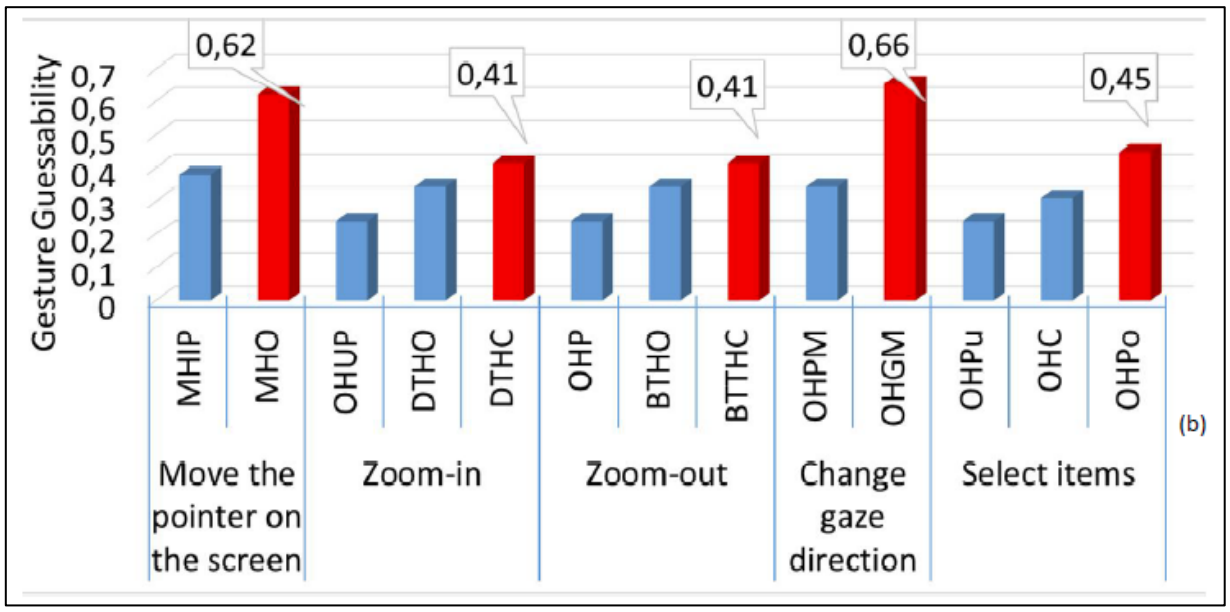


Figure 36 - Values of guessability computed for each gesture proposal. The bins highlighted in red refer to the gestures with the highest guessability and that were finally utilized in the best gesture vocabulary.

5.4. The NUI implementation

The hardware setup used for the NUI was the same as that described for the elicitation procedure with the addition of the Kinect v2 RGB-D camera that was utilized as a tracking sensor. This device was successfully employed in different domains of applicability that go from gaming (Jones et al., 2014) to ergonomics assessment (Vito Modesto Manghisi et al., 2017) to the navigation of complex 3D archaeological scenes (Fernández-Palacios, Morabito, & Remondino, 2017).

The designed interface is based on a software developed using the C# language, the Windows Presentation Foundation libraries (.NET framework), and the Microsoft Kinect for Windows SDK 2.0. One of the essential and critical requirements of the proposed interface was to develop a gesture recognition system reliable and capable of adapting to people differently shaped and sized (e.g., tall and short people, adults and children, left and right-handed users).

A control flow (Figure 37) supervised the users' navigation, where the user's activity is considered as a state machine. The clock of this state machine is event-based and triggered by the arrival of a new frame from the depth sensor. The sequence of the main actions controlled for each new-frame is the following:

1. Definition of the user that is enabled to lead the navigation of the virtual tour, hereafter we will call her/him as the user-leader.
2. Definition of the user state. For each frame, the system has to detect if the user wants to execute a gesture, and if so, which action she/he intends to trigger on the interface.
3. Triggering of the detected actions.

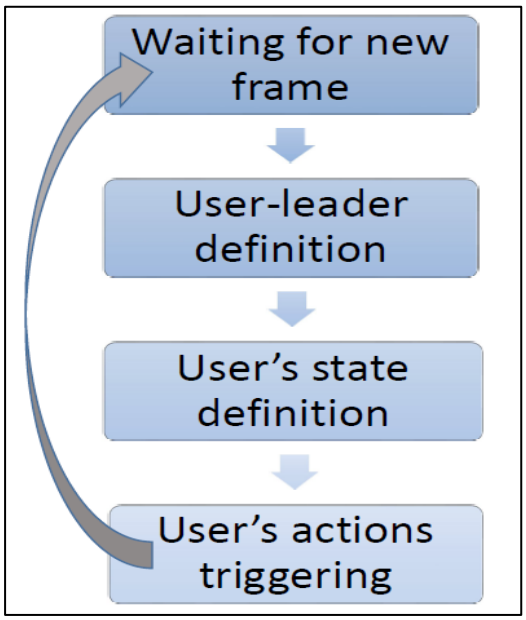


Figure 37 - Sequence of the main actions controlled for each new-frame detected by the sensor.

5.4.1. User-leader definition

Following the hypothesis that users will wander through the virtual tour in a standing position and based on the studies of the optimal position a user should occupy to maximize the accuracy and the reliability of the Skeleton Tracking algorithm for Kinect v2 (Q. Wang, Kurillo, Ofli, & Bajcsy, 2015), a "virtual area" was defined as the space located in front of the display wall, starting and terminating at 120 cm and 450 cm, respectively, from it (Figure 38). Amongst the

visitors occupying the "virtual area," the user-leader is defined as the one that is the nearest to the sensor.

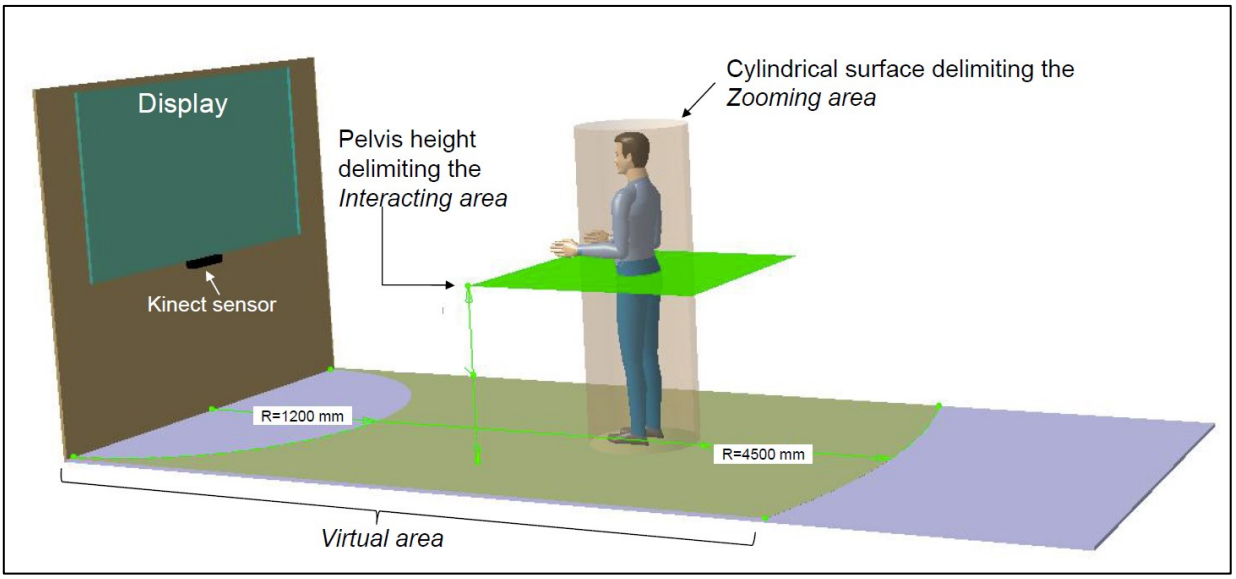


Figure 38 - Schematic of the working areas defined (with their limit dimensions) to control the user's interaction.

5.4.2. User's state definition

The state machine that detects the state of the user-leader works through 5 states: unengaged, zoom, move the pointer on the screen, change gaze direction and select items (Figure 39).

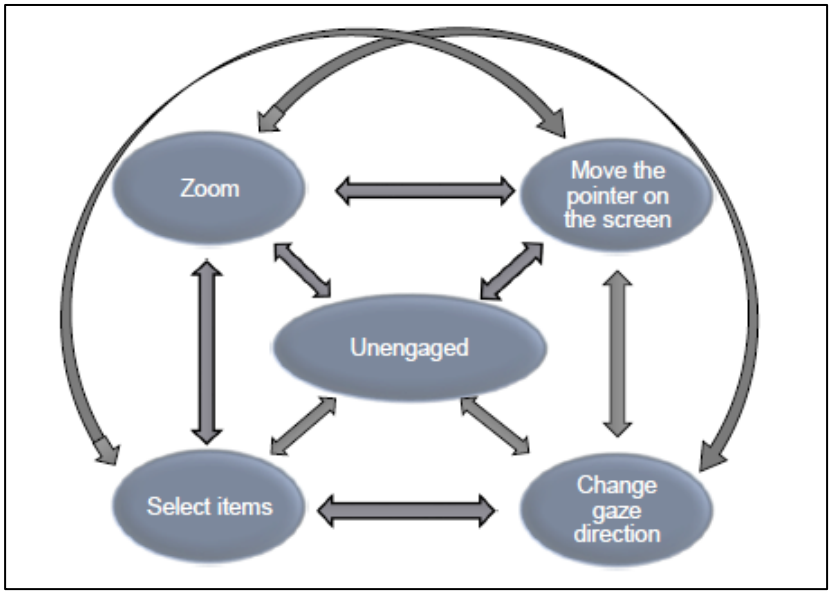


Figure 39 - The state machine that controls the user's behavior

These states are detected by evaluating, during the navigation, the state and the position of each hand of the user-leader. The Kinect v2 skeleton-tracking algorithm allows four states for each tracked hand to be identified: unknown, open, closed, and lasso (Figure 40).

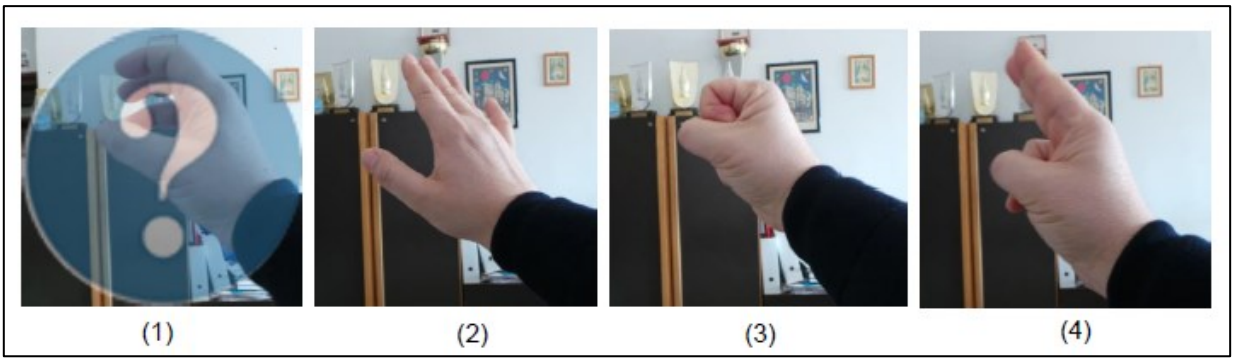


Figure 40 - The hand states detected by the Kinect skeleton tracking algorithm for the tracked hands: 1) Unknown; 2) Open; 3) Closed; 4) Lasso.

Based on the state of each hand and the combination of the states of both hands, following a specific policy described below, the state machine is identified.

To define the rules for the identification of the state machine, the following working areas were defined.

Interacting area. Fixed an ideal horizontal plane (highlighted in green, Figure 38) at the height of the user's pelvis, the interacting area is defined as the space above the plane. If both the hands of the user are below the plane, no interactions with the system are triggered. The contrary occurs once one or both the hands are in the interacting area. The height of the plane delimiting the interacting area was fixed at the level of the pelvis which is a good compromise between two opposite requirements. A too "high" interacting area has the disadvantage to increase the fatigue related to the arm movements, the so-called gorilla arm effect (Hincapié-Ramos, Guo, Moghadasian, & Irani, 2014). With a too "low" interacting area, instead, the risk of unwanted interactions due to natural movements of the hands, can be run.

Zooming area. Fixed an ideal cylindrical surface around the user with a diameter equal to the distance between the user's shoulders, the zooming area is defined as the space outside the cylinder (Figure 38). It is worthy to note that the designed interface allows the control from the right to the left hand and vice-versa to be switched in real time. To this purpose, the control-hand was defined as the hand that is kept at the largest distance with respect the floor. During the interaction, if the control-hand is kept lower than the other for more than 1 second, the control switches to the other hand. With this strategy, the user can change the control-hand in a natural manner and hence rest the tired arm, but also the system is suitable for both, left- and right-handed users.

The state machine that controls the user's behavior starts from the Unengaged state (Figure 39). In this state, the user-leader is not actively interacting with the interface. To start the interaction,

she/he has to raise one of her/his hands until reaching the interacting area (i.e., the space above the ideal horizontal plane passing through the user's pelvis). When the user is engaged, her/his state changes on the basis of the state of the control-hand and according to the following rules:

- If the state of the control-hand is open, the user's state turns to Move the pointer on the screen;
- If the state of the control-hand is closed and the state of other hand is open, or it is outside the zooming area, the user's state turns to Change gaze direction;
- If the state of the control-hand is lasso, the user's state changes to Select items;
- If hands are inside the zooming area and their state is closed, the user's state changes to Zoom.

In order to improve the reliability of the system, thus avoiding any sudden and unwanted changes of the state (between the right and the incorrect one) of the hands, the strategic approach of the most voted policy was adopted. In detail, the state and the position of each hand in each frame were stored in four queue-like buffers (i.e., two queues for the state and two queues for the position). As soon as new frames are detected by the sensor, all the buffers are updated by removing the frame on the bottom and inserting the new one on the top. Then, the hand state is identified as the most frequent state in the buffer.

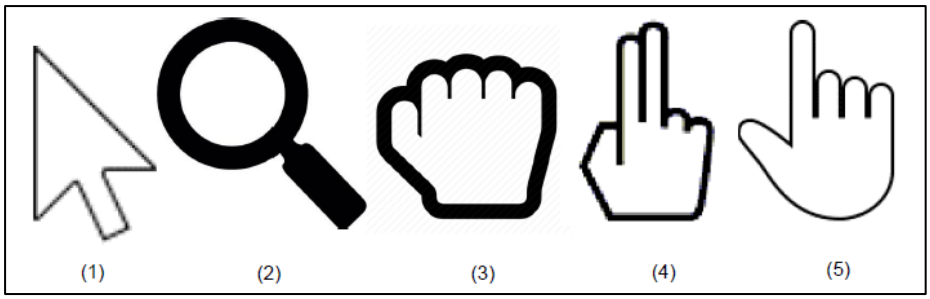


Figure 41 - Pointer icons visualized on the screen and utilized as a feedback to make aware users about their current state: 1) Move the pointer on the screen; 2) Zoom; 3) Change gaze direction; 4) Select items; 5) Hotspots pointer on-over event.

In order to make aware users about their interaction with the system, graphical cues were utilized as a feedback. The pointer icon (Figure 41) changed on the screen according to the actual state of the user. A specific pointer for Hotspots on-over event was included in addition to the pointers referring to the user's states: Move the pointer on the screen, Zoom, Change gaze direction, and Select items.

5.5. User's actions triggering

The Move the pointer on the screen action allows the user to control the pointer position on the scene in front of her/him and can be triggered by executing the MHO (Moving Hand with Open palm) gesture (Figure 34). To interact with the krpano viewer, a software simulation of the move-cursor event was utilized to trigger the actual action on the virtual tour. The pointer location is controlled by a ray casting method, where the position of the head detected by the Kinect skeleton-tracking algorithm is utilized as the projection center while the pointer position on the scene is determined by projecting (on the screen) the ray that connects the head and the control-hand (Figure 42).

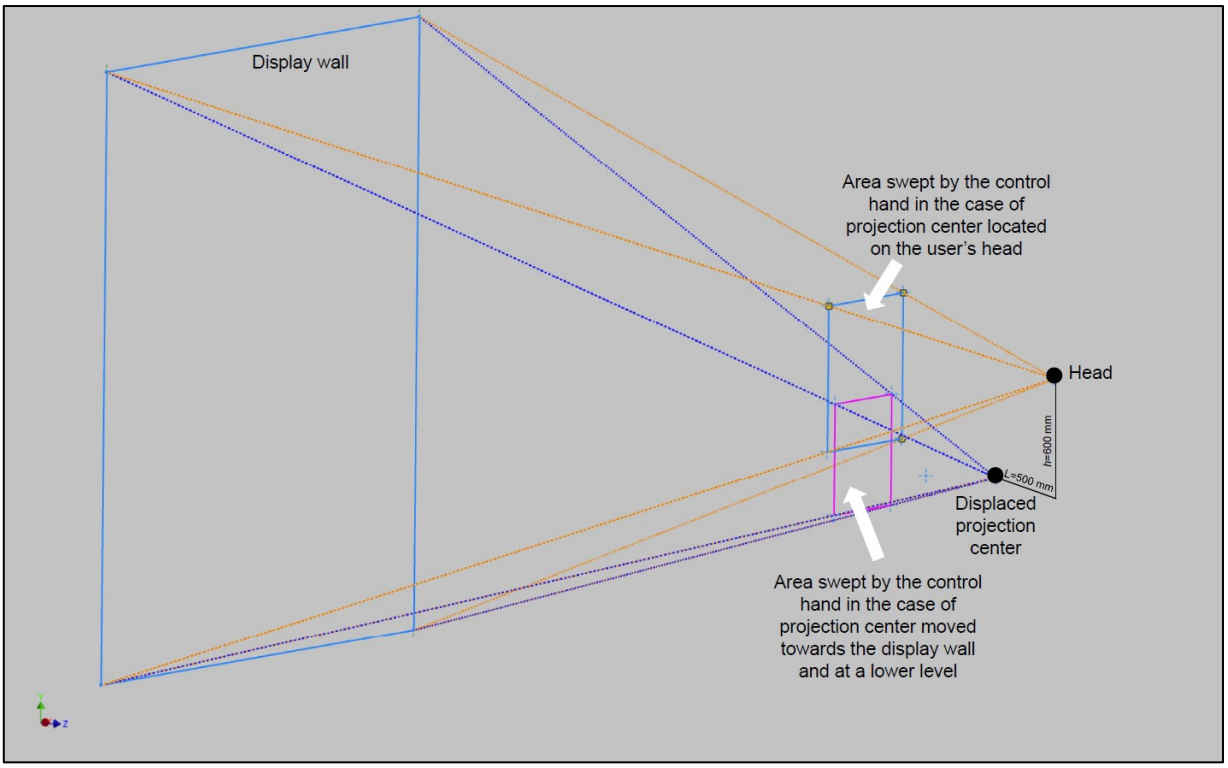


Figure 42 - Schematic of the ray casting method adopted to control the pointer location. The area (delimited by the magenta rectangle, in the case of projection center located in correspondence of the user's head) that the user's hand must sweep to roam the displayed scene becomes smaller (area delimited by the blue rectangle), when the projection center moves from the position of the head towards the display wall (by the quantity L=500 mm) and at a lower level (by the quantity h=600 mm).

Proper filtering methods were applied to minimize the jittering effects from which the Kinect skeleton-tracking algorithm suffers. Regarding the location of the hands, a median filtering approach was adopted, while regarding the head position, an accumulation method. This method stores, frame by frame, in a queue-like buffer the position occupied by the head and assumes as the actual position of the head the one computed as the mean value of all the positions stored in the buffer. As new frames are acquired, if the computed mean values change more than a fixed

threshold, then, the position of the head is updated; if, conversely, the new computed mean value differ less than a fixed threshold, then, the position of the head is hypothesized not to change. Although this approach allows a perfect alignment between the cursor position on the scene and the target of the user's gaze, presents two main limitations. The first is related to fatigue required to the user that, to control the pointer position, has to keep her/his hands raised for a long time without rest. The second is related to the interposition, between the user's eyes and the scene, of the user's control-hand. To overcome these limitations (albeit in a partial way) the projection center was moved at a lower level (h=600 mm) and towards the display wall (L=500 mm), as shown in Figure 42. Adopting this solution, the user can control the position of the pointer with smaller and at lower level movements thus reducing the fatigue required to execute them.

The Zoom actions allow the user to zoom-out and zoom-in on the scene and are triggered with the execution of the BTTHC (Bringing Together Two Hands with Clenched fists) and the DTHC (Distancing Two Hands with Clenched fists) gestures, respectively (Figure 34). The user acts as she/he is at the center of a sphere on whose internal surface the scene is stuck. Therefore, to zoom-out, she/he grabs the scene with her/his hands and reduces its dimensions by decreasing the distance between hands. Conversely, to zoom-in, she/he grabs the scene with the hands and increases the distance between them as if she/he is stretching the sphere. A software simulation of a mouse-wheel-scroll event triggers the actual action. The entity of the scrolling is proportional to the variation of the relative distance between the two hands in two subsequent frames.

The Change gaze direction action that is triggered by executing the OHGM (One Hand Grabbing and Moving) gesture, allows the user to control the direction of her/his view. The user acts as she/he is at the center of a sphere on whose internal surface the scene is stuck. Therefore, to change the direction of view she/he only has to grab the sphere and drag it wherever she/he likes. In this last action, the position of the pointer is controlled as in the Move the pointer on the screen action. A software simulation of both a mouse left-button-pressed event and a move-cursor event triggers the actual action on the virtual tour.

The Select items action allows the user to select/activate an item on the scene and is triggered by the execution of the OHPo (One Hand Pointing) gesture. To select an item (which is indicated by the Hotspots pointer on-over event) the user has to move the pointer over the corresponding hotspot and select it by pointing at it with her/his index finger. The Kinect skeleton-tracking algorithm is tailored to detect the lasso state, where, both the index and the middle fingers point. However, we found that the algorithm detects the lasso state even in the case the user points with

the only index finger. A software simulation of a mouse left-button-click event triggers the actual action on the virtual tour.

5.6. NUI testing and evaluation

A user study was carried out to evaluate the effectiveness of the proposed gesture interface and, in particular, to obtain a comparative evaluation in terms of perceived usability, user engagement/enjoyment and overall users' preferences between gesture and mouse-controlled interfaces. Given the strong legacy bias related to the use of the mouse-controlled interface, the following three hypotheses were formulated:

H_1-The mouse-controlled interface will achieve a better or at least comparable score in terms of usability;

H_2- The gesture-based interface will achieve a better score in terms of enjoyment/engagement;

H_3- The overall users' preferences will be more oriented towards the gesture-based interface.

5.6.1. Experimental procedure

We adopted two different experimental setups. In the first one, the user stood in front of the display wall at an average distance of 2 meters and navigated the virtual tour using the gesture interface. In the second experimental setup, the user kept the same position and wandered the tour by means of a wireless mouse controller using a desk as mouse support. Following Ayoub (1973), the desk upper surface was placed at 102 cm of height from the ground. Each participant tested both setups in two consecutive sessions. The order of execution was counterbalanced over users. In order to minimize any learning effect, a time of at least 15 minutes was awaited between the end of the first session and the beginning of the second one.

A total of 16 voluntary participants (13 males and 3 females, all right handed) were recruited among engineering, students, and researchers at Polytechnic University of Bari. The average age was 28.1 years (min 22 years, max 42 years, SD = 5.6 years). None of the participants ever interact-ed with a Kinect V2 sensor-based interface, 10 out of 16 had no previous experience with mid-air gesture interfaces while the other 6 had previous experience with the first version of the Kinect sensor.

Each session was supervised by an experimenter and consisted of a training- and a test-phase. In the training-phase the participant watched a video tutorial explaining the use of the interface under test and then had a free-training period of two minutes on a demonstrative tour (different

with respect to the virtual tour successively navigated). After a break of 3 minutes the test-phase started. In this phase, the user was asked to navigate freely the virtual tour of Murgia and to use each interacting metaphor at least twice. After a participant accomplished the minimally required tasks, she/he was free to interrupt the test phase. The experimenter recorded the time duration of this phase.

After each test-session, the participant was asked to estimate the time she/he spent to carry out the test and to fill in a usability satisfaction questionnaire (Barbieri et al., 2017) and a customized Intrinsic Motivation Inventory (IMI) questionnaire (the IMI questionnaire is freely available at http://selfdeterminationtheory.org/intrinsic-motivation-inventory, more details on the administered questionnaires are given in the next Section). Table VIII reports the independent and dependent variables of the experimental procedure.

Table VIII - Independent and dependent variables of the presented experiment.

INDEPENDENT VARIABLES		
Participants	16	13 males, 3 females
Interfaces	2	Gesture interface, Mouse-controlled interface
DEPENDENT VARIABLES		
Learnability		Mean of two answers on a 7 point likert scale
Interface efficacy		Mean of two answers on a 7 point likert scale
System efficacy		Mean of two answers on a 7 point likert scale
Time duration percent difference $t_{\%}$		$\frac{\text{Estimated time duration} - \text{Actual time duration}}{\text{Actual time duration}} * 100$
Interest/enjoyment		7 point likert scale
Perceived competence		7 point likert scale
Effort/importance		7 point likert scale
Value/Usefulness		7 point likert scale
Felt pressure and tension		7 point likert scale
Preferred interface		7 point likert scale
Easiest to use interface		7 point likert scale

After the participants filled in the questionnaires, the experimenter had an interview with them to gather their impressions of the navigation experience.

5.6.2. **Metrics: Administered Questionnaires**

Following Barbieri et al. (Barbieri et al., 2017), a seven point Likert scale usability satisfaction questionnaire was administered including 6 items, which is suited to catch cognitive aspects related to user satisfaction. The posed questions can be gathered in couples thus forming three

sub-groups corresponding to the following three sub-scales: learnability, interface efficacy and system efficacy.

In order to compare the interfaces, the users' subjective experience was measured by administering a post-test Intrinsic Motivation Inventory (IMI) questionnaire which is a flexible assessment tool that determines the subjects' (i) interest/enjoyment, (ii) perceived competence, (iii) effort, (iv) value/usefulness, (v) felt pressure and tension, and (vi) perceived choice while performing a given activity, thus yielding six sub-scale scores (Deci & Ryan, 2003; Ryan & Deci, 2000). This test was successfully utilized in several experiments related to intrinsic motivation and self-regulation (Deci, Eghrari, Patrick, & Leone, 1994; Plant & Ryan, 1985; Ryan, Koestner, & Deci, 1991; Ryan, 1982) as well as in studies regarding tasks execution in virtual environments (Colombo et al., 2007; IJsselsteijn, Kort, Westerink, Jager, & Bonants, 2006; Mihelj et al., 2012; Novak, Nagle, Keller, & Riener, 2014). Indeed, a customized version of the questionnaire was utilized that includes 22 questions and neglects the perceived choice sub-scale (i.e. the sub-scale (vi)).

The proposed questionnaires included two further questions: (i) the first one regarding the preferred and the easiest to use interface (gesture-based or mouse-controlled); (ii) the second one regarding the time the user thinks has spent to conduct the test. The comparison between this time and the actual time measured by the experimenter was expressed in terms of percent difference $t_{\%}$. In other words, if $t_{estimated}$ is the time estimated by the user to execute the test while t_{actual} is the actual time measured by the experimenter to carry out the test, $t_{\%}$ can be computed as:

$$t_{\%} = \frac{t_{estimated} - t_{actual}}{t_{actual}} \times 100$$

5.6.3. **Results**

The usability satisfaction questionnaire returned usability scores for the mouse interface higher than for the gesture interface (Figure 43(a)).

Paired samples t tests (Table IX) show that these differences are statistically significant for the Interface efficacy and the System efficacy sub-scales while no statistically significant differences can be seen for the Learnability sub-scale.

Table IX - Descriptive statistics and paired samples t test for the usability satisfaction questionnaire sub-scales.

Sub-scale	Interface	N	Mean	Std. D.	t(df=15)	p
Learnability	Gesture	16	6.38	0.62	-1.576	0.138
	Mouse	16	6.56	0.12		
Interface efficacy	Gesture	16	5.31	1.24	-2.573	0.021
	Mouse	16	6.38	0.81		
System efficacy	Gesture	16	5.72	1.00	-2.674	0.017
	Mouse	16	6.50	0.52		

In the IMI questionnaire, the gesture-based interface obtained, for the Interest/Enjoyment and the Value/Usefulness sub-scales, median scores significantly higher than the mouse-controlled interface counterpart (Table X).

Table X - Descriptive statistics and Wilcoxon Signed-ranks test for the IMI questionnaire sub-scales.

Sub-scale	Interface	N	Median	Min	Max	Z	p
Interest / Enjoyment	Gesture	16	30.50	26	35	-2.425	0.015
	Mouse	16	28.50	18	34		
Effort / Importance	Gesture	16	24.00	11	34	-1.337	0.181
	Mouse	16	21.00	11	32		
Perceived competence	Gesture	16	29.50	16	35	-0.427	0.669
	Mouse	16	29.00	22	35		
Pressure / Tension	Gesture	16	12.00	8	17	-1.355	0.176
	mouse	16	11.50	9	16		
Value / Usefulness	Gesture	16	18.00	14	21	-2.098	0.036
	Mouse	16	16.00	10	20		

The time spent to carry out the test was averagely underestimated by 12.08 % for the gesture-based interface and overestimated by 10.88 % for the mouse-controlled interface (Table XI). Statistically significant differences could be found between the two time estimations. Finally, users preferred the gesture-based interface (exact binomial test p=0.021) while judged the mouse-controlled interface as the easiest to use (Figure 43(b)).

Table XI - Descriptive statistics and paired samples t test for the values of t% computed for the mouse-controlled and the gesture-based interfaces.

Variable	Interface	N	Mean $t_{\%}$	Std. D.	t(df=15)	p
Percent difference	Gesture	16	-12.08%	44.91%	-2.516	0.024
	Mouse	16	10.88%	47.42%		

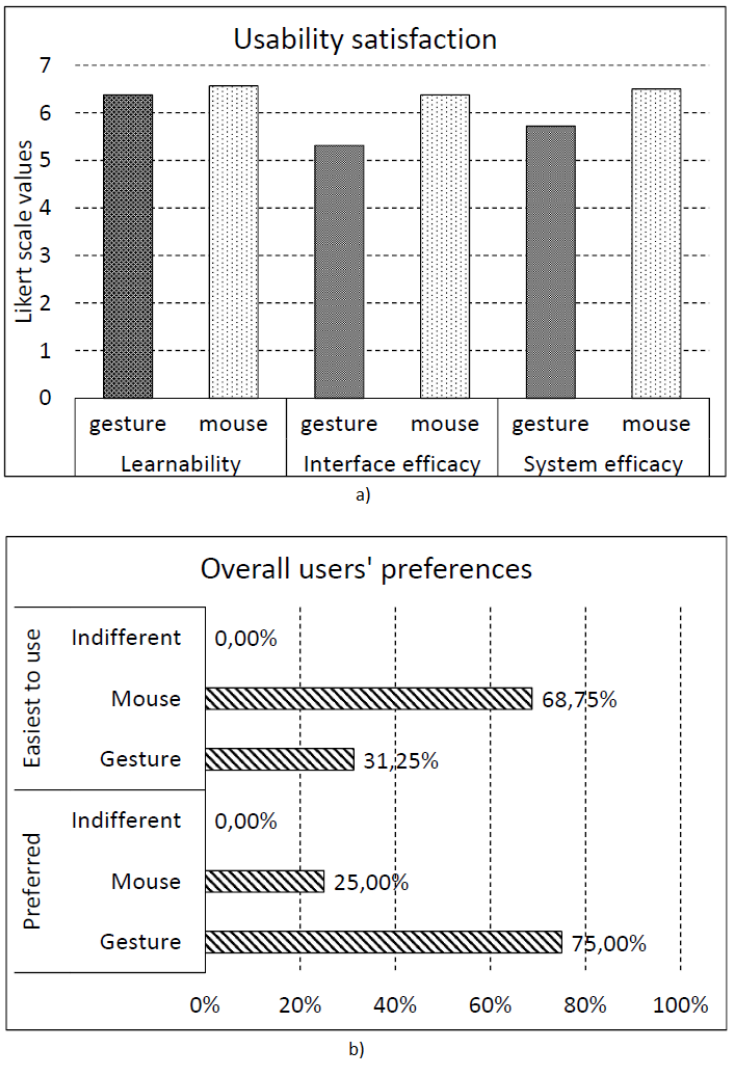


Figure 43- (a) Average values computed over the 16 participants of the usability satisfaction questionnaire scores obtained for the different sub-scales: learnability, interface efficacy and system efficacy. (b) Overall users' preferences: Preferred and Easiest to use interface.

5.7. Discussion and Conclusions

The goal pursued in the study is the development of an appealing interface capable of improving the users' engagement/enjoyment thus attracting their attention and interest towards the exploration of specific cultural heritage subjects.

Given the novelty of the proposed gesture-based interface compared to the well-known mouse-controlled one, the hypothesis H_1 was formulated as: *users evaluate the mouse-controlled interface the most usable*. Indeed, the Interface efficacy and the System efficacy sub-scales (see Table IX) of usability satisfaction questionnaire confirmed our hypothesis H_1. Furthermore, coherently with this hypothesis, the overall users' preferences indicated the mouse as the easiest to use interface (Figure 43(b)). Regarding the sub-scale of Learnability (Table IX), no statistically significant differences exist between mouse and gestures. This result leads us to conclude that the user-centric approach adopted in the design process, was capable of giving to the proposed interface an intuitiveness comparable with that obtained by the well-known mouse-controlled interface.

Regarding the Interest/Enjoyment sub-scale of the IMI questionnaire, the gesture-based interface reached a score significantly higher than the mouse-controlled interface (Table X). Similarly, the score related to the sub-scale Value/Usefulness was significantly higher for the gesture-based interface, which can be interpreted as the argument that when interacting by gestures, users give more value to their navigation experience than when interacting by mouse. Furthermore, the computed values of *t%* (Table XI) further confirm the fact that the interaction by gestures involves so much the user that she/he loses the sense of the time. These results, definitively, lead us to conclude that the hypothesis H_2*: the gesture-based interface will achieve a better score in terms of enjoyment/engagement*, holds true. For the other sub-scales of the IMI questionnaire, no statistically significantly different scores have been obtained (Table X), which implies that the interfaces gesture-based and mouse-based are practically equivalent from this point of view. From the overall users' preferences (Figure 43(b)) it appears that the preferred interface is the gesture-based one. However, the mouse-controlled interface appears to be the easiest to use. These results support the hypothesis H_3*: The overall users' preferences will be more oriented towards the gesture-based interface.*

Interestingly, during the open interviews, participants confirmed that the use of a new interface modality was challenging and enjoyable at the same time. Compared to the gesture-based interface the mouse-controlled one was boring even if most of the users judged it as the easiest to use interface. Different participants claimed also to feel the gesture-based interface more "natural" and more capable of making natural/real the virtual environment.

The results obtained in this study underline how a gesture-based interface could effectively replace a mouse-based one in VR applications. Indeed, the gesture-based interface is well accepted among users and has a learnability comparable to the one of a mouse-based interface that is commonly used in everyday tasks, furthermore it increases the sense of immersion, thus improving the effectiveness of VR applications. The lower usability scores could be easily improved by means of a short training period, thus exploiting the advantages of such an interface in the factory shop-floor, such as the capability of interacting with the CPS without using handheld devices. Anyway, the use of such interfaces in the industrial environment requires dedicated studies.

The research described in this chapter presents some limitations. First, despite our efforts to reduce fatigue and the choice of adaptive thresholds (utilized to define the areas for the interaction) based on user's anthropometric data, some users still complain about the gorilla-arm effect (Hincapié-Ramos et al., 2014). This suggests that the proposed gesture-based interface is attractive for novice users but, needs to be further improved in the case of long-lasting user-interactions. Second, a user-centric approach was adopted to define the gesture vocabulary. Generally speaking, sometime it is cumbersome to develop a reliable gesture-recognition software capable of detecting a gesture vocabulary defined via a user centric approach (Molchanov et al., 2016). For instance, this issue was encountered to recognize the gesture proposed for the referent "Select items," i.e., One Hand Pointing (OHPo, Figure 34). The skeleton-tracking algorithm included in the Microsoft Kinect SDK was not designed to detect such a specific gesture. A possible strategy that can be adopted to overcome this limitation consists in identifying this hand state by using the joints information returned by the skeleton-tracking algorithm and implementing some hand-tracking library. However, such a solution would increase the computational cost thus slowing down the response time of the user-interface and resulting in a poor user-experience. The issue of triggering the Select items referent was addressed by utilizing the detection of the lasso state. Preliminary observations confirmed, in fact, that even if the user executes the OHPo gesture, the gesture-recognition system recognizes it as a lasso gesture and consequently triggers the correct action on the interface.

Chapter 6. A general framework for mid-air gestures-interfaces design

In the previous chapter we observed the capability of a gesture-based interface to replace a mouse based one in VR applications. Furthermore, we noticed that the user-centric approach helps designing interfaces with a low cognitive load. Anyway, the carried-out study is related to the application of VR and gesture interfaces in the Cultural Heritage field, hence it is not straight forward translating such results also to the factory shop-floor. Anyway, encouraged by such results, we extended our study to propose a general framework for mid-air gesture interface design applicable to the industrial environment. In this chapter[5] we describe the proposed framework and its application to a specific use case: the design of a mid-air gesture-vocabulary for the navigation of technical instructions in digital manuals for maintenance operations.

6.1. Introduction

With the fourth industrial revolution the industrial context is experiencing a deep renewal. Also technical documentation – first and foremost instruction manuals – is involved in this process by increasingly assuming the form of digital publication and of other fruition metaphors. Different interfaces have been proposed for the navigation of such technical instructions during maintenance tasks: a laptop keyboard (Watson, Curran, Butterfield, & Craig, 2008), a touch screen smartphone and a wrist-worn controller (Henderson & Feiner, 2011), a wireless 3D mouse (M. Fiorentino et al., 2014), a speech recognition system (Schwald & De Laval, 2003b) and wireless data gloves (Witt, Nicolai, & Kenn, 2006). However, these interfaces are often not suited to the maintenance workspaces which are usually noisy, dirty and lack a reliable surface for handling a mouse (M. Fiorentino et al., 2014). In previous studies, our research group observed how, in such an environment, interfaces based on mid-air gestures might represent an effective and alternative solution to navigate instructions in maintenance applications (Michele Fiorentino, Radkowski, Boccaccio, & Uva, 2016; Michele Fiorentino, Radkowski, Stritzke, Uva, & Monno, 2013; Michele Fiorentino, Uva, Monno, & Radkowski, 2016). However, clear and

[5] At the time of writing the research presented in this chapter is under the first-round review process for publication in "IEEE Transactions on Human-Machine Systems".

well-established guidelines for the optimal design of mid-air gesture-based interfaces are, to date, not available.

General design guidelines are available (Aigner et al., 2012; Microsoft, 2014), but designing and evaluating a gesture-based interface in specific application contexts, such as maintenance workspaces, still remains an unexplored and open issue. Maintenance is a complex activity that involves mental and coordination skills, safety issues, physical effort and cognitive fatigue in uncontrolled environments. To reduce the cognitive load the number of the interface commands has to be lessened as much as possible. Gestures must be selected among those that comply with operational requirements and working environments. Moreover, mid-air interactions lead to fatigue of the upper limbs, a condition noticed since the early 70's and later termed as "the gorilla arm effect" or syndrome (English, Engelbart, & Berman, 1967). Thus, the ergonomics of mid-air gestures is a critical issue for professional applications that may last several hours a day. The interface alone, in fact, may require an unacceptable workload thus leading even to musculoskeletal disorders.

In this chapter we describe a general framework for the design of mid-air gesture-based interfaces aimed at supporting interface developers in the task of designing a vocabulary of gestures. The proposed framework adopts a user-centric approach in order to minimize both the operators' physical effort and the cognitive load, and combines traditional methods with innovative ones to investigate the optimal design of gesture-based interfaces for maintenance workspaces. We describe in parallel both the steps composing the framework and their application to the use case study. After the definition of the interface requirements, different gestures were elicited and utilized to identify, first, sets of gestures and hence, candidate vocabularies. Therefore, based on criteria related to the guessability, the ergonomics and the cognitive load, the best gesture vocabulary was determined. A validation procedure is also proposed and utilized to compare gesture vocabularies in terms of fatigue and cognitive load.

6.2. Related works

The design of an optimal gesture-based control vocabulary requires a tradeoff among various variables such as the accuracy and the speed of recognition, the intuitiveness, and the ergonomics of the gesture. Stern et al. distinguished three main approaches for gesture vocabulary design (Stern, Wachs, & Edan, 2008a): (i) the centrist or authoritarian approach, where a system developer decides which vocabulary should be used (Kirishima, Sato, & Chihara, 2005); (ii) the user-centric or consensus approach, where a group of users decides on a common vocabulary to

express a given set of commands (Munk, 2001); (iii) the individual or customized approach, where each individual defines his/her own vocabulary (Kahol et al., 2006).

The user-centric approach is commonly preferred to other approaches (Morris et al., 2010; Nielsen et al., 2004) and represents the standard methodology in natural user interfaces. For instance, the approach was successfully adopted by Piumsomboon et al. (Piumsomboon et al., 2013) for augmented reality applications. They elicited user-defined gestures and observed that users tended to adopt reversible gestures, that is, those that can be performed in the opposite direction, when commands had an opposite effect when executed.

As the performance of the Operator 4.0 become more and more demanding in terms of mental effort, it is important to provide her/him with tools facilitating her/his tasks with no additional efforts. A crucial important objective pursued by the user-centric methodology and that turns particularly useful in maintenance workspaces, is the lowering of the cognitive load and the improvement of the user experience. Wobbrock et al. (Wobbrock et al., 2005) proposed a unified approach to evaluate and maximize the intuitiveness of symbolic inputs and formalized two metrics, that is, the "guessability" - intuitiveness of a specific gesture compared to others - and the "agreement" - the consensus of a specific set of commands among users-. By developing a conflict-set rule, they also addressed the conflict set problem, that is, the problem that arises when the same gestures are proposed for more than one referent (i.e., command).

To consider both the intuitiveness of the gestures and the speed and accuracy of the recognition system, measurements of technical factors were carried out. Stern *et al.* proposed a methodology that considered both psycho-physiological measures (intuitiveness, comfort) and gesture recognition accuracy (Stern, Wachs, & Edan, 2006, 2008b). Their results, obtained with static gestures, can be used as a data repository of intuitiveness and comfort measures. Hessam *et al.* (Hessam, Zancanaro, Kavakli, & Billinghurst, 2017) focused on gesture set optimization by assessing and minimizing possible confusions deriving from the use of the same gesture for different tasks. Pereira *et al.* (Pereira, Wachs, Park, & Rempel, 2015) took into account user ratings, cognitive load and ergonomics. They created a 3-D hand gesture set for common HCI (human-computer interactions) tasks guided by user-generated gestures, with the final selection based on user ratings, estimation of postural risk, and consideration of system capabilities.

The user-centric approach presents two potential pitfalls: the *legacy* (Morris et al., 2014; Wobbrock et al., 2009), and the *performance bias* (Ruiz & Vogel, 2015). The former occurs because users' proposals are often biased by their experience with prior interfaces and technologies that have been standard for a long time. The latter because, in the elicitation phase,

the aspects related to the repetitiveness of the gesture and hence to the consequent fatigue because of repeating it many times a day – as occurs in maintenance workspaces – are not considered. Possible solutions to overcome these limitations were proposed (Morris et al., 2014) (Ruiz & Vogel, 2015). Two main approaches are currently utilized to evaluate/measure users' fatigue. The first one employs subjective ratings acquired through interviews or questionnaires, such as the CR10 Borg scale (Gunnar Borg, 1998), and the NASA-TLX (Hart & Staveland, 1988; Hart, 2006). The second approach is objective and is based on direct measurements such as intramuscular pressure and tissue oxygenation (Jensen, Laursen, & Sjøgaard, 2000), electromyography (Chittaro & Sioni, 2012) and limb position (Vito Modesto Manghisi et al., 2017; McAtamney & Nigel Corlett, 1993). Hincapié-Ramos *et al.* (Hincapié-Ramos et al., 2014) proposed a new metric to assess arm fatigue quantitatively and objectively, the consumed endurance (CE). The theory of CE, is based on Rohmert's formulation (Rohmert, 1960) that expresses the maximum amount of time that a muscle can maintain a contraction level before needing rest. Recently the American Conference of Governmental Industrial Hygienists (ACGIH, n.d.) formalized an exponential expression for the upper limb exertion time-limit fatigue.

By reviewing the state-of-the-art, we can conclude that the user-centric approach has been widely used for the designing of gesture-based vocabularies. However, it should be integrated with technical measurements to address outstanding issues such as fatigue, memorability, and legacy and performance bias.

6.3. The proposed framework

According to Blake (Blake, 2012), the proposed framework for the identification of the optimal vocabulary of gestures satisfies the following general requirements:

- low cognitive load, which means, in detail:
    a) intuitive and easy to learn gestures;
    b) wide agreement among users;
- low physical fatigue.

A schematic of the framework is depicted in Figure 44: the output of each phase is reported at the end of the corresponding block, the intermediate outputs of the sub phases inside each block are reported on the left.

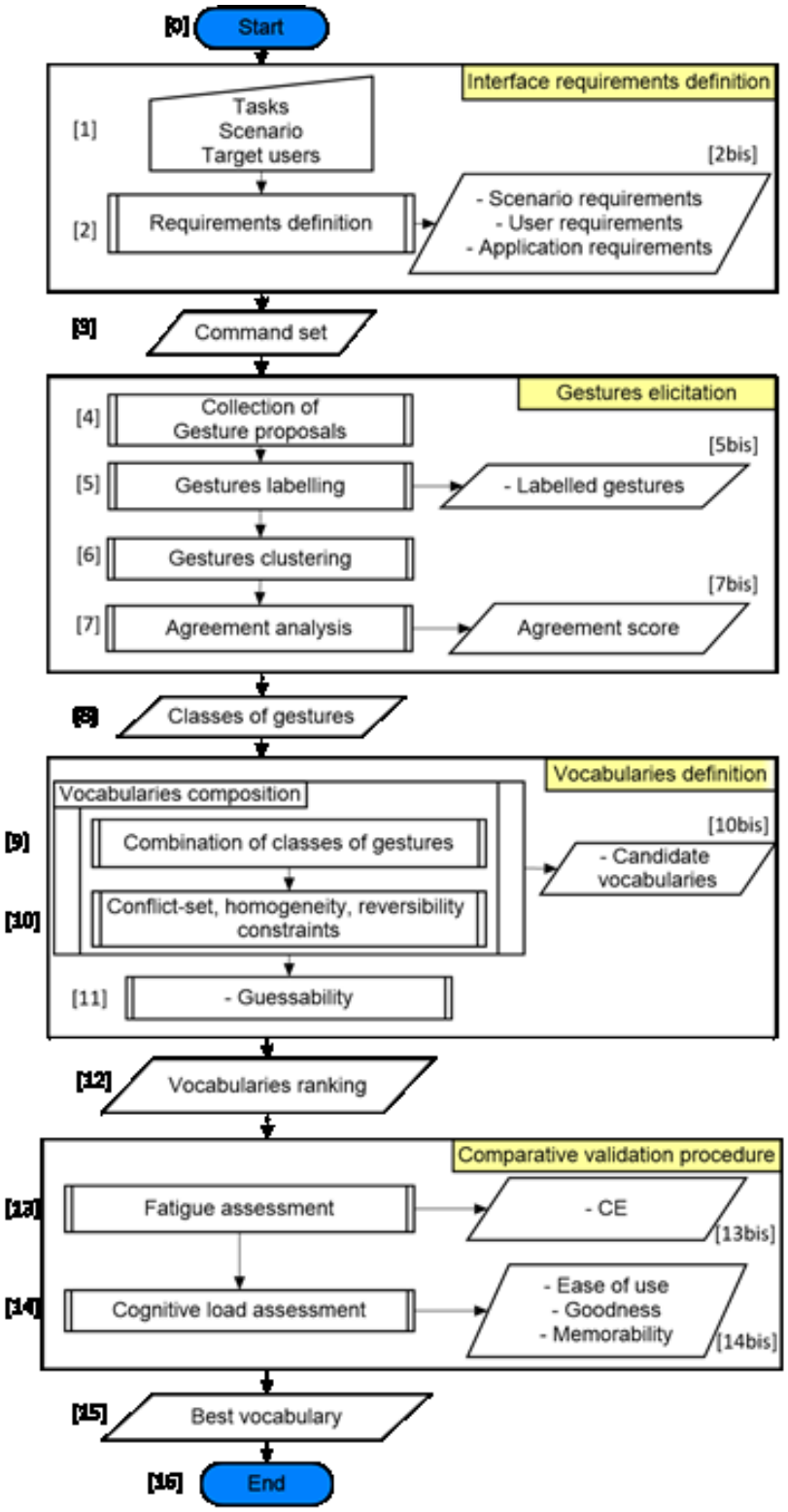


Figure 44- Flowchart with the four main phases of the framework.

The four main rectangular blocks correspond to the core phases of the framework. Each one of these phases outputs an intermediate result that constitutes the input for the next step of the design procedure. These phases are:

I. The **"Interface requirements definition"**: returns as output the set of commands most suited to the specific scenario/application;
II. The **"Gesture elicitation"**: collects spontaneous gestures proposed by the users for the interface commands and generates sets of gestures;
III. The **"Vocabularies definition"**: generates a list of candidate vocabularies of gestures and allows ranking them to choose the best one;
IV. The **"Comparative validation procedure"**, measures factors related to the ergonomics and the cognitive load of the vocabularies previously defined. The output of this step is a complete set of quantitative/qualitative indicators that guide the designer to choose thoughtfully the best vocabulary for specific applications.

In the following sections we will describe each phase by showing how it applies in the case study of designing a set of gestures for navigating technical instructions. Hereafter, we will use the following definitions:

- Referents: interface commands;
- Gesture proposals: gestures proposed by the participants during the elicitation phase to "execute" a given referent;
- Sets of gestures: gestures finally selected/identified among all the gesture proposals after implementing proper filtering and clustering processes described below;
- Vocabularies: combinations of different sets of gestures, each (set) devoted to "execute" a given referent. As in this study four referents were hypothesized, and each set of gestures executes one referent; it follows that each vocabulary includes four sets of gestures.

6.4. The case study: Interface Requirement Definition

This phase requires the definition of:

- the specific tasks to accomplish;
- the specific environment where tasks must be accomplished;
- the target users;
- the command set necessary to accomplish the tasks.

To do this, the following situation was hypothesized to take place in a maintenance workspace (Figure 44, Block [1]). A technical operator browses a maintenance instruction manual in digital format in the industrial working scenario. This environment is normally dirty, noisy and often lacks table-like supports, which means that input methods such as voice, touch and handled input devices, must be excluded. Moreover, the attention of the technician must be principally devoted to handling mechanical parts and maintenance tasks and a minimal cognitive load should be required to browse instructions (Figure 44, Block [2], Block [2bis])).

Based on these requirements, we defined a possible set of referents (Figure 44, Block [3]). In a previous study, our research group successfully organized the instructions for a maintenance procedure in a digital manual with a tree-like structure (M. Fiorentino et al., 2014). In this way, a technician can easily skip well-known details and access specific information if necessary. This kind of browsing system makes use of just four referents for the navigation of the instructions included in manuals:

- *Next.* A maintenance step is clear or completed and hence the user wants to access the following information;
- *Previous.* The user wants to come back to the previous information;
- *Go down (to a lower level)*. More detailed information is required; therefore, the current step is expanded in a more detailed sequence of sub-steps;
- *Go up (to an upper level)*. The user needs fewer details; therefore, s/he navigates through a sequence of less detailed steps. Going up to the first level, the user reaches the root node of the manual.

It is mandatory to lessen as more as possible the number of referents in order to minimize the cognitive load associated with the use of the interface. Anyway, it is worthy to note that, in general, the interface requirements definition could bring also to a number of referents higher than four, however the proposed framework remains valid.

6.5. Gestures Elicitation

The aim of this phase is to elicit the gestures users would intuitively use to trigger the interface commands. The phase includes the following four sub-phases that are detailed in the sections below (Figure 44): Collection of gesture proposals; Gestures labeling; Gestures clustering; Agreement analysis.

6.5.1. Collection of gesture proposals

To conduct the elicitation procedure, a population of 15 participants (average age 27.3 years, SD=5.11) was first recruited, including 14 males and 1 female, all right-handed. All of them hold an MS in Mechanical Engineering even as seven participants also hold a PhD in disciplines related to Mechanical Engineering. Three participants declared that they regularly used an Xbox Kinect for recreational purposes, nine utilized the device a few times, three never used it.

Following Nielsen et al. (Nielsen et al., 2004), the conscious bottom-up approach was adopted that requires an introductory explanation of referents, followed by the collection of users' proposals. To this purpose, participants were asked to watch a slideshow explaining the context

and the test execution modality. The hypothesized scenario was a maintenance workspace where the user has to navigate a digital maintenance instruction manual for the execution of maintenance operations. Each referent, — Next, Previous, Go down, and Go up — was clearly described; the participants were then asked to think aloud about the possible gestures they would propose for them.

After this introductory phase, participants made their proposals according to the following procedure. They watched an automatic sequence of slides, each of them presenting one query for one referent, and simultaneously had to execute any hand gesture that they felt was the best for that referent. Participants were asked to stay 2 meters from a screen, while a webcam on its top recorded their performance. The video was then mixed, in a picture-in-picture mode, with the synchronized capture of the screen viewed by the participant (Figure 44, Block [4]).

The sequence was a fixed series of the four referents, repeated seven times (for a total of 7 × 4 = 28 queries), and mixed according to a Latin Square design. The presentation time t_p was progressively reduced from 3 to 1 second according to the following sequence (number of gestures × duration of each gesture in seconds): 12 × 3 + 12 × 2 + 4 × 1, for a total duration of the elicitation test of 64 seconds. The strategy of reducing the time interval between two consecutive referents was adopted to evaluate how people intuitively and cognitively react to the stimuli. The adopted procedure in the elicitation phase differs from that utilized in previous studies ((Morris et al., 2014; Wobbrock et al., 2009)(Fikkert, Vet, Veer, & Nijholt, 2010; Vatavu, 2012; H. Wu & Wang, 2012), (Piumsomboon et al., 2013)) that required a smaller number of repetitions for each referent and hence a shorter duration of the elicitation phase. The rationale for this choice is to pursue two objectives:

- stressing the memorability aspect of the elicitation procedure;
- adding the 'ingredient' fatigue related to the repetition of the gestures so as to minimize the effect of the performance bias.

6.5.1. Gestures labelling

At the end of the tests, all videos were reviewed to label gesture proposals (Figure 44, Block [5]). Because the presentation time was progressively reduced, sometimes and mostly in cases where the presentation time was just 1 second, some users missed executing the gesture as requested. Thus, only 397 proposals were collected instead of 28 (number of queries) × 15 (number of participants) = 420 as expected.

By reviewing the videos, we identified six attributes characterizing each proposal (Figure 45):

1. **The number of hands**, that is, the number of hands utilized to do the gesture;
2. **The mode of hands** that can be one of the following: dominant, if the user utilizes the

hand of the dominant arm; non-dominant in the opposite case; specular if both the hands are utilized and moved symmetrically with respect to the middle sagittal plane;

3. **The direction of motion** that can be one of the following: from the right to the left side and vice versa (i.e. in the direction perpendicular to the sagittal plane), forward or backward (i.e. in the direction perpendicular to the coronal plane), top-down or bottom-up (i.e. in the vertical direction);
4. **The amplitude of the motion**. To distinguish between wide and narrow gestures an amplitude threshold A_t = 45 cm was fixed which is the mean anthropometric value of shoulder breadth (Preedy, 2012); every gesture with an amplitude A greater than A_t was classified as wide, and every gesture with $A < A_t$ was classified as narrow (Figure 45);
5. **The height of execution** which represents the zone/region where prevalently the gesture is executed. Three different regions were identified: high, that is, above the glenohumeral joint; middle, that is, in proximity of the glenohumeral joint (i.e. in a space range of ± 15 cm with respect to the horizontal plane where the joint lies); low, that is, in proximity of pelvis, that is, in a space range of ± 15 cm with respect to the horizontal plane containing the pelvis (Figure 45);
6. **The hand shape** which represents the shape assumed by the hand during the execution of the gesture. Four different shapes were identified: gestures with palm opened toward or backward to the direction of motion, gestures with thumb or index pointing toward or backward to the direction of motion.

In the case of gestures made with two hands, we applied the following strategy to establish the mode of hands (dominant, non-dominant or specular). We first identified the principal hand that performed the most significant movement (whereas the other hand acted as a reference or support) and then, checked whether it was the dominant one. Gesture proposals were then labelled according to those attributes (Figure 44, Block [5bis]).

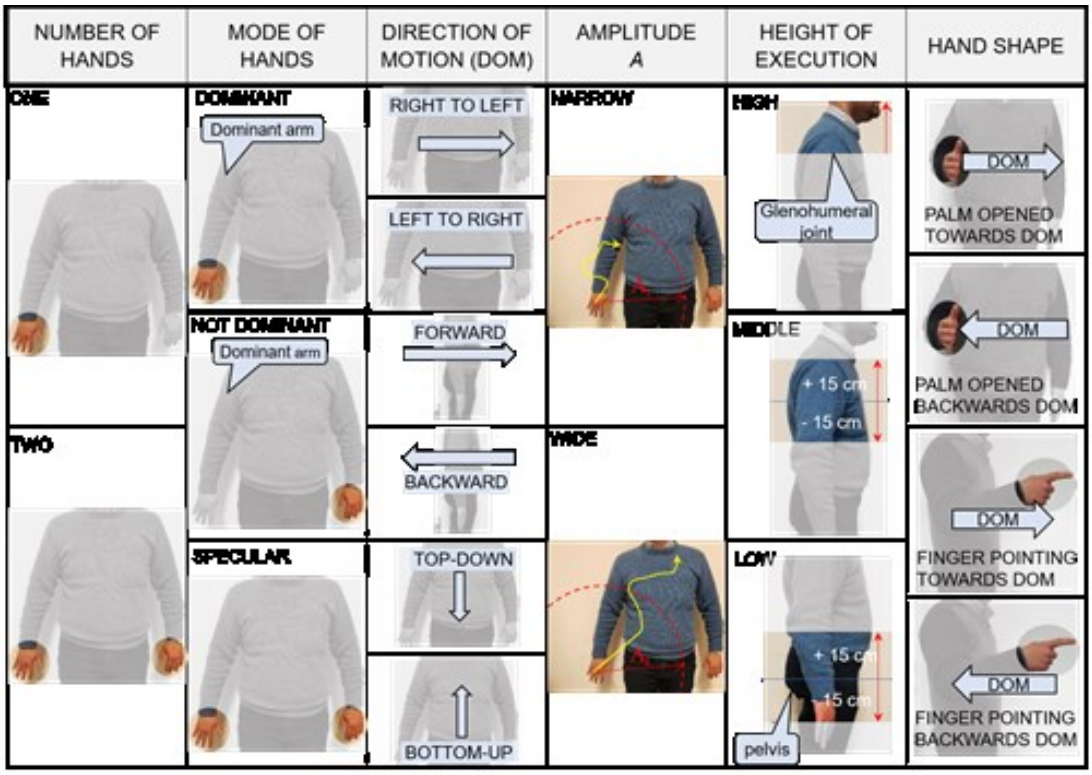


Figure 45 - Attributes used for gesture labelling.

By analyzing data with descriptive statistics, we obtained interesting information about users' preferences (Figure 46).

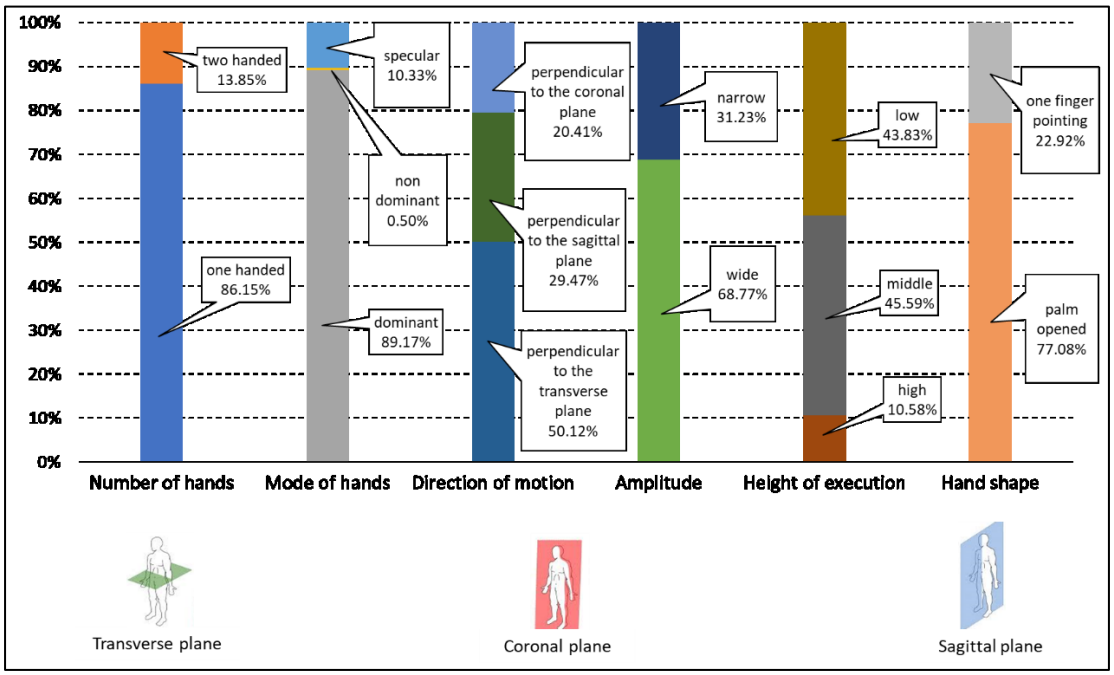


Figure 46 - Descriptive statistics of the gestures collected in the elicitation phase.

Users preferred one-handed gestures (86.15%) compared to two- handed ones (13.85%). One-handed gestures, executed with the dominant hand, were mostly performed (89.17%), followed

by specular gestures (10.33%). Non-dominant one-handed gestures were, instead, negligible (0.5%). The direction of motion perpendicular to the transverse plane was preferred (50.12%), followed by the direction perpendicular to the sagittal plane (29.47%) and by the one perpendicular to the coronal plane (20.41%). Most users preferred wide movements ($A > At$, 68.77%) rather than narrow ones ($A < At$, 31.23%). As regards the height of execution, most users preferred middle (45.59%) and low gestures (43.83%) compared to high ones (10.58%). This can be explained by the argument that people instinctively propose gestures involving low physical effort. Regarding hand shape, users proposed open palm gestures more frequently (77.08%) than gestures with thumb or index pointing (toward or backward to the direction of motion) (22.92%).

6.5.2. Gestures clustering

At the end of the labelling sub-phase, as an intermediate result each gesture was labelled by a tuple of the six attributes. We had 52 different combinations of attributes (Figure 44, Block [5bis]). In other words, all the proposals could be classified into 52 classes of gestures out of the 2 (# of levels for the attribute number of hands) × 3 (# of levels for the attribute mode of hands) × 6 (# of levels for the attribute direction of motion) × 2 (# of levels for the attribute amplitude) × 3 (# of levels for the attribute height of execution) × 4 (# of levels for the attribute hand shape) = 864 theoretical combinations. However, most of these 52 classes resulted to be similar; therefore, according to Piumsomboon et al. (Piumsomboon et al., 2013), after defining similarity rules we attempted to group similar gestures into sets. To do this, the information gain (IG), (the expected reduction in entropy caused by partitioning the classes according to a given attribute) related to each of the above six attributes was first computed (Table XII). In particular, IG was assessed by implementing the supervised Information Gain Ranking Filter inside the WEKA tool (Hall et al., 2009). Generally speaking, the higher the value of IG computed for a given attribute, the higher the capability of 'that' attribute of differentiating/distinguishing. Interestingly, the highest values of IG were found for the attributes 'Direction of motion' and 'Hand shape' (Table XII), which led us to classify the 52 gestures classes just according to these two attributes. Moreover, according to Piumsomboon et al. (2013) gestures having the palm opened toward the direction of motion were grouped with those with the palm opened toward the direction opposite to that of motion.

Table XII - Attribute ranking: the higher the information gain (IG), the better the attribute differentiation capability. in the table are highlighted in grey the attributes for which the highest values of IG were computed.

Attribute	Information gain (IG)
Direction of motion	1.361
Hand shape	0.179
Height of execution	0.068
Mode of hand	0.038
Number of hands	0.003
Amplitude	0.001

Adopting this strategy, the original 52 different classes of gestures were finally grouped into 10 classes of gestures (Figure 44, Block [6bis]) called sets of gestures, namely: right to left opened palm (RLOP), left to right opened palm (LROP), forward opened palm (FOP), forward finger pointing (FFP), backward opened palm (BOP), backward finger pointing (BFP), top-down opened palm (TDOP), bottom-up opened palm (BUOP), top-down finger pointing (TDFP), and bottom-up finger pointing (BUFP). Figure 47 summarizes these 10 sets of gestures and shows them as the result of different combinations of the two attributes 'direction of motion' and 'hand shape' (which are the attributes with the highest values of IG). The empty spaces in Fig. 3 correspond to gestures that none of the participants has proposed.

Figure 47 - Classes of gestures finally obtained after implementing the Information Gain Ranking Filter.

We encountered the conflict set problem when relating these 10 sets of gestures to the corresponding referent (Table XIII). In detail, the referent:

- *Next* was executed with 5 sets of gestures;
- *Previous*, with 6 sets;
- *Go up* with 3 sets;
- *Go down* with 4 sets.

The specific sets of gestures utilized for each referent with the corresponding number of proposals are shown in Table XIII.

Table XIII - Sets of gestures related to each referent with indicated the corresponding number of proposals.

Sets of gestures	*Next*	*Previous*	*Go up*	*Go down*
RLOP	31	31	-	-
LROP	26	27	-	-
FOP	23	3	-	-
FFP	15	-	-	-
BOP	2	8	-	-
BFP	-	31	-	-
TDOP	-	1	11	67
BUOP	-	-	69	11
TDFP	-	-	-	19
BUFP	-	-	20	1

6.5.3. Agreement analysis

The Agreement analysis (Fig. 1, Block [7]) enables to estimate the effectiveness of the set of referents chosen in the interface requirements definition phase. We calculated the Agreement rate AR to evaluate the degree of consensus each referent receives among the proposed gestures. This coefficient, originally introduced by Wobbrock et al. (Wobbrock et al., 2005) and successively updated by Vatavu and Wobbrock (Vatavu & Wobbrock, 2015), is defined as "the number of pairs of participants in agreement with each other divided by the total number of pairs of participants that could be in agreement"; this can be computed as:

$$AR(r_k) = \frac{|P|}{|P|-1}\sum_{P_i \subseteq P}\left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1};\ \text{AR} \in [0, 1]$$

where, P is the set of all proposals for the referent r_k, $|P|$ the size of the set, and $|P_i|$ the size of subsets of similar proposals included in P. The Agreement rate AR ranges in the interval [0, 1]. For a given referent r_k, the value $AR(r_k) = 0$ refers to the case where all the proposals collected for that referent r_k are different from each other, and the value 1 to the case where all the proposals (collected for r_k) are the same. The agreement rate AR for each hypothesized referent are reported in Table XIV.

Table XIV -. Agreement rate for the four referents.

Referents	Agreement Rate *AR*
Next	0.246
Previous	0.256
Go up	0.542
Go down	0.502
average *(AR)*	0.387

The encouraging values of Agreement rate *AR* obtained suggest that the cognitive load required by the hypothesized referents is acceptable and hence the set of referents does not need to be re-designed.

Interestingly, carrying out the agreement analysis on the identical gestures, without applying any similarity rule (e.g. the information gain (IG) above described), we obtain values of *AR* (Table XV), which are significantly smaller than those obtained after applying the similarity rules. This result that agrees with the findings of previous studies (Bailly, Pietrzak, Deber, & Wigdor, 2013; H. Wu & Wang, 2012) demonstrates the efficiency of the proposed similarity rules (Morris et al., 2010).

Table XV - Agreement rate calculated on the referents without applying any similarity rule.

Referents	Agreement Rate *AR*
Next	0.140
Previous	0.093
Go up	0.220
Go down	0.286
average *(AR)*	0.185

The highest values of *AR* were found for the referents "*Go up*" and "*Go down*," followed by the referents "*Previous*" and "*Next*" (Figure 44, Block [7bis]). We explain this behavior with: (i) the request of repeating the gestures for the same referent; (ii) the reduction of the presentation time t_p; (iii) the different mental cues that are triggered by the description of referents during the experiment. Indeed, differently from *next* and *previous*, the textual descriptions with words such as '*go up*' or '*go down*' suggest spatial cues that may influence the users' behaviors. This result is also consistent with the findings of Bailly *et al.* (Bailly et al., 2013) who observed that highly directional referents tend to have a high agreement rate.

6.6. Vocabularies definition

The elicitation phase returned 10 sets of gestures (Figure 47, Figure 44, Block [8]) that could be combined into many gesture vocabularies. The vocabulary definition phase aims to select among these combinations the best one that fulfills our requirement. The phase requires two sub-phases:

- composing the vocabularies;
- ranking the vocabularies.

6.6.1. Composing vocabularies

The combinatorial space of sets of gestures related to all the hypothesized referents (see Table XIII) leads to 5 (# of sets of gestures for the referent *Next*) × 6 (# of sets of gestures for the referent *Previous*) × 3 (# of sets of gestures for the referent *Go up*) × 4 (# of sets of gestures for the referent *Go down*) = 360 possible vocabularies (Figure 44, Block [9]). Applying the "conflict set" constraint ((Microsoft, 2014; Piumsomboon et al., 2013; Wobbrock et al., 2005), if two referents have the same gesture class, then assign the class to the referent with the highest number of proposals), the possible vocabularies reduce to 36 (Figure 48; Figure 44, Block [10]). For example, the class of gestures "backward opened palm" (acronym BOP, Figure 48) has been utilized for both, *Next* and *Previous* referent. The class of gestures is assigned to *Previous* (the sets of gesture removed are highlighted in red, Figure 48), because for this referent, a higher number of proposals was obtained. Then, applying the "homogeneity constraint" ((Microsoft, 2014; Piumsomboon et al., 2013; Wobbrock et al., 2005), a vocabulary is acceptable only if its sets of gestures are homogeneous and hence have the same attribute "hand shape"), we further reduce the number of vocabularies to five (Figure 48). For example, combination 1 (Table C, Figure 48) cannot be accepted as it includes sets of gestures with different attribute "hand shape" and is hence excluded (highlighted in red). For the sake of brevity, in Figure 48, Table D only the accepted combinations and one (i.e., combination 1) of the unacceptable combinations are shown. Finally, implementing the "reversibility constraint" ((Microsoft, 2014; Piumsomboon et al., 2013; Wobbrock et al., 2005), a vocabulary is acceptable only if its sets of gestures are reversible), only three vocabularies remain that are called Vocabulary A, B and C (Table D, Figure 48; Figure 44, Block [10bis]). For instance, combination II (Table D, Figure 48) cannot be accepted because the referents *Next* and *Previous* are associated to sets of gestures that cannot be thought as one the inverse (i.e., carried out in the opposite direction of motion) of the other. In other words, 'forward opened palm' is not the opposite gesture of 'left to right opened palm'. However, combination I can be accepted (Table D, Figure 48) because 'right to left opened palm' (RLOP, referent *Next*) is 'opposite' to 'left to right opened palm' (LROP, referent *Previous*) and 'bottom-up opened palm' (BUOP, referent *Go up*) is 'opposite' to 'top-down opened palm' (TDOP, referent *Go down*).

Table A. Original combinatorial space of gesture vocabularies

Referent	Number of proposals	Gesture class	Combinations
Next	32	RLOP	5
	26	LROP	
	23	FOP	
	15	FFP	
	2	BOP	
Previous	31	RLOP	6
	31	BFP	
	27	LROP	
	8	BOP	
	3	FOP	
	1	TDOP	
Go up	69	BUOP	3
	20	BUFP	
	11	TDOP	
Go down	67	TDOP	4
	19	TDFP	
	11	BUOP	
	1	BUFP	
Total combinations			**5x6x3x4=360**

Conflict set constraint

Table B. Combinatorial space after implementing the conflict set constraint

Referent	Number of proposals	Gesture class	Combinations
Next	32	RLOP	3
	26	LROP	
	23	FOP	
	15	FFP	
	2	BOP	
Previous	31	RLOP	3
	31	BFP	
	27	LROP	
	8	BOP	
	3	FOP	
	1	TDOP	
Go up	69	BUOP	2
	20	BUFP	
	11	TDOP	
Go down	67	TDOP	2
	19	TDFP	
	11	BUOP	
	1	BUFP	
Total combinations			**3x3x2x2=36**

Homogeneity constraint

Table C. Combinatorial space after implementing the conflict set and the homogeneity constraint

Ref.	Comb. 1	Comb. 2	Comb. 3	Comb. 4	Comb. 5	Comb. 6	Comb. …	Comb. 36
Next	RLOP	RLOP	FOP	FFP	RLOP	FOP		
Previous	BFP	LROP	LROP	BFP	BOP	BOP		
Go up	BUOP	BUOP	BUOP	BUB	BUOP	BUOP		
Go down	TDOP	TDOP	TDOP	TDFP	TDOP	TDOP		
Total remaining combinations								**5**

Reversibility constraint

Table D. Combinatorial space after implementing the conflict set, the homogeneity, and the reversibility constraint

Referent	Comb. I	Comb. II	Comb. III	Comb. IV	Comb. V
Next	RLOP	FOP	FFP	RLOP	FOP
Previous	LROP	LROP	BFP	BOP	BOP
Go up	BUOP	BUOP	BUB	BUOP	BUOP
Go down	TDOP	TDOP	TDFP	TDOP	TDOP
Total remaining combinations					**3**
	VOCABUL. A		*VOCABUL. B*		*VOCABUL. C*

Figure 48 - Filtering process for the identification of the optimal gesture vocabulary. After implementing: conflict set, homogeneity and reversibility constraints, the original combinatorial space of gesture vocabularies decreases from 360 (Table A) to 3 (Table D) possible vocabularies.

6.6.2. Ranking vocabularies

This sub-phase ranks the vocabularies finally obtained with the filtering process described above (Figure 44, Block [11]) to determine the best one. As ranking score we calculated the values of the *vocabulary guessability G*, which is a factor that 'measures' the intuitiveness of the vocabulary (Figure 44, Block [11bis]). In detail, the *vocabulary guessability G* appreciates the intuitiveness of the vocabulary as the sum of the intuitiveness of the single class of gestures of which it is comprised (Piumsomboon et al., 2013; Wobbrock et al., 2005, 2009) and is calculated as:

$$G = \frac{\sum_{k=1}^{N} |P_{r_k}|}{P_{TOT}} \; ; \; G \in]0, 1]$$

where, N is the number of referents included in the vocabulary, $|Pr_k|$ is the cardinality (i.e., the number of gesture proposals) of the class of the gestures related to the r_k^{th} referent, and P_{TOT} is the total number of proposals collected in the elicitation phase. G ranges in the interval]0, 1]. G is equal to 1 if all the users propose the same class of gestures for the same referent; however, it is close to zero if, for each referent, few gesture proposals are gathered. Ranking the obtained vocabularies in order of guessability G (G for Vocabulary A, 0.49; G for Vocabulary B, 0.42; G for Vocabulary C, 0.21) allows choosing the best one (Figure 44, Block [12]). In the present case study, the best gesture-vocabulary to navigate instructions in maintenance applications is Vocabulary A (Figure 49).

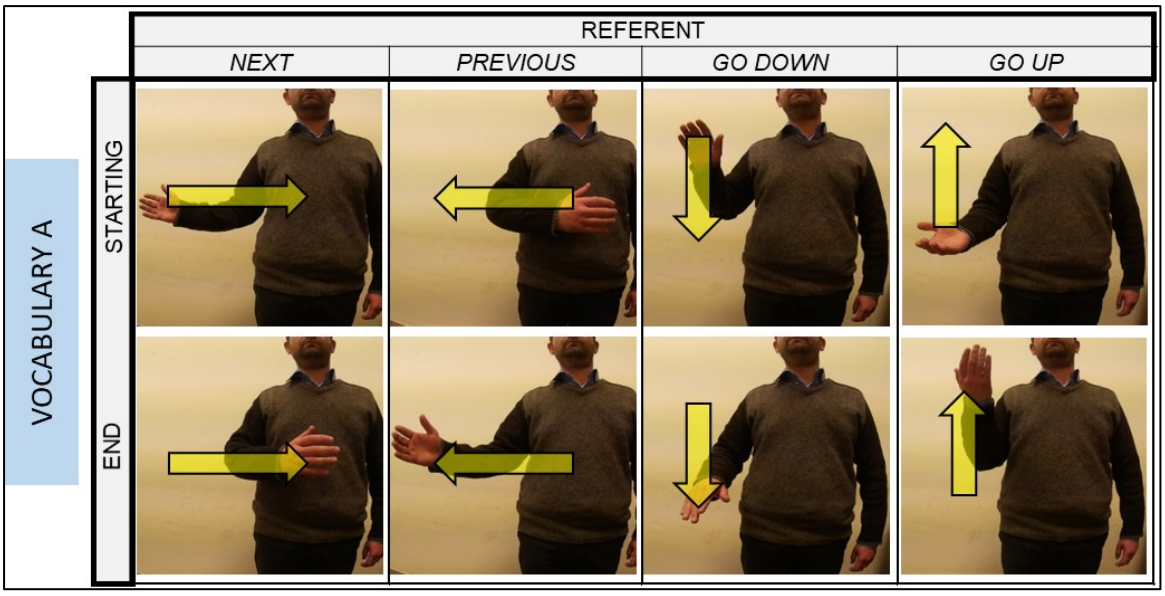


Figure 49 - The gesture vocabulary A selected as the best one.

6.6.3. Validation procedure

This phase aims to validate the ranking of vocabularies described above, by considering, in addition to intuitiveness, other factors such as the ergonomic (i.e., fatigue and ease of use) and the cognitive (i.e., goodness and memorability) ones. In other words, it allows checking whether the previously selected vocabulary is not only the most intuitive but also the most adequate from an ergonomic point of view.

With respect to our use case, the following hypotheses were formulated:

> *Hypothesis 1: Vocabulary A performs better than Vocabulary B with respect to ergonomic and cognitive factors.*
> *Hypothesis 2: Vocabulary A performs better than Vocabulary C with respect to ergonomic and cognitive factors.*
> *Hypothesis 3: Vocabulary A performs better than Vocabulary B and Vocabulary C with respect to ergonomic and cognitive factors.*

This proposed method, that we will describe below, can be extended to any number and type of vocabularies and can be used as a stand-alone procedure to compare vocabularies for the same gesture interface.

Tests on the three vocabularies were performed on a population different from the one employed for gesture elicitation: 12 participants (average age 29.25 years, SD 3.94 years), all males and right-handed. Nine of them hold an MS in Mechanical Engineering, three are undergraduate students completing their studies in Mechanical Engineering. Eight of them had no NUI experience, three had occasional experience, and one had frequent experience.

The experiments conducted to 'measure' the ergonomics and the cognitive load related to the vocabularies A, B and C included two phases: training and testing. In both of them users stood in front of a monitor and a Microsoft Kinect device, at a distance of 200 centimeters. In the training phase, a slideshow first described the gesture vocabulary under investigation. Video recordings of a person correctly executing each class of gestures were embedded in the slideshow to explain each gesture and the corresponding referent. After this initial explanation, participants were asked to execute and replicate all the gestures seen in the video. In detail, participants watched different slides showing (the writing of) different referents and they had to execute each referent, according to the instructions received in the training phase, by means of the correct gestures. By utilizing the CE workbench libraries (Hincapié-Ramos et al., 2014) an application software was developed aimed at evaluating the fatigue and the memorability. The software

showed: (i) on the user's monitor, the sequence of queries for the four referents; (ii) on the experimenter's monitor, the skeletonized scheme of the user - that serves for the computation of the consumed endurance - and the requested referent. Furthermore, the software allowed the experimenter to mark users' errors (Figure 50(a)) and to record them in memory.

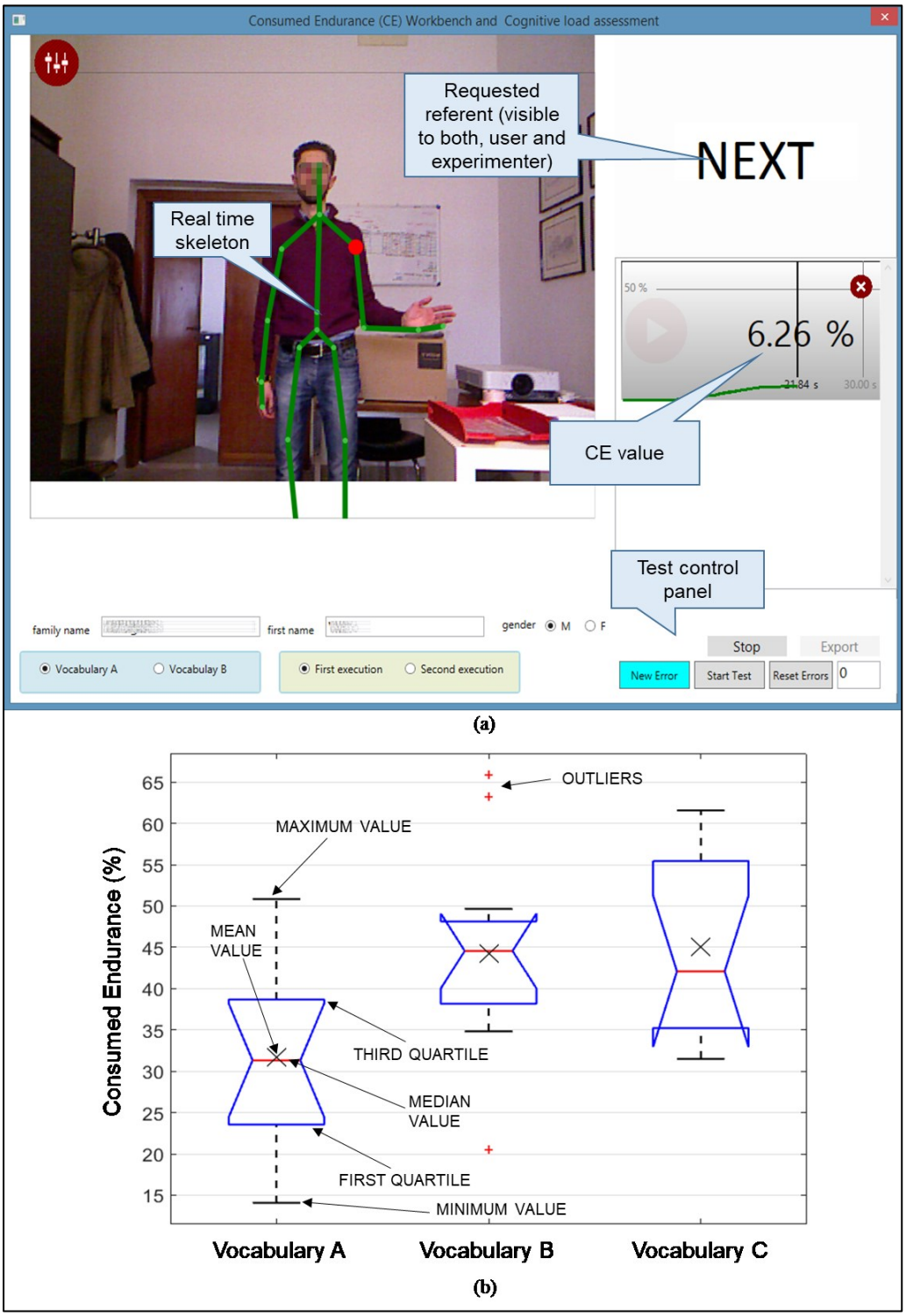


Figure 50 - (a) Screenshot of the window of the software tool developed in the study. This screenshot is visible only to the experimenter and not to the user. Participants can see only the query for the referent. (b) Fatigue assessment based on CE metrics: comparison between the three vocabularies.

If a participant missed making the appropriate gesture for the requested referent, the experimenter advised him, marked an error, and the sequence of queries restarted from the beginning. The test finished when the sequence was accomplished with no errors. The sequence presented 40 queries, i.e. 10 repetitions for each referent, arranged in Latin Square design. Each participant executed the training and the test phase for each vocabulary. To avoid any possible learning effect, the order of execution with respect to the vocabularies was counterbalanced among participants and arranged in Latin Square design while the time interval between each test was one week.

At the end of the task, each user was asked to fill: (i) a seven point (very bad = 1, very good = 7) Likert scale questionnaire to evaluate the "ease of use" of each vocabulary; (ii) a seven point (very bad = 1, very good = 7) Likert scale questionnaire to evaluate the goodness of matching/relationship between the referent and the corresponding gesture. The memorability of vocabularies was computed as the number of errors a user makes until he accomplishes a gestures sequence with no mistakes (Table XVI).

Table XVI - Tests performed in the validation phase to evaluate the ergonomics and the cognitive load related to the vocabularies A, B, and C.

Independent variables		**Description**
Participants		12 males, right-handed, average age 29.25
vocabularies		A, B, C
Dependent variables	**Type of variable**	**Description**
Fatigue	Objective	% Consumed Endurance (CE)
Ease of use	Subjective	bad=1, optimum=7
Goodness	Subjective	bad=1, optimum=7
Memorability	Objective	number of errors

The obtained results were finally submitted to statistical analysis to evaluate if the formulated hypotheses hold true or not.

The values of consumed endurance (CE) measured by means of the hardware/software system above described, were positively tested for normality using the Shapiro Wilk test (Table XVII). The Mauchly's statistic test was nonsignificant ($\chi^2(2)=1.10$, $p=0.580$), thus the assumption of sphericity was met.

Table XVII - CE samples normality tests results

Vocabulary	W statistic	p-value
A	0.9795	0.9815
B	0.9272	0.2946
C	0.9068	0.194

The ANOVA for the within-subject variable showed a significant effect ($F(2,22)=6.355$ $p=0.007<0.05$). Pairwise comparisons showed that Vocabulary A performs better than Vocabulary B (mean value 31.70% vs 44.28%, $p=0.016$) and Vocabulary C (mean value 45.10%, $p=0.023$). Thus, the validity of hypotheses 1 and 2 with respect to the CE was confirmed. Figure 50(b) depicts the boxplot of the statistical comparison.

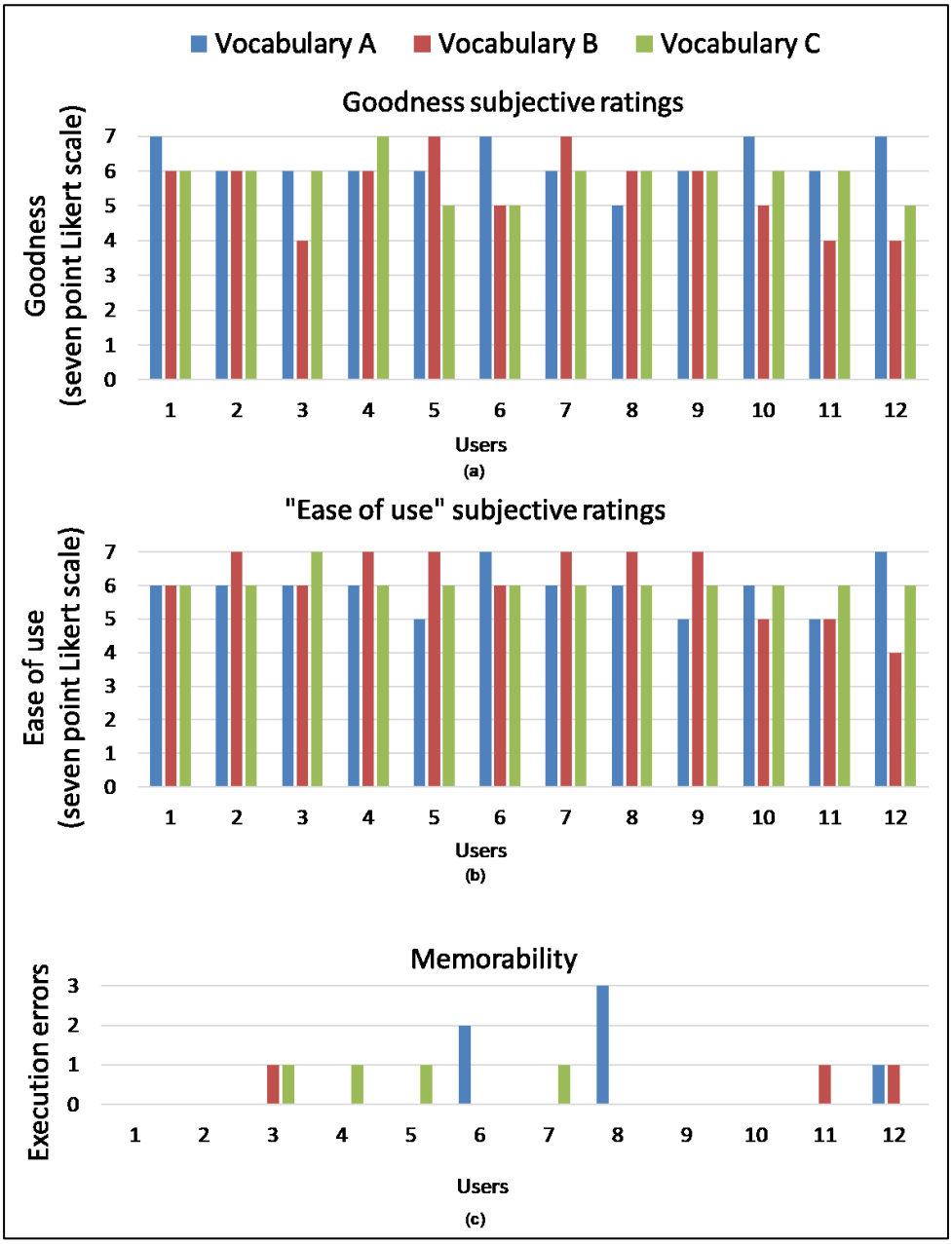


Figure 51 - Assessment of the: (a) goodness, (b) "ease of use", (c) memorability. For each user, the value reported in the Likert scale questionnaire (a, b) and the number of errors done until a gestures sequence was accomplished without mistakes (c) are diagrammed.

Regarding the “ease of use” evaluation (Figure 51(a)), the Friedman test did not confirm any difference between the vocabularies ($\chi2(2)=1.389$, $p=0.499$); therefore, none of the hypotheses was confirmed.

With respect to the “goodness” evaluation (Figure 51(b)), the Kruskal-Wallis test did not confirm any hypotheses ($\chi2(2)=4.185$, $p=0.123>0.05$).

Regarding the memorability evaluation (Figure 51(c)), the distributions of samples for all vocabularies did not show any statistically significant difference (Kruskal-Wallis, $\chi2(2)=0.179$, $p=0.914>0.05$) and did not confirm any hypotheses.

6.7. Discussion and Conclusions

The analysis of the interface requirements allowed us to identify a set of navigation commands. The elicitation phase was successfully utilized to define a consistent set of gesture proposals for each referent for navigation of technical instructions. Proper techniques of clustering and filtering were adopted to select, among all the possible solutions, three vocabularies that fulfill the design guidelines and among these three vocabularies, the best one. Regarding the ergonomics, the consumed endurance (CE) analysis was conducted to assess the fatigue related to a given vocabulary. However, CE metrics presents some limitations. The most important is represented by the noise in joint detection that requires tuning the SkeletonBufferSize parameter to guarantee the correct joint speed assessment. Furthermore, CE could not take into account other relevant fatigue elements, such as the exertion of hand muscles for the finger movements, the trunk flexion, the neck flexion or extension, and so on. However, in spite of these limitations, the results of the CE metrics are consistent with our expectations, thus confirming the adequacy of the proposed framework. For instance, the CE analysis confirmed better performances for Vocabulary A. This result can be justified with the argument that this vocabulary involves a narrow rotation of the glenohumeral joint that leads to reduced movements of the center of mass of the kinematic chain of the upper limbs which imply reduced values of the consumed endurance CE. On the contrary, Vocabulary B and C require, for some referents, lifting not only the forearm (as for vocabulary A) but also the arm. For instance, vocabulary B, referent “go up” is related to the class of gestures ‘bottom-up finger pointing’ (BUFP) that clearly requires lifting both, the forearm and the arm. The same referent in vocabulary A requires to lift up the sole forearm (Figure 49). The study includes other limitations. The adopted procedure does not take into account to what extent a given set of gestures can be easily recognized by the hardware and software systems currently utilized in the human-machine interfaces. A gesture set might be

perfect in terms of fatigue and memorability, but if it cannot be recognized accurately then it will just be frustrating to users. Nevertheless, the proposed procedure rules favor gesture recognition. Indeed, lessening the number of referents leads to reduce the chance of misrecognition (Wobbrock et al., 2009). The reversibility constraint increases the orthogonality of gesture-detection features, and hence improves detection performance. Finally, similarity rules introduce some degrees of variability in the gesture, thus facilitating the gesture recognition (Pereira et al., 2015).

Despite these limitations, the results are consistent with those reported in the literature. Users tend to reduce their mental effort as much as possible, therefore, they perform reversible gestures to execute referents that have an opposite effect, which is consistent with results obtained by Piumsomboon et al. (2013), Ooi, Wong, Tan, & Lee (2014) and Silpasuwanchai & Ren (2015). In addition, according to the findings of Piumsomboon et al. (2013), the best vocabulary found in the present study (i.e., Vocabulary A) includes the gestures "swipe left to right" for the referent "previous" and "swipe right to left" for the referent "next".

The proposed framework finally allowed to select three vocabularies A, B and C. The analysis of "ease of use" did not find any statistically significant differences among the vocabularies. No statistically significant differences between the investigated vocabularies were also found with regards to "goodness" and memorability. We believe that the inability of subjective metrics, to significantly differentiate the vocabularies is related to the fact that the investigated vocabularies are the best ones among many other possible vocabularies and are the result of a strict selection/filtering process. In other words, vocabularies A, B and C, are so good that it is difficult for the user to prefer one of the vocabularies to the others; or, equivalently, the differences between vocabularies A, B and C are so small that the adopted subjective metrics cannot distinguish them thus identifying the best one.

A very important aspect of the proposed framework is represented by its modularity. Each module of the framework, is a stand-alone unit and can be easily updated and changed according to the specific tasks to accomplish or the specific types of vocabulary to optimize. For instance, the modules of the proposed framework can be properly customized for the design of other vocabulary types that can be utilized in the industrial/maintenance context. The test case investigated includes four referents, in the case a larger number of referents is considered the number of possible vocabularies increases with a combinatorial trend. However, the constraint applied in the vocabularies composition phase certainly are adequate to manage also a large number of referents. For instance, in our case study, the conflict set constraint reduced the

number of combinations from 360 to 36, the homogeneity constraint from 36 to 5 and, finally, the reversibility constraint from 5 to 3.

In conclusion, this work proposes a general framework to design mid-air gesture-based interfaces for the navigation of technical maintenance instructions. The systematic and wide breath approach adopted in the study makes the proposed design methodology easily extendable to other industrial contexts where mid-air gestures might turn useful. In other words, the proposed framework provides useful guidelines that are valid not only in the specific investigated field, but that can be followed to identify the best gestures vocabulary even in other specific contexts and hence to properly design the required interaction interfaces.

Chapter 7. A Postural Risk Assessment Tool supporting the Healthy Operator

The application of the Human Centered approach as one of the development paradigms of the smart factory of the future provides not only considering the worker as a crucial element in the design process, but also considering her/his role in the accomplishment of everyday working activities. Consequently, it is mandatory monitoring working activities, not only to improve performance but also to keep the workers' wellbeing.

The main topic investigated during our doctoral program consisted of the development of an application for real-time monitoring of the exposure to ergonomics risk factor in the workplace. Such application aims at supporting the role of the Healthy Operator by allowing the estimation of biomechanical risk in real time and providing a direct feed back to the user that will be constantly monitored directly at work without interfering with her/his tasks.

7.1. Introduction

Despite the steady improvement in working conditions, according to the Sixth European Working Conditions Survey (Eurofound, 2015), exposure to repetitive arm movements and tiring positions still remains a common issue. Taking into account worker's health and also welfare costs, it is mandatory to apply policies aimed at minimizing risks belonging to the work-related musculoskeletal disorders (WMSDs). WMSDs include "all musculoskeletal disorders that are induced or aggravated by work and the circumstances of its performance" (WHO & others, 2003).

The factory shop-floor of the I4.0 environment will be characterized by the presence of smart sensors with computational capabilities and network connection. The capability of such sensors of being sensitive, responsive, adaptive and transparent to workers' movements allows online, real-time monitoring of work activities (Alberto, Draicchio, Varrecchia, Silvetti, & Iavicoli, 2018). On the ergonomic level one possibility consists in using available sensors to detect and report ergonomic malpractice of the individual worker and show possible improvements. A field study has shown that many workers are not aware of the ergonomic aspects of their work. Indeed, they have no references about how ergonomic postures and limb movements look like and which ones are ergonomically not recommended. Furthermore, the threshold when a certain movement

leaves the recommended area is not known. Instead, most workers follow motion sequences that they are familiar with because they seem to be comfortable or effective (Römer & Bruder, 2015). Awkward postures are not immediately recognized by workers unless they do not involve pain or discomfort, but these postures might have a negative long-term effect if repeated regularly and could cause Work related Musculoskeletal Disorders (WMSDs).

In this chapter[6] we will describe the study, the design and the development of the K2RULA, a semi-automatic biomechanical-risk evaluation software based on the Microsoft Kinect v2 depth camera, aimed at detecting awkward postures in real time, but also in an off-line analysis. We validated our tool with two experiments. In the first one, we compared the K2RULA grand-scores with those obtained with a reference optical motion capture system and we found a statistical perfect match according to the Landis and Koch scale (proportion agreement index=0.97, k=0.87). In the second experiment, we evaluated the agreement of the grand-scores returned by the proposed application with those obtained by a RULA expert rater, finding again a statistical perfect match (proportion agreement index=0.96, k=0.84), whereas a commercial software based on Kinect v1 sensor showed a lower agreement (proportion agreement index=0.82, k=0.34).

7.2. Related works

Typically, the best applicable practice to prevent WMSDs consists in the evaluation of the exposure to risk factors in the workplace and in planning an eventual ergonomic intervention as the workplace redesign. The availability of new technologies allows nowadays new interventions based on real time-monitoring able to provide real-time workbench adaptation and warning feedbacks

Many methods have been developed to evaluate the exposure to risk factors. These methods can be classified into three groups: i) self-report; ii) direct measurement, and iii) observational methods (G. Li & Buckle, 1999).

Self-reports comprehend different approaches: rating scales (G Borg, 1962; Gunnar Borg, 1998; Shackel, Chidsey, & Shipley, 1969), questionnaires or interviews (Kuorinka et al., 1987) and checklists (W. M. Keyserling, Brouwer, & Silverstein, 1992; W. Keyserling, Stetson, Silverstein, & Brouwer, 1993; Shackel et al., 1969). These approaches have been used with success for the

[6] The results of the research described in this chapter have been published in the following article:
V. M. Manghisi, A. E. Uva, M. Fiorentino, V. Bevilacqua, G. F. Trotta, and G. Monno, "Real time RULA assessment using Kinect v2 sensor," Applied Ergonomics, vol. 65, pp. 481–491, 2017.

assessment of physical workload, body discomfort, or work stress, which are difficult to measure objectively. However, subjective ratings present some drawbacks. They suffer from influences other than the task or workplace investigated (Rantanen, 1981) and are affected by intrinsic limits of subjective evaluations (Balogh et al., 2004; David, 2005). Indeed, the subjective assessments cannot give an account of small yet significant differences and, are prone to confounding variables such as the fitness of the participants, the comfort level or the state of mind. Therefore, the use of these methods could result in distorted interpretations (G. Li & Buckle, 1999).

Direct methods use data from sensors attached to the worker's body, but they are typically more expensive, intrusive, and time-consuming (Kowalski, Rhodes, Naylor, Tuokko, & MacDonald, 2012; Xu, McGorry, Chou, Lin, & Chang, 2015).

Observational methods, which are widely applied in industry, consist of direct observation of the worker during his work shift. The most common in industry are the following: the Ovako Working Posture Analyzing System (OWAS) (Karhu, Härkönen, Sorvali, & Vepsäläinen, 1981), the OCRA index (Occhipinti, 1998), the revised NIOSH equation for manual lifting (Waters, Putz-Anderson, Garg, & Fine, 1993), the Rapid Upper Limb Assessment (RULA) (McAtamney & Nigel Corlett, 1993), the Rapid Entire Body Assessment (REBA) (Hignett & McAtamney, 2000), the Loading on the Upper Body Assessment (LUBA) (Kee & Karwowski, 2001), and the European Assembly Worksheet (EAWS) (Schaub, Caragnano, Britzke, & Bruder, 2013). A detailed review of the most common observational methods can be found in (Roman-Liu, 2014) where OWAS, revised NIOSH, RULA, OCRA, REBA, LUBA, and EAWS are compared. In industrial practice, data are collected through subjective observation or estimation of body-joint angles in pictures/videos. The main advantage of these methods is that they are relatively low cost and that they do not interfere with the working process. The main disadvantage is that most techniques are very time-consuming as they require manual analysis of posture, of the repetition rate, and the forces involved during a shift. Furthermore, since the evaluation is subjective, it may be subject to intra- and inter-observer variability and data collection inaccuracy and low sampling rates remain their biggest drawbacks. The use of video-based systems and computer vision techniques partly overcame these limitations, improving accuracy and robustness of joint angle estimation (Fıglalı et al., 2015; Pinzke & Kopp, 2001).

These methods have the main disadvantage to require a field expert who performs a time-consuming analysis of the postures. The introduction of low-cost and calibration-free depth cameras, such as the Microsoft Kinect v1 sensor, provided easy-to-use devices to collect data at high frequencies, and suggested a semi-automatic approach to observational methods. Several

authors studied the accuracy of kinematic data provided by the Kinect v1 device in various application domains (Bonnechere et al., 2014; Clark et al., 2012; Clark, Bower, Mentiplay, Paterson, & Pua, 2013; Dutta, 2012; Xu et al., 2015). The results showed that Kinect v1 is accurate enough to capture human skeletons in a workplace environment. The accuracy and robustness of the provided joint positions (skeleton tracking) are promising for applications that require to fill in an ergonomic assessment grid (Diego-Mas & Alcaide-Marzal, 2014; Plantard, Auvinet, Pierres, & Multon, 2015). Patrizi, Pennestrι, & Valentini (2015) compared a marker-based optical motion capture system with a Kinect v1 for the assessment of the human posture during working tasks and the recommended weight limit in the NIOSH lifting equation. Two other works exploited Kinect v1 to compute an ergonomic score based on the EAWS method (Kruger & Nguyen, 2015; Nguyen, Kleinsorge, & Kruger, 2014).

Observational methods like OWAS, NIOSH, OCRA, and EAWS, even if supported by depth cameras user data, still require a heavy intervention by a field expert to estimate the required parameters (e.g. forces, loads, static/repetitive muscular activity etc.). The ISO standard 11228-3:2007(E) (ISO, 2007) suggests the use of a simplified method in the early stage of the analysis and, if critical conditions are detected, provides the OCRA method to be applied for additional investigation. Among the simplified methods for rapid analysis of mainly static tasks, the RULA, acronym of Rapid Upper Limb Assessment, is one of the most popular (McAtamney & Nigel Corlett, 1993). The main weakness of RULA is related to the inter-rater reliability. Robertson et al. (2009) found just "fair" inter-rater reliability of the RULA grand-score (ICC<0.5) among four trained raters. Dockrell et al. (2012) proposed an investigation of the reliability of RULA that demonstrated higher intra-rater reliability than inter-rater reliability implying that serial assessments would be more consistent if carried out by the same person. Bao, Howard, Spielholz, Silverstein, & Polissar (2009) showed that, if a "fixed-width" categorization strategy is used when classifying the angles between body segments, the inter-rater reliability grows with the amplitude of the width. Moreover, larger body parts as shoulder and elbow, allow better estimation than smaller ones, as wrist and forearm (Lowe, 2004a, 2004b).

Therefore, RULA can be effectively aided by computer processing and skeleton tracking systems. In (Haggag, Hossny, Nahavandi, & Creighton, 2013) the authors describe a framework combining the Kinect v1 with the RULA method for 3D motion analysis. The Kinect v1 skeleton tracking has also been integrated into the DHM Jack tool (Siemens, 2013), and the commercial software, Task Analysis Toolkit module (Jack-TAT), estimates, in real time, the ergonomic risk of the executed tasks. The advantages of this application of depth sensors are: the real time calculation, the portability of the device, and the reduced cost (Horejsi, Gorner, Kurkin, Polasek,

& Januska, 2013). The Kinect v1 sensor can be useful in developing ergonomic risk assessment tools, lessening the time consumption of visual-inspection assessing procedures, and removing the problem of the bias introduced by the analyst.

However, three main technical problems arose in the works using Kinect v1: the lack of wrist joints tracking, the influence of the environment lighting conditions, and the self-occlusions (in postures such as crossing arms, trunk bending, trunk lateral flexion, and trunk rotation).

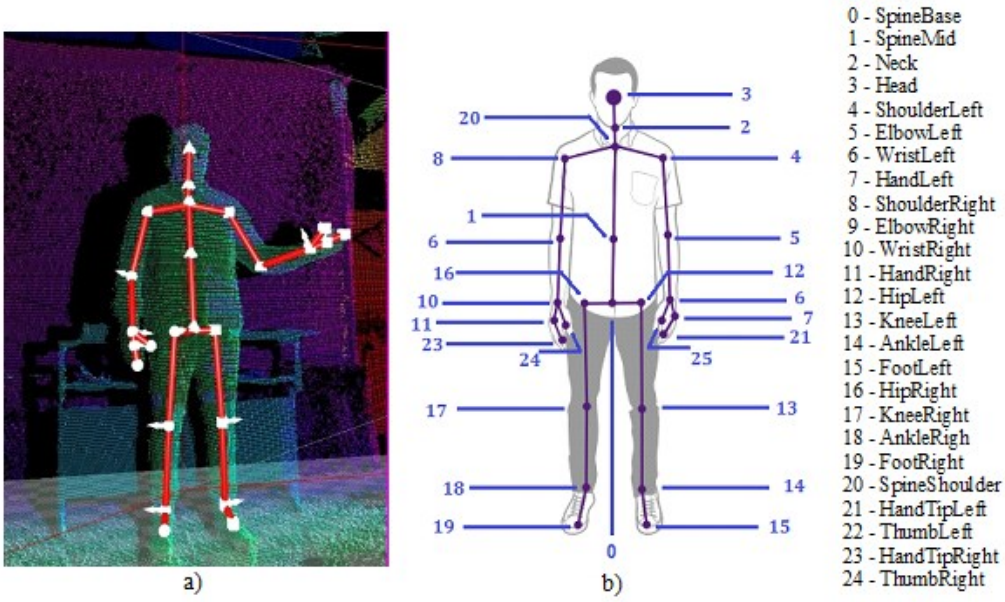


Figure 52: The skeleton returned by Kinect for Windows SDK 2.0. a) Depth map and skeleton visualized by the Microsoft Kinect Studio v2.0; b) Joints position with respect to the body as reported by Microsoft HIG (Microsoft, 2014).

The Kinect v2, presented in 2013, uses a different technology (time of-flight), and according to the specifications, it outperforms the previous version. It tracks 25 body joints including wrists (see Figure 52); it is more robust to artificial illumination and sunlight (Zennaro et al., 2015) and more robust and accurate in the tracking of the human body (Q. Wang et al., 2015). Conversely, a study (Xu & McGorry, 2015) found the non-trivial result that Kinect v1 outperforms v2 as regards average error of joint position (76 mm vs 87 mm) in seated and standing postures. Wiedemann, Planinc, Nemec, & Kampel (2015) measured the accuracy of ergonomic-relevant angles computed by Kinect v2, using a marker based motion-capture system as reference. They measured high deviations of the neck angle (-31.0°±9.1°) and of the upper body rotation along the longitudinal axis (24.0°±3.5°), while the remaining upper body inclinations and joint angles showed higher accuracies (deviation less than 7.2° in median). Furthermore, the error in the standing postures appeared to be lower than in the sitting ones. Plantard, Shum, Le Pierres, & Multon (2016) presented an interesting study on the validation of RULA grand-scores obtained

using Kinect v2 data, in both laboratory and real workplace conditions. In laboratory conditions they measured angular errors between an average value of 7.7° for the simplest case (no occlusions) and 9.2° for the worst case. They also reported RULA grand-scores correctly computed for more than 70% of the conditions.

These results feature the Kinect v2 sensor to be a promising tool for postural analyses, especially for the metrics whose calculation is based on angular thresholds that tend to minimize the effect of joint angle errors, as RULA. However, some of the results reported in the literature are controversial, since they are sensitive to the specific setup and to the postures adopted for the validation. We think that there is still a need for further tests to strengthen the knowledge. Therefore, our research question was: is it possible to effectively use the Kinect v2 data for an early screening of exposure to WMSDs risk? The typical application scenario can be derived by the ISO standard 11228-3:2007(E), e.g. the workspace is continuously monitored by a depth camera connected to an automatic RULA evaluation system and, if critical conditions are automatically detected, additional investigations (e.g. OCRA) can be carried out.

In this paper, we present the implementation of a software tool called K2RULA, a fast, semi-automatic, and low-cost tool, based on the Kinect v2. We validated the proposed tool with two experiments. In the first one, we compared the grand-scores from K2RULA with the ones obtained with data collected by a reference optical motion capture system. In the second experiment, we compared the grand-scores obtained from K2RULA, Jack-TAT and a RULA expert.

7.3. Materials and Methods

7.3.1. K2RULA software

We implemented K2RULA using C#, Windows Presentation Foundation libraries (.NET framework) and Microsoft Kinect for Windows SDK 2.0. The GUI of the K2RULA tool allows to select the video stream to be visualized (depth or infrared), and to activate a secondary window for the RBG stream (Figure 53). The button “Real Time RULA” evaluates the RULA grand-score of the current posture. Furthermore, playback control buttons allow the execution of an offline analysis of a recorded file.

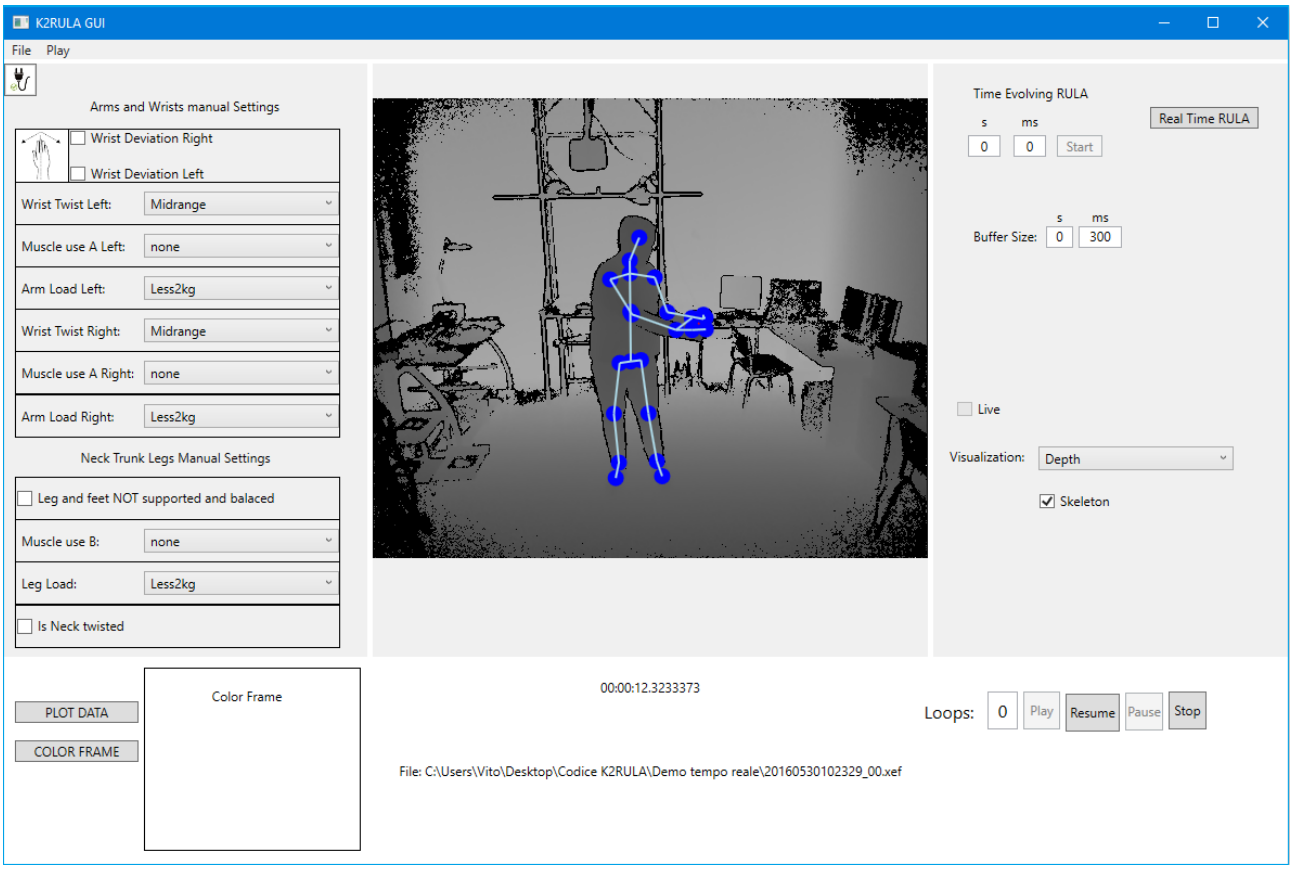


Figure 53: GUI of the K2RULA application

7.3.2. The RULA method

The RULA method consists in the fulfillment of an assessment grid, where the human body is divided into two sections

- Section A: upper arm, lower arm, and wrist;
- Section B: neck, trunk, and legs.

An expert analyst visually inspects a posture adopted by the target person during the work cycle, and then calculates the angles between the adjacent body parts for each section (i.e. shoulder and upper arm, upper arm and lower arm, trunk flexion or bending, wrist twist, neck flexion or extension). He executes her/his observations twice, once for the left side of the body and once for the right side. A score is calculated using three tables. The first two tables give the posture scores of the body segments. Each one of these scores is then corrected according to the frequency of the operations and the force load on the limbs. The third table takes as input the previous scores and returns a grand-score. An action level list indicates the intervention required to reduce the risks of injury of the operator:

- 1-2 grand-score: the posture is acceptable if it is not maintained or repeated for long periods,
- 3-4 grand-score: further investigation is needed, and changes may be required,
- 5-6 grand-score: investigation and changes are required soon,
- 7 grand-score: investigation and changes are required immediately.

7.3.3. Data retrieval

The Kinect tracking algorithm returns a hierarchical skeleton composed of joint objects (Figure 52). We calculated each joint position in real time as the average of the positions stored in a 300ms memory buffer (about 10 valid frames at 30Hz) to minimize jittering. If the sensor is not able to track a joint (e.g. occlusion), its position is inferred (inferred joints) from the surrounding joints by the Microsoft SDK.

The K2RULA algorithm requires only 19 of the 25 tracked joints. RULA parameters are trivially evaluated from geometrical angles between the joints. However, for some angles, we need additional processing.

We defined the trunk vector as the vector connecting the spinebase (from Windows SDK nomenclature) to the spineshoulder, respectively approximately corresponding to the mid posterior superior iliac spine (G. Wu et al., 2002) and the incisura jugularis (G. Wu et al., 2005).

For the *upper arms flexion/extension* we computed the angle between the trunk vector and the vector corresponding to the projection of the upper arms on the sagittal plane. The latter is evaluated as the one passing through the trunk vector and perpendicular to the straight line connecting the shoulders.

The *upper arms abduction* is evaluated with the angle between the trunk vector and the vector corresponding to the projection of the upper arms on the plane passing through the trunk and parallel to the straight line connecting the shoulders.

For the *shoulder abduction* we computed the angle between the vector connecting the spineshoulder to the neck and the vector connecting the spineshoulder to the shoulder under analysis.

To evaluate the *working position of the lower arm* with respect to the midline of the body and the side of the body, we analyzed the relative positions of the projections of the wrist, spineshoulder and shoulder on the straight line connecting the shoulders (Figure 54).

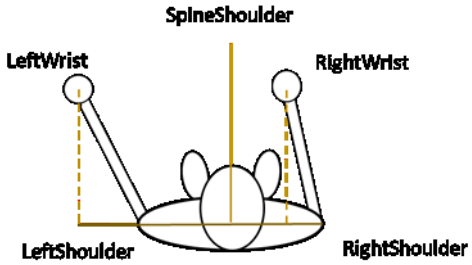


Figure 54: Lower arms working position assessment geometrical construction.

As regards the *wrist location*, we could only approximatively assess the adduction/abduction angle. We computed the angle between the vector connecting the elbow to the wrist and the vector connecting the wrist to the handtip.

The grid assessment requires taking into account the *trunk twisting and bending* state. We verified that the sensor always returns a skeleton object with the same directions for the normal to the three joints in the trunk, regardless of the twisting state of the body (Figure 52). Hence, we calculated the angles between the normal to the ankles (directed towards the outside of the body) and the normal to the trunk, directed towards the sensor (Figure 55). To detect the trunk bending state we computed the angle between the straight line passing through the hip joints and the direction normal to the horizontal plane. The trunk flexion degree is trivially assessed by the angle between the direction perpendicular to the horizontal plane and the trunk vector.

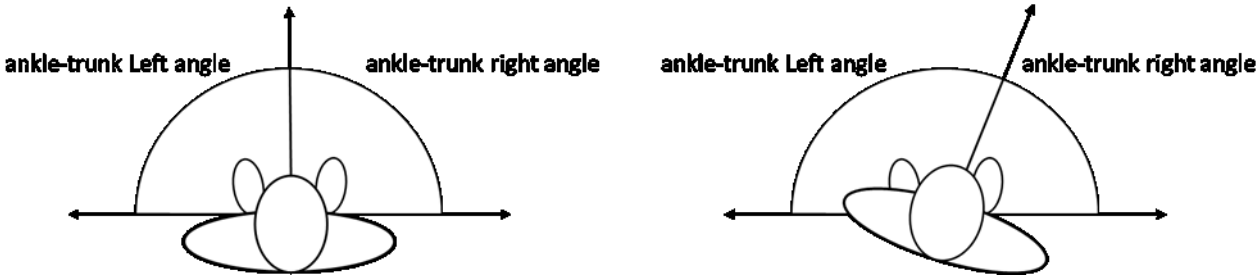


Figure 55: Trunk twisted detection scheme.

We assessed the *neck flexion/extension* computing the angle between the normal to the trunk vector in the sagittal plane and the projection in this plane of the vector connecting the spineshoulder to the head. This solution leads to an overestimation of the neck back flexion with respect to visual inspection. Therefore, we added a positive bias of five degrees in the computation of the angle on a heuristic base. We detected the *neck bending* computing the angles between the vector connecting the spineshoulder to the head and each one of the vectors connecting the spineshoulder to the shoulders.

Despite the improvements in joint detection provided by Kinect v2, the accuracy is not sufficient to detect some important parameters for some joints, such as the *wrist and neck twist*. In addition, K2RULA is not able to evaluate other factors, such as the *load on arms* and the kind of *muscle use*, that affect the RULA grand-score. As a solution, we implemented default settings, and provided a simple GUI for the operator to set them (Figure 56).

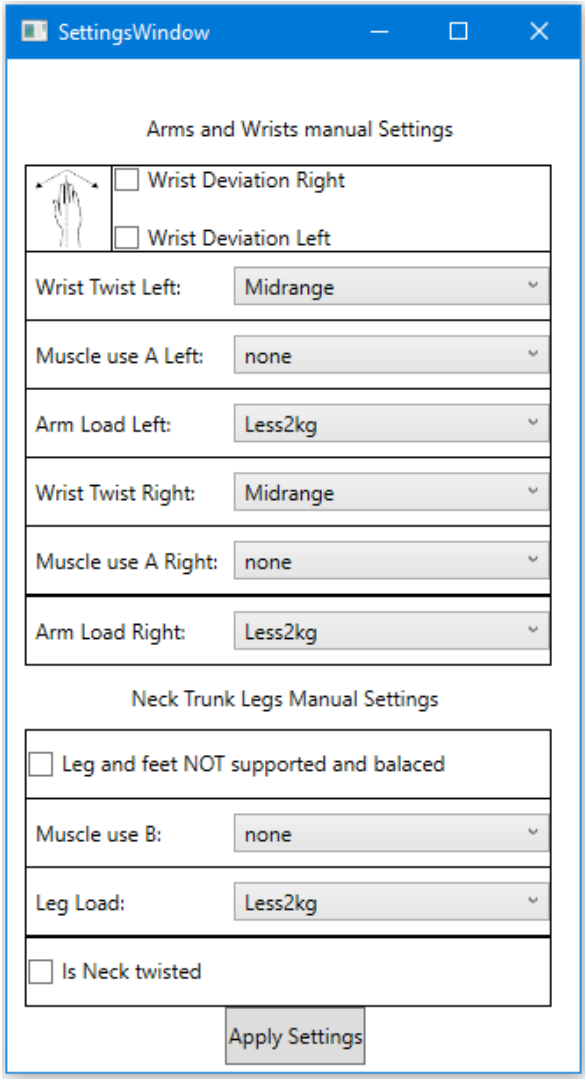


Figure 56: Window interface for manual settings and default values.

7.3.4. Functionalities

The “Real Time RULA” button activates the display of the RULA scores panel (Figure 57).

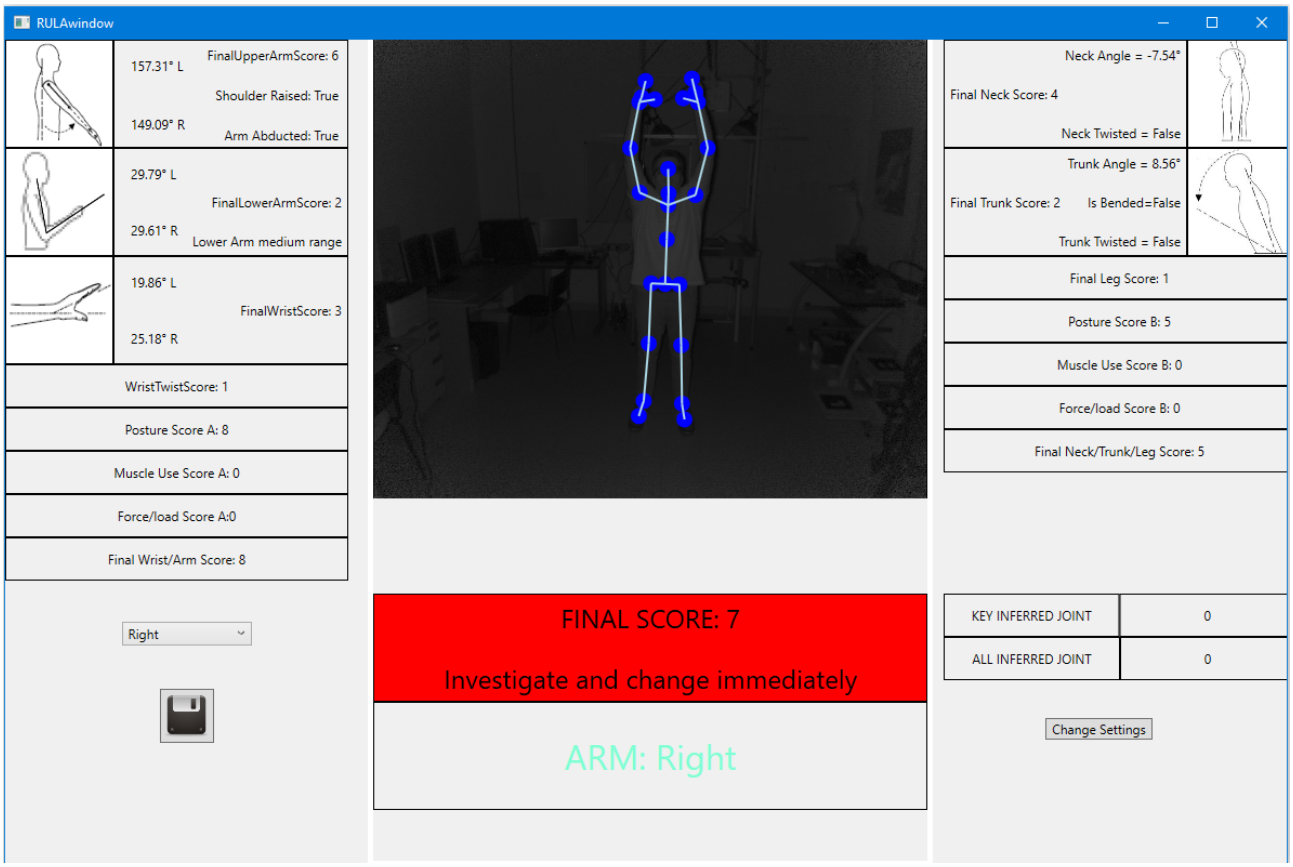


Figure 57: The RULA scores panel.

This window provides the scores of each body section for both sides, the computed angles, and the grand-score, and saves the report on a text file. The action level is visualized with a color-coded background varying from green (grand score 1-2) to red (grand-score 7). Furthermore, the inferred joints are evidenced with red circles on the skeleton to highlight the reliability of the assessed scores.

Another functionality of K2RULA is to process continuously a recorded file in the standard Microsoft format (.xef). The software calculates the grand-score for each of the frames and generates a report, exportable in a comma separated values file, while visualizing an interactive timeline plot. By clicking on one point of the graph, a pop-up label displays the RULA grand-score for that instant (Figure 58).

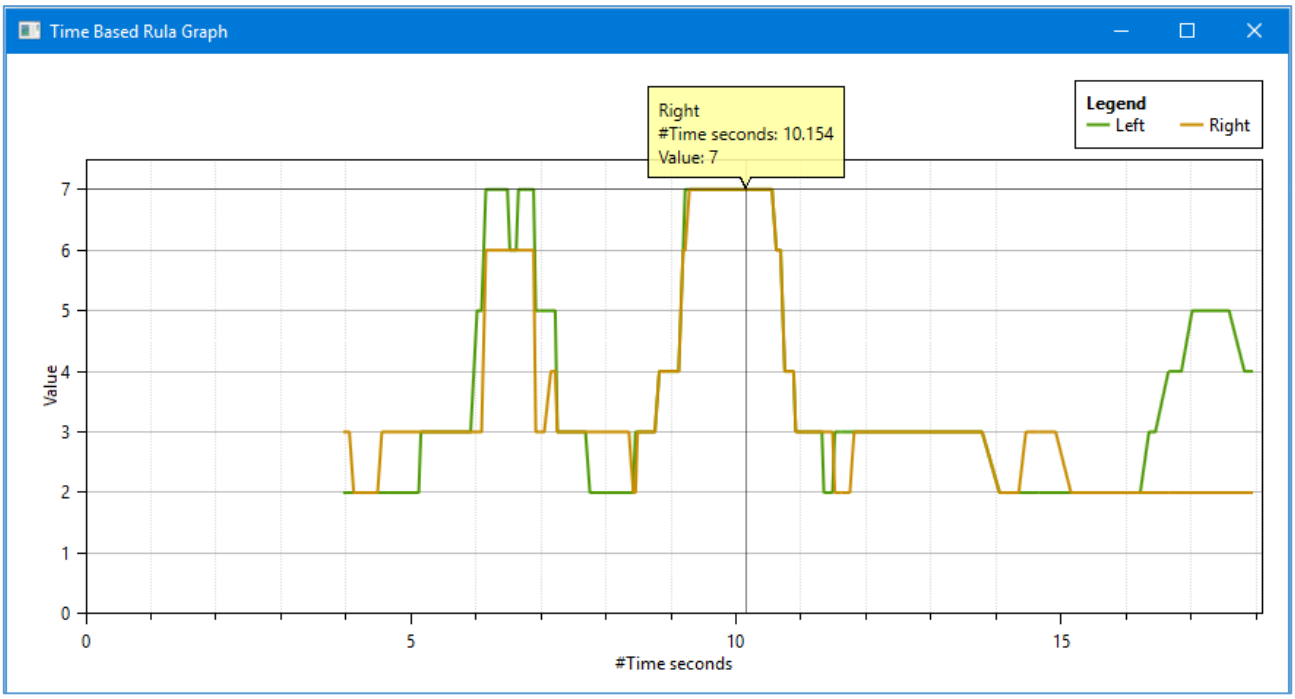


Figure 58: Grand-scores plot for an offline analysis on a recorded file: postures at seconds 6-7 and 9-11 are critical and require further analysis.

This functionality allows to continuously evaluate the working activities and to spot for critical conditions.

7.4. Experiment 1: validation with an optical motion capture system

In this experiment we studied the agreement between the K2RULA tool and a reference tracking system. We define our *hypothesis 1*: K2RULA RULA grand-scores are in accordance with an optical motion capture system.

7.4.1. Equipment

To run K2RULA, we used a Kinect v2 connected to a PC with a CPU Intel® Core™ i5-4200 2.50 GHz, 4 GB RAM, GPU NVIDIA GeForce GT 740M, OS Windows 8. The reference tracking system was a BTS SMART-DX 5000 optical motion capture systems (BTS-Bioengineering, 2016) composed by 8 infrared digital cameras, with an acquisition frequency of 100 Hz, and one PC with a CPU Intel® XEON E5640 2.67 GHz, e 3 GB RAM, OS Windows XP. We used the SMART Suite software for raw data acquisition and processing (BTS-Bioengineering, 2016).

7.4.2. **Procedure**

We selected 15 static postures: nine of them (Figure 59) from the EAWS form (IAD, 2012), and six (Figure 60) extracted from a booklet of the European campaign against musculoskeletal disorders (Colombini, Colombini, & Occhipinti, 2012).

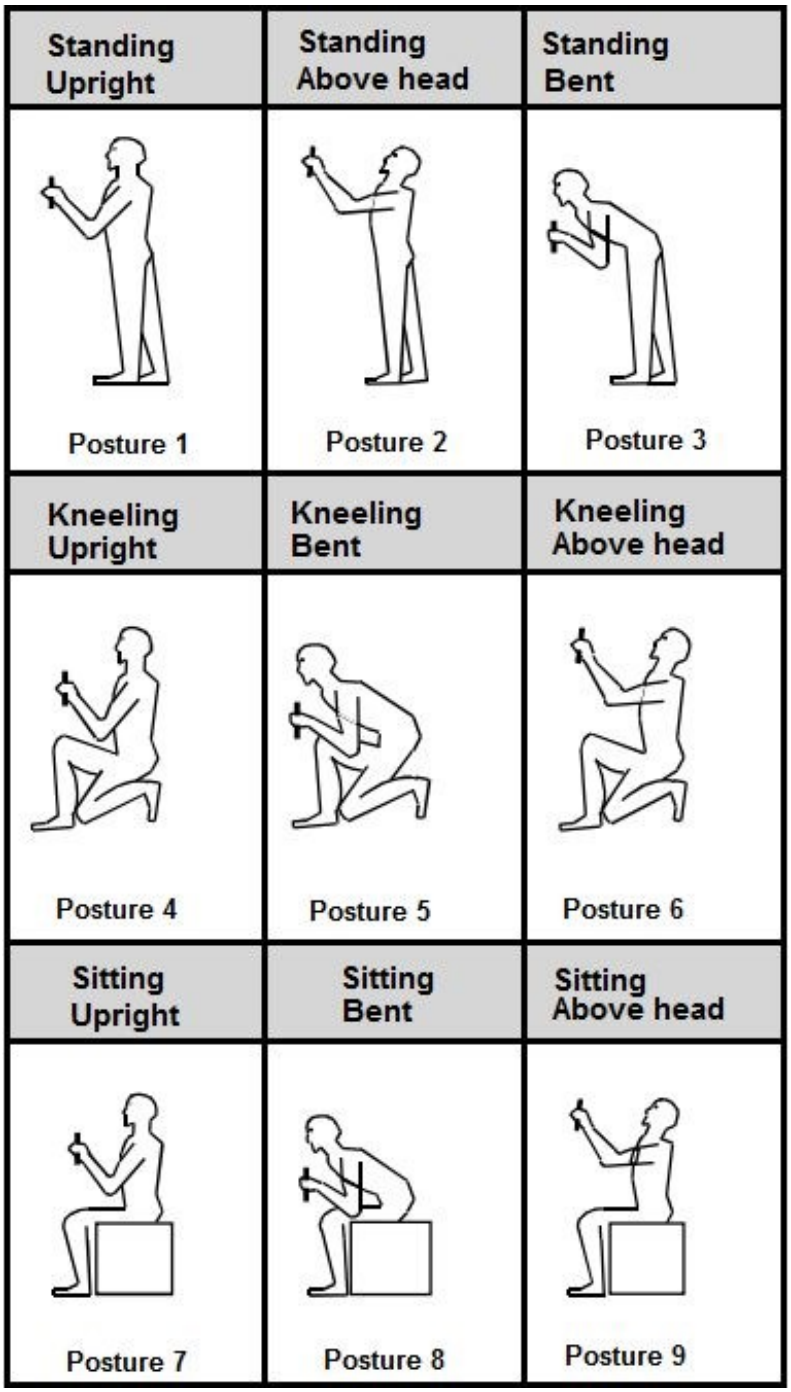


Figure 59: Postures belonging to the EAWS form v1.3.4.

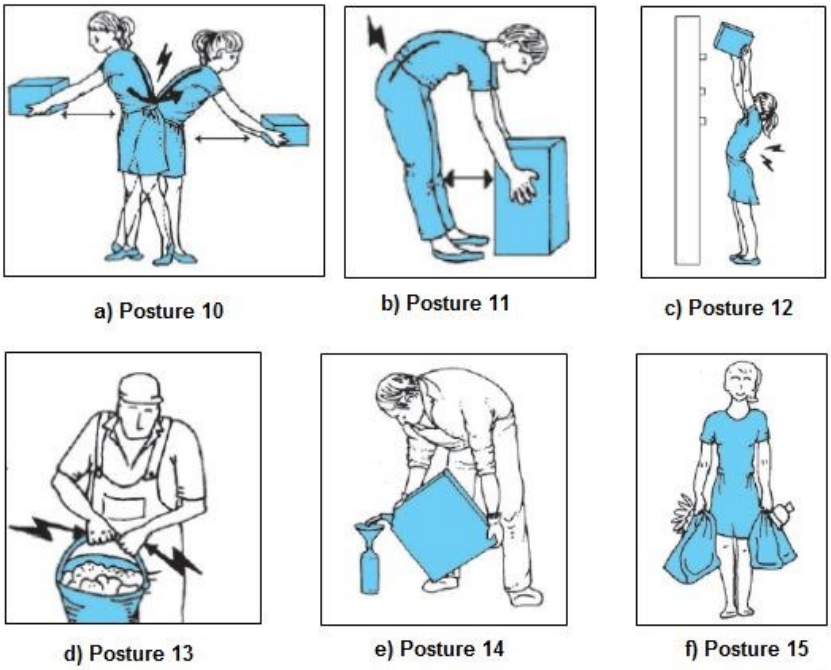


Figure 60: From image a) to e) the five most common awkward postures, in image f) the posture used as the basis for comparison.
(Source:http://www.inail.it/internet_web/wcm/idc/groups/internet/documents/document/ucm_portstg_093067.pdf)

We recruited a volunteer (male, age 26, height 180 cm, and weight 80 kg) as an actor to simulate the aforementioned postures.

We positioned eighteen markers (1.0 cm diameter reflective spheres) on anatomical landmarks, as suggested in (G. Wu et al., 2005) (see Table XVIII and Figure 61). For specific landmark choices we referred to the literature: head (Xu & McGorry, 2015), shoulders and neck (Wiedemann et al., 2015), elbows (Cutti, Paolini, Troncossi, Cappello, & Davalli, 2005; Mackey, Walt, Lobb, & Stott, 2004), wrist centers (Aguinaldo, Buttermore, & Chambers, 2007; Cutti et al., 2005), and the pelvis girdle (Ferrari et al., 2008).

We positioned the Kinect in front of the actor at a distance of about 240 centimeters and at a height of 180 centimeters from the ground. The actor was in the center of the area framed by the optical motion capture system in a laboratory with controlled lighting conditions (400lx). We recorded simultaneously the static postures with both the tracking systems and synchronized them according to the same event-based procedure as in (Xu et al., 2015).

Table XVIII: The anatomical landmarks for reflective markers positioning, the Kinect-identified joint names and their motion tracking system-based counterparts.

Body part	Anatomical landmarks	Kinect-identified joint names	Motion tracking system-based counterparts
Head	Left/Right Temporal Regions (LTR/RTR)	Head	(LTR + RTR)/2
Torso	Left/Right Medial end of the Clavicle (LMC/RMC)	(Not present)	(LMC + RMC)/2
Neck	C7	Neck	(C7 + (LMC + RMC)/2)/2
Left shoulder	Left Acromion (LA)	Left Shoulder	LA
Right shoulder	Right Acromion (RA)	Right Shoulder	RA
Left elbow	Left Lateral Humeral Epicondyle (LLHE), Left Medial Humeral Epicondyle (LMHE)	Left Elbow	(LLHE + LMHE)/2
Right elbow	Right Lateral Humeral Epicondyle (RLHE), Right Medial Humeral Epicondyle (RMHE)	Right Elbow	(RLHE + RMHE)/2
Left wrist	Left Radial Styloid (LRS), Left Ulnar Styloid (LUS)	Left Wrist	(LRS + LUS)/2
Right wrist	Right Radial Styloid (RRS), Right Ulnar Styloid (RUS)	Right Wrist	(RRS + RUS)/2
Left hip	Left Anterior Superior Iliac Spine (LASIS)	Left Hip	LASIS
Right hip	Right Anterior Superior Iliac Spine (RASIS)	Right Hip	RASIS
Sacrum	Sacrum (S)	Spine Base	S

7.4.3. Data analysis

We imported the coordinates from the optical motion capture system in a 3D CAD parametric model (Autodesk Inventor professional 2017), and we measured the required angles. We then computed the RULA grand-scores using the RULA Employee Assessment Worksheet (Hedge, 2000).

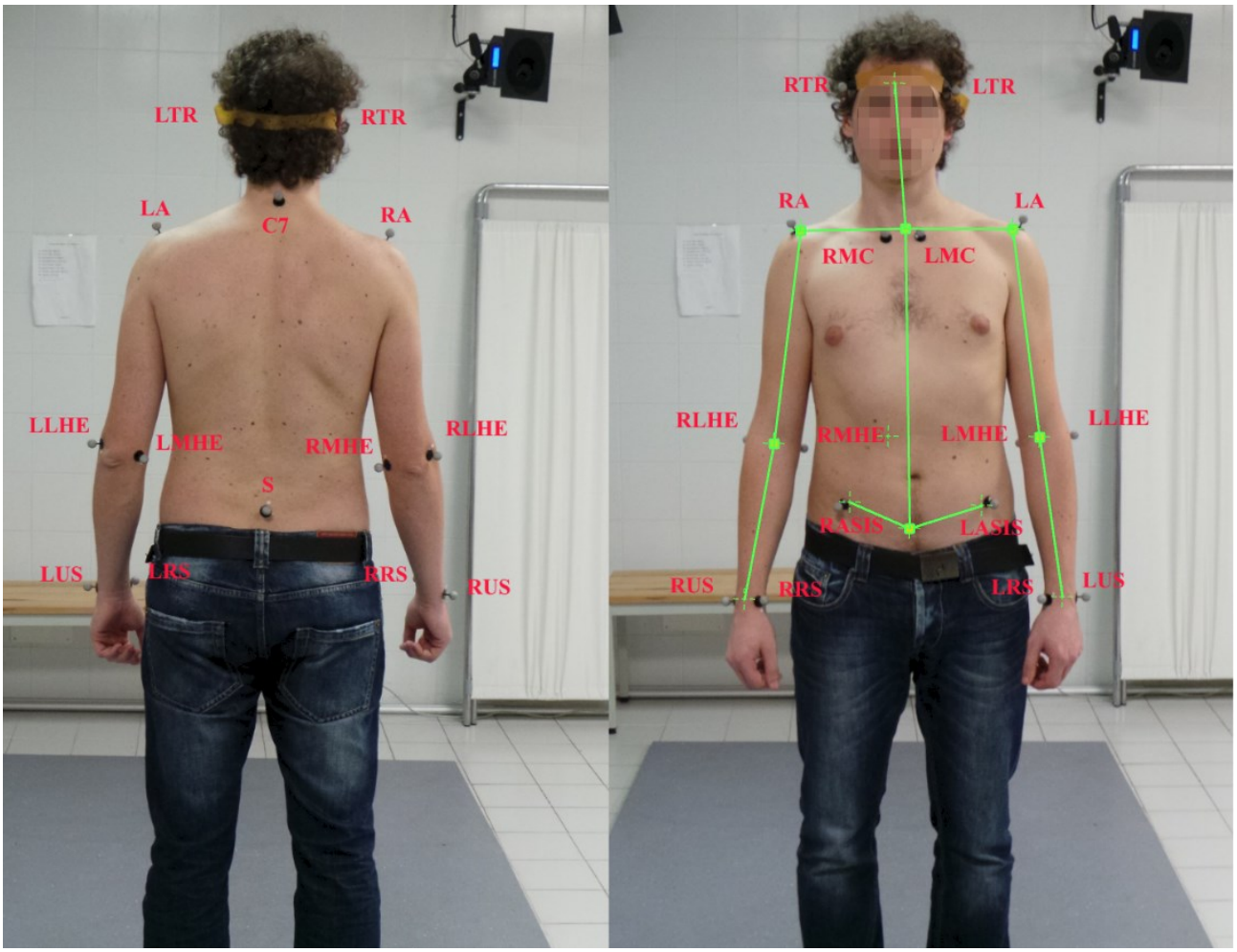


Figure 61: The anatomical landmarks for reflective markers positioning, on the right the skeleton body model generated with the 3D CAD tool is overlaid in green.

We assessed the agreement between the two systems by using two-dimensional contingency tables (Fleiss, Levin, & Paik, 2004). We computed the proportion agreement index (p_0), and the strength of agreement on a sample-to-sample basis as expressed by linear weighted Cohen's kappa.

7.5. Experiment 2: validation with RULA expert and comparison with the Jack TAT

In the second experiment, we compared the K2RULA tool with a human RULA expert and with the Jack TAT. We defined our:

- *hypothesis 2*: K2RULA grand-scores are in agreement with the ones obtained by the RULA expert;
- *hypothesis 3*: the K2RULA provides better results than the Jack Task Analysis Toolkit

7.5.1. Equipment

We collected simultaneously data with a Kinect v2, a Kinect v1, and video with a Webcam Logitech® Hd Pro C920. Two identical PCs (CPU Intel® Core™ i5-4200 2.50 GHz, 4 GB RAM, GPU NVIDIA GeForce GT 740M, OS Windows 8) ran our K2RULA and the TAT software tool version 8.0.1 (based on Kinect v1).

7.5.2. Procedure

We used the same 15 static postures of experiment 1. We recruited a RULA expert (an occupational doctor working for INAIL[7], with more than 10 years of practice) and one volunteer (male, age 28, height 170 cm, weight 72 kg) as an actor. During the experiment, we positioned the two Kinect sensors and the video camera (one above the other) in front of the "actor" as in the previous experiment in a laboratory with controlled lighting conditions (400lx). While the actor was keeping each static pose for a few seconds, we recorded each posture. We assessed the RULA grand-scores using both the K2RULA and the Jack-TAT. The RULA expert analyzed offline the recorded video of each posture and assessed the RULA grand-scores.

7.5.3. Data analysis

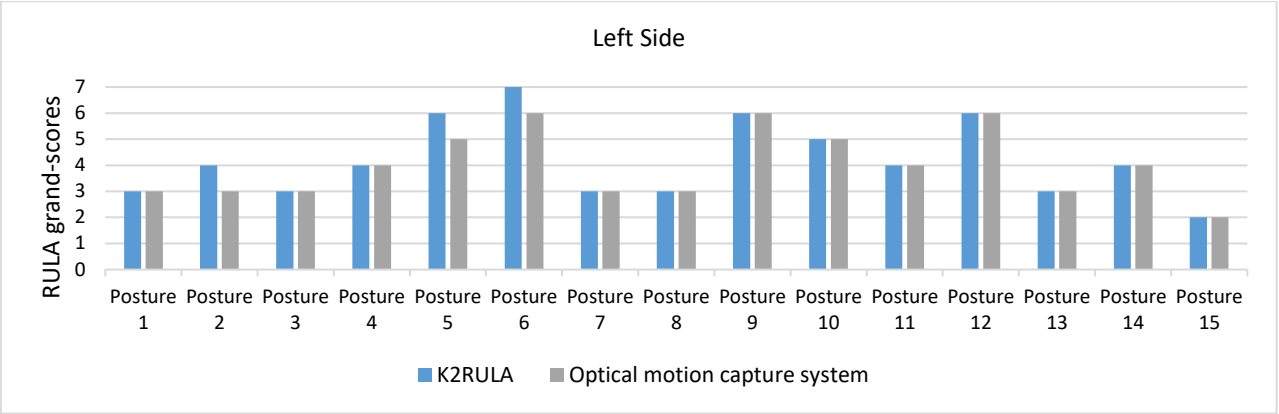
We carried out the comparison between the two Kinect based (KB) methods using as baseline the expert evaluation, as in (Diego-Mas & Alcaide-Marzal, 2014). We assessed the agreement between results as done in experiment 1.

7.6. Results

7.6.1. Experiment 1

Figure 62 shows the RULA grand-scores for the body left and right side obtained with the K2RULA and the optical motion capture system.

[7] The INAIL, the National Institute for Insurance against Accidents at Work, in Italy is the public authority that manages the mandatory insurance against occupational accidents and diseases.



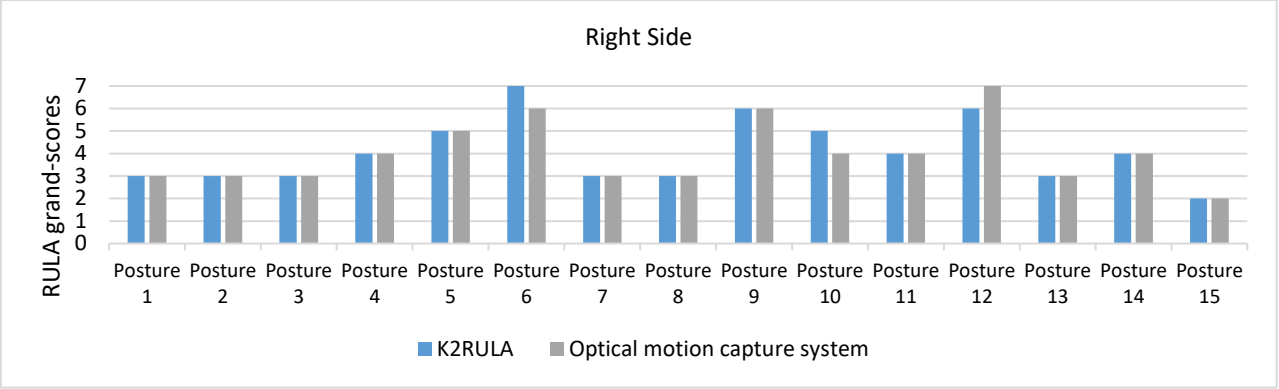


Figure 62: RULA grand-scores for the body left and right side.

These results indicate a "perfect" agreement between the two systems (see Table XIX) in the Landis and Koch scale (Landis & Koch, 1977).

Table XIX: Observed agreements between the K2RULA and the optical motion capture system, linear weighted Cohen's kappa and Z-test results.

Body side	P_o	Cohen's kappa	Agreement (Landis and Koch scale)	z (k/sqrt(var))	p value	Null hypothesis
Left	0.97	0.87	Perfect	4.38	<0.001	Reject
Right	0.97	0.87	Perfect	4.78	<0.001	Reject

To validate the statistical significance of this result, we also tested *the null hypothesis that the observed agreement is accidental*, by referring to the value of the critical ratio *z* to tables of the standard normal distribution. Rejecting the null hypothesis ($p<0.001$) for both the body left and right side, allowed us to confirm the *hypothesis 1*: K2RULA grand-scores are in accordance to the RULA assessments obtained with an optical motion capture system.

7.6.2. Experiment 2

Figure 63 shows the RULA grand-scores for the K2RULA and the Jack-TAT compared with the expert evaluation as a baseline.

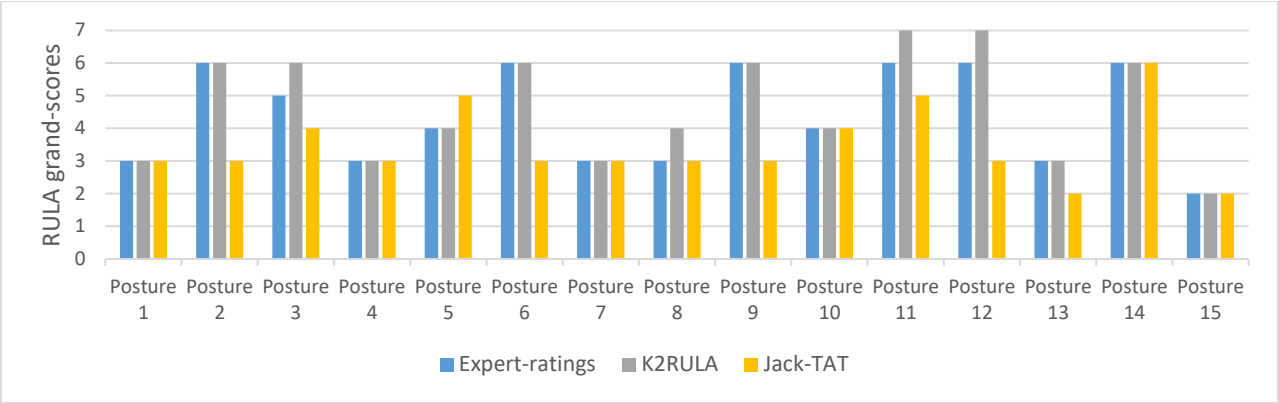


Figure 63: KB methods vs Expert evaluation.

These results indicate "perfect" agreement between the expert and the K2RULA and just "fair" agreement between the expert and the Jack-TAT (see Table XX).

Table XX: Observed agreements, linear weighted Cohen's kappa and Z-test results.

Methods	P_o	Cohen's kappa	Agreement (Landis and Koch scale)	z (k/sqrt(var))	p value	Null hypothesis
Expert- K2RULA	0.96	0.84	Perfect	3.87	<0.001	Reject
Expert- Jack	0.82	0.34	Fair	0.82	0.412	Accept

To validate the statistical significance of these results, we also tested *the null hypothesis that the observed agreement is accidental*.

Rejecting the null hypothesis ($p<0.001$) for the agreement between the expert and the K2RULA allowed us to confirm the *hypothesis 2*: K2RULA grand-scores are in agreement with the ones obtained manually by the RULA expert. On the contrary, accepting the null hypothesis ($p=0.412$) for the agreement between the expert and the Jack-TAT allowed us to confirm the *hypothesis 3*: K2RULA provides better results than the Jack Task Analysis Toolkit.

7.7. Discussion and conclusions

7.7.1. Main contributions

In the first experiment, the RULA grand-scores, returned by the two methods, were identical in 24 postures of the 30 considered. This result is in accordance with the outcomes presented by Plantard et al. (2016). The only six differences were due to detection of the arm abduction and the trunk flexion where K2RULA overestimated the grand-score (+1). The RULA assessment method, based on wide angle ranges, effectively compensates the joint position differences between the two tracking systems, indeed present as reported in the literature.

In the second experiment, the KB methods reported exactly the expert grand-scores for postures one, four, seven, eight, ten, and fourteen (Figure 63). In posture two, Jack-TAT underestimated the ergonomic risk, returning a low score for the neck. Analyzing the video frame, the neck appears back flexed. Jack-TAT was not able to detect this situation. In posture six, the operator is kneeling with outstretched hands high above the level of the shoulders. The neck and forearm have high scores, involving a high ergonomic risk. The expert and K2RULA returned the same severe grand-score. Jack-TAT gave a lower grand-score. Jack-TAT showed some problems with kneeling postures and sometimes it was not able to track the skeleton. In posture nine, the operator sits with both arms raised over shoulder height. The expert and K2RULA returned the same grand-score whereas Jack-TAT gave a lower one. Posture ten is characterized by the trunk rotation and by the left arm crossing the sagittal plane. K2RULA and the expert gave the same score for each body section. Jack-TAT in this case returned the same grand-score, but this correspondence is just accidental as Jack-TAT underestimates the arm section and overestimates the neck section. In posture eleven, the trunk is highly flexed forward. Our tool returned the highest grand-score since it detected even a small twisting and a bending of the trunk. In posture twelve, Jack-TAT did not detect the neck back flexion and underestimated the arm section.

Jack-TAT seems to underestimate the ergonomic risk returning frequently a grand-score lower than the one estimated by the expert (mean error ε = -0.933, error std. dev. σ = 1.34). K2RULA slightly overestimates the risk (mean error ε = 0.267, error std. dev. σ = 0.44). However, this overestimation is conservative and hence consistent with the goal of this tool. K2RULA showed a "perfect" agreement with respect to the expert (P_0=0.96, k=0.84). Our results showed a slightly better agreement than that obtained by Plantard et al. (2016), although their data were acquired in real work conditions.

7.7.2. Limitations of the study

We tested our tool in a laboratory set-up with controlled lighting conditions and without occluding objects. This is the best working condition for the Skeleton Tracking algorithm for Kinect v2 (Q. Wang et al., 2015) and Kinect v1 (Microsoft, 2013), and our results suffer only from the actor's body self-occlusions. We need to further investigate the behavior of our tool in a real working environment. Moreover, the hands' configuration plays a key role, thus our future research will address the hand tracking limits of the Kinect v2. We are planning to apply data fusion techniques to data gathered from the depth sensor and from low cost non-intrusive wearable devices. The availability of such data would probably allow the implementation of

methods and tools able to assess fatigue indexes more detailed than the RULA score, such as OCRA index, moving from static postures analysis to continuous measurement.

7.7.3. Conclusions and research developments

Our work led us to the successful implementation of the K2RULA tool, a real time semi-automatic RULA evaluation system based on Kinect v2. It allows to speed-up the detection of critical conditions and to reduce the subjective bias. K2RULA is able to analyze off-line data and to save the results for deeper ergonomic studies. We validated the proposed tool with two experiments, using as a baseline an optical motion capture system and a RULA expert, proving the reliability of K2RULA as a faster alternative to classical visual inspection evaluation. We also compared it with a commercial software, the Jack-TAT, based on the Kinect v1 sensor. In summary, we demonstrated in laboratory condition that:

1. K2RULA grand-scores are equivalent to the assessments obtained with an optical motion capture system;
2. K2RULA grand-scores are in perfect agreement with a RULA expert evaluation;
3. K2RULA outperforms the Jack-TAT tool, based on Kinect v1.

We can conclude that the proposed system can be effectively used as a fast, semi-automatic and low-cost tool for RULA analysis.

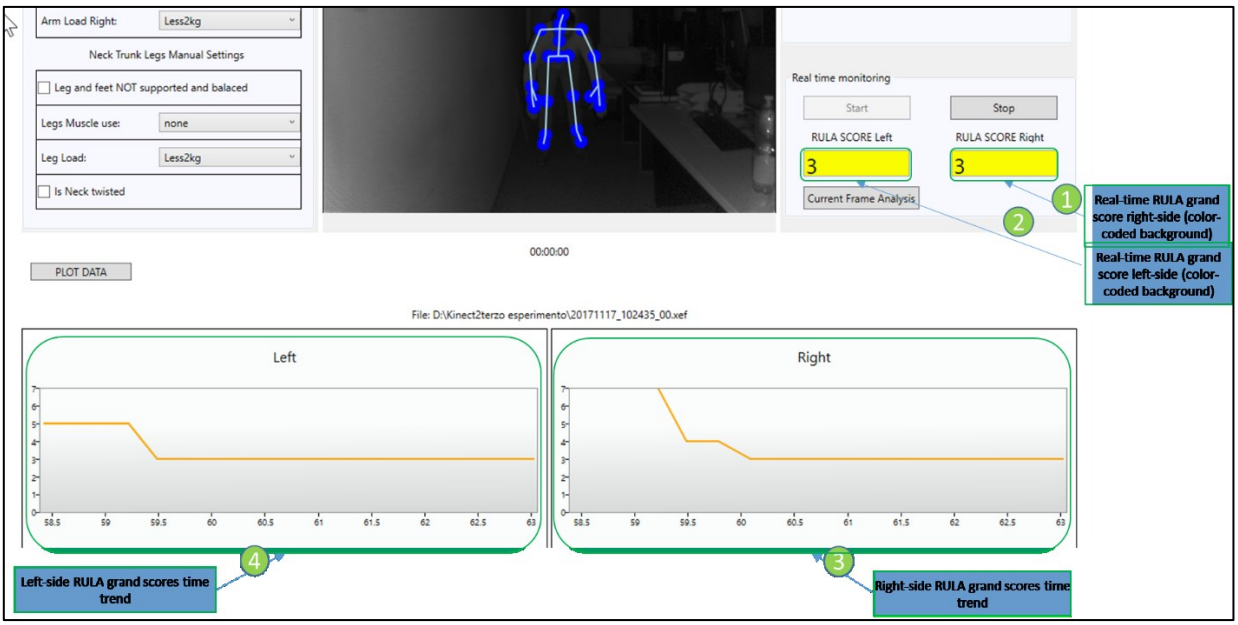


Figure 64 - The real-time feedback provided by the ERGOSENTINEL software, in the yellow boxes the real-time RULA grand score is visualized with a color-coded background.

The successful implementation of the K2RULA software allowed us to cooperate with the Italian National Institute for Insurance against Accidents at Work (INAIL). This cooperation led to

further develop the K2RULA to its new version, the ERGOSENTINEL (Figure 64). Apart from a renewed GUI, the main improvement consists of the availability of a time continuous monitoring able to provide real time warning and feed-back. In particular, the tool is able to plot the RULA grand-scores time-trend and to trigger immediate graphical and acoustic warnings in case the postural risk exceeds a fixed threshold. The real-time capabilities of this tool may represent a useful support for the role of the Healthy Operator.

The ongoing research project in collaboration with the INAIL provides to exploit the low-cost technology of the Kinect sensor to develop a reliable tool for the assessment of the residual Range of Movement (ROM) in subjects that are recovering from work accidents. Currently, this evaluation is accomplished by experts in laboratories by means of manual measures using tools such as goniometers.

Conclusion and future works

In this thesis we described the research aimed at supporting the renewed figure of the worker emerging from the scenario of the fourth industrial revolution – the Operator 4.0.

The key enabling technologies of the I4.0 have reached a level of maturity that allows their effective employment in the factory shop-floor. Nevertheless, what has already been developed in this field it is only an intermediate step towards the full realization of the factory of the future as it is described in the vision of the Industry 4.0 program.

The experiments we carried out to validate the developed solutions returned encouraging results, however, further studies are required.

The validation study of the SAR-aided MWS showed that the SAR presentation mode is significantly better than paper one in completion times (reduction of 20.3% in overall completion time) and error rates (83.3% fewer errors with the SAR mode). The subjective evaluation confirmed good user acceptance of SAR technology for conveying technical instructions. However, in this study, we did not consider as a variable the experience of operators, as well as the learning effect due to the repetition of a procedure at different times, or the possible fatigue of operators using SAR. In a follow-up study, it would be interesting, to evaluate the effect of these variables on user performance with SAR. Thus, we plan to test our SAR system in our industrial partners' facilities and carry out some pilot studies in a real scenario with actual workers.

The study about text legibility evidenced that some of the novel indices proposed had a better correlation with text legibility than those used before in the literature. Anyway, we used just one text color and one text height to explore the possibility to use these indices to classify backgrounds according to their texturization. In a follow-up study, we should validate those indices with other text settings.
Even if technology advancement in OST displays would make legibility less and less an issue, our study could be used as a starting point for the optimization of the Graphical User Interfaces also for novel OST devices, like Microsoft HoloLens. In future experiments, we could analyze a greater collection of industrial backgrounds and rank them according to the best indices found in this work: FR6, GSD, and TY. This ranking could be used to find specific solutions for

legibility for every level of background texturization. In future studies we also plan to evaluate the computational cost of the image processing, and its influence on real time performances.

Our research concerning gesture interfaces returned two main results. The first one points out that users tend to rate well-known interaction modes more usable than novel ones. The second one highlights that, by applying a user centric approach it is possible to lessen the cognitive load and the physical effort involved by a gesture interface, thus making it as acceptable as a classical one.

At the same time, the novelty of such interfaces and their inherent transparency enhances the user experience. This result is particularly relevant for VR application for training in the industrial scenario, where immersion may play a key role in the effectiveness of the training experience.

Our effort to design and develop an effective tool for real-time assessment of postural risk to support the role of the Healthy Operator was successful. Indeed, we validated the K2RULA software with two experiments. In the first one, we compared the K2RULA grand-scores with those obtained with a reference optical motion capture system and we found a statistical perfect match according to the Landis and Koch scale (proportion agreement index = 0.97, k = 0.87). In the second experiment, we evaluated the agreement of the grand-scores returned by the proposed application with those obtained by a RULA expert rater, finding again a statistical perfect match (proportion agreement index = 0.96,

k = 0.84). Anyway, we tested our tool in a laboratory set-up under the best working condition for the Skeleton Tracking algorithm for Kinect v2 (Q. Wang et al., 2015) and Kinect v1 (Microsoft, 2013), and our results suffer only from the actor's body self-occlusions. We need to further investigate the behavior of our tool in a real working environment. Moreover, the hands' configuration plays a key role, thus our future research will address the hand tracking limits of the Kinect v2. We are planning to apply data fusion techniques to data gathered from the depth sensor and from low cost non-intrusive wearable devices. The availability of such data would probably allow the implementation of methods and tools able to assess fatigue indexes more detailed than the RULA score, such as the OCRA index, moving from static postures analysis to continuous measurement.

As already stated, at the moment, we are observing just at an intermediate step towards the fulfillment of the modifications involved by the I4.0 program. The complete accomplishment of this program requires a dedicated effort. Among the various ingredients needed for the success, the ability to efficiently integrate the key enabling technologies, by allowing cooperation among know-hows belonging to various fields of research and industry, represents

a crucial aspect. Indeed, the factory of the future will be a macro cyber physical system, resulting from a deep integration and cooperation processes.

The applications described in this work could be the object of such an integration process too.

For instance, our IAR Framework for P&ID enhanced comprehension described in chapter 4 could be integrated with the method developed by Neges, Wolf, & Abramovici (2017), that proposes the synchronization of engineering data, including the P&ID, via a dynamic graph-based model. With the integration of the two frameworks, it would be possible to both visualize and manipulate in real time, the P&ID, and the related technical information. Indeed, a future step of this research, that we already planned, is that of displaying dynamic information like Key Performance Indices of the plant directly from the P&ID.

Apart from the use of other types of sensors, as described before, the K2RULA tool could be integrated into a smart Manual Working Station. Such an MWS, provided with an on-board processing unit, a wireless network system, and the SAR technology described in chapter 2, could acquire data from the K2RULA. These data could be used both for real time surveillance of ergonomics and for the optimization of the production process for the operator wellbeing, to identify the optimal solution for minimizing the risk exposure of the worker and for achieving a global balancing of the workload. Indeed, data gathered from our tool could feed a model developed for minimizing the exposure risk of workers involved in repetitive manual tasks thus solving the Job Rotation Scheduling Problem, as suggested by Digiesi, Facchini, Mossa, & Mummolo (2017). Further study could also evaluate the effectiveness of conveying ergonomics information both with a SAR system and an HWD system integrated with the K2RULA tool. Our tool could also be used in the training process to make the operator aware of the postural risks s/he is subject to during her/his daily tasks. In a future study we could evaluate the effectiveness of our tool in this application scenario.

The studies carried out in this work buttress our choice to develop applications based on IAR, HMI and ergonomics supporting the role of the Operator 4.0, by enhancing her/his performance and by ensuring the psychophysical wellbeing. The continuous technological innovation will most likely lead to changes in the way these technologies are used in industry and the problems associated with them. In any case, considering the results achieved, their use appears essential to support the operator's commitment to the production scenario of the smart factory of the future.

Acknowledgements

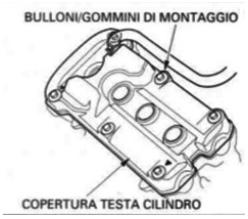
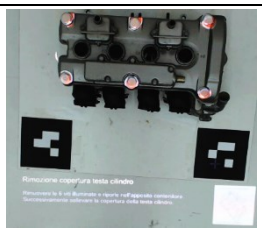
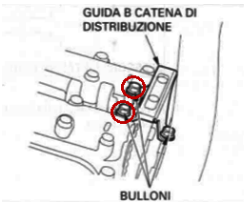
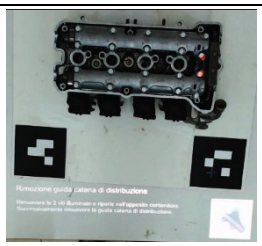
At the end of this work, I have I would like to thank all those who contributed to its realization:

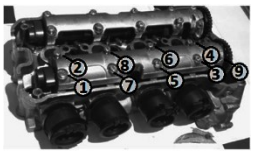
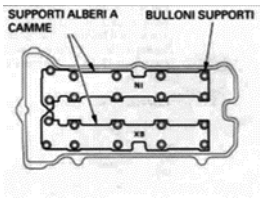
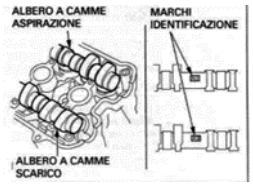
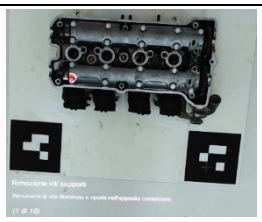
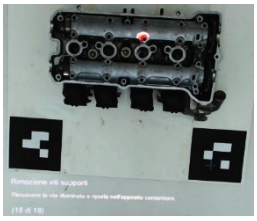
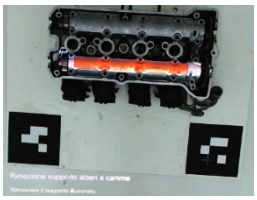
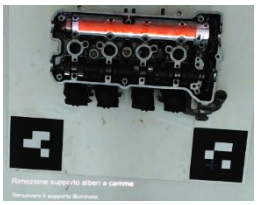
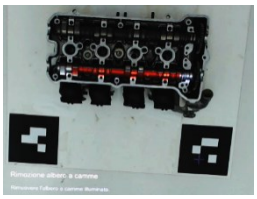
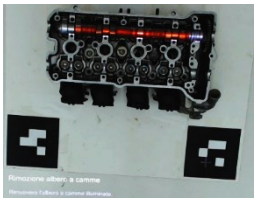
- The National Institute for Insurance against Accidents at Work (INAIL) whose expertise contribution was useful in the validation of the K2RULA tool;
- The pool of partner companies in the VirtualMurgia project: Ali6, GEM, WPS;
- The spin off APIS of the Politecnico di Bari, that developed the software framework of the prototype described in Chapter 2, following our guidelines;
- The start-up Idea75 and the Casillo group that cooperated at the realization of the AR framework for P&ID enhanced comprehension;
- All the people that were part of VR3Lab in these three years: Antonio Boccaccio; Michele Gattullo; Gianpaolo F. Trotta, Alessandro Evangelista and Saverio Debernardis. As friends, they supported me facing the difficulties of my doctoral experience, as colleagues they were plenty of useful suggestions;
- All my cohautors;
- All the undergraduates supported in these years.

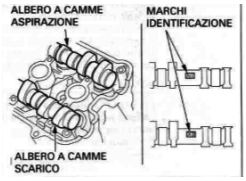
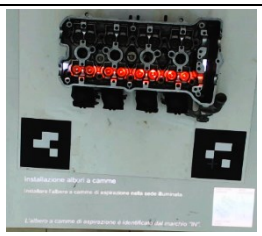
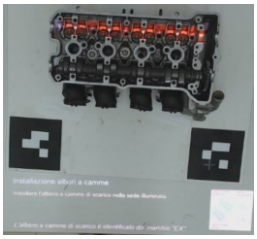
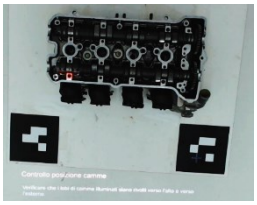
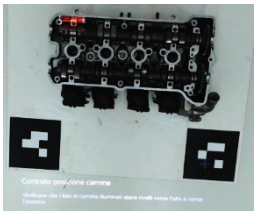
During this three years' experience I was smartly guided by my Professors and Tutors: Giuseppe Monno, Antonio E. Uva, Michele Fiorentino, and Vitantonio Bevilacqua. I would like to express my gratitude towards them. They were both scientific and life tutors to me.

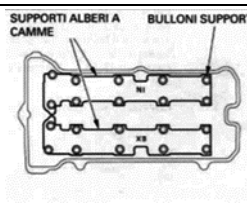
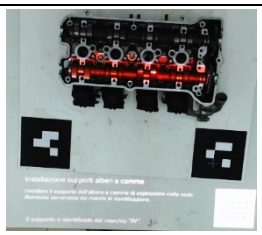
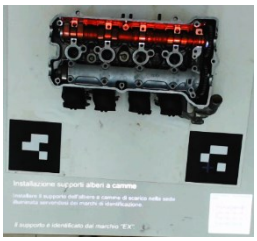
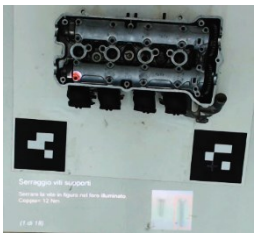
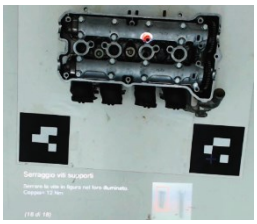
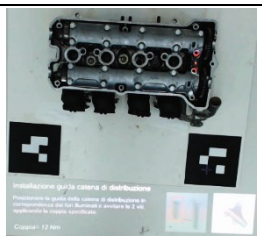
Finally, I thank all my friends, my family and my wife that both buttressed and tolerate me during all the time.

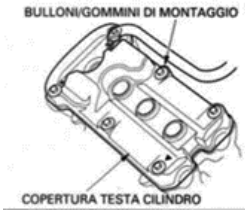
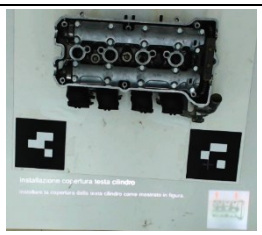
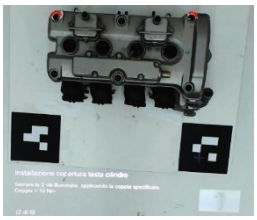
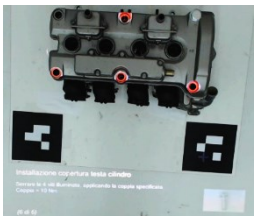
Appendix A

	Paper mode		SAR mode	
	Textual content	**Visual cues**	**Textual content**	**Visual cues**
Task 1	• Remove the 6 screws of the cylinder head cover and place them in the container; • Lift the cylinder head cover.		• Remove the 6 illuminated screws and place them in the container; • Lift the cylinder head cover.	
Task 2	• Remove the 2 screws of the distribution chain guide, and store them in the container; • Remove the distribution chain guide.		• Remove the 2 illuminated screws and place them in the container; • Lift the cylinder head cover.	

	Paper mode		SAR mode	
	Textual content	**Visual cues**	**Textual content**	**Visual cues**
Task 3	• Remove the screws of the intake camshaft holder, identified by the mark IN, in the following order and store them in the container: 1-8-3-9-2-4-7-5-6; • Remove the screws of the exhaust camshaft holder, identified by the mark EX, in the following order and store them in the container: 8-9-3-2-1-7-5-4-6; • Remove the intake camshaft holder, identified by the mark IN. • Remove the exhaust camshaft holder, identified by the mark EX. • Remove the intake camshaft, identified by the mark IN. • Remove the exhaust camshaft, identified by the mark EX.	  	• Remove the illuminated screw and place it in the container. (1 of 18); • Remove the illuminated screw and place it in the container. (18 of 18); • Remove the illuminated camshaft holder; • Remove the illuminated camshaft holder; • Remove the illuminated camshaft; • Remove the illuminated camshaft.	     

	Paper mode		SAR mode	
	Textual content	**Visual cues**	**Textual content**	**Visual cues**
Task 4	• Install the camshafts on the cylinder head with the cam lobes # 1 facing upwards as shown in the figure. • NOTE: Install the camshafts in the correct position using the identification marks: • "IN" – intake camshaft • "EX" – exhaust camshaft • Install the intake camshaft. • Install the exhaust camshaft. • Check that the bearings are aligned with their seats.		• Install the intake camshaft in the illuminated seat. The intake camshaft is identified by the mark "IN"; • Install the exhaust camshaft in the illuminated seat. • The exhaust camshaft is identified by the mark "EX"; • Check that illuminated cam lobes face upwards and outwards; • Check that illuminated cam lobes face upwards and outwards;	   

	Paper mode		SAR mode	
	Textual content	**Visual cues**	**Textual content**	**Visual cues**
Task 5	• Install the camshaft holders in the correct position using the identification marks: • "IN" – intake camshaft holder • "EX" – exhaust camshaft holder • Install the intake camshaft holder. • Install the exhaust camshaft holder. • Install the screws of the intake camshaft holder, in the following order: 1-8-3-9-2-4-7-5-6; • Install the screws of the exhaust camshaft holder, in the following order: 8-9-3-2-1-7-5-4-6.	 	• Install the intake camshaft holder in the illuminated position using the identification mark: • "IN" – intake camshaft holder • Install the exhaust camshaft holder in the illuminated position using the identification mark; • "EX" – exhaust camshaft holder • Insert the screw in the illuminated hole (1 of 18); • Insert the screw in the illuminated hole (18 of 18).	   
Task 6	• Install the distribution chain guide; • install the two screws of the distribution chain guide.		• Place the distribution chain guide in correspondence with the illuminated holes and install the two screws.	

	Paper mode		SAR mode	
	Textual content	**Visual cues**	**Textual content**	**Visual cues**
Task 7	• Install the cylinder head cover. • Install the special screws of the cylinder head cover • Caution: Install first the screws in the holes marked by "Δ."		• Install the cylinder head cover as in figure; • Insert the screw in the two illuminated holes. (2 of 6); • Insert the screw in the four illuminated holes. (6 of 6);	  

Appendix B

Gesture proposals description

For the Move the pointer on the screen referent, two different gestures were proposed (Figure 34):

- Moving Hand with Open palm (MHO): Users keeping open the palm, moved the hand to control the pointer position on the screen;
- Moving Hand with Index Finger Pointing (MHIP): Users kept the index finger pointing at the pointer position on the screen while moved their hand to control the pointer position.

For the Zoom-in referent three different gestures were proposed ((Figure 34):

- One Hand Un-Pinching (OHUP): Two fingers of the dominant hand un-pinch, just as for touch interfaces;
- Distancing Two Hands with Open palms (DTHO): The user moves her/his hands in a plane approximately parallel to the frontal plane. Palms are kept open while the user distances them increasing their relative distance;
- Distancing Two Hands with Clenched fists (DTHC): The user moves her/his hands in a plane approximately parallel to the frontal plane. Fists are kept clenched while the user distances them increasing their relative distance.

For the Zoom-out referent there were three different gesture proposals (Figure 34):

- One Hand Pinching (OHP): Two fingers of the dominant hand pinch just as for touch interfaces;
- Bringing Together Two Hands with Open palms (BTTHO): The user moves her/his hands in a plane approximately parallel to the frontal plane. Palms are kept open while the user brings together hands reducing their relative distance.
- Bringing Together Two Hands with Clenched fists (BTTHC): The user moves her/his hands in a plane approximately parallel to the frontal plane. Fists are kept clenched while the user brings together hands reducing their relative distance;

For the Change gaze direction referent two gestures were proposed (Figure 34):

- One Hand Grabbing and Moving (OHGM): The user clenches her/his fist and changes the direction of view by dragging the scene with her/his hand;
- One Hand Pointing and Moving (OHPM): The user points at the scene with the index finger and changes the direction of view by moving the scene with her/his index finger.

For the Select items referent there were two different gesture proposals (Figure 34):

- One Hand Pushing (OHPu): The user points at the target on the display by positioning the cursor on it, then she/he selects the target by pushing her/his hand toward it;

- One Hand Pointing (OHPo): The user points at the target on the display by positioning the cursor on it, then she/he selects the target by pointing at it with the index finger;
- One Hand Clicking (OHC): The user points at the target on the display by positioning the cursor on it, then she/he selects the target by “clicking” on it.

Appendix C

The *Agreement Rate AR* is defined as "the number of pairs of participants in agreement with each other divided by the total number of pairs of participants that could be in agreement".
In detail, the *AR* can be computed as:

$$AR(r_k) = \frac{|P|}{|P|-1}\sum_{P_i \subseteq P}\left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1};\ AR \in [0, 1]$$

where, P is the set of all proposals for the referent r_k, $|P|$ the size of the set, and $|P_i|$ the size of subsets of similar proposals included in P. The Agreement rate *AR* ranges in the interval [0, 1]. For a given referent r_k, the value $AR(r_k) = 0$ refers to the case where all the proposals collected for that referent r_k are different from each other, the value 1 to the case where all the proposals (collected for r_k) are the same. The agreement rate *AR* for each of the five hypothesized referents (Table C1) was computed by implementing the AGATE tool (AGreement Analysis Toolkit, http://depts.washington.edu/aimgroup/proj/dollar/agate.html).

Table C1. Agreement rate computed for the five hypothesized referents

Referents	Agreement Rate *AR*
Move the pointer on the screen	0.512
Zoom-in	0.325
Zoom-out	0.325
Change gaze direction	0.532
Select items	0.320
average *(AR)*	0.403

It is worthy to note that values of the agreement rates too much low indicate a high cognitive load which requires to redesign the commands set. However, implementing the Variation between agreement rates statistics (V_{rd} statistics), - which is a statistical significance test for comparing two or multiple agreement rates calculated from proposals elicited from the same participants (i.e., repeated measures design) (Vatavu & Wobbrock, 2015) -, we found that the agreement rates related to all the hypothesized referents have statistically significant differences with respect to zero: (*AR*(*Move the pointer on the screen*) = 0.512, $V_{rd}(1) = 208.000$, $p = .001$; *AR*(*Zoom-out*) = 0.325, $V_{rd}(1) = 132.000$, $p = 0.001$; *AR*(*Zoom-in*) = 0.325, $V_{rd}(1) = 132.000$, p

= .001; *AR*(*Select items*) = 0.320, $V_{rd}(1) = 130.000$, $p = .001$); *AR*(*Change gaze direction*) = 0.532, $V_{rd}(1) = 216.000$, $p = .001$)).

Bibliography

Aach, T., Kaup, A., & Mester, R. (1995). On texture analysis: Local energy transforms versus quadrature filters. *Signal processing*, *45*(2), 173–181.

ACGIH. (n.d.). Upper Limb Localized Fatigue: TLV(R) Physical Agents 7th Edition Documentation. Report number 7DOC-782; ACGIH; Cincinnati, Ohio, 2016.

Aguinaldo, A. L., Buttermore, J., & Chambers, H. (2007). Effects of upper trunk rotation on shoulder joint torque among baseball pitchers of various levels. *Journal of Applied Biomechanics*, *23*(1), 42.

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., Zhao, S., et al. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111*, *2*.

Albarelli, A., Celentano, A., Cosmo, L., & Marchi, R. (2015). On the Interplay between Data Overlay and Real-World Context using See-through Displays. *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter* (pp. 58–65). ACM.

Alberto, R., Draicchio, F., Varrecchia, T., Silvetti, A., & Iavicoli, S. (2018). Wearable Monitoring Devices for Biomechanical Risk Assessment at Work: Current Status and Future Challenges—A Systematic Review. *International journal of environmental research and public health*, *15*(9), 2001.

Aleksy, M., Vartiainen, E., Domova, V., & Naedele, M. (2014). Augmented reality for improved service delivery. *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on* (pp. 382–389). IEEE.

Alvarez, H., Aguinaga, I., & Borro, D. (2011). Providing guidance for maintenance operations using automatic markerless augmented reality system. *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on* (pp. 181–190). IEEE.

Andaluz, V. H., Castillo-Carrión, D., Miranda, R. J., & Alulema, J. C. (2017). Virtual Reality Applied to Industrial Processes. *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (pp. 59–74). Springer.

Andersen, R. S., Madsen, O., Moeslund, T. B., & Amor, H. B. (2016). Projecting robot intentions into human environments. *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on* (pp. 294–301). IEEE.

Arroyo, E., Hoernicke, M., Rodrıguez, P., & Fay, A. (2016). Automatic derivation of qualitative plant simulation models from legacy piping and instrumentation diagrams. *Computers & Chemical Engineering*, *92*, 112–132.

AutoCAD Plant 3D. (n.d.). Retrieved from https://www.autodesk.com/products/autocad-plant-3d/overview, last accessed 2018/03/02s://www.autodesk.com/products/autocad-plant-3d/overview, last accessed 2018/03/02

Ayoub, M. (1973). Work place design and posture. *Human Factors*, *15*(3), 265–268.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, *21*(6), 34–47. doi:10.1109/38.963459

Bailly, G., Pietrzak, T., Deber, J., & Wigdor, D. J. (2013). Métamorphe: augmenting hotkey usage with actuated keys. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 563–572). ACM.

Balogh, I., Ørbæk, P., Ohlsson, K., Nordander, C., Unge, J., Winkel, J., Hansson, G.-Å., et al. (2004). Self-assessed and directly measured occupational physical activities—influence of musculoskeletal complaints, age and gender. *Applied ergonomics*, *35*(1), 49–56.

Bao, S., Howard, N., Spielholz, P., Silverstein, B., & Polissar, N. (2009). Interrater reliability of posture observations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(3), 292–309.

Barbieri, L., Bruno, F., & Muzzupappa, M. (2017). Virtual museum system evaluation through user studies. *Journal of Cultural Heritage*, *26*, 101–108.

Benko, H., Ofek, E., Zheng, F., & Wilson, A. D. (2015). Fovear: Combining an optically see-through near-eye display with projector-based spatial augmented reality. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 129–135). ACM.

Besbes, B., Collette, S. N., Tamaazousti, M., Bourgeois, S., & Gay-Bellile, V. (2012). An interactive augmented reality system: a prototype for industrial maintenance training applications. *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 269–270). IEEE.

Billinghurst, M., Clark, A., Lee, G., & others. (2015). A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, *8*(2-3), 73–272.

Bimber, O., & Raskar, R. (2006). Modern approaches to augmented reality. *ACM SIGGRAPH 2006 Courses* (p. 1). ACM.

Blake, J. (2012). *Natural User Interfaces in .Net*. Manning Publications.

Bonnechere, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., Sholukha, V., et al. (2014). Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry. *Gait & posture*, *39*(1), 593–598.

Bordegoni, M., Ferrise, F., Carrabba, E., Di Donato, M., Fiorentino, M., & Uva, A. E. (2014). An application based on Augmented Reality and mobile technology to support remote maintenance. *Conference and Exhibition of the European Association of Virtual and Augmented Reality* (Vol. 1, pp. 131–135).

Borg, G. (1962). *Physical performance and perceived exertion. Studia Psychologica et Paedagogica*. Series altera, Investigationes XI.

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Champaign, IL, US: Human kinetics.

Boulanger, P. (2004). Application of augmented reality to industrial tele-training. *Computer and Robot Vision, 2004. Proceedings. First Canadian Conference on* (pp. 320–328). IEEE.

BTS-Bioengineering. (2016). BTS SMART DX 5000, http://www.btsbioengineering.com/products/kinematics/bts-smart-dx/. last accessed (Oct, 20, 2016). Retrieved from http://www.btsbioengineering.com/products/kinematics/bts-smart-dx/.

Carrozzino, M., & Bergamasco, M. (2010). Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of Cultural Heritage*, *11*(4), 452–458.

Chen, S. E. (1995). Quicktime VR: An image-based approach to virtual environment navigation. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (pp. 29–38). ACM.

Chen, Z., & Li, X. (2010). Markless tracking based on natural feature for augmented reality. *Educational and Information Technology (ICEIT), 2010 International Conference on* (Vol. 2, pp. 2–126). IEEE.

Chittaro, L., & Sioni, R. (2012). An electromyographic study of a laser pointer-style device vs. mouse and keyboard in an object arrangement task on a large screen. *International Journal of Human-Computer Studies*, *70*(3), 234–255. doi:http://dx.doi.org/10.1016/j.ijhcs.2011.11.005

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences*, *86*(23), 9631–9635.

Clark, R. A., Bower, K. J., Mentiplay, B. F., Paterson, K., & Pua, Y.-H. (2013). Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. *Journal of biomechanics*, *46*(15), 2722–2725.

Clark, R. A., Pua, Y.-H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait & posture*, *36*(3), 372–377.

Colombini, D., Colombini, C., & Occhipinti, E. (2012, April). I disturbi muscolo-scheletrici lavorativi. *Milano, Ed. INAIL*. Retrieved from http://www.inail.it/internet_web/wcm/idc/groups/internet/documents/document/ucm_portstg_093067.pdf

Colombo, R., Pisano, F., Mazzone, A., Delconte, C., Micera, S., Carrozza, M. C., Dario, P., et al. (2007). Design strategies to improve patient motivation during robot-aided rehabilitation. *Journal of neuroengineering and rehabilitation*, *4*(1), 3.

Cutti, A. G., Paolini, G., Troncossi, M., Cappello, A., & Davalli, A. (2005). Soft tissue artefact assessment in humeral axial rotation. *Gait & posture*, *21*(3), 341–349.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, *6*(4), 258–266.

David, G. (2005). Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational medicine*, *55*(3), 190–199.

De Weck, O. L., Ross, A. M., & Rhodes, D. H. (2012). Investigating relationships and semantic sets amongst system lifecycle properties (ilities).

Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., & Uva, A. E. (2014). Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE transactions on visualization and computer graphics*, *20*(1), 125–139.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of personality*, *62*(1), 119–142.

Deci, E. L., & Ryan, R. M. (2003). Intrinsic motivation inventory. *Self-Determination Theory*, *267*.

Di Donato, M., Fiorentino, M., Uva, A. E., Gattullo, M., & Monno, G. (2015). Text legibility for projected Augmented Reality on industrial workbenches. *Computers in Industry*, *70*, 70–78.

Diego-Mas, J. A., & Alcaide-Marzal, J. (2014). Using Kinect™ sensor in observational methods for assessing postures at work. *Applied ergonomics*, *45*(4), 976–985.

Digiesi, S., Facchini, F., Mossa, G., & Mummolo, G. (2017). A RULA-Based Optimization Model for Workers' Assignment to an Assembly Line. *XVII International Scientific Conference on Industrial Systems (IS'17)* (pp. 8–13).

Dockrell, S., O'Grady, E., Bennett, K., Mullarkey, C., Mc Connell, R., Ruddy, R., Twomey, S., et al. (2012). An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children's computing posture. *Applied ergonomics*, *43*(3), 632–636.

Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2016). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *The International Journal of Advanced Manufacturing Technology*. doi:10.1007/s00170-016-9164-5

Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2017). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *The International Journal of Advanced Manufacturing Technology*, *89*(5-8), 1279–1293.

Dünser, A., Grasset, R., & Billinghurst, M. (2008). *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand.

Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied ergonomics*, *43*(4), 645–649.

Elia, V., Gnoni, M. G., & Lanzilotto, A. (2016). Evaluating the application of augmented reality devices in manufacturing from a process point of view: An AHP based model. *Expert systems with applications*, *63*, 187–197.

English, W. K., Engelbart, D. C., & Berman, M. L. (1967). Display-selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, (1), 5–15.

Eurofound. (2015). First findings: Sixth European Working Conditions Survey. Eurofound. doi:10.2806/59106

Fernández-Palacios, B. J., Morabito, D., & Remondino, F. (2017). Access to complex reality-based 3D models using virtual reality solutions. *Journal of Cultural Heritage*, *23*, 40–48. doi:https://doi.org/10.1016/j.culher.2016.09.003

Ferrari, A., Benedetti, M. G., Pavan, E., Frigo, C., Bettinelli, D., Rabuffetti, M., Crenna, P., et al. (2008). Quantitative comparison of five current protocols in gait analysis. *Gait & posture*, *28*(2), 207–216.

Fikkert, W., Vet, P. van der, Veer, G. van der, & Nijholt, A. (2010). Gestures for Large Display Control. In S. Kopp & I. Wachsmuth (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction*, Lecture Notes in Computer Science (Vol. 5934, pp. 245–256). Springer Berlin Heidelberg. doi:10.1007/978-3-642-12553-9_22

Fiorentino, M. (2016). HMD testbed - augmented reality test. (Sourceforge, Ed.). Retrieved from https://sourceforge.net/projects/hmdtexttester/

Fiorentino, M., Debernardis, S., Uva, A. E., & Monno, G. (2013). Augmented reality text style readability with see-through head-mounted displays in industrial context. *Presence: Teleoperators and Virtual Environments*, *22*(2), 171–190. doi:10.1162/PRES_a_00146

Fiorentino, M., Radkowski, R., Boccaccio, A., & Uva, A. E. (2016). Magic Mirror Interface for Augmented Reality Maintenance: An Automotive Case Study. *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 160–167). ACM.

Fiorentino, M., Radkowski, R., Stritzke, C., Uva, A. E., & Monno, G. (2013). Design review of CAD assemblies using bimanual natural interface. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *7*(4), 249–260.

Fiorentino, M., Uva, A. E., Gattullo, M., Debernardis, S., & Monno, G. (2014). Augmented reality on large screen for interactive maintenance instructions. *Computers in Industry*, *65*(2), 270–278. doi:10.1016/j.compind.2013.11.004

Fiorentino, M., Uva, A. E., Monno, G., & Radkowski, R. (2016). Natural interaction for online documentation in industrial maintenance. *International Journal of Computer Aided Engineering and Technology*, *8*(1-2), 56–79.

Fite-Georgel, P. (2011). Is there a reality in Industrial Augmented Reality? *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on* (pp. 201–210). Basel, Switzerland: IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6162889

Fleiss, J. L., Levin, B., & Paik, M. C. (2004). The Measurement of Interrater Agreement. *Statistical Methods for Rates and Proportions* (pp. 598–626). John Wiley & Sons, Inc. doi:10.1002/0471445428.ch18

Foxlin, E., Altshuler, Y., Naimark, L., & Harrington, M. (2004). FlightTracker: A novel optical/inertial tracker for cockpit enhanced vision. *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality* (pp. 212–221). doi:10.1109/ISMAR.2004.32

Fraga-Lamas, P., Fernández-Caramés, T. M., Blanco-Novoa, Ó., & Vilar-Montesinos, M. A. (2018). A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. *IEEE Access*, *6*, 13358–13375.

Fıglalı, N., Cihan, A., Esen, H., Fıglalı, A., Çesmeci, D., Güllü, M. K., & Yılmaz, M. K. (2015). Image processing-aided working posture analysis: I-OWAS. *Computers & Industrial Engineering*, *85*, 384–394.

Gabbard, J. L., Swan, J. E., & Hix, D. (2006). The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoperators & Virtual Environments*, *15*(1), 16–32.

Gabbard, J. L., Swan, J. E., Hix, D., Kim, S.-J., & Fitch, G. (2007). Active text drawing styles for outdoor augmented reality: A user-based study and design implications. *2007 IEEE Virtual Reality Conference* (pp. 35–42). IEEE.

Gattullo, M., Uva, A. E., Fiorentino, M., & Monno, G. (2015). Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting. *IEEE Transactions on Visualization and Computer Graphics*, *21*(5), 638–651. doi:10.1109/TVCG.2014.2385056

Gorecky, D., Campos, R., Chakravarthy, H., Dabelow, R., Schlick, J., & Zühlke, D. (2013). MASTERING MASS CUSTOMIZATION–A CONCEPT FOR ADVANCED, HUMAN-CENTERED ASSEMBLY. *Academic Journal of Manufacturing Engineering*, *11*(2).

Gorecky, D., Schmitt, M., Loskyll, M., & Zühlke, D. (2014). Human-machine-interaction in the industry 4.0 era. *Industrial Informatics (INDIN), 2014 12th IEEE International Conference on* (pp. 289–294). Ieee.

Guerra, J. P., Pinto, M. M., & Beato, C. (2015). Virtual reality-shows a new vision for tourism and heritage. *European Scientific Journal, ESJ*, *11*(9).

Haggag, H., Hossny, M., Nahavandi, S., & Creighton, D. (2013). Real Time Ergonomic Assessment for Assembly Operations Using Kinect. *Computer Modelling and Simulation (UKSim), 2013 UKSim 15th International Conference on* (pp. 495–500). doi:10.1109/UKSim.2013.105

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), 10–18.

Haralick, R. M., Shanmugam, K., & others. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610–621.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 904–908). San Francisco, USA: Sage Publications. Retrieved from http://pro.sagepub.com/content/50/9/904.short

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, *1*(3), 139–183.

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, *30*(2), 87–93.

Havard, V., Baudry, D., Louis, A., & Mazari, B. (2015). Augmented reality maintenance demonstrator and associated modelling. *Virtual Reality (VR), 2015 IEEE* (pp. 329–330). IEEE.

Hedge, A. (2000). RULA Employee Assessment Worksheet. *Ithaca, NY: Cornell University*.

Henderson, S., & Feiner, S. (2011). Exploring the benefits of augmented reality documentation for maintenance and repair. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(10), 1355–1368. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5620905

Hessam, J. F., Zancanaro, M., Kavakli, M., & Billinghurst, M. (2017). Towards optimization of mid-air gestures for in-vehicle interactions. *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 126–134). ACM.

Hignett, S., & McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied ergonomics*, *31*(2), 201–205.

Hill, A. L., & Scharff, L. F. (1999). Readability of computer displays as a function of colour, saturation and background texture. *Engineering psychology and cognitive ergonomics.*, *4*, 123–130.

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14 (pp. 1063–1072). Toronto, Ontario, Canada: ACM. doi:10.1145/2556288.2557130

Horejsi, P., Gorner, T., Kurkin, O., Polasek, P., & Januska, M. (2013). Using kinect technology equipment for ergonomics. *Modern Machinery (MM) Science Journal*, *389*.

Horejš*ι*, P. (2015). Augmented reality system for virtual training of parts assembly. *Procedia Engineering*, *100*, 699–706.

Hou, L., Chi, H.-L., Tarng, W., Chai, J., Panuwatwanich, K., & Wang, X. (2017). A framework of innovative learning for skill development in complex operational tasks. *Automation in Construction*, *83*, 29–40.

IAD. (2012). EAWS (Ergonomic Assessment Worksheet - section 4) url http://ergo-mtm.it/wp-content/uploads/2013/09/EAWS-form-v1.3.4-EN.pdf last accessed 10 september 2015. (I. M. Directorate, Ed.). Retrieved from http://ergo-mtm.it/wp-content/uploads/2013/09/EAWS-form-v1.3.4-EN.pdf

IJsselsteijn, W. A., Kort, Y. de, Westerink, J., Jager, M. de, & Bonants, R. (2006). Virtual fitness: stimulating exercise behavior through media technology. *Presence: Teleoperators and Virtual Environments*, *15*(6), 688–698.

ISO. (2007). *System of standards for labor safety. Ergonomics. Manual handling. Part 3. Handling of low loads at high frequency*. International Organization for Standardization.

Jankowski, J., Samp, K., Irzynska, I., Jozwowicz, M., & Decker, S. (2010). Integrating text with video and 3d graphics: The effects of text drawing styles on text readability. *Proceedings*

of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1321–1330). ACM.

Jensen, B. R., Laursen, B., & Sjøgaard, G. (2000). Aspects of shoulder function in relation to exposure demands and fatigue–a mini review. *Clinical Biomechanics*, *15*, S17–S20. Retrieved from http://www.sciencedirect.com/science/article/pii/S0268003300000541

Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., et al. (2014). RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-camera Units. *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14 (pp. 637–644). Honolulu, Hawaii, USA: ACM. doi:10.1145/2642918.2647383

Jung, T., Gross, M. D., & Do, E. Y.-L. (2002). Annotating and sketching on 3D web models. *Proceedings of the 7th international conference on Intelligent user interfaces* (pp. 95–102). ACM.

Kahol, K., Tripathi, K., & Panchanathan, S. (2006). Documenting motion sequences with a personalized annotation system. *IEEE Multimedia*, *13*(1), 37–45.

Kantowitz, B. H. (1987). 3. Mental workload. *Advances in psychology* (Vol. 47, pp. 81–121). Elsevier.

Karhu, O., Härkönen, R., Sorvali, P., & Vepsäläinen, P. (1981). Observing working postures in industry: Examples of OWAS application. *Applied Ergonomics*, *12*(1), 13–17.

Kee, D., & Karwowski, W. (2001). LUBA: an assessment technique for postural loading on the upper body based on joint motion discomfort and maximum holding time. *Applied Ergonomics*, *32*(4), 357–366.

Keyserling, W. M., Brouwer, M., & Silverstein, B. A. (1992). A checklist for evaluating ergonomic risk factors resulting from awkward postures of the legs, trunk and neck. *International Journal of Industrial Ergonomics*, *9*(4), 283–301.

Keyserling, W., Stetson, D., Silverstein, B., & Brouwer, M. (1993). A checklist for evaluating ergonomic risk factors associated with upper extremity cumulative trauma disorders. *Ergonomics*, *36*(7), 807–831.

Kim, S., & Dey, A. K. (2009). Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 133–142). ACM.

Kirishima, T., Sato, K., & Chihara, K. (2005). Real-time gesture recognition by learning and selective control of visual interest points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(3), 351–364.

Klinker, G., Creighton, O., Dutoit, A. H., Kobylinski, R., Vilsmeier, C., & Brügge, B. (2001). Augmented maintenance of powerplants: A prototyping case study of a mobile AR system. *isar* (p. 124). IEEE.

Koehl, M., Schneider, A., Fritsch, E., Fritsch, F., Rachedi, A., & Guillemin, S. (2013). Documentation of historical building via virtual tour: the complex building of baths in Strasbourg. *Proceedings of the XXIV International CIPA Symposium on Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Strasbourg, France* (pp. 2–6).

Kowalski, K., Rhodes, R., Naylor, P.-J., Tuokko, H., & MacDonald, S. (2012). Direct and indirect measurement of physical activity in older adults: a systematic review of the literature. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 1.

Kruger, J., & Nguyen, T. D. (2015). Automated vision-based live ergonomics analysis in assembly operations. *{CIRP} Annals - Manufacturing Technology*, *64*(1), 9–12. doi:http://dx.doi.org/10.1016/j.cirp.2015.04.046

Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on* (pp. 3–12). IEEE.

Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H., Biering-Sørensen, F., Andersson, G., & Jørgensen, K. (1987). Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Applied ergonomics*, *18*(3), 233–237.

Kwiatek, K., & Woolner, M. (2009). Embedding interactive storytelling within still and video panoramas for cultural heritage sites. *Virtual Systems and Multimedia, 2009. VSMM'09. 15th International Conference on* (pp. 197–202). IEEE.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision research*, *27*(7), 1165–1177.

Leykin, A., & Tuceryan, M. (2004). Automatic determination of text readability over textured backgrounds for augmented reality systems. *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 224–230). IEEE Computer Society.

Li, G., & Buckle, P. (1999). Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*, *42*(5), 674–695.

Li, X., Yi, W., Chi, H.-L., Wang, X., & Chan, A. P. (2018). A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction*, *86*, 150–162.

Liverani, A., Amati, G., & Caligiana, G. (2006). Interactive control of manufacturing assemblies with Mixed Reality. *Integrated Computer-Aided Engineering*, *13*(2), 163–172.

Livingston, M. A. (2006). Quantification of visual capabilities using augmented reality displays. *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on* (pp. 3–12). IEEE.

Livingston, M. A., Gabbard, J. L., Swan, J. E., Sibley, C. M., & Barrow, J. H. (2013). Basic perception in head-worn augmented reality displays. *Human factors in augmented reality environments* (pp. 35–65). Springer.

Loch, F., Quint, F., & Brishtel, I. (2016). Comparing video and augmented reality assistance in manual assembly. *Intelligent Environments (IE), 2016 12th International Conference on* (pp. 147–150). IEEE.

Lorenz, M., Ruessmann, M., Strack, R., Lueth, K. L., & Bolle, M. (2015). Man and machine in industry 4.0: How will technology transform the industrial workforce through 2025. *The Boston Consulting Group*.

Lowe, B. D. (2004a). Accuracy and validity of observational estimates of shoulder and elbow posture. *Applied ergonomics*, *35*(2), 159–171.

Lowe, B. D. (2004b). Accuracy and validity of observational estimates of wrist and forearm posture. *Ergonomics*, *47*(5), 527–554.

Mackey, A. H., Walt, S. E., Lobb, G., & Stott, N. S. (2004). Intraobserver reliability of the modified Tardieu scale in the upper limb of children with hemiplegia. *Developmental Medicine & Child Neurology*, *46*(4), 267–272.

Manghisi, V. M., Fiorentino, M., Gattullo, M., Boccaccio, A., Bevilacqua, V., Cascella, G. L., Dassisti, M., et al. (2017). Experiencing the Sights, Smells, Sounds, and Climate of Southern Italy in VR. *IEEE computer graphics and applications*, (6), 19–25.

Manghisi, V. M., Uva, A. E., Fiorentino, M., Bevilacqua, V., Trotta, G. F., & Monno, G. (2017). Real time RULA assessment using Kinect v2 sensor. *Applied ergonomics*, *65*, 481–491.

McAtamney, L., & Nigel Corlett, E. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics*, *24*(2), 91–99. Retrieved from http://www.sciencedirect.com/science/article/pii/000368709390080S

Microsoft. (2013). Microsoft Developer Network. Natural User Interface for Kinect for Windows, url:http://msdn.microsoft.com/en-us/library/hh855352. aspx, last accessed

(May, 10, 2016). Retrieved from http://msdn.microsoft.com/en-us/library/hh855352.aspx

Microsoft. (2014). Human Interface Guidelines v2.0, http://download.microsoft.com/download/6/7/6/676611B4-1982-47A4-A42E-4CF84E1095A8/KinectHIG.2.0.pdf. Retrieved from http://download.microsoft.com/download/6/7/6/676611B4-1982-47A4-A42E-4CF84E1095A8/KinectHIG.2.0.pdf

Mihelj, M., Novak, D., Milavec, M., Ziherl, J., Olenšek, A., & Munih, M. (2012). Virtual rehabilitation environment using principles of intrinsic motivation and game design. *Presence: Teleoperators and Virtual Environments*, *21*(1), 1–15.

Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016). Online Detection and Classification of Dynamic Hand Gestures With Recurrent 3D Convolutional Neural Network. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Moloney, J. (2006). Augmented reality visualisation of the built environment to support design decision making. *Information Visualization, 2006. IV 2006. Tenth International Conference on* (pp. 687–692). IEEE.

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m. c., & Wobbrock, J. O. (2014). Reducing Legacy Bias in Gesture Elicitation Studies. *interactions*, *21*(3), 40–45. doi:10.1145/2591689

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding users' preferences for surface gestures. *Proceedings of graphics interface 2010* (pp. 261–268). Canadian Information Processing Society.

Mortara, M., Catalano, C. E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., & Petridis, P. (2014). Learning cultural heritage by serious games. *Journal of Cultural Heritage*, *15*(3), 318–325. doi:https://doi.org/10.1016/j.culher.2013.04.004

Munk, K. H. (2001). Development of a gesture plug-in for natural dialogue interfaces. *International Gesture Workshop* (pp. 47–58). Springer.

Nakai, A., Kajihara, Y., Nishimoto, K., & Suzuki, K. (2017). Information-sharing system supporting onsite work for chemical plants. *Journal of Loss Prevention in the Process Industries*, *50*, 15–22.

Navab, N. (2004). Developing killer apps for industrial augmented reality. *IEEE Computer Graphics and applications*, *24*(3), 16–20.

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

Nee, A. Y., & Ong, S.-K. (2013a). Virtual and augmented reality applications in manufacturing. *IFAC proceedings volumes*, *46*(9), 15–26.

Nee, A. Y., & Ong, S.-K. (2013b). Virtual and augmented reality applications in manufacturing. *IFAC proceedings volumes*, *46*(9), 15–26.

Neges, M., Wolf, M., & Abramovici, M. (2017). Enabling Round-Trip Engineering Between P&I Diagrams and Augmented Reality Work Instructions in Maintenance Processes Utilizing Graph-Based Modelling. *International Conference on Intelligent Systems in Production Engineering and Maintenance* (pp. 33–42). Springer.

Nguyen, T. D., Kleinsorge, M., & Kruger, J. (2014). ErgoAssist: An assistance system to maintain ergonomie guidelines at workplaces. *Emerging Technology and Factory Automation (ETFA), 2014 IEEE* (pp. 1–4). doi:10.1109/ETFA.2014.7005258

Nielsen, M., Störring, M., Moeslund, T., & Granum, E. (2004). A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction*, Lecture Notes in Computer Science (Vol. 2915, pp. 409–420). Springer Berlin Heidelberg. doi:10.1007/978-3-540-24598-8_38

Novak, D., Nagle, A., Keller, U., & Riener, R. (2014). Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. *Journal of neuroengineering and rehabilitation*, *11*(1), 64.

Occhipinti, E. (1998). OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*, *41*(9), 1290–1311.

Ohshima, T., Kuroki, T., Yamamoto, H., & Tamura, H. (2003). A mixed reality system with visual and tangible interaction capability-application to evaluating automobile interior design. *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality* (p. 284). IEEE Computer Society.

Okuma, T., Kurata, T., & Sakaue, K. (2004). A natural feature-based 3D object tracking method for wearable augmented reality. *Proc. of Advanced Motion Control (AMC'04)*, 451–456.

Olwal, A., Gustafsson, J., & Lindfors, C. (2008). Spatial augmented reality on industrial CNC-machines. In I. E. McDowall & M. Dolinsky (Eds.), *The Engineering Reality of Virtual Reality 2008*. SPIE-Intl Soc Optical Eng. doi:10.1117/12.760960

Ong, S., Yuan, M., & Nee, A. (2008). Augmented reality applications in manufacturing: a survey. *International journal of production research*, *46*(10), 2707–2742.

Ooi, B., Wong, C., Tan, I., & Lee, C. (2014). Towards Natural Gestures for Presentation Control Using Microsoft Kinect. In W. Ooi, C. M. Snoek, H. Tan, C.-K. Ho, B. Huet, & C.-W. Ngo (Eds.), *Advances in Multimedia Information Processing – PCM 2014*, Lecture Notes in Computer Science (Vol. 8879, pp. 258–261). Springer International Publishing. doi:10.1007/978-3-319-13168-9_28

Orlosky, J., Kiyokawa, K., & Takemura, H. (2014). Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review*, *18*(2), 20–31.

Paelke, V. (2014). Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment. *Emerging Technology and Factory Automation (ETFA), 2014 IEEE* (pp. 1–4). IEEE.

Palmarini, R., Erkoyuncu, J. A., Roy, R., & Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, *49*, 215–228.

Patrizi, A., Pennestrı, E., & Valentini, P. P. (2015). Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics. *Ergonomics*, 1–8.

Pereira, A., Wachs, J. P., Park, K., & Rempel, D. (2015). A user-developed 3-D hand gesture set for human–computer interaction. *Human factors*, *57*(4), 607–621.

Petkov, N., & Westenberg, M. A. (2003). Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition. *Biological cybernetics*, *88*(3), 236–246.

Pinzke, S., & Kopp, L. (2001). Marker-less systems for tracking working posturesâ€"results from two experiments. *Applied Ergonomics*, *32*(5), 461–471. doi:http://dx.doi.org/10.1016/S0003-6870(01)00023-0

Piumsomboon, T., Clark, A., Billinghurst, M., & Cockburn, A. (2013). User-Defined Gestures for Augmented Reality. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013*, Lecture Notes in Computer Science (Vol. 8118, pp. 282–299). Springer Berlin Heidelberg. doi:10.1007/978-3-642-40480-1_18

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of personality*, *53*(3), 435–449.

Plantard, P., Auvinet, E., Pierres, A.-S. L., & Multon, F. (2015). Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin. *Sensors*, *15*(1), 1785–1803.

Plantard, P., Shum, H., Le Pierres, A.-S., & Multon, F. (2016). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*, *30*, 1e8.

Platonov, J., Heibel, H., Meier, P., & Grollmann, B. (2006). A mobile markerless AR system for maintenance and repair. *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on* (pp. 105–108). IEEE.

Preedy, V. R. (Ed.). (2012). *Handbook of Anthropometry*. Springer US. Retrieved from https://books.google.it/books?id=4nak4tYX8WIC

Pullambaku, M., & Tsering, T. (2016). Multi-user and immersive experiencesin education: the Ename 1290 game, based on the use of MicrosoftKinect and Unity.

Ramirez, H., Mendivil, E. G., Flores, P. R., & Gonzalez, M. C. (2013). Authoring software for augmented reality applications for the use of maintenance and training process. *Procedia Computer Science*, *25*, 189–193.

Rantanen, J. (1981). Risk assessment and the setting of priorities in occupational health and safety. *Scandinavian journal of work, environment & health*, 84–90.

Rea, M. S. (2000). The IESNA lighting handbook: reference & application.

Regenbrecht, H., Baratoff, G., & Wilke, W. (2005). Augmented reality projects in the automotive and aerospace industries. *Computer Graphics and Applications, IEEE*, *25*(6), 48–56.

Reinfeld, K. (2016). krpano https://krpano.com/ last (accessed genuary 2017). Retrieved from https://krpano.com/ last accessed genuary 2017

Ren, G., Li, C., O'Neill, E., & Willis, P. (2013). 3d freehand gestural navigation for interactive public displays. *IEEE computer graphics and applications*, *33*(2), 47–55.

Ridgway, K., Clegg, C. W., Williams, D., Hourd, P., Robinson, M., Bolton, L., Cichomska, K., et al. (2013). The factory of the future. *Government Office for Science, Evidence Paper*, *29*.

Robertson, M., Amick, B. C., DeRango, K., Rooney, T., Bazzani, L., Harrist, R., & Moore, A. (2009). The effects of an office ergonomics training and chair intervention on worker knowledge, behavior and musculoskeletal risk. *Applied Ergonomics*, *40*(1), 124–135.

Rohmert, W. (1960). Ermittlung von Erholungspausen für statische Arbeit des Menschen. *European Journal of Applied Physiology and Occupational Physiology*, *18*(2), 123–164.

Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for MSD risk assessment. *Applied ergonomics*, *45*(3), 420–427.

Römer, T., & Bruder, R. (2015). User centered design of a cyber-physical support solution for assembly processes. *Procedia Manufacturing*, *3*, 456–463.

Romero, D., Bernus, P., Noran, O., Stahre, J., & Fast-Berglund, Å. (2016). The operator 4.0: human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. *IFIP International Conference on Advances in Production Management Systems* (pp. 677–686). Springer.

Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., & Gorecky, D. (2016). Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies. *INTERNATIONAL CONFERENCE ON COMPUTERS & INDUSTRIAL ENGINEERING (CIE46)* (pp. 1–11).

Ruiz, J., & Vogel, D. (2015). Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15 (pp. 3347–3350). Seoul, Republic of Korea: ACM. doi:10.1145/2702123.2702583

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, *43*(3), 450.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54–67.

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and emotion*, *15*(3), 185–205.

Sand, O., Büttner, S., Paelke, V., & Röcker, C. (2016). smart. assembly–projection-based augmented reality for supporting assembly workers. *International Conference on Virtual, Augmented and Mixed Reality* (pp. 643–652). Springer.

Scharff, L. F., & Ahumada, A. J. (2002). Predicting the readability of transparent text. *Journal of vision*, *2*(9), 7–7.

Scharff, L. F., Ahumada Jr, A. J., & Hill, A. L. (1999). Discriminability measures for predicting readability. *Electronic Imaging'99* (pp. 270–277). International Society for Optics and Photonics.

Schaub, K., Caragnano, G., Britzke, B., & Bruder, R. (2013). The European assembly worksheet. *Theoretical Issues in Ergonomics Science*, *14*(6), 616–639.

Schneider, M., Rambach, J., & Stricker, D. (2017). Augmented reality based on edge computing using the example of remote live support. *Industrial Technology (ICIT), 2017 IEEE International Conference on* (pp. 1277–1282). IEEE.

Schwald, B., & De Laval, B. (2003a). An augmented reality system for training and assistance to maintenance in the industrial context.

Schwald, B., & De Laval, B. (2003b). An augmented reality system for training and assistance to maintenance in the industrial context. *Journal of WSCG*, *11*(1-3). Retrieved from https://otik.uk.zcu.cz/handle/11025/1662

Schwerdtfeger, B., Pustka, D., Hofhauser, A., & Klinker, G. (2008). Using laser projectors for augmented reality. *Proceedings of the 2008 ACM symposium on Virtual reality software and technology* (pp. 134–137). ACM.

Shackel, B., Chidsey, K., & Shipley, P. (1969). The Assessment of Chair Comfort. *Ergonomics*, *12*(2), 269–306. doi:10.1080/00140136908931053

Siemens. (2013). Jack and Process Simulate Human, http://www.plm.automation.siemens.com/en_gb/products/tecnomatix/manufacturing-simulation/human-ergonomics/jack.shtml last accessed (May, 10, 2016). Retrieved from http://www.plm.automation.siemens.com/en_gb/products/tecnomatix/manufacturing-simulation/human-ergonomics/jack.shtml

Silpasuwanchai, C., & Ren, X. (2015). Designing Concurrent Full-Body Gestures for Intense Gameplay. *International Journal of Human-Computer Studies*, (0). doi:http://dx.doi.org/10.1016/j.ijhcs.2015.02.010

Siltanen, S. (2012). *Theory and applications of marker-based augmented reality*. VTT.

Smparounis, K., Mavrikios, D., Pappas, M., Xanthakis, V., Viganò, G. P., & Pentenrieder, K. (2008). A virtual and augmented reality approach to collaborative product design and demonstration. *Technology Management Conference (ICE), 2008 IEEE International* (pp. 1–8). IEEE.

Solomon, J. A., Pelli, D. G., & others. (1994). The visual filter mediating letter identification. *Nature*, *369*(6479), 395–397.

Stern, H. I., Wachs, J. P., & Edan, Y. (2006). Human factors for design of hand gesture human-machine interaction. *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on* (Vol. 5, pp. 4052–4056). IEEE.

Stern, H. I., Wachs, J. P., & Edan, Y. (2008a). Optimal consensus intuitive hand gesture vocabulary design. *Semantic Computing, 2008 IEEE International Conference on* (pp. 96–103). IEEE.

Stern, H. I., Wachs, J. P., & Edan, Y. (2008b). Designing hand gesture vocabularies for natural interaction by combining psycho-physiological and recognition factors. *International Journal of Semantic Computing*, *2*(01), 137–160.

Styliani, S., Fotis, L., Kostas, K., & Petros, P. (2009). Virtual museums, a survey and some issues for consideration. *Journal of cultural Heritage*, *10*(4), 520–528.

Subakti, H., & Jiang, J.-R. (2016). A marker-based cyber-physical augmented-reality indoor guidance system for smart campuses. *2016 IEEE 18th International Conference on High-Performance Computing and Communications, IEEE 14th International Conference on Smart City, and IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 1373–1379). IEEE.

Tallig, G., Zender, R., & Runge, M. (2017). Framework-Based Augmented Reality Learning Scenario in Automotive Education. *Proceedings of DeLFI and GMW Workshops*.

Tan, W. C., Chen, I.-M., Pan, S. J., & Tan, H. K. (2016). Automated design evaluation on layout of Piping and Instrumentation Diagram using Histogram of Connectivity. *Automation Science and Engineering (CASE), 2016 IEEE International Conference on* (pp. 1295–1300). IEEE.

Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., & Nishio, S. (2008). An information layout method for an optical see-through head mounted display focusing on the viewability. *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 139–142). IEEE Computer Society.

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative Effectiveness of Augmented Reality in Object Assembly. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03 (pp. 73–80). Ft. Lauderdale, Florida, USA: ACM. doi:10.1145/642611.642626

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. *Conference on Human Factors in Computing Systems - Proceedings* (pp. 73–80). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0037699657&partnerID=40&md5=ce5f76e50aa13426b5ce46ae9d2b8630

Tegeltija, S. S., Lazarevi, M. M., Stankovski, S. V., osi, I. P., Todorovi, V. V., & Ostoji, G. M. (2016). Heating circulation pump disassembly process improved with augmented reality. *Thermal Science*, *20*(suppl. 2), 611–622.

Thanedar, V., & Höllerer, T. (2004). Semi-automated placement of annotations in videos. *UC, Santa Barbara, Tech. Rep., Nov*.

Thomas, B. H. (2009). Augmented Reality Visualisation Facilitating The Architectural Process. *Mixed Reality in Architecture, Design and Construction* (pp. 105–118). Springer.

Tzafestas, S. (2006). Concerning human-automation symbiosis in the society and the nature. *Int'l. J. of Factory Automation, Robotics and Soft Computing*, *1*(3), 6–24.

Van Krevelen, D., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, *9*(2), 1.

Vatavu, R.-D. (2012). User-defined Gestures for Free-hand TV Control. *Proceedings of the 10th European Conference on Interactive Tv and Video*, EuroiTV '12 (pp. 45–48). Berlin, Germany: ACM. doi:10.1145/2325616.2325626

Vatavu, R.-D., & Wobbrock, J. O. (2015). Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15 (pp. 1325–1334). Seoul, Republic of Korea: ACM. doi:10.1145/2702123.2702223

W3C. (2012). W3C Consortium Contrast (Minimum). Retrieved from https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-contrast.html

Wang, Q., Kurillo, G., Ofli, F., & Bajcsy, R. (2015). Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. *Healthcare Informatics (ICHI), 2015 International Conference on* (pp. 380–389). IEEE.

Wang, X., Ong, S. K., & Nee, A. Y. (2016). A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, *4*(1), 1–22.

Wang, X., Ong, S., & Nee, A. (2016). Real-virtual components interaction for assembly simulation and planning. *Robotics and Computer-Integrated Manufacturing*, *41*, 102–114.

Waters, T. R., Putz-Anderson, V., Garg, A., & Fine, L. J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, *36*(7), 749–776.

Watson, G., Curran, R., Butterfield, J., & Craig, C. (2008). The Effect of Using Animated Work Instructions Over Text and Static Graphics When Performing a Small Scale Engineering Assembly. In R. Curran, S.-Y. Chou, & A. Trappey (Eds.), *Collaborative Product and Service Life Cycle Management for a Sustainable World*, Advanced Concurrent Engineering (pp. 541–550). Springer London. doi:10.1007/978-1-84800-972-1_51

Webel, S., Becker, M., Stricker, D., & Wuest, H. (2007). Identifying differences between CAD and physical mock-ups using AR. *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (pp. 1–2). IEEE Computer Society.

Weber, K. H. (2016). *Engineering verfahrenstechnischer Anlagen: Praxishandbuch mit Checklisten und Beispielen*. Springer-Verlag.

WHO, W. H. O., & others. (2003). Protecting Workers' Health Series no. 5, Preventing musculoskeletal disorders in the workplace, 2003.

Wiedemann, L., Planinc, R., Nemec, I., & Kampel, M. (2015). Performance evaluation of joint angles obtained by the Kinect v2. *Technologies for Active and Assisted Living (TechAAL), IET International Conference on* (pp. 1–6). IET.

Witt, H., Nicolai, T., & Kenn, H. (2006). Designing a Wearable User Interface for Hands-free Interaction in Maintenance Applications. *Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops*, PERCOMW '06 (pp. 652–655). Pisa, Italy: IEEE Computer Society. doi:10.1109/PERCOMW.2006.39

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005). Maximizing the Guessability of Symbolic Input. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05 (pp. 1869–1872). Portland, OR, USA: ACM. doi:10.1145/1056808.1057043

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined Gestures for Surface Computing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09 (pp. 1083–1092). Boston, MA, USA: ACM. doi:10.1145/1518701.1518866

Wu, G., Helm, F. C. T. van der, Veeger, H. E. J. (DirkJan), Makhsous, M., Roy, P. V., Anglin, C., Nagels, J., et al. (2005). {ISB} recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand. *Journal of Biomechanics*, *38*(5), 981–992. doi:http://dx.doi.org/10.1016/j.jbiomech.2004.05.042

Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., et al. (2002). {ISB} recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. *Journal of Biomechanics*, *35*(4), 543–548. doi:http://dx.doi.org/10.1016/S0021-9290(01)00222-6

Wu, H., & Wang, J. (2012). User-Defined Body Gestures for TV-based Applications. *Digital Home (ICDH), 2012 Fourth International Conference on* (pp. 415–420). doi:10.1109/ICDH.2012.23

Wuest, H., Engekle, T., Wientapper, F., Schmitt, F., & Keil, J. (2016). From CAD to 3D Tracking—Enhancing & Scaling Model-Based Tracking for Industrial Appliances. *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on* (pp. 346–347). IEEE.

Xu, X., & McGorry, R. W. (2015). The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. *Applied ergonomics*, *49*, 47–54.

Xu, X., McGorry, R. W., Chou, L.-S., Lin, J., & Chang, C. (2015). Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking. *Gait & Posture*, *42*(2), 145–151. doi:http://dx.doi.org/10.1016/j.gaitpost.2015.05.002

Zar, J. H. (1972). Significance testing of the Spearman rank correlation coefficient. *Journal of the American Statistical Association*, *67*(339), 578–580.

Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S., & Menegatti, E. (2015). Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications. *Multimedia and Expo (ICME), 2015 IEEE International Conference on* (pp. 1–6). IEEE.

Zhu, J., Ong, S., & Nee, A. (2014). A context-aware augmented reality system to assist the maintenance operators. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *8*(4), 293–304.