Università degli Studi di Salerno

Dipartimento di Fisica "E.R. CAIANIELLO"

Facoltà di Scienze Matematiche Fisiche e Naturali

Corso di Dottorato in Scienze e Tecnologie dell'Informazione,
dei Sistemi Complessi e dell'Ambiente – XIII ciclo

Tesi di Dottorato in

COMPUTER VISION METHODS APPLIED TO PERSON TRACKING AND IDENTIFICATION

Il Tutor:
  Prof. Andrea F. Abate

Il Coordinatore:
  Prof. Roberto Scarpa

Candidato:
  Fabio Narducci
  Matricola 8886200013

Anno Accademico 2013/2014

Declaration of Authorship

I, Fabio NARDUCCI, declare that this thesis titled, 'Computer Vision Methods applied to Person Tracking and Identification' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: *Fabio Narducci*

Date: 24/02/2015

i

"Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house"

Henri Poincaré

UNIVERSITY OF SALERNO

Abstract

Faculty of Mathematical, Physical and Natural Sciences
Sciences and Technologies of Information, Complex Systems and Environment

Doctor of Philosophy

Computer Vision Methods applied to Person Tracking and Identification

by Fabio NARDUCCI

Computer vision methods for tracking and identification of people in constrained and unconstrained environments have been widely explored in the last decades. Despite of the active research on these topics, they are still open problems for which standards and/or common guidelines have not been defined yet. Application fields of computer vision-based tracking systems are almost infinite. Nowadays, the Augmented Reality is a very active field of the research that can benefit from vision-based user's tracking to work. Being defined as the fusion of real with virtual worlds, the success of an augmented reality application is completely dependant on the efficiency of the exploited tracking method. This work of thesis covers the issues related to tracking systems in augmented reality applications proposing a comprehensive and adaptable framework for marker-based tracking and a deep formal analysis. The provided analysis makes possible to objectively assess and quantify the advantages of using augmented reality principles in heterogeneous operative contexts. Two case studies have been considered, that are the support to maintenance in an industrial environment and to electrocardiography in a typical telemedicine scenario. Advantages and drawback are provided as well as future directions of the proposed study. The second topic covered in this thesis relates to the vision-based tracking solution for unconstrained outdoor environments. In video surveillance domain, a tracker is asked to handle variations in illumination, cope with appearance changes of the tracked objects and, possibly, predict motion to better anticipate future positions.

An experimental vision-based tracker is proposed in this work. It is an hybrid software solution that fuses an Haar-like cascade classifier with an optical flow tracker to locate, track and predict the position of people in the scene. Preliminary results are also provided to highlight the performances and the problems to be addressed. The third and last subject faces with the recent (increasing) trend of adding biometric features to mobile devices. The sudden diffusion of mobile devices has led to a drastic improvement of their computing power. Nowadays, mobile phones are not only used for mere communication but they enable applications such as internetting, receiving and sending emails and storing (sensitive) documents. This opens to a wide range of potential risks for privacy. Based on this observation, an experimental proposal for ubiquitous iris recognition running on mobile devices is explored in this work of thesis. It exploits spatial histograms for iris matching offering a good trade-off between accuracy and computational demand. Preliminary results confirm the feasibility of such an approach to mobile platforms.

Acknowledgements

Contents

List of Figures

List of Tables

Abbreviations

AR	**A**ugmented **R**eality
MR	**M**ixed **R**eality
VR	**V**irtual **R**eality
HMD	**H**ead **M**ounted **D**isplay
LKT	**L**ucas **K**anade **T**omasi
FOV	**F**ield **O**f **V**iew
DOF	**D**egree **O**f **F**reedoms
CAD	**C**omputer **A**ided**D**esign
MOTA	**M**ultiple **O**bject **T**tracking **A**ccuracy
MOTP	**M**ultiple **O**bject **T**racking **P**recision
TPR	**T**rue **P**ositive **R**ate
FPR	**F**alse **P**ositive **R**ate

To Elisa
the only one who can give me genuine happiness. . .

Chapter 1

Introduction

1.1 Context and motivations

Computer vision methods for tracking and identification of people in constrained and unconstrained environments have been widely explored in the last decades. According to Moeslund and Granumby, the three major application areas for tracking of people are: control, surveillance and analysis [1]. The control area relates to applications where person tracking is used to provide functionalities like human-computer interaction or to customise the application while using. The surveillance area covers applications where one or more subjects are being tracked over time and possibly monitored for special actions. The third class of applications relates to studies of motion aimed at inferring knowledge suitable for several different purposes, e.g., attention analysis or diagnostic of orthopaedic patients.

The identification of a person interacting with computers represents another important task for automatic systems in the area of information retrieval, automatic banking, control of access to security areas, buildings, and so on [2].

Although they address separated issues, the computer vision-based methods designed for tracking or identification of a user are often problem of pattern recognition at

1

different level of difficulties. Moreover, systems designed to identify and recognise people necessarily ask for a method to detect and track them first. In video surveillance systems, the problem of detecting and tracking people is one of the most critical to address.

This work of thesis covers the aforementioned topics focusing the attention on restricted problems. It investigates on computer-vision solutions in three different domains, that are:

- performance analysis of augmented reality environments;

- iris recognition on mobile platforms;

- unconstrained outdoor video surveillance systems.

1.2 Person tracking

Computer vision-based *Person Tracking* is a wide specialised branch of an even bigger research field commonly defined as *object tracking*[3]. Applications of object tracking are incredibly numerous. Some of them are:

- motion-based recognition: human identification based on gait, automatic object detection, etc.

- automated surveillance: monitoring a scene to detect suspicious activities or unlikely events

- video indexing: automatic annotation and retrieval of videos in multimedia databases

- human-computer interaction: gesture recognition, eye gaze tracking for data input to computers, etc.

- traffic monitoring: real-time gathering of traffic statistics to direct traffic flow

- vehicle navigation: video-based path planning and obstacle avoidance capabilities

Generally speaking, a definition of tracking can be *the problem of estimating the trajectory of an object in the image plane as it moves around a scene.* In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Going deeper and depending on the domain, a tracker can also provide object centric information, such as orientation, area, or shape of an object. This information can be used to estimate the current pose of an object or, especially in the domain of mixed reality, to infer the point observed in a real environment.

At different level of complexity, every computer vision-based tracking method is split in two main steps:

1. build some model of what has to be tracked;

2. use what is known about where the object was in the previous frame(s) to make predictions about the current frame and restrict the search;

3. repeat the two subtasks, possibly updating the model.

When facing with the problem of object tracking by computer vision, a big amount of factors can increase the complexity of the proposed solution. From a technical point of view, this domain is rich and challenging because of the need to segment rapidly changing scenes in natural environments involving non-rigid motion and (self) occlusion. Loss of information caused by projection of 3D world on 2D image, noise in images, complex object shapes/motion, real-time processing area are also requirements (rather impossible to list all of them) that make object tracking difficult.

Application fields of computer vision-based tracking systems are almost infinite. Gavrila [4] in 1998 surveyed the possible domains for vision-based tracking systems,

TABLE 1.1: Application domain for people tracking.

Virtual reality	Interactive virtual worlds
	Games
	Virtual studios
	Character animation
	Teleconferencing
	(e.g., film, advertising, home-use)
Smart surveillance systems	Access control
	Parking lots
	Supermarkets, department stores
	Vending machines, ATMs
	Traffic
Advanced user interfaces	Social interfaces
	Sign-language translation
	Gesture driven control
	Signaling in high-noise environments
	(airports, factories)
Motion analysis	Content-based indexing of sports video footage
	Personalized training in golf, tennis, etc.
	Choreography of dance and ballet
	Clinical studies of orthopedic patients
Model-based coding	Very low bit-rate video compression

grouped in table 1.1. At the time of the above mentioned survey, the augmented reality was not particularly studied and only experimental proposals were present in literature. Nowadays, the augmented reality is a very active field of the research and can be considered fully integrated in the application domains that exploit user tracking to work. Indeed, the tracking system is the essence of any augmented reality application. Being defined as the fusion of real with virtual worlds, the success of an augmented reality application is completely dependant on the efficiency of the exploited tracking method.

Demonstrating that a tracking method performs successfully is another open problem. Literature is full of possible measures for evaluating the tracking performance, typically with the comparison against ground truth, considering the target presence and position. These methods often require a considerable amount of annotation, with the consequence that the amount of videos with ground truth is often limited

up to this point [5]. Anyway, no standards or common guidelines exist in this context. This makes also difficult comparing existing solutions between each other in order to infer quantitative and objective evaluations about strengths and weaknesses of the compared methods.

Based on this point of view, the first topic discussed in this work of thesis covers the issues related to the **tracking systems in augmented reality applications** and how to quantitatively analyse them. Formal analyses of augmented reality applications are made difficult by the fact that factors like ergonomic, interaction paradigm and others significantly impact on the overall performance of the proposed approach. Therefore, becomes hard to isolate the contribution of each factor to the final conclusions. The work presented in this study designs and discusses a possible framework for a formal analysis of augmented reality application that makes possible to objectively assess and quantify the advantages of using AR principles in a specific operative context.

Another topic covered in this work of thesis relates to the **tracking issues in video surveillance system**s. Video surveillance is an attractive research field for vision-based tracking methods. As long as the assumptions constrain the possible conditions of the system, it could be relatively easy to ensure a reliable vision-based tracker. Unfortunately, especially in outdoor environments, the issues to address related to the tracking of people is really challenging due to the weak constraints to be assumed. Literature presents a big amount of proposals but tracking people *in the wild* scenarios still remains an open problem. The work presented in this thesis relies on an hybrid method that combines an Haar-like cascade classifier and optical flow trackers to address the problem of tracking people in unconstrained outdoor environments.

1.3   Biometric identification on mobile

The worldwide diffusion of latest generations mobile devices, namely smartphones and tablets, represents the technological premise to a new wave of applications for which reliable owner identification is becoming a key requirement. Mobile devices, mobile phones, PDAs etc. are considered as an essential tool in human being's everyday life. They are not only used for mere communication such as calling or sending text messages; however, these devices are also used in applications such as internetting, receiving and sending emails and storing (sensitive) documents. [6]. Security can be approached by means of biometrics (face, iris or fingerprint) by exploiting high resolution imaging sensors typically built-in on this class of devices, possibly resulting in a ubiquitous platform to verify owner identity during any kind of transaction involving the exchange of sensible data. Face is probably the first biometric proposed for person-authentication on mobile devices (Song-yi et al.[7]). Fingerprints [8] have been also proposed to improve security of mobile devices. Biometric characteristics represent an interesting choice in mobile device context because they have the advantage that, unlike passwords, PINs, tokens etc., they cannot be stolen or forgotten and they do not need to be typed, thus excluding the risk to be hacked. Among the aforementioned biometrics, iris is known for its inherent invariance and accuracy. The iris has been proposed as a highly reliable and stable biometric identifier for person authentication/recognition about two decades ago. Since then, most work in the field has been focused on segmentation and matching algorithms able to work on pictures of whole face or eye region typically captured at close distance, while preserving recognition accuracy. However, only a few works explored this topic on mobile devices.

The work presented in this thesis aims at proposing a comprehensive vision-based method for iris authentication on mobiles by means of spatial histograms. *Spatial histograms* consist of marginal distributions of an image over local patches and they can preserve texture and shape information of an object simultaneously. Based on

this observation, an experimental application on mobile platform (Android operating system) has been proposed and discussed in this thesis. Being a comparison of histograms, the proposed recognition method based on spatial histogram matching asks for a very low computational power. Thanks to its low computing load, it results particularly suited to iris recognition applications on mobile devices. Preliminary results confirm the validity of the proposal.

1.4   Organisation of the thesis

Covering different topics and in order to help the reader, here an overall view of the contents of the document is provided. The current work of thesis consists of six chapters, that are organised as follow:

▶ Chapters 2 3 4 relate to the tracking of the user's point of view in Augmented Reality applications. Respectively, the first chapter is an introduction to the issues and describes a proposal for a cheap and efficient marker-based tracking method. The second chapter focuses the attention on two case studies for which a prototype AR application was designed and developed. The third chapter presents a comprehensive and exhaustive formal analysis of advantages and disadvantages of using the prototype systems in the selected case studies. Conclusions and future directions are provided.

▶ Chapter 5 relates to a novel approach to iris recognition by means of spatial histograms running on mobile platforms.

▶ Chapter 6 describes a contribution to an outdoor unconstrained video surveillance system, in development at the SociaLab of the *Universidade da Beira Interior* in *Covilhã* (Portugal). In particular, a computer-vision tracking is proposed, illustrating the method and the preliminary results achieved.

Chapter 2

Tracking in Augmented Reality

2.1 Introduction

In this chapter the problems related to the user's tracking in augmented reality applications are discussed. It is important to specify that this work of thesis does not pretend to carefully cover the principles and the approaches of Augmented Reality (AR) nowadays available in literature. AR has become really popular over the years and a survey of current proposals and the state-of-the-art is not one of the main focuses of this work. For this reason, the following section 2.2 serves just as an introduction to the proposed study. The tracker will be carefully described in section 2.5. For a deep formal experimentation aimed at demonstrating the applicability of the proposed approach in different contexts, refer to Chapter 4.

2.2 Augmented Reality systems

The most reliable definition of Augmented Reality dates back to the 1997 by Azuma et al. who say that *"Augmented reality (AR) is a variation of virtual environments*

8

(VE), or virtual reality as it is more commonly called. VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects coexisted in the same space"[9].

Behind this definition there are, indeed, many issues to address that make the augmented reality a very hard solution to be implemented. Real time rendering, visualisation devices, tracking systems, occlusion, freedom of movements in wide 3D space, wired vs wireless technology are just a small set of all hardware and software aspects that impact on the success of a solution of augmented reality for a specific purpose. Moreover, ergonomics and user friendliness are other two typical aspects that influence the design of any AR applications. The application field also introduces other constraints that imposes particular design choices rather than others. Trying to summarize the main components that are crucial for the proper usage of a solution based on augmented reality, these are the following:

1. the *tracking method*: it is in charge of tracking the point of view of the user in the real space;

2. the *visualisation device*: that allows the user to see the augmenting virtual contents;

3. the *interaction paradigm*: that allows the user to interact with the application.

Augmented reality has many applications. First used for military, industrial, and medical applications, it has also been applied to commercial and entertainment areas. Archaeology [10] Architecture [11], Education [12], Gaming [13] and much more have all studied over the years and several approaches have been proposed.

As anticipated, the visualisation device represents a crucial factor in every augmented reality application. The user should see the real world (with the virtual

contents added to the scene) as much as possible similar to the way he/she sees through the eyes. In this context, a significant role is covered by the Head Mounted Display (HMD). In figure 2.1 have been shown the state-of-the-art in terms of HMDs available in the market or in developing stage. Compared to old devices, nowadays they are able to provide a very wide field of view, little weight and a comfortable experience. Most of them feature CPU and storage resources that make them perfect for unconstrained augmented reality solutions. Being a mere technological aspect, in this work of thesis they only have been collected and briefly illustrated. Ergonomics aspects aside, the technical specifications offered by the HMD do not significantly affect the experimental study proposed in Chapter 4. For this reason here it is not provided a focused comparison between them nor a discussion about the strengths and weaknesses of each.

Realtime-rendering is nowadays made feasible by high performance CPU and

FIGURE 2.1: A collection of some of the most recent HMD available.

graphic card that are able to render virtual objects with an excellent level of realism. For this reason, it is no more considered as a challenging aspect like happened in the past. To this regard, it is important to make a clarification about the concept of Augmented Reality and Mixed Reality (MR). Although they are often used like

synonyms of each other, they relate to slight different contexts. When facing with Augmented Reality, the challenge is to properly track the user's point of view so that the real scene can be augmented with virtual information (like labels, arrow, signals and so on). In Mixed Reality, the goal is to render virtual objects that co-exist with the real scene in such a way that they are indistinguishable from each other. In this perspective, the computational power becomes a challenging aspect that has seriously taken into account while designing an MR application. Since this study relates to augmented reality, the processing power has been considered as a negligible factor.

Conversely, the interaction paradigm is a particularly sensitive topic in augmented reality[14]. In typical working environments, the user is free to move in a varyingly wide space. In such scenario, common I/O devices, i.e., keyboard and mouse, do not provide a sufficient level of freedom of movement to the user. New paradigms have to be introduced when thinking at this aspect. In this work, as described in the proper sections in chapter 3, different approaches were implemented. They consist in optical finger-based interaction, gesture recognition as well as speech recognition and synthesis.

Since tracking and registration are the most critical issues for AR applications[15], this study started exactly from this aspect in order to quantitatively and qualitatively demonstrate if augmented reality solutions, featuring optical marker-based tracker, can be efficiently adopted in operative scenarios. The sections that follow explain in detail the optical tracking method developed during this work of thesis. Details about the case study, the scenarios and the developed prototype are provided in Chapter 3. Chapter 4 is completely focused on the experimentations conducted in reasonable operative scenarios.

2.2.1   Tracking methods and technologies

Accurate registration and tracking between computer-generated objects and real world objects are the key challenges for developing an augmented reality application. When a user moves his/her position/viewpoints, the virtual objects must remain aligned with the position and orientation of real objects. According to a categorization proposed by Zhou et al.[16], augmented reality tracking techniques can be:

- *sensor based*: sensor based tracking techniques are based on sensors that are placed in an environment. Active sensors are used in sensor based tracking, which are then used to track the position of camera movement. Sensors may be optical, magnetic, inertial, acoustic or ultrasonic. Each sensor has its own advantages and disadvantages. The selection of a sensor depends on different factors including accuracy, calibration, cost, environmental temperature and pressure, range and resolution.

- *vision based*: vision based tracking techniques used image information to track the position and orientation of a camera. In vision based tracking, computer vision methods are used to calculate the camera pose relative to the real world objects. Early vision based tracking used fiducial markers in prepared environments. Currently, vision based tracking research is based on marker-less augmented reality for tracking the proper physical positions.

- *hybrid*: Azuma et al. [9] described that none of existing techniques gives complete solution for outdoor tracking, so hybrid tracking techniques have been developed which combine several technologies.

Rabbi et al. presented a very exhaustive survey about the possible tracking technologies and methods used in augmented reality [17]. Despite the advances made in augmented reality tracking, there are still some issues in this area that need to be

overcome. The problem is involved in the modelling of complex 3D spatial models and the organization of storage and querying of such data in spatial databases as these databases need to change quite rapidly with real environment as real environments are dynamic. Similarly, drastic motions often lead to tracking failure and recovery is time consuming with a temporary loss of real-time tracking abilities. Currently, no single tracking technique provides best solution for the orientation and pose tracking in an unprepared environment. Sensor-based tracking sensors (i.e., optical, magnetic, inertial, acoustic or ultrasonic) are often used. Optical tracking sensors are sensitive to optical noise, occlusion, computationally very costly and slow. Tracking is difficult in optical domain when multiple similar objects are present in the scene. The magnetic sensors are disturbed in the presence of electronic devices nearby. Magnetic tracking sensors also suffer in terms of jitter, accuracy degrades when their distance increases from the source and are sensitive to electromagnetic noise. The update rate for acoustic system is low as sound travels relatively slow. The speed of sound in air can change due to the change of temperature or humidity in the environment, which can affect the efficiency of the tracking system. In inertial tracking system, problems can occur due to small friction between the axis of wheel and bearing.

Marker based tracking can be used in indoor prepared environment. Due to the limited range, marker-based tracking cannot be used in large scale outdoor environments. Model-based tracking extends the tracking range by using natural features for tracking but these techniques lack robustness and have high computational cost. Model-based tracking requires heavy computation for generating models of complex environment.

Hybrid tracking provides a robust and accurate tracking but it is costly and involves computational difficulties. Providing a fast and accurate tracking, with minimal efforts in setup and low cost, is the main challenge in augmented reality systems. Tracking in large factory that satisfies industrial demands is also another typical issue as well as continuous tracking in wide-area workplace.

TABLE 2.1: Comparison of different tracking sensors [17].

Sensor Tracking	Accuracy	Sensitivity	Cost	DOF	Advantages	Disadvantages
Optical Sensors	Accurate	Light	Cheap	3/6DOF	High update rate and resolution	optical noise and occlusions
Magnetic Sensors	Less accurate	electromagnetic noise	Chep	6DOF	High update rate no occlusions	Small working volume distance affects accuracy
Acoustic Sensors	Less Accurate	Temperature humidity pressure	Cheap	3/6DOF	No distorion	Slow occlusion Ultrasonic noise
Inertial sensors	Accurate	Friction	Cheap	1/3 DOF	No reference No prepared environment	conservation error
Hybrid techniques	Accurate	Depend on sensor used	Costly	6DOF	Accurate and stable	Depend on sensor used

According to what Rabbi et al. surveyed, it results that vision-based tracking is the most active area of research in AR [16]. It benefits from the advantages of not requiring expensive sensors to track the user's point of view. Since 2000, improvements in pattern-based tracking have opened to new paradigm of AR systems based on the usage of fiducial markers to detect and track the point of view of the user acquired by a camera. Hirokazu Kato, at the end of 1999, presented a complete framework to design and develop augmented reality applications called *ARToolkit* [18]. Since then, tracking approaches based on it and further improvements of the framework, have been increasingly studied. In fact, it is possible to find in literature a very large variety of proposals that are all based on the same idea of inferring the position of the acquiring camera by detecting and recognising fiducial markers in the scene. The approach proposed in this study relies on the just mention strategy of using fiducial marker to augment the scene observed by the user. The focus was on computer aiding system in augmented reality of which the section that follows immediately after offers a general overview.

2.3 Computer-based aiding systems

Computer-based aiding systems supporting decisions and comprehension of problems have been studies since 1970s [19, 20]. Computer Aided Design(CAD) systems

represent for sure the first attempts to help the work of designers by consciously using the limited computational power of computers available at that time. They ran on very expensive workstations placing *the engineer into a subordinate position by the mere fact that his time is less valuable than that of a computer*[21]. Nowadays, the computational power of CPUs and video cards has incredibly grown making possible to render complex 3D models on real time with an impressive level of realism. Over the years, virtual reality has gained an increasing attention due to its infinite possibilities of simulating virtual worlds. The simulations can be so accurate to be almost indistinguishable from the reality thus implying the applicability of such methods to a broad range of applications. Literature is full of proposals in the field of design and assembly [22–24], maintenance [25, 26], learning [27], medical sciences [28–30] and for therapy [31, 32] just to mention a few. A significant implication of virtual reality applications is the fact that the user loses any link to the real world. Although this could be consider a negligible aspect, the truth is that it significantly affects the efficacy of exploiting them. While learning something in a virtual environment, the trainee does not realise sizes and weights of simulated objects as well as the real feeling of working with them. Haptic sciences have been proposed to solve this issue introducing the sense of touch in virtual simulations. The user can experience the physical contact with virtual objects receiving feedbacks from the virtual world that are not just visual. In such context, industrial and medical field are the most studied and investigated over the recent years. Khan et al. [33] presented a quantitative and qualitative study of using haptics in industrial environments while in [34, 35] have been proposed how to effectively reproduce haptic feedbacks in such research area. Interacting and controlling robots with haptic feedbacks is another interesting research field [36–38]. As just mentioned, medical simulated environments [39, 40] are also an attractive field of research for haptic sciences. Simulation of clinical breast [41], elbow spasticity [42] and clinical evaluation of colonoscopy [41] are just few examples of proposals available in the literature. Haptic devices have been considerably improved over the years but they suffer for the fact that human skin

receptors have an incredible high sensitivity to vibrations that are almost impossible to simulate.

Although virtual reality and haptic simulations represent effective solutions, which address diverse critical problems, they can not entirely simulate the feelings that a user experiences when interacting with real worlds and objects. This created the right conditions to introduce the augmented reality as the technology to assist and train users in several real working environments without asking him/her to wear devices aimed at reproducing feelings of real world. Education and training are stimulant research areas where augmented reality has been demonstrated helpful. *In a rapidly changing society where there is a great deal of available information and knowledge, adopting and applying information at the right time and right place is needed to main efficiency in both school and business settings.* (Kangdon Lee [43]). Santos et al. [44] surveyed 87 research articles covering the topics of augmented reality learning experiences designed for pre-school, grade school and high school education. They proposed a quantitative study of student performances based on an effect size analysis(a quantitative meta-analysis method that measures the strength of a phenomenon.[45]). By their analysis they obtained a mean effect size of 0.56 to student performance. Although the result is rather limited, it must be taken into account the variability of possible ways to use AR, as well as, differences in experimental design. Quantitative analysis of augmented reality system is rather difficult to perform due to the big number of variables to consider. This make also difficult comparing own proposals with others in literature. Chapter 3 discusses in detail this topic introducing and critically analysing the method proposed in this work of thesis. Mau-Tsuen et al. [46] dealt with topics of physical-virtual immersion and real-time interaction in augmented reality learning environments. Human-computer communication in augmented reality often can not rely on common I/O devices (e.g., mouse and keyboard). When the assumption is that the user is free to move in a working environment (and not sitting in front of a desktop platform) new interaction

paradigms become necessary. During the work of thesis, this topic has been also covered by investigating on computer vision methods to resolve real/virtual occlusions by hands [47] and providing finger tips detection and recognition to interact with AR systems [48].

2.4 Tracker Intersense IS900

Before introducing the vision-based tracker proposed in this study, it is interesting to briefly discuss the tracker Intersence IS900. It is an ultrasonic tracking system consisting of some bars (called SonyStrip) that feature some ultrasonic microphones (called SonyDisc). The Sonidisc are able to record the sounds transmitted by one or more sensors worn by the user. By doing this, the system can track simultaneously many body parts like the head, hands and feet. It is particularly suited for motion capture and it offers a wide volume of coverage (which can be enlarged by adding extra bars). In figure 2.2 is depicted the typical setting of this kind of system. At
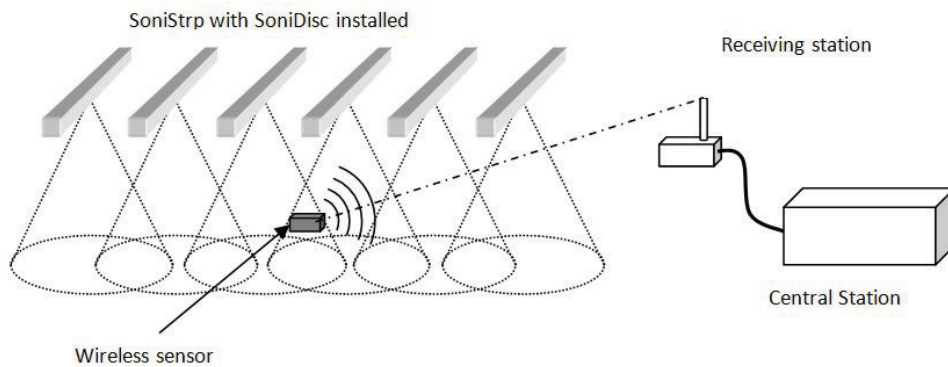
FIGURE 2.2: A typical setup of a IS900 tracking system.

first stage of the work presented in this thesis, it was considered a very good solution for the considered application domains (more in industrial field rather than in

telemedicine). Unfortunately, first attempts to use it for augmented reality application revealed critical technological problems of such a tracking system. It must be said that this system was designed and engineered for virtual reality applications. This non-negligible consideration has a significant impact on the possibility of exploiting it for augmented reality purposes. In a typical VR scenario, the user loses any reference with the real world. This implies that, even if the system produces a little error in estimating the pose, the user can not realise the misalignment between his/her current position and the position estimated by the system. In augmented reality, any tracking error (even if really short) implies an improper visualisation of the augmenting contents. Consequently, the user perceives the misalignment so vanishing the effect that virtual and real co-exist in the same space, which is the main goal of the augmented reality. During the preliminary test, the tracker IS900 has revealed to be non-sufficiently robust for the augmented reality requirements. Tracking error of few centimeters in position estimation as well as 1 degree in rotation estimation produced really poor results. In augmented reality scenarios, errors in rotational estimation have a drammatic impact on the augmentation of the scene. Looking at figure 2.3 and considering just $\alpha = 1$ degree of error and a distance from the observed (and augmented) point of $d' = 0.50 metres$ and $d'' = 1 metres$, the position of the augmented point ($d \tan(\alpha)$) appears shifted of 8 millimetres in the first case and of 1.7 centimetres in the second case. An error of 8 millimetres can be accepted when augmenting an outdoor environment where the level of precision is not very strict. However, when the object to be augmented is full of features that are really close to each other (like a panel of a device that is full of led indicators and buttons) is not ensured that an error of 8 millimetres can be tolerated. Those were the reasons that led to investigate on other tracking solutions, cheaper and non strictly sensor-dependant. Taking into account the trends of augmented reality of recent years [16], the immediate choice was investigating on computer vision methods that have the advantage of being supported by the most recent approaches and algorithms, are customisable, versatile and, when carefully designed, easy to scale.
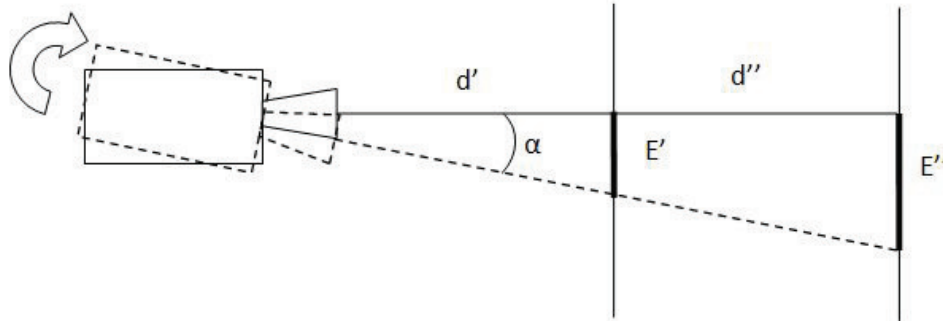
FIGURE 2.3: The effect of error rotation estimation.

2.5 The proposed vision-based tracker

The vision-based tracking system proposed in this study was built around a known C library for marker-based tracking called ARToolkit [49]. ARToolKit uses computer vision techniques to calculate the real camera viewpoint relative to a real world marker. It grants a high and real-time precise tracking even in poor light conditions (by a proper setting of parameters). As for any other similar library, it is able to track the user's point of view until the camera correctly frames at least one marker. If one marker is only partially framed, the system will not recognise it involving a tracking fault. However, even problems like strong reflections, high camera noise, extreme shadows and so on could determine a failure. To overcome these limitations it is sufficient that the markers are printed on a low reflecting surface and avoiding too bright perpendicular lights over them. By this way the tracking system results more reliable. A marker (see figure 2.4) in ARToolkit is a special pattern (simply printed on paper by common ink-jet printers) that is used to infer the position of the camera in the real world.

A marker has its own specific rules that are:

1. a perfect square;
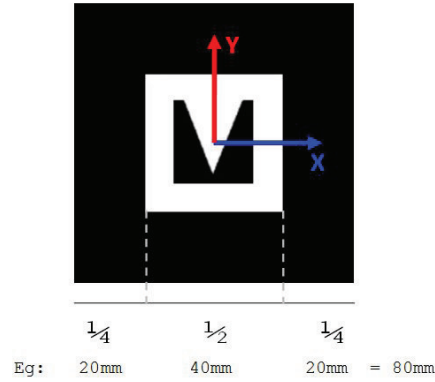
FIGURE 2.4: A typical marker of Artoolkit and its properties.
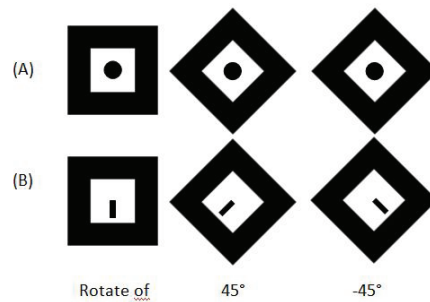
FIGURE 2.5: Ambiguity of pose estimation in presence of symmetric markers. The marker with the dark circle is completely symmetric making impossible to estimate its current rotation (A). A non-symmetric marker ensure a proper estimation of the pose regardless the perspective angle of view.

2. high contrast of the contours;

3. the pattern has to be asymmetric in order to prevent ambiguous pose estimation (see figure 2.5);

4. if using more than one marker, the similarity between each possible pair of them should be close to zero.

Once a marker is detected, the problem of projecting the position and rotation of the marker in world coordinates to camera coordinates has to be solved. The

transformation matrix T_{cm} that represents the relationship between the marker $\mathbf{M}$ and the camera $\mathbf{C}$ can be expressed as:

$$
\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_M \\ Y_M \\ Z_M \\ 1 \end{bmatrix} = T_{CM} \begin{bmatrix} X_M \\ Y_M \\ Z_M \\ 1 \end{bmatrix}
$$

Generally speaking, after thresholding of the input image, the regions whose outline contour can be fitted by four line segments are extracted. Parameters of these four line segments and coordinates of the four vertices of the regions found from the intersections of the line segments are stored for later processes. The regions are normalized and the sub-image within the region is compared by template matching with patterns that were given the system before to identify specific user ID markers. The detected region, in camera coordinates, represents the location of the marker in the 2D video frame. The problem is mapping these coordinates to the world coordinates and it is done by calculating the transformation matrix T_{cm}. The problem of getting the transformation matrix T_{cm} is achieved by *a priori* knowledge about the distortion factors of the camera and its intrinsic and extrinsic parameters (obtained by calibration algorithms, see AppendixB B).

The camera parameter consists of a perspective projection matrix and distortion parameters. The perspective projection matrix consists the following matrix:

$$
\begin{bmatrix} f & s & x & 0 \\ 0 & af & y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
$$

where of f is the field of view of the camera, a is the aspect ratio, s is the skew factor and the vector (x, y) represents the image center.
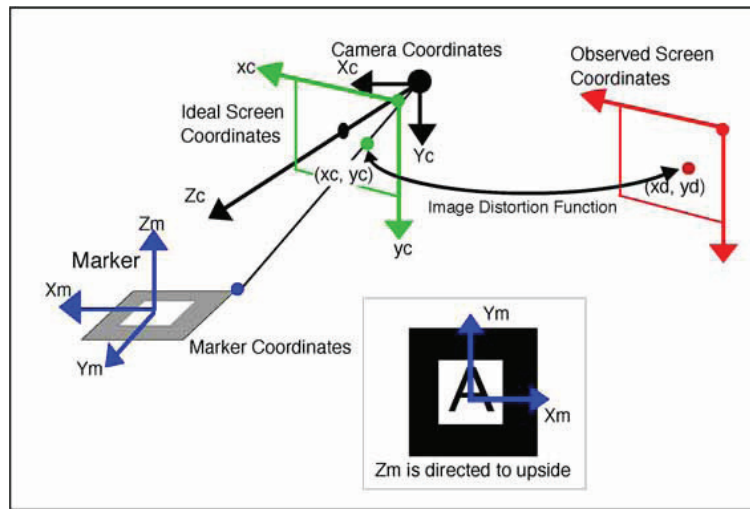
FIGURE 2.6: The coordinates system of ARToolkit.

2.5.1 The workflow

The workflow of any ARToolkit-based tracking system (and for any marker-based tracker in general) consists of the following steps (see figure 2.7):

1. First the live video image is turned into a binary (black or white) image based on a lighting threshold value. This image is then searched for square regions.

2. ARToolKit finds all the squares in the binary image, many of which are not the tracking markers.

3. For each square, the pattern inside the square is captured and matched again some pre-trained pattern templates.

4. If there is a match, then ARToolKit has found one of the AR tracking markers. ARToolKit then uses the known square size and pattern orientation to calculate the position of the real video camera relative to the physical marker.

5. A 3x4 matrix is filled in with the video camera real world coordinates relative to the card (see figure 2.6. For details, refer to Appendix A). This matrix

is then used to set the position of the virtual camera coordinates. Since the virtual and real camera coordinates are the same, the computer graphics that are drawn precisely overlay the real marker.
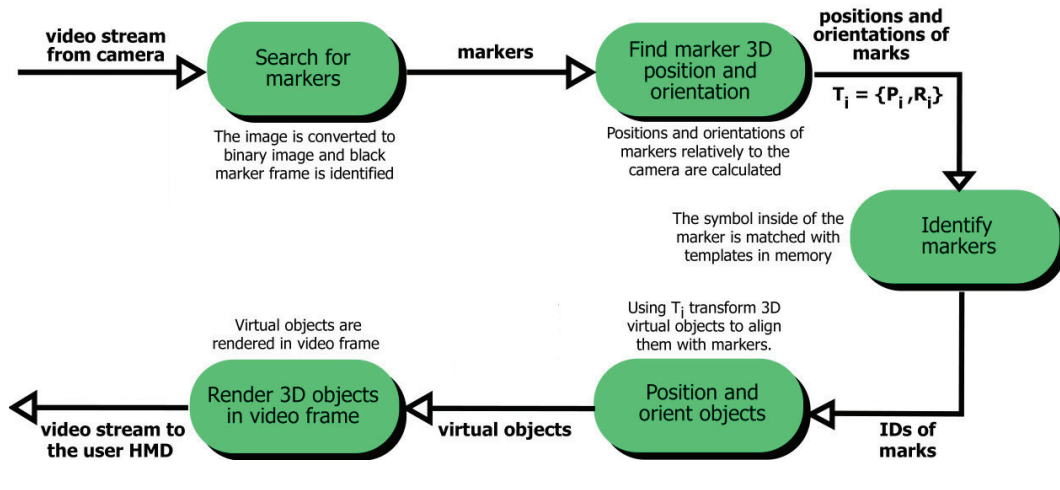
FIGURE 2.7: The ARToolkit workflow.

2.5.2 The multi marker tracking

When the tracking is based on a single marker, its active tracking volume results very narrow. Moreover, if the tracking reference point is completely balanced on a single marker, the marker itself should be big enough to grant a good detection and it must stay always visible in the field of view of the camera. In a desktop environment, where the user does not make ample movements, it could be considered feasible using a single marker. However, if the user needs to interact with the augmented environment, the probability of occlusions of the markers by user's hands is really high. This is the main reason that led this work of thesis to move to a multi-marker approach where the user's tracking does not exclusively depend on a single marker but on two or more of them. This solution ensures a wider and scalable coverage tracking volume. A set of different markers defines only a single 3D point (see figure 2.8). The relative disposition of each marker from the shared

point is well-known. So, when at least one single marker of the entire set is framed, the system is able to track the user. The multi-marker approach also impacts on the tracking accuracy. When the system tracks only a marker the augmenting point could result very unstable. But, if more than one marker contribute to the computation of a single augmenting point, this one will result more stable. That is due to the fact that the estimation error computed on a single marker is averaged over the estimations made for all correctly detected marker at each time. In this manner, the average error results lower than the error on each single marker, thus making the tracker more stable and accurate also reducing the effects of jittering. To ensure a good quality of the tracking it is important that the cameras and the virtual counterparts are correctly co-registered. A prior camera calibration (see appendix B B) routine is provided by the Artoolkit and generally it is sufficient to prevent most of the tracking inefficiency due to the camera lens distortions. However it could happen that real and augmented objects do not appear correctly aligned even after the co-registration step. For this reason the proposed system gives the user the possibility to manipulate the tracking values by means of a specific calibration interface. It allows the user to achieve a more accurate co-registration between the real camera and its virtual counterpart in charge of render the required graphics. It is possible to manage the tracking values by adding displacements to the position/rotation tracking values or changing the virtual camera FOV as well as to modify the marker detection threshold. This task is done only at the first system start-up unless strong changes in lighting which would lead to a different precision's detection. As for other low level system functions like this, the corresponding GUI is accessible by mouse and keyboard.

During the experimentations (carefully described in Chapter 4)the adopted multi-marker solution resulted very accurate with regard to positions. Even at long distances from the camera the computation of the position of any marker resulted very stable. Instead for the rotation's computation a significant amount of tracking failures have been observed. It has been observed that tracking failures happens when

a rapid increasing in rotational velocity of the head is registered. Rotational velocity of the head (and the related acceleration) can be really high instead of the translational velocities. However, high rotational speed happen for a very short amount of time sometimes below the tenths of seconds. Although any consumer camera does not provide sufficient refresh rate to face with such an issue, the really short duration of such event makes possible to prevent the tracking failure by predicting the possible position by a simple and fast damping function. The damping function provides to damp the stronger peaks in rotation and also in the estimation of the position when a tracking failure occurs. Moreover, it also reduce the phenomena of jitter and vibrations that are typical in marker-based tracking because related to the image noise and distortions:

$$f(x') = \left\{ \begin{array}{ll} x - k(|x - x'|) & if\, x > x' \\ x + k(|x' - x|) & otherwise \end{array} \right\}$$

where variable $k \in [0, 1]$ is a threshold that acts as damping factor. It has to be choose carefully because if it is too close to zero the function would produce a high perceived tracking latency.

2.5.3 Tracking issues

A well known problem of vision tracking is the control of lighting. Changing levels of light and limited contrasts disable correct registration. The ARToolKit requires relatively large black surfaces, which printed out with some laser printers tend to reflect the light, giving highlights in the video image. Anyway, this problem can be overcome by printing the markers on special surface that are able to reduce the problems of reflection of light. The problem of occlusion is considerably harder to face with. Even though the method proposed in this work of thesis exploits an approach based on multi-marker set, it does not entirely resolve the problem of occlusions. Tags need to be fully within camera view to be detected. Thus occlusion by the
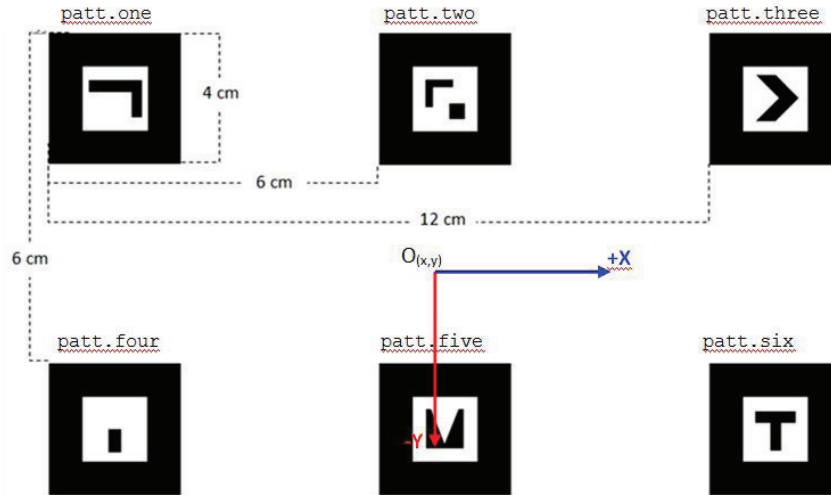
FIGURE 2.8: A possible configuration of multi-marker set. Each markers contribute the localisation of a common origin $O_{(x,y)}$ in marker coordinates.

interactors' bodies or by stacking objects makes tagged objects virtually absent. If one marker overlaps with another marker, the overlapped one will not be detected. Furthermore the camera field determines the interaction space and limits it. Marker tags need to be visible and thus may interfere with aesthetics and intuitive legibility of objects. The looks of a tangible interface, "simulated" with tags, often differ from the intended product and may distract users and evaluators from the general idea. The required tag size (for detection) also limits the smallness of registered objects. And over time tags will deter, fade or get dirty, thus endangering long term usage of tags (this is less of a problem when prototyping). While for large devices optical markers and the camera may be hidden inside the device, this is not possible for small devices (e.g. a handheld with many buttons). A disadvantage of using the ARToolKit is that it only eases the registration process and the creation of events. Interpretation of events and output of interaction (system response) must be carefully implemented in standard programming languages, requiring some programming experience. Necessary is also programming of basic position calculations.

As it is used mostly for Augmented Real-ity, the toolkit eases detection of markers and overlaying an image at the appropriate point in an AR display.

Camera lenses and distortion also represent another critical factor that impact on the quality of the tracking provided. Using a camera with a narrow FOV ensures that distortions in peripheral region of the image are relatively negligible but, conversely, it does not grant that a sufficient amount of markers are visible in the video frame. On the other hand, using wide-view cameras resolve the problem of the visibility of the marker but wide-camera are affected, in general, by high distortion factors. The resolution also play a key role. By using high resolution cameras it is possible to significantly reduce the size of the markers in the scene but at the cost of a higher computational cost needed to process the frame grabbed from the camera. While designing the marker-based tracker presented in this work of thesis, all this factors have been taken into account. Experiment results (discussed in section 4.1) demonstrate that, though the issues related to the marker-based trackers, augmented reality systems with vision-based user's tracking are indeed a feasible solution that can significantly improve the performance of users in erogenous fields.

Chapter 3

Case studies

3.1 Outline

Augmented reality has been exploited and demonstrated to be feasible in several application contexts. In this thesis, two different fields are taken into account with the main intention of assessing the adaptability of the tracking method proposed in 2.5.1. The first application field relates to a typical industrial scenario where a non-skilled technician tries to maintain a complex industrial rack assisted by virtual aids provided in augmented reality. The second scenario refers to telemedicine research field where an augmented reality system was built around the proposed tracker and used to support untrained users in performing an ECG test. Since the work of thesis is focused on the tracking method, the benefits of exploiting augmented reality will be evaluated in terms of accuracy of the tracker and the efficacy of completing some tasks in a correct manner. Aspects strictly related to ergonomics or user-friendliness of the systems are not considered, even if they play an important role in terms of acceptability in operational fields. For these reasons, only short subjective evaluation about the proposed prototype is shown in this section. Conversely, a deep investigation about the tracking system is discussed in Chapter 4. It provides a critical

discussion about quantitative and qualitative evaluations thus showing advantages
and drawbacks of using an optical image-based tracker in augmented reality appli-
cations.

Following sections introduce the case studies considered during this work. An
overview is therefore proposed as well as a brief description of the architecture of
the systems. At the end of each of them, a short subjective evaluation is presented.

3.2 Servicing and repair for industrial equipment

Maintenance and repair operations represent an interesting and opportunity-filled
problem domain for the application of augmented reality. The majority of activities
in this domain are conducted by trained maintenance personnel applying established
procedures to documented designs in relatively static and predictable environments.
These procedures are typically organized into sequences of quantifiable tasks target-
ing a particular item in a specific location. These characteristics and others form a
well-defined design space, conducive to a variety of systems and technologies that
could assist a mechanic in performing maintenance.

In such a context, here is described how to exploit the optical tracker (discussed
in section 2.5) in the design of augmented reality applications tailored to industrial
requirements. The intention of the proposed prototype is assisting an untrained
technician in reducing the effort of performing large sequences of maintenance tasks
in complex systems. The proposed system exploits AR to enhance items localization
in ordinary maintenance sequences by means of context-based instructions, visual
aids, virtual labels, 2D and 3D graphics, and animated virtual tools (figure 3.1 shows
an example of augmented scene). The information is transformed to be visualized
onto the real world exploiting a tracked head-mounted display. Figure 3.1 shows an
industrial rack with some virtual augmenting visual aids that help the technician to

understand the embedded equipment and how to interact with them during ordinary maintenance procedures.

FIGURE 3.1: An example of augmenting an industrial equipment with a virtual screwdriver that suggest the screws to be unmounted during a specific step of the maintenance procedure.

3.2.1 Related works

Scientific literature presents a big number of studies covering the topic of augmented reality applied to industrial contexts. In 2002, the project ARVIKA [50] fostered the research and the development of AR technologies for production and service in the automotive and aerospace industries, for power/process plants, for machine tools and production gear. Klinker et al. [51] presented an AR system for the inspection of power plants at Framatome ANP, while Shimoda et al. [52] presented an AR solution in order to improve efficiency in nuclear power plants (NPP) maintenance interventions and to reduce the risk of human error. Mendez et al. [53] developed a virtual (AR based) filter to reveal hidden information that is behind an occluding object, to enhance data of interest or to suppress distracting information. Pentenrieder et al. [54] showed how to use AR in automotive design and manufacturing,

to analyze interfering edges, plan production lines and workshops, compare variance and verify parts. Still et al. [55] proposed an augmented reality system for aiding field workers of utility companies in outdoor tasks such as maintenance, planning or surveying of underground infrastructure, exploiting geographical information system. More recently, De Crescenzio et al. [56] described AR based interfaces as valuable tools in preventing manufacture errors in the aviation field. Whatever the context considered, tracking precisely and reliably the user point of view (POV) with respect to six degrees of freedom is of paramount importance for co-registering virtual objects with the surrounding environment. Over the years different technologies have been proposed for this purpose (magnetic, ultrasonic, inertial, computer vision based, hybrid, etc.), each with advantages and disadvantage. However, to this date, none of them can be considered as a general solution, whereas each approach can be suited to a particular domain (indoor/outdoor usage, small/wide/scalable operating volume, presence/absence of ferromagnetic materials or electromagnetic fields, etc.). Computer vision, in both the marker-based [52, 57] and marker-less [58, 59] variants, is generally recognized as the only tracking methodology enabling non-invasive, accurate and low cost co-registration between virtual and real [60]. For this reason, the AR system used as test bed described in this study exploits a multi-marker tracking method that resulted capable to deliver highly accurate and robust tracking on the target environment featuring mission critical rack sized equipment rich of small components and operating in the vicinity of strong electromagnetic fields. Another relevant aspect to be considered concerning industrial AR applications is represented by the computing/visualization hardware required by the system to operate. The growing diffusion of new-generation smartphones/tablets promise to deliver cheaper and more usable [61] AR platforms [62, 63]. This is probably true if the interaction is always mediated by the touch-screen, but when the interaction also implies a contact with physical environment, the user is forced to hold the device with one hand while operating with the other hand behind the device's screen. In this case, a prolonged working session is likely to become a stressful experience.

Concerning AR system evaluation in real world scenarios it is useful to remark that although several AR application have been developed, most of them only aim at demonstrating the feasibility of the approach in the field of interest and/or the system performance in terms of tracking error or real-time operations. There are very few rigorous experimental evaluations of such systems [64]. One reason for this lies in the objective difficulties related to the typically more spatially-complex physical environment in which AR systems can be applied: the users rarely just sit in front of a PC; they can move from one place to another, they talk, they move parts of their body in order to explore and interact [65]. Consequently, experiments have to be carefully planned and implemented in order to avoid disturbances deriving from users' freedom of motion and interaction. Goldie et al [66] presented a formal study for human way finding performance in traversing a maze. They compared the performance of 20 participants supported by augmented reality map with a no augmented environment. The authors verified that the participants involved traversed the maze in AR treatment in an amount of time significantly shorter than in no augmented treatment. The work by Henderson and Feiner [67] represents perhaps the most complete and accurate experimental evaluation of a Augmented Reality system. They involved six participants, recruited from a recent class of graduates, on an experimental maintenance procedure of 18 elementary steps. The quantitative analysis performed together with the interpretation of subjective answers to a questionnaire provided to users allowed them to draw interesting conclusions about the feasibility of their approach and the real advantages of AR over more traditional computer-based aid systems. Quantitative analysis is particularly useful to objectively assess AR system's strengths and drawbacks but a good qualitative analysis can be also helpful [68]. Wang and Dunston [69] also discuss the benefits of augmented reality in collaborative design presenting a formal qualitative analysis based on NASA Task Load Index [70]. Though the last mentioned studies provide a valuable insight of the main advantages and issues related to AR technologies in a working context, they have been conducted on a very small number of casual users rather than subjects

with experience of real operations. This choice probably reduces the chances of a negative bias in the evaluation (for an expert is common be accustomed to specific well-established procedures), but it also reduces the value of the results achieved. Starting from this analysis, two experimental trials and the test bed were designed for testing the impact of AR-based maintenance in a real industrial scenario. In this process, the lessons learned from the state of the art was exploited to select a reliable set of quantitative and qualitative measurements to assess the maturity of this technology.

3.2.2 Method

The schematic view of the proposed system is shown in figure 3.2. The proposed architecture integrates precise multiple marker based tracking, two modalities for scene augmentation and a not-instrumented finger based interface to provide effective and reliable visual aid during maintenance operations. It is designed around three main components. The Augmented Reality Engine (MRE) is in charge of user's head tracking, scene augmentation/rendering and servicing procedures management. The Finger Tracking module captures fingers position enabling the human-computer interaction, while the Database contains the working environment setup, the virtual contents and the maintenance procedures required for the system to work. To start the assisted servicing procedure, the user has to wear a video-based see-through HMD, a backpack enclosed notebook and a few fingertips caps required for contactless interaction.

A set of six 4x4 cm sized markers provides an optimal tracking coverage of approximately 60x60x60 cm with an equivalent co-registration error within 2 mm, which is below the size of most small parts. As the relative position of each marker with respect to the absolute reference system is known, when more than one marker is
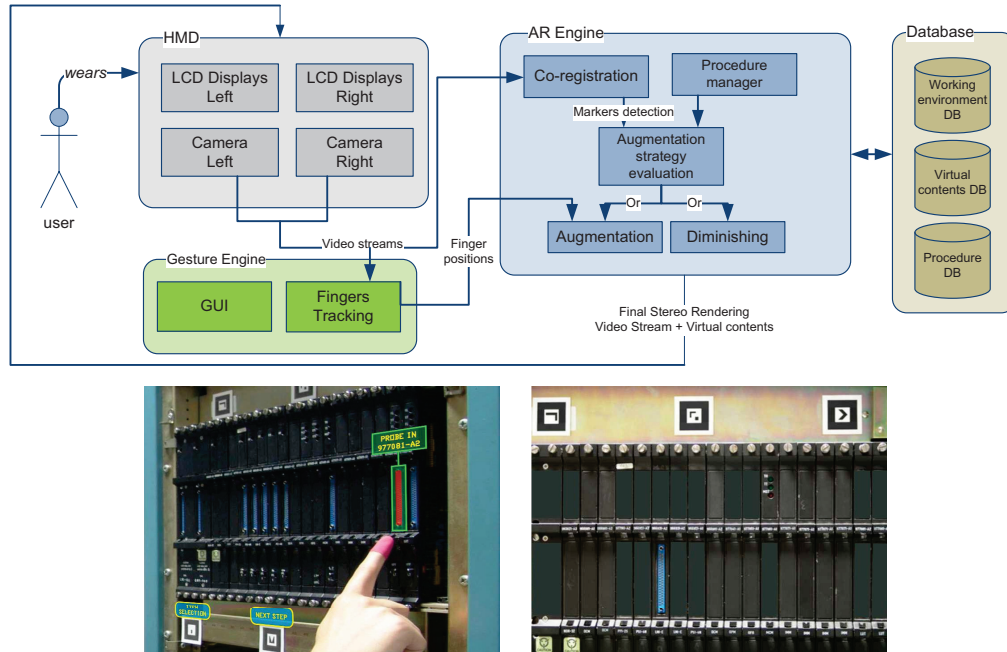
FIGURE 3.2: The overall system architecture of the system prototype (top). An example of human-computer interaction by fingertips tracker (bottom,left). An example of diminished reality where the system deletes from the real scene rather than adding virtual aids (bottom, right).

recognised each approximated estimate the position/orientation of the camera (relative to a particular marker) contributes in reducing the overall co-registration error through a weighted average strategy based on the quality and number of visual features recognised. To this regard it has to be remarked that the rotational component of camera tracking has a greater impact on augmentation accuracy compared to the positional component (for details refer to Chapter 4). In fact, even a degree of error may produce a visible misalignment between virtual-to-real as the distance from the tracked point increases. To minimize this effect the origin of absolute reference system (respect to which is expressed the position of any virtual content) is deliberately located in the geometric center of the marker-set to further reduce the rotation co-registration error of all objects falling within it. Additionally, to reduce unwanted camera shaking tracking data are smoothed by simple smoothing functions.

3.3   Telemedicine supported by augmented reality

Telemedicine refers to the use of telecommunications and information technologies for the delivery of medical services where is needed [71, 72]. For many applications, the correct usage of medical device at the point of need is essential to provide an appropriate service, but there are some practical situations that may require untrained or inexpert people to interact with medical devices and patients. Some examples are telemedicine services on transportation (e.g. aircraft [73–75], boats [76], trains, etc.), application during military actions [77], on islands or remote areas [78, 79], some emergency applications [80, 81], but also home care telemedicine supported by family members [82–84], elderly care [85, 86], operators training and so on. In these cases, untrained or improvised (but necessary) actors can, involuntarily, use medical instruments in an inappropriate manner and/or make improper connection between the patient and the medical device seriously invalidating the telemedicine service. In these scenarios, tailored applications of Augmented Reality can offer a valid support by guiding non-trained people to a correct usage of medical devices at the point of need.

The objective of the case study explored in this work of thesis is the analytical assessment of the benefits of exploiting augmented reality principles to estimate how untrained users with limited or no knowledge can interact with an electrocardiography (ECG) device and properly placing ECG electrodes on patient's chest. In many clinical activities, 12-lead electrocardiogram is an essential medical investigation. On the other hand, it should carefully carried out. Misplaced ECG electrodes can cause changes in ECG recordings, which could have an impact on clinical decisions [87]. Incorrect electrode cable connections, reversal of electrodes, inadequate placement are common source of error (changes the true ECG morphology) and can conceal or simulate different pathology such as, myocardial ischemia or infarction, arrhythmias, ventricular hypertrophy [88–90](it is also worth mentioning that untrained service providers are one of the key barriers to implementation of telemedicine services).

FIGURE 3.3: The possible scenarios of usage of the proposed telemedicine proto-type. The user wears the HMD and observes the ECG device. From the system he/she receives visual and audio feedbacks to guide him/her during the setup (left). The thorax of a real patient augmented of virtual aids to a proper placing of the electrodes (right).

3.3.1 Related works

Recently, there is a growing interest about AR in medicine. The main applications of AR are in the field of surgery, rehabilitation and teaching/training. Interventional medicine, surgery [91, 92], laparoscopy and other procedures (e.g. needle biopsy [93]) can be assisted by integrating preoperative and intraoperative anatomic and functional data improving the visual perception of the surgeon [62, 94–96]. Obviously, surgery AR applications require very accurate registration and camera calibration [97]. AR and virtual reality have long since found use in rehabilitation and particularly in neurorehabilitation, by guiding and aiding the patient to perform therapy [98, 99]. Teaching and training of students or physicians can take great advantage by AR, which can be further enriched with direct haptic and auditory feedback [100–102]. Superimposition in real time of images from US, CT or MR scans can also help in learning [62].

Compared to purely medical applications, there are only very few examples of AR applications in telemedicine. Among them there are systems of virtual reality supporting distance teaching of minimally invasive surgery and systems for interactive

telemedicine in the operating theatre [103, 104]; some low-cost peripherals to support telehealth, visualization, education and collaborative systems [105]; some applications of distance training for the restoration of motor function, supported by virtual reality [106].

3.3.2 Method

The developed prototype guides the user to perform a predefined sequence of simple tasks on the ECG device (e.g. connect cables, press buttons, check indicator lights, etc.) and on the patient (e.g. connect electrodes in a specific position, etc.), driving user's attention step-by-step to the relevant item. A flow-chart of activities was created to describe all possible sequences of simple tasks necessary to carry out an electrocardiographic test using that particular ECG device. The developed AR application drives the user to perform each predefined task by presenting text, graphics and audio messages to him/her . Specifically designed sets of markers were attached to the ECG device and to the patient: this allows the AR engine to evaluate on real-time the 3D pose of these objects with respect to the user. This permits to indicate or highlights a specific point or a part of the objects (e.g. a button to be pressed) by superimposing opportune signs to the current scene. The operators worn a Head Mounted Display (HMD) coupled to a webcam in order to see the virtual contents that augment the current scene (future developments will involve the use of a single Tablet or Smartphone). Figure 3.4 presents a general scheme of the proposed AR application.

The Augmented Reality engine is in charge of user's head tracking and scene augmentation/rendering (for details about the tracking refer to section 2.5). A constellation of three markers was associated to the ECG device and a different set of three to the patient's thorax (see figure 3.5. The portable ECG device considered in this application was a Cardiette Microruler 12/1. The three markers were arranged in line
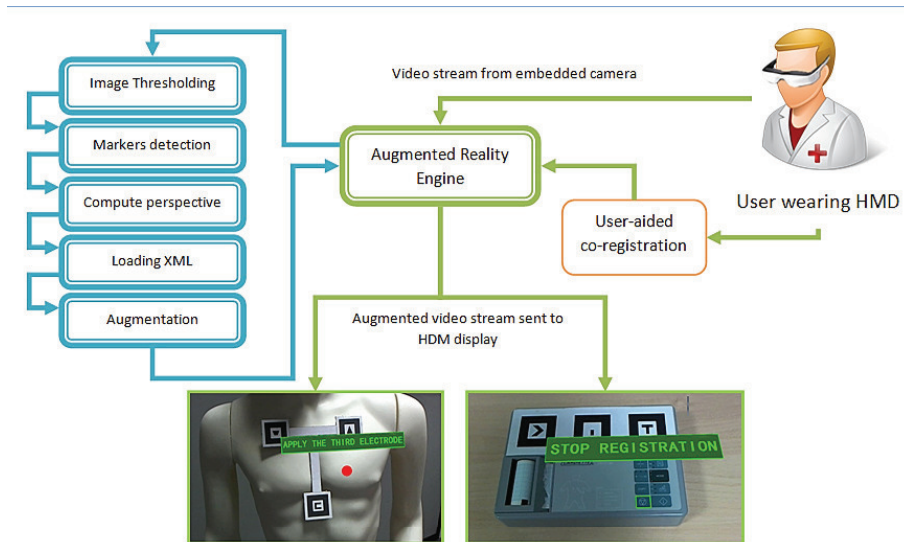
FIGURE 3.4: The architecture of the prototype system developed in telemedicine research.
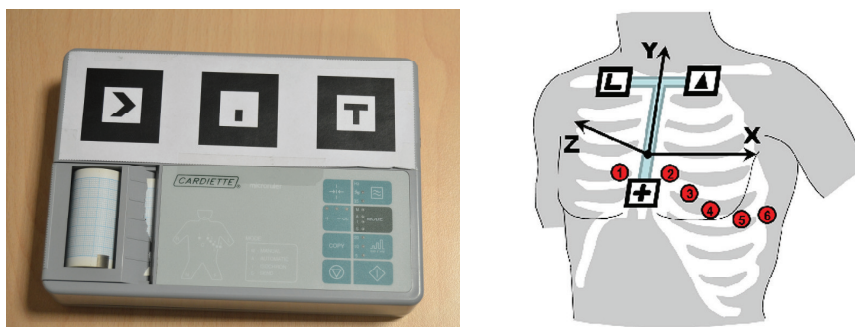
FIGURE 3.5: The constellations of markers used to detect and track the point of view of the user while working with the ECG device (left) and while placing the electrodes on the patient's thorax (right).

and horizontally aligned, spaced from each other by 2 cm and attached to the top of the front part of the ECG device. The geometrical localization of the following parts of the ECG device was accurately measured with respect to central point of the markers costellation (located in the geometrical center of the middle marker) and then recorded in XML file: the main switch (placed on the left lateral side of the device), the connector of the patient cable (placed on the right lateral side), all the buttons (eight in total, placed on the main panel), all the LEDs (twelve in total, placed on the main panel), the accessible parts of the chart printer (placed on the main panel).

For the augmentation of the patient's thorax, three markers were arranged on the tips of a T-shaped structure. The T-structure was made flat and rigid, its horizontal and vertical segment measured 10 cm (empirically chosen according the mean size of adult's thorax [107]). This structures was inspired by a well-known ideal triangle (called Einthoven triangle [108]) that is considered as a reference point for the proper positioning of eletrodes for ECG tests. The mentioned triangle and the corresponding T-Structure markerset designed for this purposed are illustrated in figure 3.6.
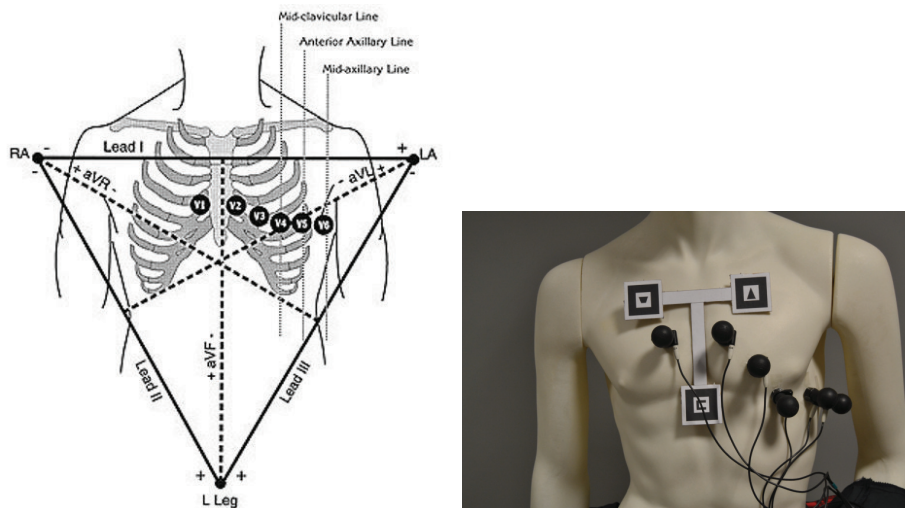
FIGURE 3.6: The Einthoven triangle (left) and the markerset arranged on a T-Structure specifically designed for the case study.

The T-structure was applied with stickers onto the patient's thorax, taking care to align the vertical segment along the sternum and positioning the upper side of the lower marker (corresponding in figure 3.5(right) to the marker with a plus sign) in correspondence of the xiphoid process (the lower end of the sternum), which can be easily recognizable by touching. The anatomical landmarks for the standard precordial electrode positions were fixed in accordance with the current ECG international standards [109]. The precordial electrode positions are: V1 and V2 at the fourth intercostals space to the right and left sternal border, respectively; V4 at the fifth left intercostal space in the mid-clavicular line; V3 midway between V2 and V4; and V5 and V6 at the horizontal level of V4 in the anterior and midaxillary lines, respectively. Coordinates x, y and z were obtained for each of the precordial electrodes with respect to the reference system fixed to the T-shape marker set and opportunely coded in the XML file. The x-, y- and z-axis correspond to the latero-lateral, cranial-caudal and dorsal-ventral patient's anatomical axes, respectively (for details refer to [110]).

3.4   Preliminary results

A subjective evaluation of the prototype systems was carried out to assess the level of acceptability by the users. Before analysing the tracking method, it was considered important to briefly assess the impact on potential users of the technological choices proposed in this work. For this reason, 10 people in each case study was involved. Participants were selected among those who did not have any previous knowledge or experience of augmented reality applications and devices. Each participant was asked to wear the HMD and was explained the intention of the testing session. After a first approach with the system, was asked to perform a simple sequence of tasks in augmented reality. At the end of each session a brief interview took place.

The main observations were that many participants reported a good confidence

feeling during system's usage. Some of them also reported an operational advantage in performing the proposed tasks in terms of time required to accomplish them. These very preliminary results did not represent a strict evaluation of the efficacy of using augmented reality principles in the selected case studies. However, they represented an encouraging starting point that made reasonable to start a deep investigation about one of the first issues to address in augmented reality, that is the tracker. As mentioned in the outline section 3.1, the main objective of subjective analysis proposed here was motivating a deeper investigation on such field. The critical and formal analysis provided for the tracking system in Chapter 4 represents, in fact, the attempt of demonstrating how a robust and reliable tracking system can dramatically impact on users performance and the acceptability of a fully functional augmented reality system.

Personal contributions

Parts of the topics and experimental results discussed in this chapter are from the following research papers of which I am co-author:

1. Augmented reality aided maintenance for industrial applications. In *Eurographics Italian Chapter Conference*[111]

2. Biometrics Empowered Ambient Intelligence Environment. In *Accademia Peloritana Dei Pericolanti Journal. (in course of publication)*[112].

3. Photorealistic Virtual Exploration of an Archaeological Site. *Accademia Peloritana Dei Pericolanti Journal (in course of publication)*[113].

4. An image based approach to hand occlusions in mixed reality environments. In Virtual, *Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, Springer.[47]

5. Mixed reality environment for mission critical systems servicing and repair. In *Virtual, Augmented and Mixed Reality. Systems and Applications.*Springer [48].

6. Telemedicine supported by augmented reality: an interactive guide for untrained people in performing an ecg test. *Biomedical engineering online*[110].

7. Raffaele Vertucci, Fabio Narducci. Mixed Reality in Advanced Maintenance and Logistics. In *Polaris Innovation Journal*, 2014.

Chapter 4

Experiments

4.1 Objectives

It has been proved that augmented reality *can be means of enhancing, motivating and stimulating learners understanding of certain events* [114]. However, this consideration cannot be separated from well-known AR drawbacks, e.g., the limited field of view, the cameras, the displays resolutions and the weight of the HMD, which may have a significant negative impact on the usability of the whole system. Objectively assess all variables that impact on the success of an augmented reality application is a non-trivial task. Advantages and disadvantages of augmented reality could be related to a big amount of factors. Just to mention a few of them, the reliability of the tracker, the interaction paradigm, the rendering of virtual objects, the video resolution of the adopted visualization system, the ergonomic and much more are all involved when evaluating the benefits of augmented reality in a specific operative context.

Considering these issues, it is rather difficult to investigate on each of those factors and then making the proper associations to infer dependencies and draw final conclusions. Therefore, this work focuses the attention on how to estimate and formally

quantify the benefits of implementing AR systems based on the proposed optical tracker (described in section 2.5) in the selected case studies described in Chapter3. The experiments were designed in order to answer this questions:

- Is the optical tracker able to efficiently detect and track the point of view of the user?

- Can augmented reality approach be adopted in the selected case study?

- Does the augmented reality prototype improve user's performance?

- Is it possible to quantify the effort made by users by observing their behaviour while working?

4.2   Testing design and settings

In the previous section the main objectives of the experimentation have been introduced. However, at this point is important to make a consideration about the differences of the case studies and the typical user profiles. In the telemedicine context, the user is asked to work with an ECG device as well as to place six electrodes on human chest. It can be reasonably assumed that the whole interaction (and then the testing session) is carried by the user while sitting on a chair. This significantly limits the range of movement of the user, consequently relaxing the constraints for the tracker. In the industrial case study the user is supposed to stand in front of a rack, like in normal maintenance operation happens. This condition introduces new difficulties in terms of accuracy and robustness expected from the tracking method. The experimental sessions were designed keeping in mind these aspects. Therefore, two testing environments where designed aimed at analysing different aspects, that is one for telemedicine and the second one for the industrial case study. In both cases, participants to the tests were selected among people who did not take part to

any kind of similar experimentation before and with no previous knowledge about the kind of operations to perform. This assumptions ensures that the skills of the participants about one or more topics covered during the experimentation can not affect in any way the analysis of the results obtained.

4.3 Telemedicine prototype

Two separated experimental tests were designed: the first on a life-size mannequin and the second on a real patient. The experimental tests were carried out involving two groups of 10 people each (14 men and 6 women in total, mean age of 31.3 and median age of 30.5) with no medical expertise, no experience of ECG test and who never used an ECG device. At each tester was asked to wear the HMD (a Silicon Micro Display ST1080 featuring a Logitech C910 webcam) and, after familiarizing with the system for few minutes, the experimental trial was started. The first group of 10 people carried out complete ECG recordings interacting with the ECG device and placing electrodes on the mannequin. The second test was performed on a volunteer, male adult acting as patient who lay supine on a table and breathed normally to resemble a practical case. Similarly, the second group of 10 untrained people was asked to perform complete ECG recordings only supported by the AR application.

The ability of performing the required operations, the time necessary to complete the ECG recording, the positioning errors of the electrodes and the user's judgement were collected for all testers. In particular, the three spatial components of the distance between an expected (true) location of an electrode and its actual placement were recorded as the error committed by the untrained user. This information was used to quantitatively assess the efficacy and the clinical acceptability of the developed AR application as well as the precision of the proposed tracker.

4.3.1 Positioning errors for precordial electrodes

Bar plots in figure 4.1 show the averages and standard deviations of the X, Y and Z components of errors (computed on all testers) in placement of the precordial electrodes on the mannequin (figure 4.1(left)) and patient's chest (figure 4.1(right)). Averages can be considered as an indication of the accuracy, while standard deviations as a measure of the precision (repeatability) achieved during the trials. In addition and in a more concise way, table 4.1 reports errors measured as distance between the expected and the occurred position for each of the precordial electrodes V1-V6. The average, standard deviation and maximum of these displacement errors are available for both tests (mannequin and real patient).
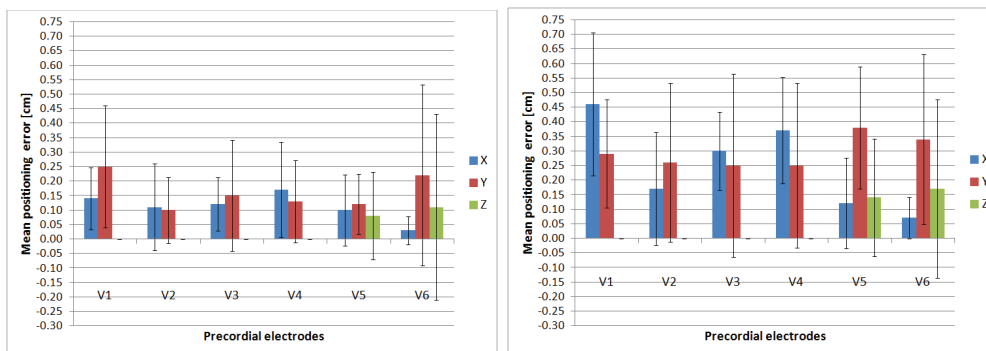
FIGURE 4.1: Averages (bars) and Standard Deviations (segments) of the errors (distance from the expected location reported in centimetres for each of the x, y and z axes of the space) in the placement of the precordial electrodes on the mannequin (left) and on the patient's chest (right) by the ten untrained users. V1, V2, V3, V4, V5 and V6 are the six precordial locations.

For the first test, the average errors in electrode positioning on the mannequin resulted less than or equal to 3 mm, while the standard deviation less than 5 mm. Involvement of the real patient instead of the mannequin led to a slight increase of the errors committed. This can be reasonably due to thorax motion caused by breathing or other little movements of the patient. Nevertheless, even in the test of the real patient the mispositioning of the electrodes resulted less than 7 mm on

TABLE 4.1: Positioning errors for precordial electrodes

Test		V1	V2	V3	V4	V5	V6
Mannequin	**Average error [cm]**	0.30	0.18	0.23	0.24	0.22	0.27
	SD of the error [cm]	0.10	0.10	0.13	0.19	0.11	0.41
	Max error [cm]	0.45	0.40	0.45	0.54	0.41	1.42
Patient	**Average error [cm]**	0.40	0.27	0.44	0.46	0.49	0.69
	SD of the error [cm]	0.7	0.20	0.20	0.22	0.18	0.28
	Max error [cm]	0.50	0.73	0.78	0.81	0.79	1.56

Average, standard deviation and maximum of the displacement errors done by the ten users of the developed AR application in placing precordial electrodes (V1-V6) during the tests on mannequin and real patient. All data are expressed in centimeters. The displacement errors were computed as the square root of the sum of the square of the three Cartesian components (x, y, and z) of the displacement vector.

average and reached a maximum of 16 mm on V6, these data support the effectiveness of the AR procedure and the clinical acceptability of the recorded ECG. As a matter of fact, electrode malposition exceeding 25 mm is associated with potentially significant ECG changes [87]. Taking into account this threshold value and observing the results achieved during both tests, the average errors in electrode positioning resulted reasonably acceptable and comparable with placement errors usually made by technicians and nurses in an emergency care department [115]. This supports the clinical validity of the acquired ECG waveforms by means of the developed AR system. It is worth noting that in both tests only for V5 and V6 significant variations of positioning were registered on the z-axis. This is inherent to the specific positions of these two electrodes and their relative positioning with respect to the markers on the thorax. When the user directs his gaze towards V5 or V6, the plane on which lies the marker (x- and y-axes represented in figure 4.2(d)) results significantly angled and then the 3D pose errors increase [116] . This mainly occurs because of the natural curvature of the human thorax. Indeed, many studies have shown that the accuracy of positioning virtual object on the real scene mainly depends on camera distance and viewing angle with respect to the markers and also on other factors (e.g. size of the marker, focal length, field of view, pixel resolution, etc).
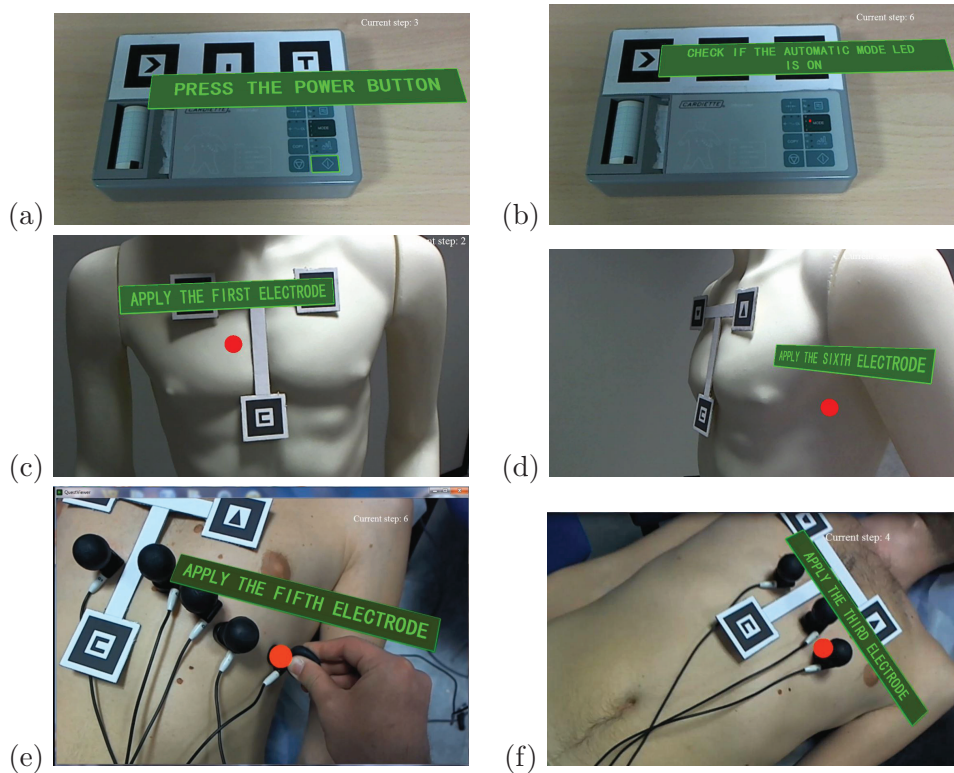
FIGURE 4.2: Examples of augmented real scenes: the power button of the ECG device is surrounded by a light green rectangle, while audio and text invite the user to press it (a); a little red spot indicates the LED to be checked by the users (b); the location of the V3 precordial electrode on the mannequin is highlighted with a red circle (c); the location of the V6 precordial electrode is highlighted with a red circle (d) please note the inclination with respect to the plane of the markers; the location of the V3 precordial electrode on a real patient is highlighted with a red circle (e); placement of the V5 precordial electrode on the patient's thorax, the location of V5 is highlighted with a red circle (f).

The execution time over both tests was on average of 8 minutes (including 3 minutes of ECG recording in the automatic mode) that can be considered acceptable for practical purposes.

4.4   Industrial prototype

The objective of this study is to objectively assess the advantages and disadvantages of Augmented Reality applied to the field of industrial maintenance and training. To this aim, the AR prototype was compared to the most common way of performing maintenance, i.e. consulting technical manuals. The research was focused on engineers, selected among specialized technicians and not, in the field of naval or land systems. All potential participants were firstly screened by a simple interview to understand their level of confidence in AR environments, computer-aided systems and industrial maintenance procedures. In order to provide a fair comparison between the two experiments, we discarded all engineers who had a significant experience of AR systems as well as the maintenance experts. This "recruiting" process gave us the opportunity to consciously select 40 participants. An important aspect common in comparative testing that is worth to note, is the common finding of improved user performance in the second treatment, independently from the intrinsic difficulties faced in first one. For this reason, rather than involving all participants to both testing sessions, they were properly divided into two groups. Twenty users were asked to perform twelve common maintenance steps using the proposed AR environment (see table 4.2).

The other twenty users were asked to perform the same sequence of tasks aided by a conventional electronic manual. Each participant was also asked to respond to a NASA TLX (Task Load IndeX [70]) questionnaire in order to have a further feedback on the easy of use of the prototype as well as an estimation of the effort to accomplish the required sequence of tasks.

The sequence of steps involved in the experiments was randomly built from a set of elementary tasks among the most typical in maintenance procedures. The list of twelve tasks and their order in the testing procedure. Figure 4.3 shows the industrial rack on which the testing sessions were performed and the points of interests of each

TABLE 4.2: List of elementaty tasks of the testing maintenance procedure

Task	Description
T1	Unscrew four fixing screws
T2	Check if led is ON
T3	Unplug the 41 pins connector
T4	Check is led if OFF
T5	Change the state of the switch
T6	Screw four fixing screws
T7	Change the state of the switch
T8	Plug the 41 pins connector
T9	Unscrew two fixing screws
T10	Change the state of the switch
T11	Check if led is ON and if it is yellow
T12	Screw two fixing screws

task considered. Concerning the order of execution of maintenance steps, it is worth to note that the selected sequence is built of random steps with no any correlation to each other. In other words, it was chosen to do not simulate any known maintenance procedure. Although most of the participants to the experiments had no particular prior knowledge about maintenance, they were engineers in the field of naval or land systems. Thus, choosing a random sequence of elementary maintenance steps ensures that no memory effect on maintenance procedures can affect the statistical significance of the results. The same procedure was used in both experiments in order to allow a fair comparison and to easily discuss the obtained results.

After choosing the target users and defining the testing maintenance procedure, it was defined the experimental trials in both treatments and how to collect data to be statistically analysed. To this purpose, each experimental session has been divided into four main steps.

1. *Introduction to the test.* As remarked before, each partic-ipant took part in only one treatment. In order to put them at ease, we informed all participants
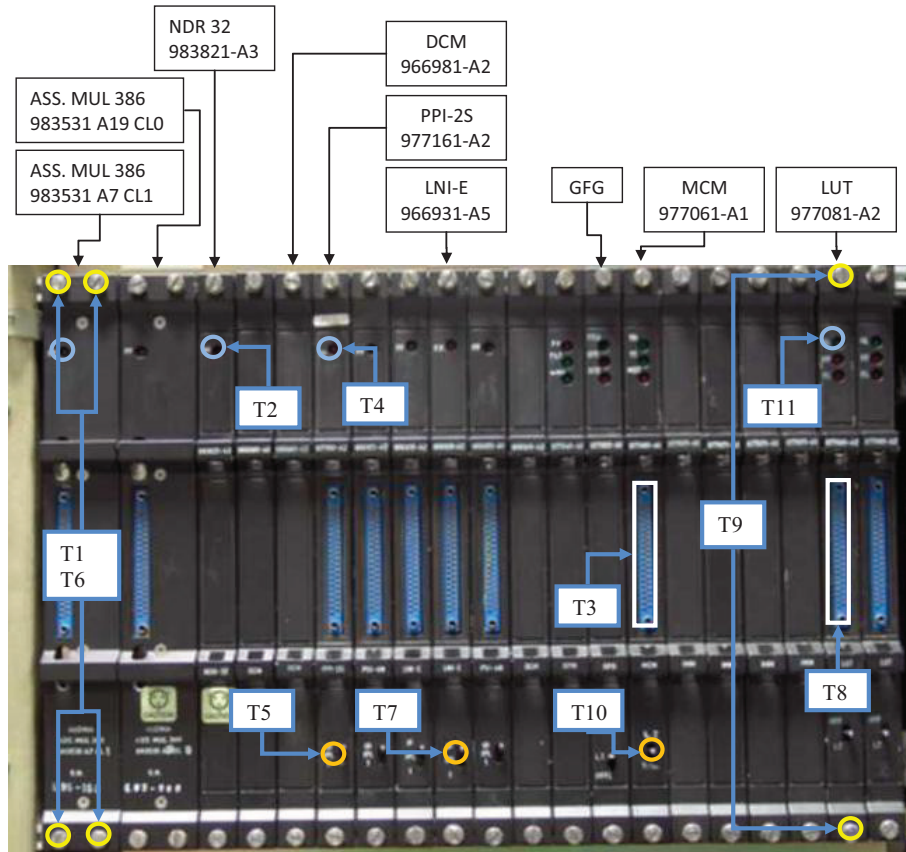
FIGURE 4.3: Front panel of the rack involved in the experiments. Each label shows the point of interest of the corresponding task described in table 4.2.

in both treat-ments about the purposes of the testing and the challenges to face with. No details were revealed about the kind of data collected during the test to avoid the risk of participants performing the requested tasks in a biased manner, affecting the significance of the data captured and their statistical significance.

2. *Training session.* After introducing the purpose of the test session, the participant was invited to perform a typical step of maintenance not related to the real testing procedure. For what concerns the AR experiment, all subjects

were invited to wear the HMD equipped with tracking camera and earphones and to explore the system and the user-machine interface provided. In ExpTM treatment, to be able to track movements of the head, we asked all subjects to wear a non-obtrusive cap including the same camera user for image-based tracking in AR treatment. In both experiments we also gave the opportunity to briefly examine the rack and the related tools.

3. *Experiment.* During each testing session, the system collected two types of data: 6DOF data from the image-based head tracking and the completion time of each maintenance step. All users in both experiments were standing in front of the rack, were free to look around and to interact with the environment in the preferred way. A synthesized speech accompanied the current task for a more involving experience. From the beginning to the end of each test, none of the participants was helped in any way. They were left alone, observing their behavior from a proper distance.

4. *Questionnaire and Discussion.* After the testing ses-sion, the subjects were asked to complete the NASA TLX rating questionnaire. Moreover, all partic-ipants were free to express doubts or recommendations to improve the system tested during a de-briefing session that followed each trial.

4.4.1 Data post-processing

Once that testing sessions were completed, a statistical comparison was performed. To achieve a fair analysis, tracking data were post-processed by normalising them and filtering out misleading captures due to unforeseen events. In the following lines, the most relevant issues arisen during experiments and the counter measures adopted are briefly reported. The webcam used for optical tracking captures the scene at 30 fps. Although this is a reasonably good capturing frequency, it cannot ensure sharp capture of images during rapid movements, particularly for brief and quick

rotations on pitch and yaw axes, which are much more common for the head than horizontal/vertical translations. The consequences of this hardware constraint is, possibly, a visible image blurring that compromise markers detection and therefore tracking reliability. To cope with this issue, rapid tracking failures were annotated. By analysing users' movement records, we noted that the average duration of these tracking failures was below 0.5 seconds. In these cases, a linear interpolation resulted to be effective to restore motion continuity. Another kind of less frequent tracking failures were reported whenever participants moved sideways with respect to the working area and the marker set, leading the camera-to-markers line of sight close to be tangential to the marker plane. In these rare occasions were reported tracking failures for a few seconds up to 15 seconds with isolated groups of valid tracking samples. In this case, considering the wider time window, valid tracking samples were used as anchors for a spline interpolation of the motion curve. Concerning the measurement of completion time for each step of the procedure, this was computed considering the time lapse from the beginning of the current task to the system signalling the user to proceed to the next step. In a not negligible number of instances, vocal commands were not pronounced loud enough by the user. In these cases, the system asked them to repeat the command one or even two more times to progress in the procedure. In other rare instances, participants forgot the vocal command required to go on. Since the main focus of the study is not on the usability of the human-computer interface, the impact of this aspect was filtered from tracking data as well as from the completion time measurement.

Even though all subjects were monitored at a distance, some of them sometimes let the screwdriver drop while trying to tighten or loosen a screw. In such occasions, in order to do not significantly alter the data recorded, the participant was helped collecting the tool from the ground and providing to properly cut the time spent and the useless movements in the meanwhile from the time of completion for that step and from the tracking data. Finally, the presence of one outlier in both treatments implied a reduction of the populations thus leading each to 19 samples for a total

number of 38 participants (instead of the initial number of forty selected participants).

The first treatment, that for the sake of brevity will be referred as ExpAR (experiment in augmented reality) in the rest of the discussion, consisted in performing the testing maintenance procedure by means of the augmented reality visual and audio aids. The subjects involved in this treatment experimented an augmented view of the scene by means of labels, graphics and virtual tool co-registered to the real objects helping them to quickly localise the intervention point. The second treatment, which from now on will be in the rest of this study will be shorten in ExpTM (experiment with technical manual), refers to the trials in which the subjects were asked to perform the testing maintenance sequence by following instructions displayed on a computer monitor, positioned beside the working area and showing a step-by-step guide to the procedure based on the technical manual. To this regard, the disparity between the nature of augmented reality visual aids and the graphical layout of a conventional manual was taken into account. In order to reduce this gap, the contents to be shown on the support monitor for ExpTM was designed adopting a layout close the one featured by the augmenting contents in ExpAR. This ensured a fairer comparison between the two experiments in terms of understanding of the procedure and identification of the element on which operating.

4.4.2 Task completion timing

A two-tailed Student test was performed on total completion times for each participant between the two treatments. Before performing the t-test, it was firstly checked if the populations were normally distributed in the treatments by running a Shapiro-Wilk [117] test on both of them. Recalling that the null hypothesis H_0 of the test is that the population is normally distributed, with a criterion of significance of $\alpha = 0.05$ the *p-value* was 0.1003 in ExpAR and 0.1741 in ExpTM. This result allowed to accept the null hypothesis, namely, that both populations

were normally distributed. This result also allowed to perform the *t-test* on total completion times with the null hypothesis H_0: *there is no significant difference in terms of completion times for the two treatments.* Using $\alpha = 0.05$ as criterion of significance, the statistical test allowed to discard the null hypothesis. In other words, there was a significant difference in total completion times between the treatments ($|T_0| = |-4.3407| > T_{(0.025,36)} = 2.0281$, $p = 0.0001$). The boxplot in figure 4.4 shows that mean completion times in ExpAR are shorter than in ExpTM. This finding is also confirmed by performing a one-tailed *t-test* where null
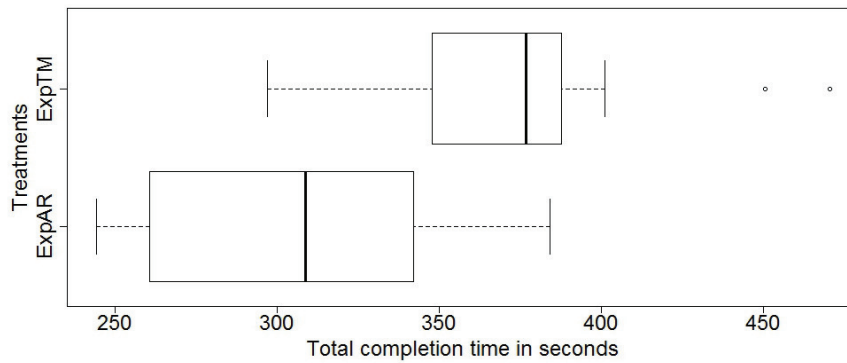
FIGURE 4.4: Completion time for ExpAR and ExpTM treatments.

hypothesis H_0 is that *the mean time of completion in ExpAR is not significantly shorter than in ExpTM.* Using $\alpha = 0.05$ as criterion of significance, H_0 was rejected ($T_0 = -4.3407 < T_{(0.05,36)} = 1.6883$, $p = 5e^{-5} \ll 0.05$) proving that subjects in ExpAR completed the procedure in a time significantly faster than in ExpTM (mean time in ExpAR is 82.67% of mean time in ExpTM).

During the experimental trials all completion times were recorder of each step of the testing procedure. Although the statistical result confirmed a reduction of completion time for the procedure in ExpAR compared to ExpTM treatment, a deeper statistical study on completion times per single task was performed. This allowed to draw some considerations about what kind of maintenance tasks were more difficult

to perform in ExpAR. Not all maintenance tasks had the same level of interaction, as locating a led is different than unscrewing a component. During ExpAR trials indeed, most of the subjects experienced evident difficulties in all operations involving the use of a screwdriver. In order to objectively evaluating this aspect, both two-tailed and one-tailed Mann-Whitney test [118] were performed. This kind of statistical test was preferred to the Student test because, according to results of Shapiro-Wilk test, most of the populations were not normally distributed. In table 4.3 are summarized the results obtained by the Mann-Whitney test on completion times per single task between the two treatments. In gray are underlined all tasks for which the p-value was higher than the significance threshold of 0.05. By looking more closely, all tasks that passed the two-tailed test also allowed to reject the null hypothesis in the corresponding one-tailed test with H_0: *the mean completion times in ExpAR is significantly less than that in ExpTM*. This ensures that for these tasks, subjects in ExpAR reported a mean time of completion smaller than that in ExpTM. In particular for task T_{10}, a value of $W = 0$ tells that all participants in ExpAR completed the task in an amount of time shorter than all subjects in ExpTM. Concerning tasks for which the test failed (i.e., $T_1, T_6, T_8, T_9, T_{12}$), it looks evident that all of them required a kind of interaction that suffered from the vision through the HMD in ExpAR (also confirmed by results of interviews reported in section 4.4.4. The problem of aiming and properly pointing the tip of the screwdriver at the screw slot is strongly hardware-dependent. The possible causes of this problem are:

1. the lack of confidence with AR environments;

2. the webcam resolution that should be enhanced;

3. the lack of stereoscopic vision that affects movements coordination and distances perception.

TABLE 4.3: Mann-Whitney test on completion times per single task

Index	Description	Two-tailed	One-tailed
T1	Unscrew	W = 107	
		p=0.1754	p=0.0877
T2	Check led	W = 20	
		p=3e-06	p=1.5e-06
T3	Unplug	W = 82	
		p=0.0042	p=0.0021
T4	Check led	W = 20	
		p=1.5e-07	p=7.7e-08
T5	Change switch state	W = 74	
		p=0.0014	p=0.0007
T6	Screw	W = 178.5	
		p=0.9651	p=0.4825
T7	Change switch state	W = 55	
		p=0.0001	p=6.1e-05
T8	Plug	W = 220	
		p=0.2582	p=0.8770
T9	Unscrew	W = 230	
		p=0.1542	p=0.9271
T10	Change switch state	W = 0	
		p=1.5e-07	p=7.3e-08
T11	Check led	W = 95	
		p=0.0119	p=0.0059
T12	Screw	W = 218	
		p=0.2836	p=0.8646

4.4.3 Analysis of Motion

All subjects in both treatments were standing in front of the rack and were free to make every kind of movement. Even though this decision impacted on the continuity of data tracking, it gave to each user the possibility of behaving in the most natural way. After correcting short tracking failures due to this choice (as described in section 4.4.1) some processing on the head tracking data was performed to draw conclusions on aspects related to translational and rotational components. Firstly, the movements performed by each participant in terms of head translations were

analysed. The combined plot in Figure 4.5 shows the average movements of each (referred to the head) for the three axes in ExpAR and ExpTM. The black curve shows the mean movement on each axis among all users. Comparing the sideways user movements (X axis), recorded in both treatments, it can be easily noted a relevant difference between the average values. Subjects in expAR treatment exhibited a more stationary behaviour, standing in front of the rack for most of the time. Participants in ExpTM, instead, exhibited a much wider range of movement, mainly due to the need of consulting the electronic manual placed beside the working area. This is also confirmed by the different mean standard deviation values among both treatments, that is 8.21 for ExpAR and 12.86 in ExpTM, a value 1.56 times the former one. It should be taken into account that the post-processing of the tracking data was also performed in order to reduce this gap between the treatments. This means that the range of movement in ExpTM is even bigger than that sampled and showed in the figures. Concerning the movements recorded on Y-axis (normalised according to the mean height of subjects) a similar behaviour in both treatments was observed. In this case, mean standard deviation values are 5.30 and 4.16 for ExpAR and ExpTM respectively (standard deviation for ExpAR is 1.27 times the value measured for ExpTM). This result is not surprising, since the test did not ask participants to kneel down during none of the tasks of the procedure. The reduced field of view experimented through the HMD, could explain the slightly higher mean standard deviation in expAR compared to ExpTM as well as for movements on Z axis. Similar results have been found by analysing backward/forward movements (Z axis). Here again users were more stationary in ExpTM than in ExpAR, likely because they performed the procedure without wearing any visualization device. The wider field of view of human eyes allowed subjects to look the entire rack without moving backward/forward. In fact, even in this case the mean standard deviation of all participants is 9.40 in ExpAR compared to 7.52 in ExpTM (standard deviation for ExpAR is 1.24 times the value measured for ExpTM).

Translational components of captured data enable a significant comparison among

(a) Mean X Translation Movements in ExpAR and ExpTM treatments (position in cm)
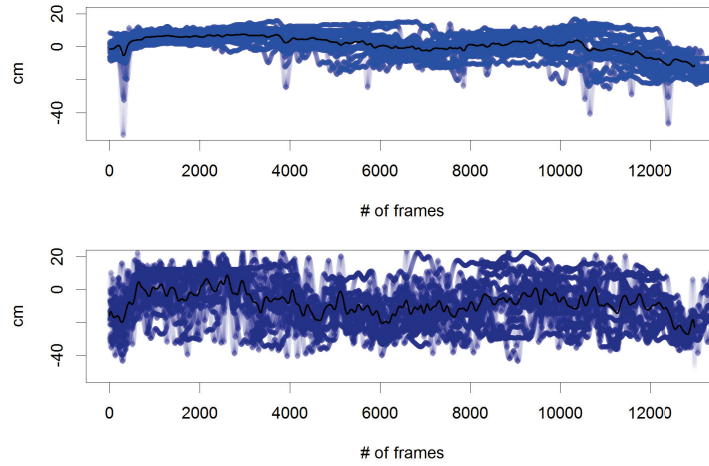
(b) Mean Y Translation Movements in ExpAR and ExpTM treatments (position in cm)
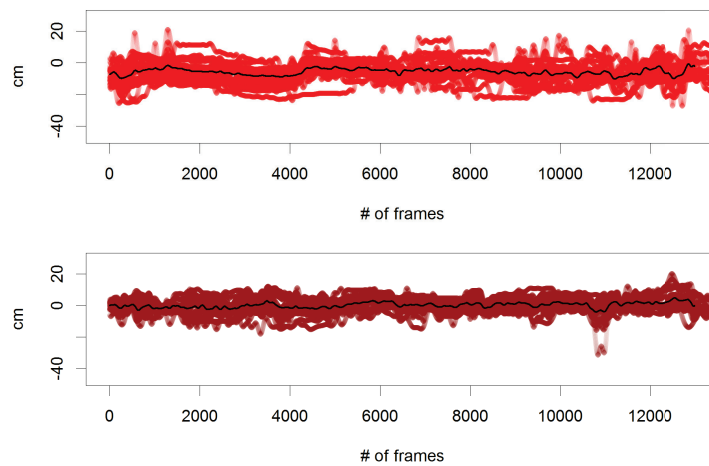
(c) Mean Z Translation Movements in ExpAR and ExpTM treatments (position in cm)
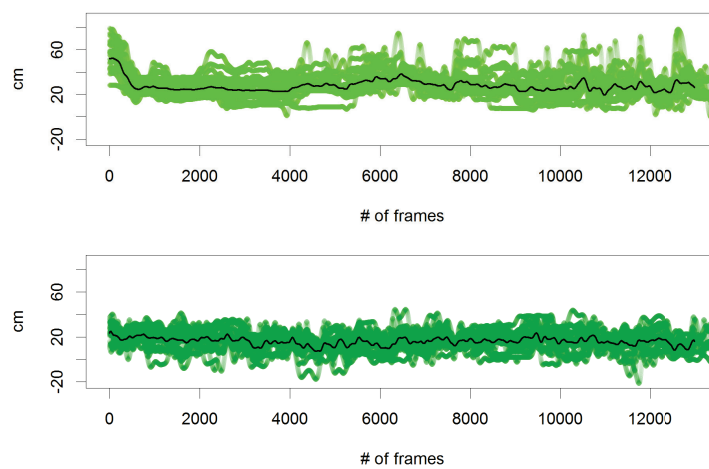
FIGURE 4.5: Aggregated plots of mean users' movements on each translational axis. Each pair of plots shows participants' movements in ExpAR treatment (top) and in ExpTM treatment (bottom) on XYZ axes respectively.

the dynamic behaviours of all subjects in both treatments. Performing a similar analysis solely on the rotational component would be rather misleading, indeed. However, by combining rotational and translational components of head tracking, it is possible to infer an approximated estimation of the user's attention. Of course, this kind of analysis does not take into account the movements of eyes (an eye tracker would provide a finer estimation of the users' fixation points). Since not focused on a precise study of the point observed did not, this approach has proved to be meaningful enough to the selected case studies. In figure 4.6 a unified plot per each treatment illustrates how much the observed point on the rack changes in ExpAR compared to ExpTM. In each plot all subjects' behaviour were overlapped per treatment. Therefore, dense regions in the plot mean that most of the partic-
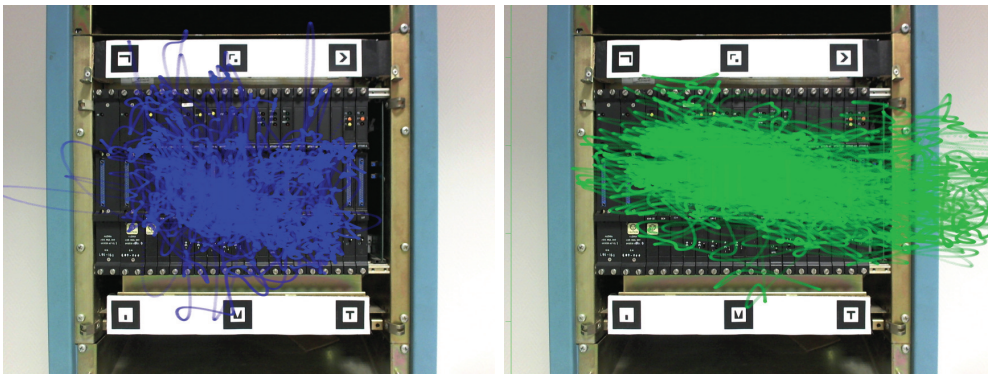
FIGURE 4.6: Normalized plots of mean observed point during maintenance procedure in ExpAR treatment (left) and in ExpTM treatment (right). By observing these plots, it looks evident that the observed point in ExpR is well confined in the servicing region on the rack, while participants' dynamic behaviour in ExpTM treatment is characterized by much more ample motions on the X axis (movements from left to right and vice versa).

ipants concentrate their attention in that area. In the former, we can see that the attention is mainly focused on the center of the rack. Ignoring extreme and repeated movements on left and right beyond the rack surface (due to the need of understanding the maintenance step to perform), the high density in the middle of the rack in ExpTM is a clear indication of a big amount of pass with the look compared to the

most controlled behaviour in ExpAR.

The aspects just discussed so far serve as an introduction to the analysis of translational/rotational velocity and the length of the estimated path covered by sight in the two treatments. Movement analysis is useful to infer some considerations about the behaviour of all subjects during the maintenance procedure in both treatments. Velocity analysis, instead, includes a higher power of expression because it allows drawing assumptions on physical and mental demand with the increasing of the time needed to perform the maintenance task. Considering that tracking data was recorded as well as the time needed to complete every single task of the testing procedure, linear and angular velocities were approximately estimated by dividing the space covered in each pair of consecutive samples for the elapsed time. In figures 4.7 and 4.8 the computed velocities are shown by means of density plots that help to better understand the mean velocity peak in each treatment per translation or rotation axis and the distribution around that peak. By observing each pair of density plots, it is possible to note a similar mean velocity on translational Y-axis and on pitch rotation in ExpAR and ExpTM treatments. On others, the mean velocity in ExpAR is always lower than the corresponding in ExpTM. To evaluate if all observed velocities were non-statistically significant, a distinct running of a two-tailed Student-test per pair of observation between two treatments was performed. Before showing the results of the different running of the test, it results useful to remember the critical value of the t-distribution calculated according to the probability of $\alpha = 0.05$ and 36 as degree of freedom that is T(0.025, 36) = 2.0281. Starting from the observation made on the mean translational velocity on Y-axis and the mean pitch rotational velocity, two different running of a *t-test* with the null hypothesis that *there is not a significant difference in mean velocity between both treatments* was performed. Using $\alpha = 0.05$ as significance criterion, in both cases the null hypothesis was accepted with $p = 0.4903$ ($T_0 = |-0.6969| < 2.0281$) for Y translational velocity and $p = 0.2610$ ($T_0 = |1.1419| < 2.0281$) for pitch velocity. The first result is not really surprising. Since the testing procedure in both treatments
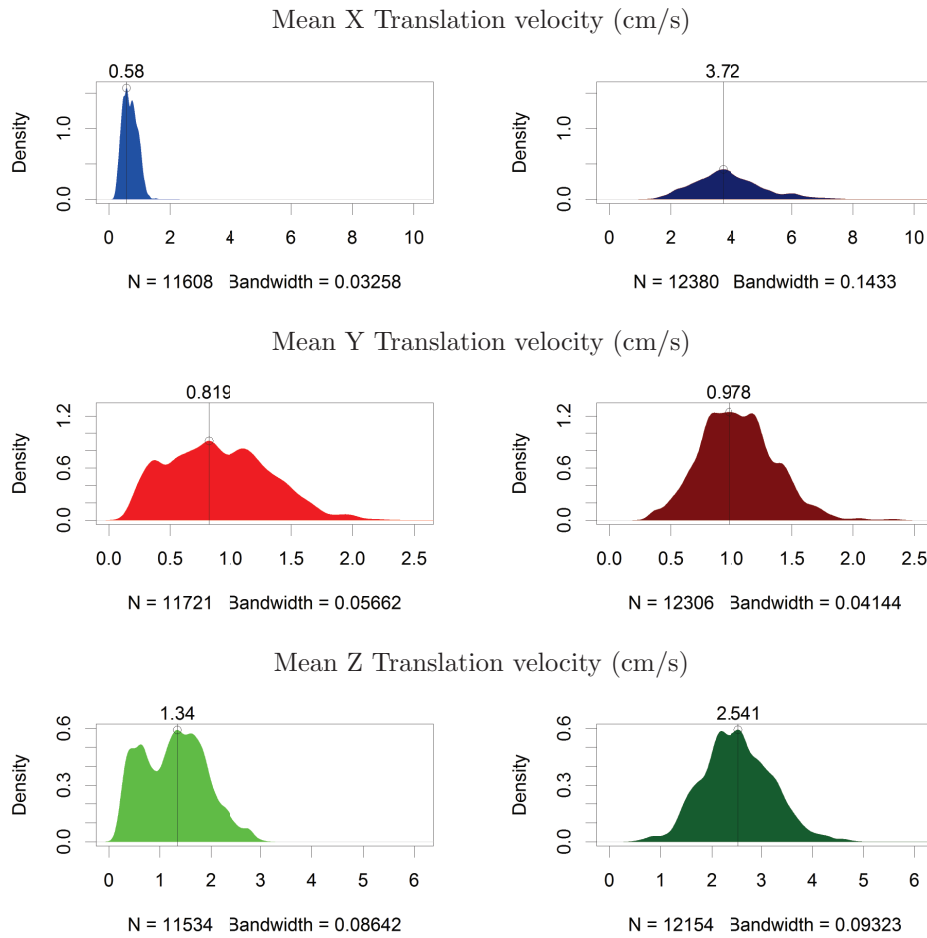
FIGURE 4.7: Density plots of mean movement velocity in both treatments (ExpAR on the left and ExpTM on the right). By observing each pair of plots, the most significant differences in mean value of velocity is related to the translational velocity on X-axis.

did not ask participants to kneel down, the insignificance between translational Y velocity confirmed the observation already made. Subjects in both treatments did not perform significant movement on Y-axis during the whole session. As expected, this produced the same results in terms of insignificance of velocity on that axis. The non-significance of pitch rotations also confirmed the observation made above about the difference between the two treatments concerning the visualization system. In ExpAR the augmenting contents displayed on the HMD helped participants
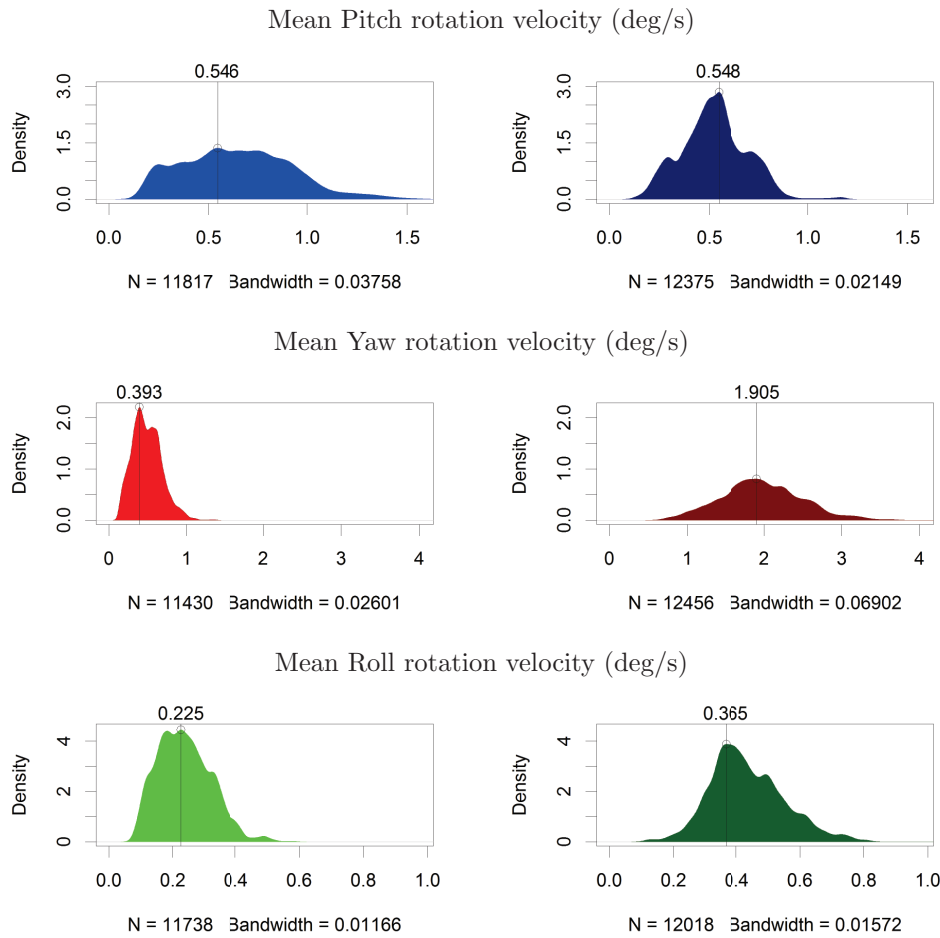
FIGURE 4.8: Density plots of mean rotation velocity in both treatments (ExpAR on the left and ExpTM on the right). By observing each pair of plots, the most significant differences in mean value of velocity is related to the rotational velocity on Y-axis.

to immediately localise the point of interest asking negligible pitch rotation of the head. In ExpTM, the wide natural field of view of the eyes allowed subjects to localise each point of intervention without performing pitch rotations by the head but simply looking up/down by the eyes. In all other cases, the results of the *t-test* allowed to reject the null hypothesis. In other words, the mean velocity in ExpAR is significantly different from the corresponding in ExpTM. Table 4.4 summarises the results from the set of different running of the *Student test* on the sampled tracking

TABLE 4.4: Comparison of mean translation velocity and rotation velocity between treatments

Mean translation velocity (cm/s)			
Axis	**ExpAR**	**ExpTM**	**Comparisons**
X	0.58	3.72	**ExpTM is 6.4 times that of ExpAR** $p = 2.81e - 11$ $T_0 = \lvert -9.4371 \rvert > 2.0281$
Y	0.82	0.98	**ExpTM is 1.2 times that of ExpAR** $p = 0.4903$ $T_0 = \lvert -0.6969 \rvert < 2.0281$
Z	1.34	2.54	**ExpTM is 1.9 times that of ExpAR** $p = 6.2e - 05$ $T_0 = \lvert -4.5281 \rvert > 2.0281$

Mean rotational velocity (deg/s)			
Axis	**ExpAR**	**ExpTM**	**Comparisons**
Pitch	0.55	0.55	**ExpTM is 1 time that of ExpAR** $p = 0.2610$ $T_0 = \lvert 1.1419 \rvert < 2.0281$
Yaw	0.39	1.90	**ExpTM is 4.9 times that of ExpAR** $p = 2.6828e - 09$ $T_0 = \lvert -7.8399 \rvert > 2.0281$
Roll	0.22	0.36	**ExpTM is 1.6 times that of ExpAR** $p = 0.0001$ $T_0 = \lvert -5.1671 \rvert > 2.0281$

data. Velocity reported on translational X and Z-axis for ExpAR resulted significantly lower than the corresponding in ExpTM. Concerning X axis mean velocity, subjects in ExpTM moved on their left and right at a velocity 6.4 times higher than in ExpAR. Even without considering the significance of other mean velocity between the treatments, the implication of this observation result clear: the longer the time needed to complete a complex maintenance procedure, the greater the physical load experienced in ExpTM compared to ExpAR. However, it is worth to note that this result does not take into account the added physical load introduced by the HMD used in ExpAR. To this regard, section 4.4.4 discusses in more details the aspects related to physical/mental load and other common parameters in the two treatments

to assess strengths and weaknesses of both.

Tracking data were also exploited to infer the space covered by the estimated point of view during both treatments. In particular, the intention was to evaluate the space covered on the surface of the rack. To this aim, it was approximately estimated as the summation of the differences between every pair of consecutive samples in post-processed projection XY data (i.e., the Cartesian plane corresponding to the surface of the rack). The boxplot in figure 4.9 shows the comparison between two treatments. The average difference in terms of space covered is particularly evident in the boxplot. A running of a *t-test* on the processed data also confirms as just observed. In fact, with H_0 that there is not a significant difference in mean value of space covered between two treatments and using $\alpha = 0.05$ as criterion of significance, the null hypothesis is completely rejected ($T_0 = |-8.2755| > 2.1009$, $p = 1.5079e^{-07}$). Without running a one-tailed t-test and observing the boxplot shown in figure 4.9(right), it looks clear that the space covered by the estimated point of attention on the surface of the rack during ExpAR is significantly shorter than the corresponding one in ExpTM. In the bar plot, subjects in both treatments were arranged in ascending order of space covered. Comparing the shortest path followed by the best subject in ExpTM (the second bar of the first pair on the left) to the longest path covered by the worst subject in ExpAR (the first bar of the last pair on the right), the length of the latter resulted 40% shorter than the length of the former one.

4.4.4 User evaluation

Once the testing procedure was completed, each participant was asked to fill a NASA TLX questionnaire [70]. NASA-TLX is a multi-dimensional scale (consisting of six twenty-point rating scales) designed to obtain workload estimates from one or more operators while they are performing a task or immediately afterwards. To evaluate
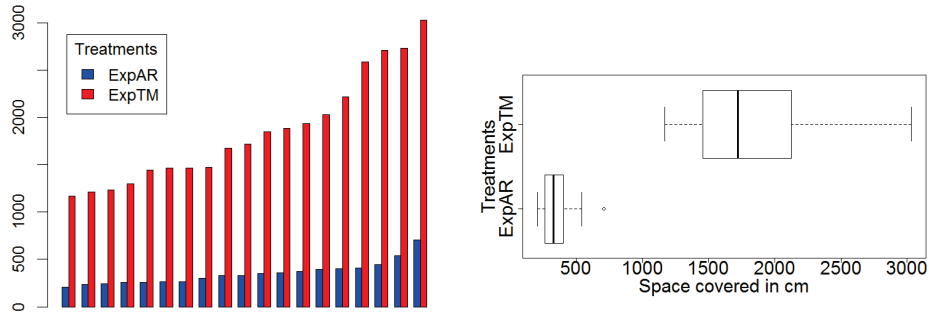
FIGURE 4.9: Barplot (left) and boxplot (right) of space covered in ExpAR treatment (blue) and ExpTM treatment (red). Participants in both treatments have been sorted in ascending order referring to the length of the path followed by the estimated observed point.

TABLE 4.5: Statistical significance for each NASA TLX index in two experiments.

Rating category	W-value	p-value	Significant
Mental demand	28.5	8.7e-06	Yes
Physical demand	240.5	0.0808	No
Temporal demand	59	0.004	Yes
Performance	105	0.0272	Yes
Effort	142	0.2632	No
Frustration	57	0.0003	Yes

the overall workload the raw task load index analysis was exploited [119]. The average scores are shown in figure 4.10 and the statistical analysis on them has been summarised in table 4.5. Considering that samples came from scores of a questionnaire and considering that most of the populations were non-normally distributed, different running of two-tailed Mann-Whiteny test on each pair of rating in the two treatments were performed. In the null hypothesis there is no significant difference in average rating between two treatments using $\alpha = 0.05$ as criterion of significance. Discussing the details and the implications of the analytical evaluation of the results is not trivial and difficultly allow making generalisations. Individual rating systems typically require a very big number of samples to avoid any bias in the results. With big population of samples a strict analytical study may be consider sufficient to draw
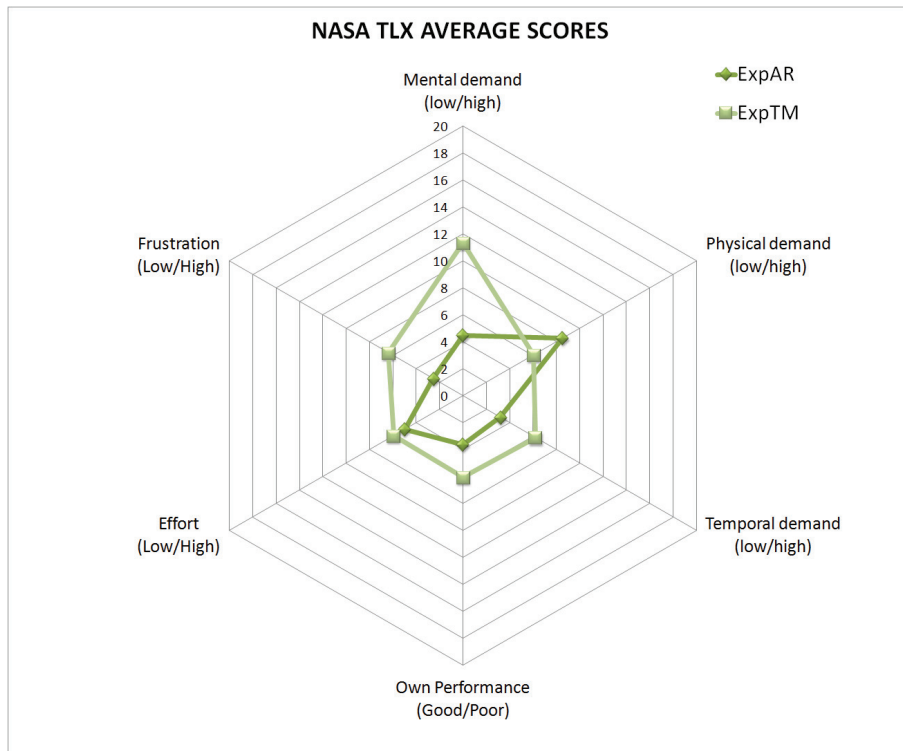
FIGURE 4.10: Hexagonal plot of mean rate to NASA TLX indices in both treatments. The ExpAR hexagon is almost totally inscribed in the bigger one from ExpTM treatment. This implies that, in mean, participants in ExpAR reported a better ranking to each index that is inscripted in ExpTM. The only one that is higher in ExpAR compared to ExpTM is the physical demand. This index is particularly sensible to the HMD and its ergonomic features.

conclusions. In this case, considering the limited size of the population, the numerical answers to the questionnaire have been combined with considerations made after the subsequent informal talk.

The informal talk was intended to create a more colloquial condition where subjects felt free to express their doubts or opinions about the testing session experienced. The main aim here is to draw some conclusions based on the analytic interpretation of the questionnaire results, eventually weighted by considerations coming from the informal talks recording, to provide a more complete picture. By observing the visual presentation of the scores in both treatments, the hexagon formed by six mean

values of ExpAR is almost completely subscripted in the bigger one obtained from mean scores in ExpTM. Only one point, i.e., the physical demand, of the questionnaire in ExpAR resulted higher than the corresponding in ExpTM. Looking carefully in Table 4.5, results revealed that even if physical demand in ExpAR is higher, the null hypothesis has been accepted ($p = 0.0808$). This means that the average score on physical demand is not significantly different from that in ExpTM. Concerning all other indexes, the average score in ExpAR resulted always smaller than the corresponding one in ExpTM. However, only for four of them the null hypothesis was completely rejected (i.e., mental demand, temporal demand, performance and frustration). Remembering that the main objective of the testing sessions was to assist non-skilled technicians in properly completing an unknown maintenance procedure, the result reported about the significance of the difference in mental demand is encouraging. It proves, indeed, that by using ExpAR there is a real added value in reducing the mental effort needed to understand the procedure and to locate the point of intervention. Informal talks also confirmed these conclusions. Most of the participants in ExpAR positively accepted the visual augmentation of the working area considering it as a very useful aid. Of course, the opinions gathered from subjects involved were not always free of criticisms about the mixed reality paradigm. About the frustration index, intended as a measure of how much stressing and discouraging was the sequence of tasks to be performed, it also reported a significant difference in favour of ExpAR compared to ExpTM. About temporal demand, that is the feeling of pressure of the user on completing each task, the average score in ExpAR resulted smaller (better) than that in ExpTM. Although the null hypothesis for temporal demand have been rejected, it was interesting to note that a significant number of subjects in ExpAR treatment expressed a negative opinion about the time management in each task. Each task of the procedure was accompanied by a synthesised speech providing information about the task to be performed rather than showing a long label in augmented reality that might be difficult to read. Interestingly, the subjects mentioned above considered this decision as a weakness of

the system. The debriefing revealed that even if they understood in advance the kind of intervention to be performed, they waited until the end of the synthesised speech before starting to perform the task. For this reason, they considered this as a negative aspect that affected the time needed to complete the procedure.

Considering that in expAR the total time of completion was significantly smaller than that in ExpTM (see figure 4.4, the issue raised from the temporal demand index can be considered as a double benefit. On one hand it suggests improving the interaction paradigm. On the other hand, it demonstrates that total completion time would have been even smaller in ExpAR, further reducing the time needed to complete the sequence of tasks, compared to ExpTM. Concerning the performance index, most subjects judged their performance as of a good level in both treatments. Even if the difference is significant ($p = 0.0272$), the informal talks suggest that participants were as much satisfied by their performance in ExpAR as in ExpTM.

The score registered for the physical demand index reveals an interesting aspect that appears in contradiction with the considerations about the general benefits related to usage of augmented reality. From the analysis of linear and angular velocities, indeed, it was observed that subject in ExpAR reported mean velocities significantly smaller than that reported in ExpTM. The conclusion is that with the increasing of the procedure complexity, the physical load in ExpTM grows faster than in ExpAR. This conclusion is apparently in contrast with the higher physical demand score measured in ExpAR. Also In this case, the informal talk was particularly useful to understand the reasons that led to this result. Six of ten participants in ExpAR complained about the discomfort of the HMD. In particular, they reported that the HMD was too heavy, since the weight of the visualization device is totally centered on the user's nose. Recalling the details of the architecture of the system, a Silicon Micro display was chosen for its good display resolution. Since the designing of the experiments, the unbalanced weight of this display on user's nose and its limited ergonomics in this sense were known. To address this issue, an elastic band surrounding the user's head was used to reduce the weight of the display. Unfortunately, the

camera mounted on the HMD made the visualisation device a little heavier. As a consequence, the display was still judged uncomfortable to wear, accounting for a higher physical demand value in ExpAR and an effort rating equivalent to that in ExpTM. These findings suggest to experiment other types of HMDs to evaluate how much the users' ratings would be affected by this gear. We believe that a device like the Oculus Rift DK2, expanded with a matching video see-through option, represent an interesting design of wearable visualization system with a good trade-off in terms of performance compared to the costs.

4.5 Conclusions

A comprehensive evaluation study on the application of Augmented Reality was presented. The AR prototypes were designed around the optical image-based tracker proposed in this thesis in two different case studies, that is industrial and telemedicine fields.

In the industrial case study, the experiments performed on two guidance solutions (one exploiting AR and the other relying on conventional computer based manuals) enabled both quantitative and qualitative analysis of the data collected from the working sessions operated by skilled and non-skilled engineers. Based on these results obtained, it appears evident the potential of augmented reality in the field of maintenance and assistance. Participants in AR treatment reported a significant shorter time of completion of the maintenance procedure compared to those performed with current traditional methods. Moreover, the space covered in terms of head movements also resulted significantly different in AR compared to traditional technical manual treatment. Though this results could be considered not surprising, they have been obtained in a real industrial environment by a significant higher number participants compared to the most recent works available in literature, daily committed to work hands-on on complex electronic equipment with a strong request

of effective and efficient intervention motivated by the uninterrupted nature of the air surveillance routine.

In the field of telemedicine, the study has highlighted the possibility of using Augmented Reality to support untrained user while performing an ECG-test. It demonstrates that other home-care solutions, where an help may be request to interact with medical devices or emergency applications (e.g. use of a defibrillator by an inexpert user), can take advantage from AR principles and approaches. However, some issues and question should be fully addressed before moving to clinical applications. While no particular difficulties arose in using AR with medical devices (mainly because they are rigid and their 3D geometry known and fixed), problems occur when "augmenting" a generic, alive, human body: large anatomical variations and patient's motion can be accurately taken into account. It is worth to remember that even if some parts of the human body can be roughly considered rigid, change in proportions should be accounted, e.g., between adults of infants, but also between different body types. Some parts of the body are in continuous motion such as the chest during breathing. However, in the conducted trials, with the patient breathing normally, this phenomenon did not significantly affect the results. Being essentially a proof of concept, the study proposed in this work of thesis does not pretend to address all these problems nor to provide solutions to them.

Augmented reality approach aside, the results reported have to be considered just a starting point toward a really usable AR guidance systems. Many usability and ergonomic aspects still have to be taken into account, particularly concerning the design and the engineering of see-through HMDs. Recent low-cost, wide field-of-view and light HMDs proposed by electronics giants (Samsung, Sony, etc.) and aggressive startups (Oculus) seem encouraging in this sense.

The experimental analysis conducted on the case studies revealed a positive impact of augmented reality with optical image-based tracking proposed in this study. Existing head-tracking systems all suffer from various limitations, such as latency, cost, accuracy, or drift [120]. Providing an accurate estimation of the position and

rotation of the point of view of the user in the space is a non-trival task. Factors like low computational consumption, fault-tollerance, ease of installation and so on make hard finding a tradeoff that meets all of them. The results obtained during the experimentations revealed that the optical tracker proposed in this study resulted robust in estimating translational movements. Conversely, the results revealed that high rotational velocities due to rapid movements of human's head are hard to control by using common camera lenses. The continuous improvements in this direction, with the enhancements of the refresh rate of the cameras, makes reasonable to suppose that this could significantly improve the performance of the proposed tracker. However, it must be admitted that single sensor tracking systems are generally non-sufficient to provide a complete and reliable 6DOF tracker. Nowadays, gyroscopes and digital compasses can efficiently and accurately estimate the rotational movements but they do not provide any information about translations. Hybrid approaches, obtained by the combination of two or more tracking systems, seem to properly address all of these issues. They allow to benefit from the strengths of each of them discarding the weaknesses. For this reason, the future direction of this study aims at integrating the proposed marker-based tracker with inertial sensors thus trying to overcome the problems that led to tracking failures during the experimentations. By increasing the efficacy of the tracking, the practical advantages of AR technology could be even greater than what the study presented in this work of thesis suggests.

Personal contributions

Parts of the topics and experimental results discussed in this chapter are from the following research papers of which I am co-author:

1. An image based approach to hand occlusions in mixed reality environments. In Virtual, *Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, Springer.[47]

2. Mixed reality environment for mission critical systems servicing and repair. In *Virtual, Augmented and Mixed Reality. Systems and Applications.*Springer [48].

3. Telemedicine supported by augmented reality: an interactive guide for untrained people in performing an ecg test. *Biomedical engineering online*[110].

4. Raffaele Vertucci, Fabio Narducci. Mixed Reality in Advanced Maintenance and Logistics. In *Polaris Innovation Journal*, 2014.

5. Measuring the Benefits of Augmented Reality in Industrial Environment. Submitted to *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (**under review**)[121].

Chapter 5

Iris detection and recognition

5.1 Outline

This chapter discusses a comprehensive proposal for iris-based person authentication on mobile devices. According to the most recent statistics about the worldwide diffusion of mobile communication technologies, 4.55 billion people (accounting for 63.5% of the global population) are expected to use a mobile phone in 2014, with a fraction of 1.75 billion using a smartphone. Market experts estimate smartphone adoption will continue at fast pace and it is likely to reach 50% of the mobile users in the next three years, while in Europe and North America this figures are considered rather conservative as mobile phone usage is already close to ubiquitous. Along with the aforementioned smartphones diffusion, a great numbers of applications has been released and downloaded to perform commercial transaction or mobile banking by means of token-based authentication systems. As smartphones are representing the first real form of mass ubiquitous computing, reliable assessment of owner's identity is becoming a key factor for a wide range of present and future applications for which using a password might not be the most appropriate choice. Biometrics could reasonably represent a more practical and reliable way to perform

person authentication by exploiting the sensors typically embedded in this class of devices (multiple cameras, microphone, accelerometers, gyros, digital compass, fingerprint sweep-scanner, etc.). Face is perhaps the first biometric proposed for person-authentication on mobile devices[7], eventually combined with voice [122], but the growing imaging quality featured by the last generations of smartphones and tablets provides the technological premise to perform iris recognition as well. Iris, indeed, is inherently characterized by high discriminating power and stability over time [123] and it therefore represents a good candidate for the aforementioned applications, though its small size (compared to the whole face) makes capturing the intricate texture details responsible for its saliency a challenging task. To this regard, the last generation of smartphones and tablets features advanced micro cameras capable to capture images at a resolution of up to 12 Megapixels or even full HD video, while keeping noise levels much lower than in the past for daylight shots and acceptable performance in indoor condition.

Considering these specifications, iris acquisition and matching on mobiles seems a feasible option that deserves an in-deep investigation. Firstly, a short overview of iris recognition method is proposed with a special attention to the reason that motivated the work presented in this thesis. After that, following sections discuss in details the proposed method from the step of iris segmentation to the recognition. Inspired by the trend of recent years, the proposed technique exploits the concept of homogeneity and separability for iris segmentation [124] and the spatial histograms (also called spatiogram) for iris matching recognition.

5.1.1 Background

has been explored as a biometric firstly in the seminal work by J. D. Daugman [125] who assessed the statistical independence of two coded patterns originated from different eyes and described a method for iris localization based on integro-differentials operators, the use of 2D Gabor filters to extract iris texture features

and a statistical-based approach to iris codes matching. In 1996, Wildes patented a non-invasive system for iris recognition and compared that to the one from Daugman [126]. These pioneering works were mostly focused on achieving maximum accuracy in iris recognition under controlled conditions including specific enrolment protocols for the user to undergo.

The following studies on this biometric have progressively targeted from the one side iris segmentation and from the other side its recognition under less predictable acquisition conditions, involving uncontrolled lighting and environmental factors which can result in noisy iris images (for instance due to strong reflections over the cornea surface, blur, low contrast, etc.). To this regard [127] proposed an approach to represent iris features by means of wavelet transform zero crossing. This descriptor is translation, rotation, and scale invariant and very slightly affected by variations in illumination and noise levels. In [128] the Haar wavelet transform is used to optimize the dimension of feature vectors to 87 bits, to the aim of reduce processing time without affecting accuracy of recognition. The authors combined this descriptor with a method of initializing weight vectors and a method of determining winners for recognition in a competitive learning neural network, resulting in a method accurate even for "real world" applications. In 2007, Proença and Alexandre [129] presented the results of the NICE.I contest for the performance evaluation of recognition algorithms on noisy iris images, that resulted in a measurable robustness improvement of the state of the art, particularly for iris segmentation. More recently, Bowyer[130] resumed the results of the NICE.II Iris Biometric Competition, focused on performance in feature extraction and matching, arguing that *since the top-ranked algorithms seem to have relatively distinct technical approaches, it is likely that a fusion of the top algorithms would result in further performance improvement.* In this line of research, Jeong et al [131] presented a new iris segmentation method that can be used to accurately extract iris regions from non-ideal quality iris images while Shin et al. [132] proposed an integrated iris recognition method that discriminates the left or right eye on the basis of the eyelash distribution and specular reflection

and exploits iris region color and texture information to achieve a reliable classification. Among the first to investigate the possibility to optimize iris segmentation and recognition algorithms specifically for mobile phones Jeong et.al., proposed a method of extracting the iris code based on Adaptive Gabor Filter in which the Gabor filter's operating parameters depends on the amount of blurring and sunlight in captured image. To the aim of improving the robustness of iris recognition on mobile phones in various environments, Cho et al. [133], in 2005, presented a pupil and iris localization method exploiting not only information of the pupil and iris, but also the characteristics of the eye images. In [134] a pupil and iris localization is based instead on detecting dark pupil and corneal specular reflection by changing brightness and contrast value. On a similar line of research, Park et al. [135] presented a recognition method for mobile phones based on corneal specular reflections while Kang [136] proposed to pre-process iris through an automatic segmentation of pupil region, pupil and eyelids detection to remove the most noise from the iris image and improve recognition performance. Finally, as the mobile computing power is approaching that of low-end desktop computers, combine face and iris recognition on mobile system have been proposed, proving that multi-biometrics person authentication on mobiles can be a feasible option [137, 138].

5.1.2 Segmentation of the iris

The iris segmentation algorithm used in this study is ISIS [124] and it is composed by four stages: iris image pre-processing; pupil localization; image linearization and limbus localization, which are described in the following lines. Homogeneity and separability are important concepts used for the selection of the best circle, which is a candidate to represent the pupil. The pupil can be considered as a circular region with a homogeneous distribution of pixels. Some approaches are based on the assumption that the darker region within the image is the pupil. Nevertheless, this is not always true because the pupil may changes its appearance in relation to

the lighting. The proposed scoring function is based on the gray tones histogram of the pupil. Each circular region receives a score according to the homogeneity degree of the pixels it contains. Let H be the region histogram, s_H will be the maximum number of occurrences of the same value in that region:

$$s_H = \max_i [H(i)] \, /_{i=1}^{255} H(i)$$

As for the limbo, also the outline of the pupil has a zone in which it passes from a dark color to a lighter one; of course by analysing an iris dark in color, this transition is more subtle. Therefore it was defined an index of separability. Given a candidate circle C with center $c = (c_x, c_y)$ and radius

rho in the image I, the Cartesian coordinates are given by:

$$x_C(\rho, \theta) = c_x + \rho \cos \theta$$
$$y_C(\rho, \theta) = c_y + \rho \sin \theta$$

where $\theta \in [0, 2\pi]$. Considering the circle C_{IN}, internal to C, with radius $\rho_1 = 0.9\rho$ and the circle C_{OUT}, external to C, with radius $\rho_2 = 1.1\rho$; measuring the difference of gray tones, on the edge of the circle for each angle θ_i, using an operator similar to the Daugman's integro-differential operator, structured as follows:

$$D(i) = I(x_C(\rho_2, \theta_i), y_C(\rho_2, \theta_i)) - I(x_C(\rho_1, \theta_i), y_C(\rho_1, \theta_i))$$

where $i = 1, ..., 360$ represents the discrete value of the angle and then the index within the gradient vector h; while $\theta_i = i\pi/180$ is the same angle, in radians. At the pupil, we expect a high and constant value for D, i.e. an average high and low variance $\theta(D)$. Based on these observations, the index of separability can be defined as:

$$S_D = \frac{\bar{D}}{\theta(D) + 1}$$

By analysing the polarized image of eye in the horizontal direction, (see Figure 5.1), it is possible to accurately define the region of separation between the iris and the sclera, which corresponds to limbo. However, the captured iris image contains some information that is not useful for the localization of the pupil. This is the case for features such as the pores of the skin, eyelashes and eyelids, which negatively interfere with edge detection. The first stage of IS_{IS} eliminates interference through an enhancement filter. A square window W of size $k \times k$, is passed over the entire image pixel by pixel; for the window W, a histogram H_W is calculated and the value with the highest occurrence is replaced in a central position. A *canny* filter is applied
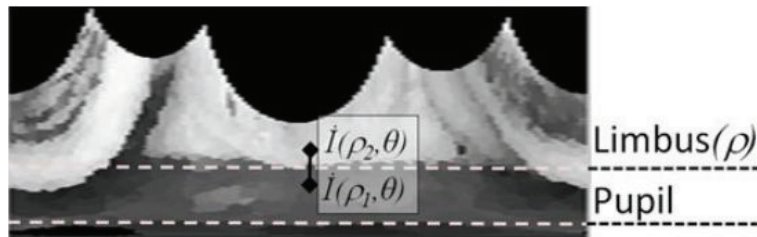
FIGURE 5.1: The regions of separation of iris in polar coordinates.

to the image resulting from the previous stage to locate the pupil. The canny filter is applied with ten different thresholds $t_h = 0.05, 0.010, 0.015, ..., 0.055$ (some examples are shown in figure 5.2). For each of the ten images the connected components are identified. Components that contain a number of pixels above a threshold th_C are included in a list L. Then, the algorithm of Taubin [139] is applied to each element of the list to calculate the corresponding circle. The circles that are not completely within the image are eliminated from the list L, in order to obtain a final list L_C. To identify the pupil, the circles of the list L_C undergo a procedure based on the homogeneity and separability criteria. For each circle the value: $s = s_H + s_D$ is calculated. Finally, the circle C_{max} with the highest s_{max} value is considered as the circular shape that best approximates the pupil.

At this stage, the pixels with the ρ_{max} distance from the center of the localized pupil are searched. Then, the image is transformed from Cartesian coordinates to
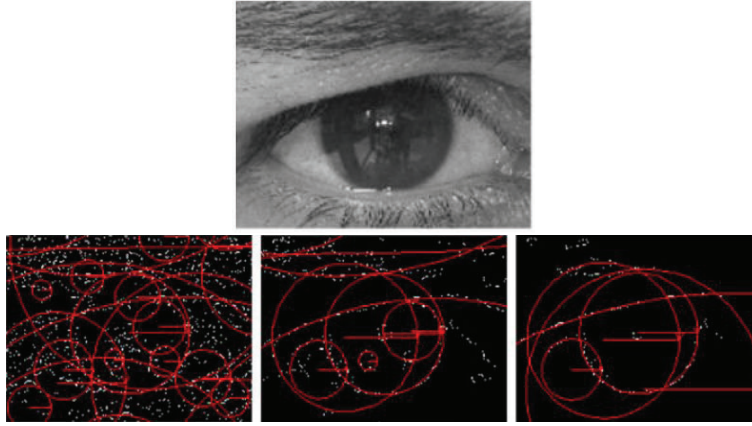
FIGURE 5.2: Canny edge detection filter applied on enhanced images at thresholds of 0.05, 0.15, 0.35 (from left to right).

polar coordinates, in order to obtain a new image $\dot{I}$. This transformation is made to locate the boundary between the sclera and the iris. Image I is further filtered with a median filter. If R is a row of the image, for each pixel P contained in R we consider the neighbourhood including *2q+1* pixels (i.e. himself, q previous pixels and q following pixel). Then the neighbourhood pixels are sorted and the pixel P takes the median value. For each column, which is located beyond ρJ and the corresponding position on the 'horizontal axis of i and θi, the following weighted difference is calculated pixelwise:

$$\Delta(\rho_j, \theta_i) = \varphi(I, \rho_j, \theta_i) \bullet (I(\rho_j + \delta, \theta) - I(\rho_j - \delta, \theta_i))$$

where

$$\varphi(i, \rho_j, \theta_i) = \left\{ \begin{array}{c} 1 \quad if \dot{I}(\rho_j + \delta, \theta_i) - \dot{I}(\rho_j - \delta, \theta_i)) > 0 \\ and \min(\dot{I}(\rho_j + \delta, \theta), \dot{I}(\rho_j - \delta, \theta_i)) > \epsilon_G \\ 0 \qquad otherwise \end{array} \right\}$$

It is worth noting that the pupil occupies the lower part ρ_J of the polarized image $\dot{I}$, followed by the iris and sclera. The sign of the difference is relevant as it is expected

that the sclera is brighter than the iris. This indicates that you should look for changes with a positive sign, which represent the transition region between iris and sclera. In the previous formula: the first inequality imposes a positive gradient; the second inequality excludes the pixels of the border between the pupil and iris, as it requires the darkest pixel in the pair to have a gray level greater than a threshold $\epsilon \in [0, 255]$. The area on the limb is composed of points that maximize the weighted difference for each column θ_i in I.

Figure 5.3 shows the whole process that, starting from an image containing an eye, leads to the segmentation of the iris from the representation in polar coordinates.

5.2 Spatial histograms for iris matching

For a given discrete function $f : x \leftarrow v$ where $x \in X$ and $v \in V$, a histogram of f counts the number of occurrences for each element in the range of f. In particular, the histogram is $h_f : v \leftarrow Z^*$, where $v \in V$ and Z^* is the set of positive integers, and $h_f(v)$ is the number of elements $x \in X$ such that $f(x) = v$. The histogram h_f can be seen as a binary function $g_f(x, v)$, where $g_f(x, v) = 1$ if $f(x) = v$ and $g_f(x, v) = 0$ otherwise. The moment of zero order of g on the dimension v is:

$$h_f(v) = \sum_{x \in X} g_f(x, v)$$

Histograms are relevant to iris segmentation because they discard all the information about the domain, so as to create the respective invariant for the one by one transformations of domain of the original function. A limited amount of information regarding the domain, can be extrapolated by means of higher order moments to the binary function g, where the *i-th* order moment is given by:

$$h_f^{(i)}(v) = \sum_{x \in X} x^i g_f(x, v)$$
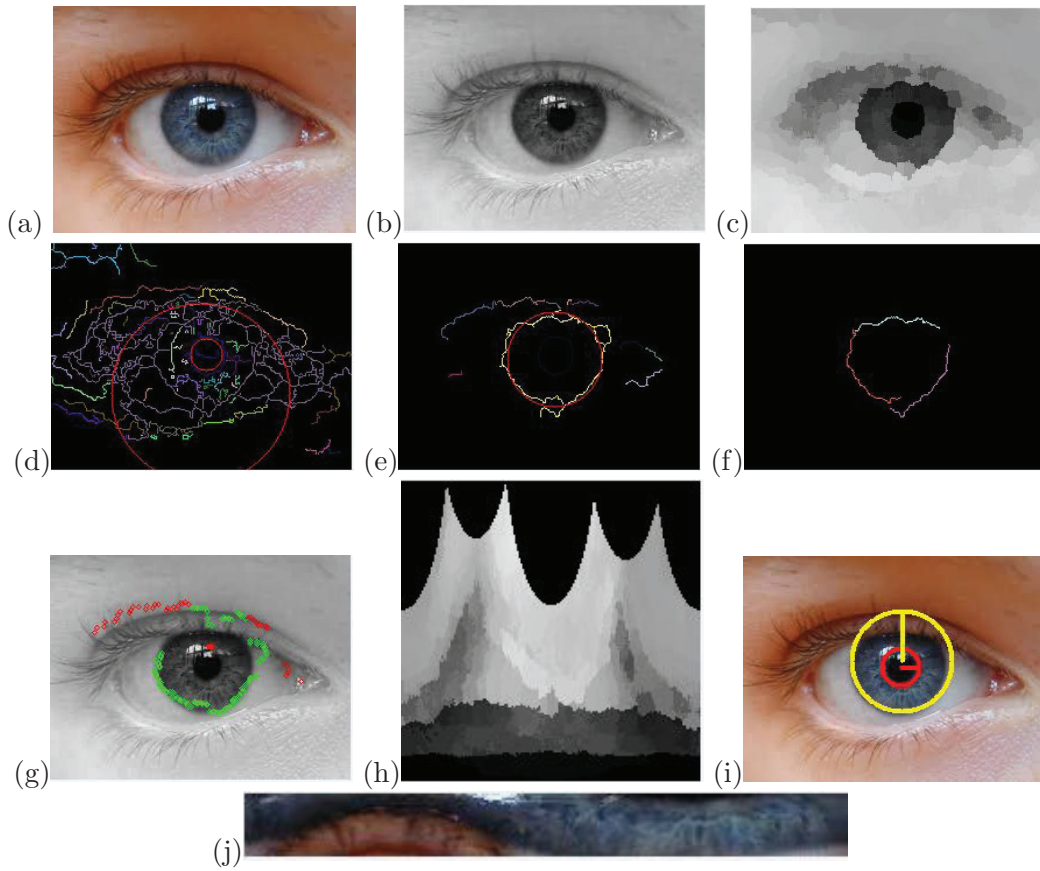
FIGURE 5.3: Partial result images obtained during the segmentation process. (a) The original image containing an eye. (b) the grayscale converted image. (c) The grayscale image posterization. (d-e-f) Consecutive rounds of Canny edge detector to look for the largest connected components. (g) The localisation of the frontier. (h) The representation in polar coordinates of the eye. (i) The localization of the iris and pupil. (j) The iris extracted from the eye discarding sclera and pupil.

To define this concept, we use the term spatial histogram or simply spatiogram; this because it captures not only the occurrences of information relating to the range of the function, such as the histogram, but also the information relating to the spatial domain. The *k-th* order spatiogram defines a tuple containing all the moments up to k:

$$\langle h_f^{(o)}(v), ..., h_f^{(k)}(v)\rangle$$

A histogram is thus only the zero-order moment of a spatiogram [140, 141]. The spatiogram can be also seen as a geometric model that allows arbitrary transformations and more specific models such as: translation, similarity, etc. As histograms, the spatiograms are efficient to calculate the differences between the correspondences of the images, without precisely calculate the geometric transformation between them. However, even spatiograms as more specific models, retain information about the geometry of the region of image.

The spatiograms, differently from the comparison of co-occurrences between arrays, capture the global position of the pixels instead of the relation between their pairs. An image is a two-dimensional map $I : x \leftarrow v$ of pixel $x = [x, y]^T$ with v values. The pixel value may represent an arbitrary value: gray tones, colors, or the result of a preprocessing (quantization, the color transformation of the space, etc..). The second order spatiogram of the image can be represented as:

$$h_I^{(2)}(b) = \langle n_b, \mu_b, \Sigma_b \rangle \text{ with } b = 1, ..., B$$

Where n_b is the number of pixels whose values are represented by the *b-th* bin, μ_b is the mean vector and Σ_b are the covariance matrices. $B = |V|$ is the number of bins in the spatiogram. Given h and *h'* as the enrolled spatiogram and the currently acquired one respectively, the similarity between them can be calculated as the following weighted sum:

$$\rho(h, h') = \sum_{b=1}^{B} \psi_b \rho_n (n_b n_b')$$

For a zero-order spatiogram $\psi_b = 1$. For a second order spatiogram, ψ_b can be seen as the probability that x_b is calculated by a Gaussian distribution described by multiplying the probability in the reverse direction:

$$\psi_b = \eta exp \left\{ -\tfrac{1}{2} (\mu_b - \mu_b')^T \hat{\Sigma}_b^{-1} (\mu_b - \mu_b') \right\}$$

Where η is the Gaussian normalization constant, while $\Sigma_b^{-1} = (\Sigma_b^{-1} + (\Sigma_b'^{-1})$ is a covariance matrix. It should be noted that the values of the summation are the average of the two Mahalanobis distances[142], the first one calculated between x and x' and the other one calculated between x' and x.

In order to assess the performance of the proposed algorithm, two public datasets were taken as reference, i.e., UBIRIS[129] and UPOL [143] (actually a subset of UPOL composed by 173 elements as probe and 173 as gallery and a subset of UBIRIS including 114 elements as probe and 228 elements as gallery). Although performing below the state-of-the-art, first experiments showed a good rate of successfully recognised irises. The performance on both dataset were compared and results are shown in figure 5.4 by using ROC curves (Receiver Operating Char-acteristic) and their related CMS curves (Cumulative Match Characteristic). Spatiograms performed slightly better on UPOL dataset than on UBIRIS. On greater values of GAR, the two curves are very similar to each other. For what concerns the CMS, reversed behaviour was reported. On UBIRIS it performed clearly better because the CMS value quickly raises to 0.7 for rank 1 compared to rank 9 on UPOL.
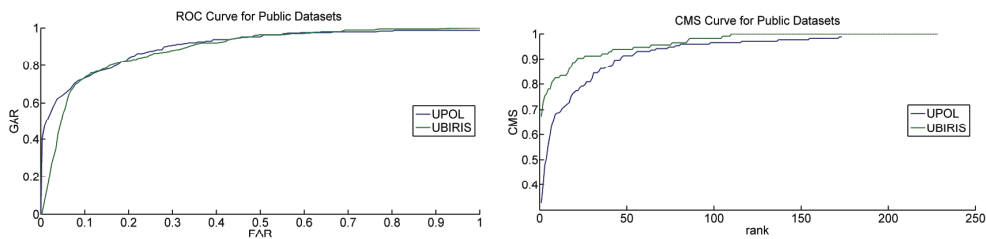
FIGURE 5.4: ROC and CMS curves comparison related to UPOL and UBIRIS datasets

5.2.1 Implementation on Android platform

The method of iris segmentation and recognition by means of spatial histrograms have been designed in order to be implemented on mobile platform. In particular, it was implemented in order to run on Android 4 (and later versions) compliant devices. The application consists of two functionality:

- iris segmentation by IS_{IS}

- iris registration and recognition by means of spatiograms

The modular software architecture is divided in:

1. *native code*: it represent the kernel of the application. It executes routine specifically written in C/C++ language as well as it calls routine from OpenCV (a well-known C/C++ open source library for computer vision);

2. *middleware*: that wraps the native code to the top layer of the architecture;

3. *graphical user inteface*: that was specifically designed to be adapted to several displays size.

The main programming language for Android systems is Java. By using the AndroidSDK and the IDE Eclipse, it is relatively easy to develop scalable and versatile application. The GUI is completely programmed by the XML markup language. It makes the design of the interface particularly light and efficient as well as easy to maintain. A screenshot of the application while running is shown in figure 5.5. The middleware is also a software module written in Java language but it is able to communicate with the lower layer of code that it is written in C/C++. That is made possible thank to the JNI (Java Native Interface) that is a programming framework that enables Java code running in a Java Virtual Machine (JVM) to call and be
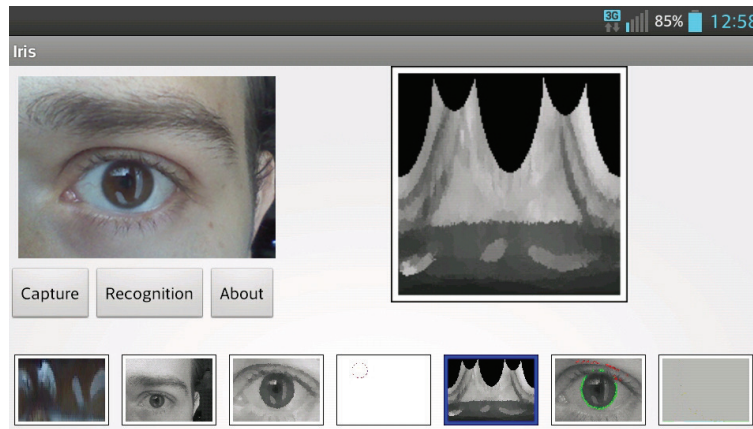
FIGURE 5.5: A screenshot of the iris segmentation that shows all partial images obtained during the execution of the algorithm.
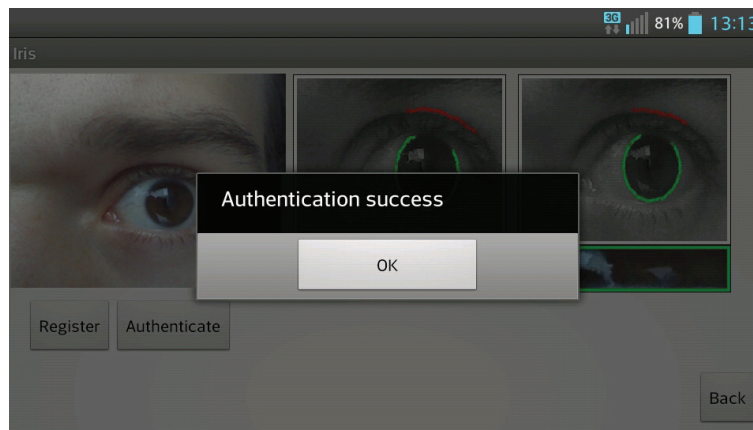
FIGURE 5.6: A screenshot of the recogition GUI that shows a succesful recognition of the irises provided.

called by native applications (programs specific to a hardware and operating system platform) and libraries written in other languages such as C, C++ and assembly. It represent a key aspect of programming for Android platform because it enables re-using of code written for other platform or purposes without the need of a drastic conversion and migration to Java programming principles. In fact, the lower layer of the architecture consists of several routines that were available only in C/C++. In order to develop the application discussed in this work of thesis, the just mentioned

routines have been partially customized for the new platform. Below, short details about the code of the application are provided.

The segmentation routine consists of four main tasks that are:

1 - Preprocessing e iris localisation: in this task an image is acquired by the camere of the mobile device. The function *cvLoadImage()* grabs a frame from the camera stream. Immediately after, the image is converted in *grayscale* (it improves the recognition of the regions in the image discarding information on colors that, moreover, increase the computational cost of the processing). The next step is the localisation of the eye by means of an Haar-like cascade classifier, trained so that can extract the best candidate regions that are signed as "eye" [144]. The function called is *cvHaarDetectObjects*. Follows an pass of enhancement that aims at improving the probability of correcty detect and segment the iris discaring other components like pupil, sclera, eye leads and so on.

```
1  /*load image*/
2  IplImage *input=cvLoadImage("/mnt/sdcard/eye.jpg",-1);
3
4  //grayscale transformation
5  compImg=cvCreateMat(height,width,CV_8UC3);
6  gray=convertImage(image,weight);
7
8  //execution of the cascade classifier to detect regions of interest
9  cascade = (CvHaarClassifierCascade*)cvLoad( cascade_name, 0, 0, 0 );
10 eye = cvHaarDetectObjects( gray, cascade, storage, 1.1, 2, 0, cvSize(20, 20)
      );
11
12 //Enhancement step
13 enhance= enhanceFilter(gray,enhanceStep);
```

2 - Pupil localisation: The Canny edge detector is applied to the imaged obtained so far.The OpenCV library *cvCanny()* is exploited to this purpose. After,

the pipeline aims at looking for connected components by executing the Taubin algorithm on the edges detected on the previous step. The pupil is then localised. In normal operative condition, the pupil is a little circle inscribed inside another one, bigger, that represent the iris. This method tries to look for this condition in the image, selection pairs of circles where one contains the other in its middle.

```
1  //Canny edge detector
2  cvCanny(enhance,canny,w*0.4,w,3);
3
4  //Connected components
5  components = connectedcomponent(canny);
6
7  //Taubin algorithm on all the extracted components by the edge detector
8  //This iterative process, looks for all circles allocated in the stack
9  //components until it discover that one circle is inscribed into another.
10 while (temp=(Stack*)stackPop(components)){
11      if (temp->size>compDim){
12        c=circleFitByTaubin(temp->first,temp->size);
13        if (c.radius){
14          iris=(iris *) malloc(sizeof(iris));
15          iris->pupil=c;
16          queueAdd(ccomp,iride);
17          if (verboseLevel&_circleDisplay){
18            cvCircle(compImg,cvPoint(c.x,c.y),c.radius,CV_RGB(255,0,0),1,8,0);
```

3 - Linearization of the image: it a pupil is found, the current step converts the image from Cartesian coordinates to polar by a function call *cartesian2Polar()*.

```
1  //if a pupil has been found it has a radious greater than 0
2  if (c.radius){
3  found.pupil=c;
4
5  //conversion from Cartesiane to Polar coordinates
6  polar=cartesian2Polar(enhance,cvPoint(c.x,c.y),0);
```

4 - Localisation of the limbus: in this step, the algorithm looks for the frontier region that separates the pupil from the iris and from the sclera. A function *find-Frontiers()* is called. Once the frontier has been found, the algoritm can extact from the image in polar coordinates only the region of pixels of the image that correspond to the iris. A scaling function produce a standard dimension of the iris exracted of 192×32 pixels.

```
1 //iris extraction from polar image
2 extracted=extSector(image, cvPoint(found.iris.x,found.iris.y),
      found.pupil.radius, found.iris.radius,0, 2*pi, 1, 1/found.iris.radius);
3
4 //Scaling
5 IplImage* iris_resized =
      cvCreateImage(cvSize(192,32),found.extracted->depth,found.extracted->nChannels);
6 cvResize(found.extracted, ris_resized, CV_INTER_LINEAR);
```

Once an iris is properly segmented, the recognition phase follows. The recognition has been implemented by generating the spatial histograms for the segmented iris that are then compared to a spatiogram of the iris of the same subject in order to test if they match or not. The function that converts the segmented image into a spatiogram is the following:

```
1 void getSpaziogram (char *path, double h[8],double mu[2][8] ,struct Matrice
      sigma[8]):
```

where the parameter *path* represent the extracted iris to be converted into a spatiogram. The paramater *h* is the histogram while the vector *mu* represent the centroids. The parameter *sigma* is the variance. *h*, *mu* e *sigma* are the paramters that represent the spatiogram of the image passed as input to the function.

The matching is then performed by comparing two spatiogram. The function is the following:

```
1 double compareSpaziogram(double h[8],double mu[2][8],struct Matrix
      sigma[8],double h1[8],double mu1[2][8],struct Matrix sigma1[8])
```

it receive in input the information from the function *getSpatiogram()* by simply comparing the vectors *h*, *mu* and *sigma* and returns a double $\in [0, 1]$ representing the level of similarity of two compared spatiograms. This information in other terms, represent the *distance* between two irises. If it results greater than a threshold, which represent the tolerance of the method, empirically set to 0.95 it means that the two irises come from the same subject, otherwise the recognition fails.

Personal contribution

Parts of the topics and experimental results discussed in this chapter are from the following research papers of which I am co-author:

1. Fast iris recognition on smartphone by means of spatial histograms. In *Biometric Authentication*, Springer, 2014. [138]

2. Ubiquitous iris recognition by means of mobile devices. *Pattern Recognition Letters*[141].

3. Biometrics Empowered Ambient Intelligence Environment. In *Accademia Peloritana Dei Pericolanti Journal.* (*in course of publication*)[112].

Chapter 6

Person Tracking in Video Surveillance

6.1 Outline

Efforts in biometrics are being held into extending robust recognition techniques to *in the wild* scenarios. Nonetheless, and despite being a very attractive goal, human identification in the surveillance context remains an open problem. This chapter discusses the collaboration of six months at the SociaLAB of the *Universidade da Beira Interior* in *Covilhã*(Portugal) for biometric video surveillance systems.

The biometric video surveillance systems it called "*Quis-Campi*" and it is expected to effectively bridge the gap between surveillance and biometric recognition while having a minimum amount of operation restrictions. The acronym of this proposal comes from the Latin and summarizes its goals: "Quis" stands for "Who is" and "Campus" refers to a delimited space. Hence, this project aims at researching and developing a biometric recognition system able to work in completely covert conditions. The main idea is that whenever a subject enters a QUIS-CAMPUS, it is automatically recognized, using multiple biometric sources and without requiring

any active participation from the subject side. The system is supposed to be completely automatic, attained by combining human detection and tracking, further enhanced by a PTZ camera that delivers data with enough quality to perform biometric recognition.

This chapter briefly describes the system which the SociaLAB is currently working on and concentrates the attention on the problem addressed during the PhD of which this work of thesis summarises the obtained results.

6.1.1 The system Quis-Campi

Since the systems is currently in a development phase, just a short overview is provided. In Figure 6.1 is depicted the main software architecture of the system. Essentially, it is divided into three layers that are:

- Hardware control

- Scene understanding

- Biometric recognition

Concerning the *hardware components*, a wide-view camera is mounted outdoor (approximately 5 meters above the ground), and pointing towards a parking lot. Although this kind of camera offers a more complete overview of the scenery, it does not provide enough quality for recognition methods to work at the distances the driveway ranges from (15 to 35 meters). To provide recognition methods with reasonable quality data, a PTZ camera is added to the system. This way, pointing and zooming over a specific region allows acquiring a detailed view of detected subjects. The two cameras are strongly related to each other so that it is possible to project a point in the 2D coordinate of PTZ camera to the wide camera and vice versa.

At the *scene understanding* layer, the system has two main modules that are (1)people detection and tracking, (2) facial landmark detection.
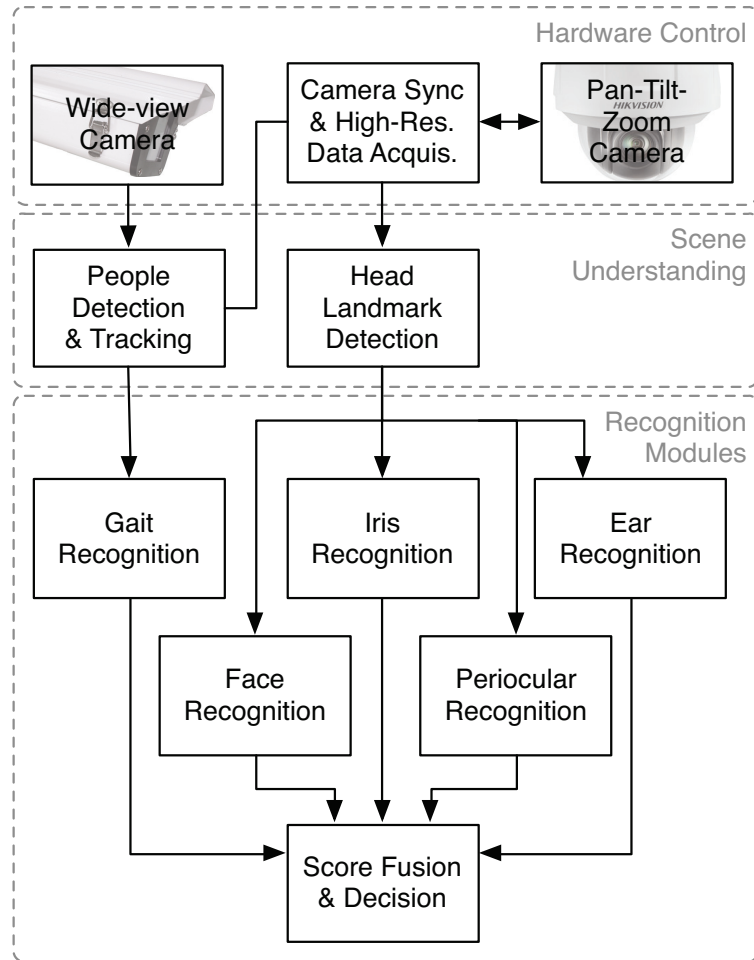
FIGURE 6.1: Working diagram of the proposed system, and the three layer architecture: hardware control (top), scene understanding (middle) and recognition modules (bottom)..

The first module is responsible for locating the people as they enter the scene, and tracking them until they are no longer visible. It takes as input the video feed from the wide-view camera, and has three main steps: background subtraction, upper-body detection and tracking. Further to that, the second module of the scene understanding layer is applied to that closer view, identifying which facial landmarks are visible, thus deciding the weight of each recognition module. For *recognition*

purposes the proposed system relies on a multi-modal biometric approach that combines face, iris, periocular, ear shape and gait information. Figure 6.2 shows how the system *Quis-campi* aims at using a pair of cameras (Wide/PTZ) for unconstrained biometric recognition in outdoor environment.
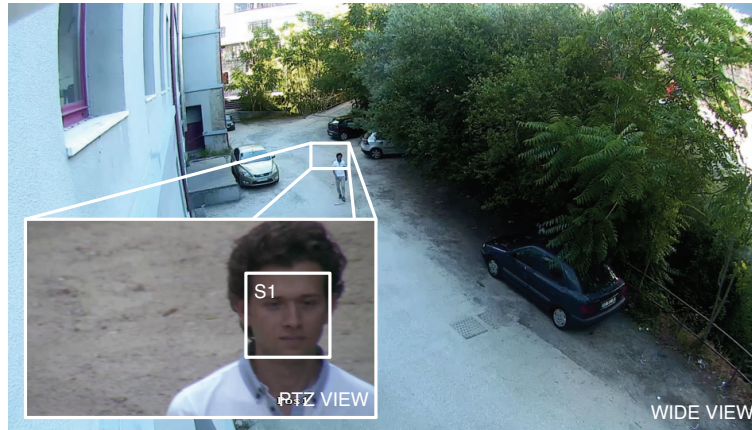
FIGURE 6.2: Illustration of the acquired data for both cameras.

6.1.2 Upper Body detector

In this section will be presented the contribution provided to the tracking component of the *Quis-Campi* architecture. Starting from the results obtained with the background subtraction, the next problem to address was that of detecting and tracking people, discarding every other moving object in the scene.

The specific nature of the problem, suggested to explore the possibility of using trained Haar-like cascade classifiers in order to detect people in the scene. The available choices were therefore detecting: (1)the whole body shape of people in the stream; (2) the legs only; (3) the upper body part only. Table 6.1 summarises the considerations arose during a preliminary study [145]. Moreover, person tracking in outdoor unconstrained environment is non-trivial several reasons, some of them are the following:

TABLE 6.1: Comparison of Cascade Classifiers for people detection.

	Full Body Detection	**Upper-Body Detection**	**Lower-Body Detection**
Features	The entire body's shape has lot of variations	Has fewer movements and is more stable and simpler	Has complex but stable and classifiable dynamics
Difficulties	Not easy to track but reliable	Relatively easy to detect and track	Non trivial but feasible
Drawbacks	Movements of legs and complexity of clothing	High false-alarm rate due to the simplicity of the shape	Legs have to be clearly visible in the stream
Purposes	Versatile	Head/Arms detection and tracking	Movement analysis

- Monocular RGB camera is really challenging

- People in crowd occlude each other making them hard to detect and track

- When people are far from the camera, the very low resolution of the crop might be hard to process.

- Lights/shadows transitions

- The position of the camera might play an interesting role. It affects the probability of having good samples for detecting, tracking and estimating.

| Wide Camera Stream | $\rightarrow$ | Upper Body Detector | $\rightarrow$ | LKT Tracker |

FIGURE 6.3: The work flow of the proposed tracking method

Based on the considerations above, the most reliable solution for people detection has revealed to be the Upper Body Detector. When counting people in crowd, the easiest and quickest way of doing that is by enumerating the heads that are visible. It is easier that the entire body shape is non-visible when more than one person is walking in the covered area rather than occlusions of the heads. However, upper body detection is computationally demanding. While developing a real time system,

the impact of the research of head in the stream can significantly impact on the performance of the system. The trick was exploiting the background subtraction mask that, reducing the research space, makes possible to run a cascade classifier on real time without introducing a high delay.

Although such improvements, it has been revealed particularly error-prone in outdoor scenarios, where lighting condition may suddenly change, occlusions produce detection failures and so on. In addition, it is strongly shape-dependant. For these reasons it was used only as a starting point for detecting and tracking people in the scene. The upper body detector is only in charge of looking for a possible Head-/Shoulder area. Once it recognises a head/shoulder, it stops the execution in that region because, on the same are of the image, a LKT tracker (Lucas Kanade Tomasi tracker [146], that is an optical-flow method for detecting and tracking features) is launched. In figure 6.3 it is simply showed the work flow of the proposed method. Until the LKT tracker is able to track the current head, it refresh the position and frame after frame keep track of the position of the person in the scene. The 2D position (x_W, y_W) in the coordinates system of the wide camera are then projected into the vector (x_{PTZ}, y_{PTZ}) expressed in coordinate system of the PTZ camera. This information is then used to correctly point the objective of the PTZ camera and, by zooming the region o interest, grab the biometric data suitable for a biometric recognition.

The algorithm is presented at the end of the section. Basically, the *MAIN()* is responsible for grabbing the stream from the camera. As soon as a new frame is available, a background subtraction mask is generated. The background mask is then combined with the original frame in order to recover RGB data from the original stream exclusively for moving objects in the scene. A run of the Upper Body detector detects heads in the scene. For each of them, a separated thread is started. Each thread is a running of the LKT tracker that looks for features and tracks them. All thread share the information about the update position of the head so that the main process can run the upper body detector only on regions that are not currently

FIGURE 6.4: Illustration of the preliminary results obtained by the proposed tracker: (left) sample image acquired with the wide-view camera; (middle) foreground regions attained by background subtraction; (right) people tracking module results. The blue dots highlights the point tracked over the time.

tracked. The LKT tracker, when loses some features, try to recover refreshing the feature list. If the tracked feature decreases under a threshold, it returns and removes the lock on the region of the image. By doing that, the upperbody detector can try again to look for a head in that region and launch a new thread on it.

```
1 MAIN()
2     Stream = videoStreamFromCamera();
3     for each frame \in stream do
4         mask = backgoundSubtraction(frame);
5         merged = multiply(frame, mask);
6         ub_Areas = upperBodyDetection(merged);
7         for each ub in ub_Areas do
8             keyPoints = LKkeypoints(ub)
9             LKT(keyPoints)
10        end
11    end
12 end
13
14 LKT(Keypoints)
15     while (1)
16     numFeatures = track(keypoints)
17     ub = updateUBArea();
18     if numFeatures < threshold
19         generateNewFeatureClose(ub)
20 end
```

TABLE 6.2: Tracking performance in our surveillance scenario, when using LKT. Performance metrics are Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), True Positive Rate (TPR), False Positive Rate (FPR) and mismatch (MIS).

Scenario	MOTA	MOTP	TPR	FPR	MIS
S1	0.940	0.600	0.970	0.030	0
S2	0.800	0.590	0.900	0.100	0
S3	0.745	0.336	0.862	0.138	0
S4	0.589	0.288	0.792	0.202	3

6.2 Preliminary results

A preliminary test to assess the reliability of the proposed method for tracking was performed. Four simple different scenarios with increasing level of difficulties (from S1, the simplest, to S4 the hardest) were considered during the testing. Here follows a brief description of each one of them:

S1: A single person starting to slowly walk away from a point nearby the camera; high quality background subtraction mask (absence of noise); no significant changes in lighting conditions;

S2: A single person starting to quickly walk away from a point nearby the camera; background subtraction mask with some noise; no significant lighting conditions;

S3: A single person slowly walking towards the camera from the distance; background subtraction mask with significant noise (wind and variable lighting conditions);

S4: three people to slowly walk away from a point nearby the camera; good background subtraction (little noise as consequence of the wind); one person crosses the path of other two.

Tracking performance was evaluated by referring to the CLEAR MOT metrics [147]. As reported in the table 6.2, a good level of accuracy (MOTA) was achieved in scenario S1 as well as for the precision (MOTP). The false positive (FN) and true negative (TN) rate are ignorable in this first experimental trial. For scenarios S2 and S3, even though they are different from each other, the results obtained in terms of MOTA and FN/TN rates were very similar, confirming an encouraging level of performance of the tracking algorithm. The difference between the two scenarios was significant with respect to the MOTP. A considerable loss of precision was reported for scenario S3. This result is strongly related to the fact that, at the beginning, the person is very far from the camera. The implication of this is that the omega shape (that is the head/shoulder area) is hard to be clearly detected at such distances. Once the head starts to be detected, the number of pixels in the region of interest is very short negatively impacting on the performance of the tracker. This test was really useful to suggest an alternative processing for people far from the camera. The omega shape at long distance is hard to detect but, reasonably, the whole body shape should be easier to detect. In that case, once a tracker detects the shape of a person, his/her head will appear on the top of the selected area. Scenario S4, which is the most meaningful case, reported an encouraging result in term of accuracy and the rate of true positive (TP) remains high. However, an increased rate of false positive was observed and, in addition to this, some mismatches (mainly related to the path of one person that crosses the path of other two). Although the loss of precision, it was possible to assert that the tracking method achieved good level of performances. Moreover, considering that it is at a very early stage, the results obtained opens to other investigations and combination of image processing methods to improve the quality of the whole pipeline that starts from the image background and to arrive to the prediction of movement.

Figure 6.5 shows the performance of the tracking method when more than one person is in the field of view of the camera. The square region represent the first time the Upper Body detector detects a head. Since that moment, on the selected regions

FIGURE 6.5: The image, obtained by overlapping several frames grabbed over the time in a single image, shows how the tracker is able to keep track of more than one people.

a LKT tracking thread starts. The coloured dotted lines shows the tracker at work. Observing carefully the first two people (from the left) are tracked continuously until they are at a distance from the camera so big to make impossible to track the features. Conversely, the third person (on the right in the image) is no longer tracked after he/she crosses the path of the other two. Since then, it is no longer able to detect and track him/her.

As anticipated in the introduction section, the current chapter mainly describes the contribution provided to the project *Quis-campi*. The systems is still at a developing stage and several improvements have to be done. At the time of writing, the tracking solution proposed was integrated in the prototype currently available and it revealed to be suitable to partially cover some of the ambitious amount of project goals.

Personal contributions

Parts of the topics and experimental results discussed in this chapter are from the following research papers of which I am co-author:

1. Quis-Campi: Extending In The Wild Biometric Recognition to Surveillance Environments. *to be submitted*.

2. Biometrics in Visual Surveillance : A survey. *being prepared for submission*.

Appendix A

Perspective geometry

Projective geometry is a branch of mathematics that deals with the relationships between geometric objects and their projections. While Euclidean geometry describes the world as it is, projective geometry describes the world as it appears or as it is seen. Euclidean geometry deals with properties of objects that are invariant under rigid motion, e.g. lengths, angles and parallelisms, whereas projective geometry deals with the properties that are invariant under perspective projections, e.g. incidences and cross-ratios.

If we were to close one eye and draw the world on a glass plate as we see it, we would notice that the distances between the points on a drawing differ from the true distances in the world, and the angles between the lines differ from the angles in the real world. Euclidean geometry is not able describe this, whereas projective geometry is. In this appendix, we focus on the aspects of projective geometry that are essential in many application areas of computer vision, including augmented reality.

Homogeneous coordinates

With Cartesian coordinates, we can conveniently present transformations in Euclidean space with matrix operations. For example, we can calculate a rotation of a rigid object by multiplying the Cartesian coordinates with an adequate matrix. Cartesian coordinates, however, are unable to perform transformations of projective space with matrix operations. For instance, it is impossible to present a translation in matrix form using Cartesian coordinates. Homogeneous coordinates make matrix calculations possible in projective space just as Cartesian coordinates do in Euclidean space. Homogeneous coordinates simplify calculations, as all the necessary transformations (and series of transformations) and their actions on points are presentable using matrices and matrix multiplications. This enables the computational efficiency required for realtime computer graphics systems such as OpenGL and DirectX.

Homogeneous coordinate vectors of a $point x = (x_1, x_2, x_3, ..., x_n)$ in Cartesian coordinates are all vectors $X = (x_1', x_2', x_3', ..., x_n', w)$ such that

$$\mathbf{X} = (x_1'/w, x_2'/w, ..., x_n'/w),$$

and vice versa, each point in projective space $X \neq 0, X = (x_1, x_2, x_3, ..., x_n)$ has a corresponding projection point in Euclidean space

$$x = (x_1/w, x_2/w, ..., x_n/w)$$

In projective space, scaling is unimportant and therefore in homogeneous coordinates

$$aX \equiv X \forall a \neq 0$$

Thus, a point in homogeneous coordinates is always equivalent to a representation with the last coordinate equal to one.

$$X = (x_1', x_2', x_3', ..., x_n', x_{n+1}') \equiv \frac{1}{x_{n+1}} X = (\frac{x_1'}{x_{n+1}'}, \frac{x_2'}{x_{n+1}'}, ..., \frac{x_n'}{x_{n+1}'}, \frac{x_{n+1}'}{x_{n+1}'}) =$$
$$(x_1, x_2, ..., x_n, 1)$$

Furthermore, the point $X = (X_1, X_2, X_3, ..., X_n, 0)$ correspond to a point at infinity in the direction of the line passing through $0 \in \Re^n$ and $(x_1, x_2, x_3, ..., x_n)$. The zero vector $0 = (0, 0, 0, ..., 0) \in \Re^{n+1}$ is non defined in projective space.

For consistency, the representations with the last coordinate equal to one are often used for homogeneous coordinates. The division operation to get the last coordinate equal to one is called homogeneous divide. It maps the vector to the real plane.

For simplicity, the projection to a lower dimension (e.g. perspective projection) is often first calculated in the original dimension. The dimensions are then reduced one by one by performing successive homogeneous divisions. For example, mapping the perspective projection into the image space is the projection $\Re^2 \rightarrow \Re^4, X = (X, Y, Z, 1) \rightarrow x = (x, y)$.

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ y' \\ z' \\ w' \end{pmatrix} \equiv \begin{pmatrix} \frac{x'}{w'} \\ \frac{y'}{w'} \\ \frac{z'}{w'} \\ \frac{w'}{w'} \end{pmatrix} \text{ and further } \begin{pmatrix} \frac{x'}{w'} \\ \frac{y'}{w'} \\ \frac{z'}{w'} \end{pmatrix} \equiv \begin{pmatrix} \frac{x'}{z'} \\ \frac{y'}{z'} \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \text{ which is }$$

$$\text{equal to } \begin{pmatrix} x \\ y \end{pmatrix}.$$

Geometric transformation

In AR as in 3D computer vision in general, perspective projections play an essential role. To be able to analyse a camera image, we need to understand how world coordinates are transformed into image coordinates, and this can be explained using projective geometry. Geometric transformations form a hierarchy of subsets

$$\text{Projective} \supset \text{Affine} \supset \text{Similarity} \supset \text{Linear (Euclidean geometry)}$$

where the transformation groups become smaller and less general and the corresponding spatial structures become more rigid and have more invariants as we go down the hierarchy. Homogeneous coordinates provide a framework for geometric operations in projective space. Euclidean geometry is a special case of projective geometry with more restrictions. Thus, it is possible to use homogeneous presentation in Euclidean geometry as well, if the operations are restricted to Euclidean ones. Accordingly, homogeneous presentation can also be used in affine and similarity transformations. Thus, all geometric transformations and their combinations can be presented with matrix multiplications using homogeneous coordinates. In a projective transform, only collinearity, cross-ratios and incidences remain invariant. Affine transformations also preserve parallelism and the ratios of areas.Similarity transforms preserve angles and length ratios. Euclidean transformations preserve angles, lengths and areas (see table A.1.
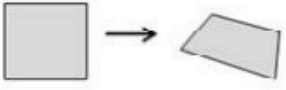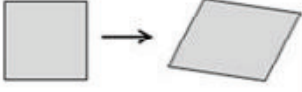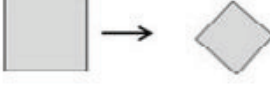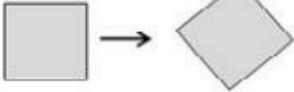
Transform	DOF	Invariants	
Perspective (Projective)	8DOF	Collinearity, Cross-ratios, Incidences	
Affine	6DOF	Parallelism, Ratios of areas	
Similarity	4DOF	Angles, Length ratios	
Linear (Euclidean)	3DOF	Angles, Lengths, Areas	

FIGURE A.1: Properties of different transformation spaces.

Starting from the discussion with linear transformations it propagates to affine and projective transformations. In the following, are also defined specific transformation types, such as rotations, translations and scaling, which are commonly needed in augmented reality.

Linear Transformation

A mapping $L : R^n \rightarrow R^n$ is called *linear transformation* if

$$L(u + v) = L(u) + L(v) \text{ and } L(cu) = cL(u)$$

Linear transformation $x' = L(x)$ can be presented in matrix form $x' = Lx$, where L is an *nxm* matrix. Matrix operations are easily expanded to homogeneous space. Any linear tranformation presented by matrix A can be expanded to homogeneous form

$$\begin{pmatrix} A & 0 \\ 0^T & 1 \end{pmatrix}, \text{ where } A \in \mathbb{R}^{NxN} \text{ and } 0 \text{ is zero vector, } 0 \in \mathbb{R}^N.$$

For example, scaling, rotation and shear are linear transformation. Rotation with an angle α around the x-axis is presented by a rotation matrix

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \text{ and } x' = R_x x$$

Similarly, rotations with an angle β around the y-axis and the angle γ around the z-axis are:

$$R_y = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \text{ and } R_z = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0a & 1 \end{bmatrix}.$$

Rotation around an arbitrary axis can be split into rotations around coordinate axes, and rotations around coordinate axes can be concatenated to present an arbitrary rotation. The rotation matrix (and the angles) depend on the rotation order and, for example, $R = R_z R_y R_x$ is

$$
R = \begin{bmatrix}
\cos\beta\cos\gamma & \cos\gamma\cos\alpha\sin\beta - cos\gamma\sin\alpha & \cos\gamma\sin\beta\cos\alpha + \sin\alpha\sin\gamma \\
\cos\beta\sin\gamma & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\gamma\sin\beta\cos\alpha - \cos\gamma\sin\alpha \\
-\sin\beta & \sin\alpha\cos\beta & \cos\alpha\cos\beta
\end{bmatrix}
$$

when written out. The written out forms are seldom used; it is more convenient to calculate separate rotations and multiply them. Sometimes people use the terms yaw, pitch and roll for rotations. The yaw is a right-handed rotation about the z-axis. The pitch is a right-handed rotation about the new (once rotated) y-axis. The roll is a right-handed rotation about the new (twice rotated) x-axis. Let φ, θ, ψ be the yaw, pitch and roll angles respectively, then the total rotation is

$$
R = R_x(\varphi)R_y(-\theta), R_z(-\psi).
$$

Occasionally, we may find it more convenient to represent the rotation by giving an arbitrary rotation axis and rotation angle around it, instead of dividing it into rotations around the coordinate axes. The rotation around an axis represented by a unit vector $u = (u_1, u_2, u_3)$ by an angle ω is

$$
R =
$$
$$
\begin{bmatrix}
u_1^2 + (1 - u_1^2)\cos\omega & u_1 u_2(1 - \cos\omega) - u_3\sin\omega & u_1 u_3(1 - \cos\omega) + u_2\sin\omega \\
u_1 u_2(1 - \cos\omega) + u_3\sin\omega & u_2^2 + (1 - u_2^2)\cos\omega & u_2 u_3(1 - \cos\omega) - u_1\sin\omega \\
u_1 u_3(1 - \cos\omega) + u_2\sin\omega & u_2 u_3(1 - \cos\omega) + u_1\sin\omega & u_3^2 + (1 - u_3^2)\cos\omega
\end{bmatrix}
$$

The rotation here is clockwise about the axis defined by **u** (right-hand rule). This can be presented with the Rodrigues formula in the form

$$R = I + [u]_x \sin \omega + (1 - \cos \omega) [u]_x^2,$$

where $[u]_x$ is a so-called cross product matrix, i.e., $[u]_x v = u \times v$.

The rotation matrix is an orthogonal (even orthonormal) matrix and thus it preserves angles and lengths. Furthermore, by the definition of orthogonality, the inverse of a rotation is its transpose

$$R^{-}1 = R^T$$

We can present scaling as a matrix operation using homogeneous coordinates

$$x' = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix},$$

where s_x, s_y and s_z are the scaling factors in the direction of coordinate axis.

Affine tranformation

Translation by a vector $t = (t_1, t_2, t_3)$, $x' = x + t = (x_1 + t_1, x_2 + t_2, x_3 + t_3)$ can be presented in matrix form only using homogeneous coordinates. Translation $x' = x + t$ is equivalent to multiplication by homogeneous transformation matrix $\mathbf{T}$,

$$T = \begin{bmatrix} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inverse operation to translation is translation by the opposite vector $\mathbf{-t}$, and the inverse of the translation matris is

$$T^{-1} = \begin{bmatrix} 1 & 0 & 0 & -t_1 \\ 0 & 1 & 0 & -t_2 \\ 0 & 0 & 1 & -t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Affine transformation A combines linear mapping and translation. A is an affine transformation if

$$x' = A(x) = L(x + t),$$

where $\mathbf{L}$ is a linear mapping and $\mathbf{t}$ is a translation vector. In the matrix representation we have $A = LT$ where $\mathbf{A}$ is now an affine transformation matrix, $\mathbf{L}$ is a linear transformation matrix and $\mathbf{T}$ is the corresponding translation matrix.

Affine transformation can be concatenated but not commutated

$$A1(A2(x)) = (A1(A2))(x)$$
$$A1(A2(x)) \neq A2(A1(x)) \text{ in general}$$

The non-commutative law means that the order of transformations is significant, and in AR application we need to do them in the correct order. The concatenation property allows us to multiply series of transformation matrices into one matrix. Affine transformations preserve lines and parallelism (i.e. parallel lines and planes remain parallel). They also preserve ratios of length, area and volume and the degree of a polynomial. Intersecting lines and planes are also transformed into intersecting lines and planes, but they do not preserve angles and shapes.

Let $\mathbf{L}$ be a 3×3 linear mapping and $\mathbf{t}$ a 3×1 translation vector, then

$$A = \begin{bmatrix} L_{3\times 3} & t_3 \\ 000 & 1 \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & l_{13} & t_1 \\ l_{21} & l_{22} & l_{23} & t_2 \\ l_{31} & l_{32} & l_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is an affine transformation.

Appendix B

Camera calibration

In this appendix, we discuss camera calibration, which is essential for computer vision applications including AR. We start the discussion with camera calibration and then review linear methods and non-linear optimization methods used for it. More profound surveys on camera calibration can be found in computer vision literature (e.g. [148]), to mention just one.

Camera calibration methods

Camera calibration means finding out the camera-dependent parameters for a scene model. Camera calibration includes, at least, approximating the intrinsic camera parameters and distortion functions. Sometimes it also includes finding out the extrinsic camera parameters (i.e. camera pose). Small differences in the camera assembly or physical variations in the individual lenses affect the parameters. Thus, for applications that require high accuracy, each camera needs to be calibrated individually, even if the cameras are of an identical model. Camera calibration methods are based on the detection of known control points. Differences between measured and calculated control point coordinates are used to construct an approximation of

the distortion functions and to estimate the intrinsic parameters. *Metric cameras* are cameras specifically designed for photogrammetric tasks. They have a robust mechanical structure, well-aligned lenses with low distortion and a lack of autofocus and other functionalities that may, uncontrollably, change the internal geometry of the camera. For metric cameras, the manufacturer usually provides exact intrinsic parameters as well as distortion coefficients, which simplifies the calibration process. However, most augmented reality applications are designed for use with ordinary (non-metric) cameras. We can divide *non-metric cameras* into professional (high-quality) cameras and consumer (low-cost) cameras (e.g. USB cameras and camera phones). Nonmetric cameras may have some of the features of the metric cameras but not all of them. In addition, autofocus, zoom lenses, image stabilizers, etc. can reduce the potential accuracy of a given camera. Augmented reality applications designed for the mass market (e.g. games) are often based on low-end cameras, with various distortions and calibration results clearly affecting the visual quality of augmentation. Thus, camera calibration is an essential stage of augmented reality application development.

Calibration pattern

Calibration methods commonly assume that the system can obtain an accurate set of correspondences between known world and image points. Calibration is often done using a known calibration pattern, calibration rig, e.g. a chequerboard pattern, but calibration is also possible using random features. The system can obtain these correspondences using, for example, a predefined planar calibration pattern, e.g. a chequerboard with a known pattern size (see Figure B.1). It knows the exact position of each corner point in the pattern and can then use corner detection and/or line detection methods to find the corresponding points from the image. The advantage of a single planar calibration is that it is extremely easy to produce. However, there are certain limitations with a single planar calibration pattern. The
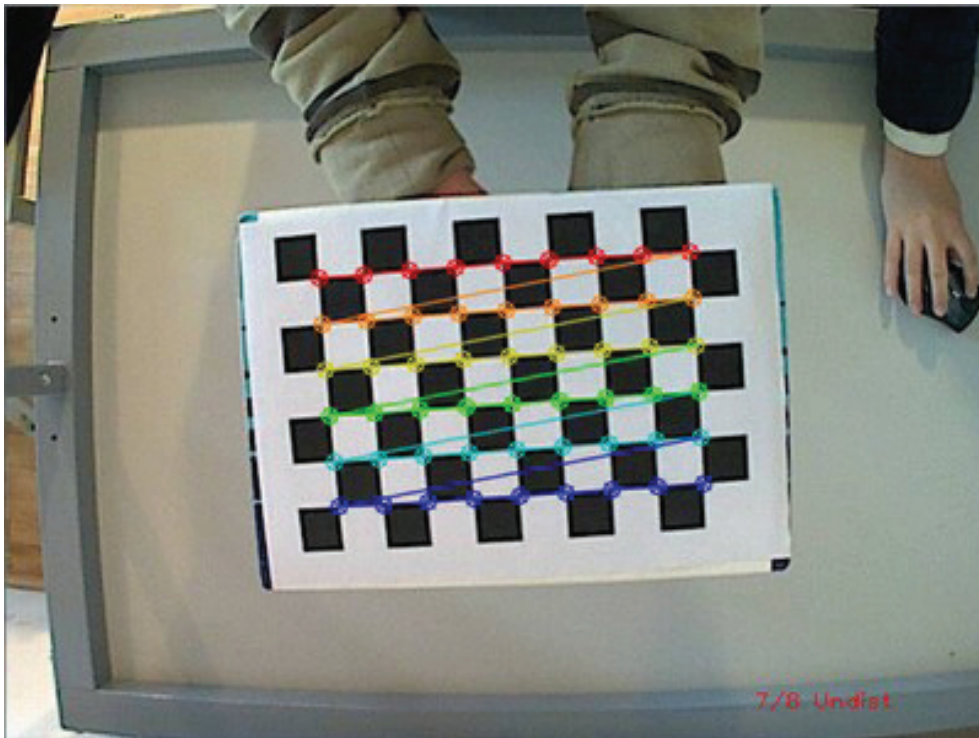
system can overcome these using a multi-planar calibration object, e.g. two (or more) perpendicular planes consisting of chequerboard patterns of known sizes. This kind of calibration pattern is relatively simple to produce. Another solution is to use a 3D calibration pattern of known formation, e.g. detectable points at the end of sticks pointing in different directions. A 3D calibration pattern is more complex to produce as it is not printable. It is also possible to implement calibration without a calibration rig using feature detection. In feature-based calibration, features are detected and tracked, and the system calculates the camera movement and parameters based on the behaviour of the features.

Calibration using a special calibration rig is easier to implement than feature-based calibration and is thus more common. For example, in AR applications the widely

used ARToolKit [49] provides a calibration tool with which the user prints a chequer-board pattern and moves it around in front of the camera. Similar calibration tool-boxes are also available for Matlab and OpenCV. Although feature-based calibration is more difficult to implement than methods using a calibration rig, feature-based methods are more user friendly as the calibration can be automated. For success-ful auto-calibration, the system needs a relatively large number of well-distributed points and at least one image must have a roll angle that is significantly different from the others. However, the common user is unfamiliar with camera calibration and its requirements. The system may therefore need to guide the user to move the camera in an appropriate way. Depending on the type of application and sys-tem assembly, it may be more convenient to solve the extrinsic parameters in the calibration process or at least the initial pose of the camera. For stationary cam-eras, it is convenient to solve the camera position and pose once with high accuracy rather than use valuable processing capacity to solve it at run-time. For rotating cameras the system only needs to discover the rotation for each frame in run-time, and for moving cameras it needs to calculate the pose for each frame. For fixed lens cameras, the focal length is static and the system needs to solve it only once. For zoom cameras, the focal length changes, and even with the same zoom step (depending on the mechanical accuracy) it may vary. In the case of the zoom-lens camera, the system therefore needs to approximate the focal length every time the zoom factor is changed. The principal point may also change with zooming cam-eras. A change in image resolution (digital zoom) does not require new calibration for intrinsic parameters, just scaling.

Intrinsic and extrinsic parameters

There are different approaches to calculating the intrinsic and extrinsic parameters for a specific camera setup. *Extrinsic camera parameters* are those define the loca-tion and orientation of the camera reference frame with respect to a known world

Object Coordinates (3D)
↓
World Coordinates (3D)⋆
↓
Camera Coordinates (3D)⋆
↓
Image Plane Coordinates (2D)⊗
↓
Pixel Coordinates (2D, int)⊗
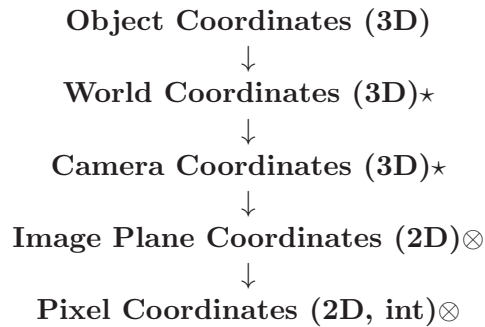
reference frame. *Intrinsic camera parameters* are the parameters necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame. The workflow is shown below. The steps marked with ⋆ use the extrinsic parameters and those marked with ⊗ the intrinsic ones.

The extrinsic parameters are those that identify uniquely the transformation between the unknown camera reference frame and the known world reference frame. Typically, determining these parameters means:

- finding the translation vector between the relative positions of the origins of the two reference frames.

- finding the rotation matrix that brings the corresponding axes of the two frames into alignment (i.e., onto each other)

Using the extrinsic camera parameters, it can be found the relation between the coordinates of a point $\mathbf{P}$ in world P_w and camera P_c coordinates (see figure B.2).

The intrinsic parameters are those that characterize the optical, geometric, and digital characteristics of the camera:

- the perspective projection (focal length f).

- the transformation between image plane coordinates and pixel coordinates.
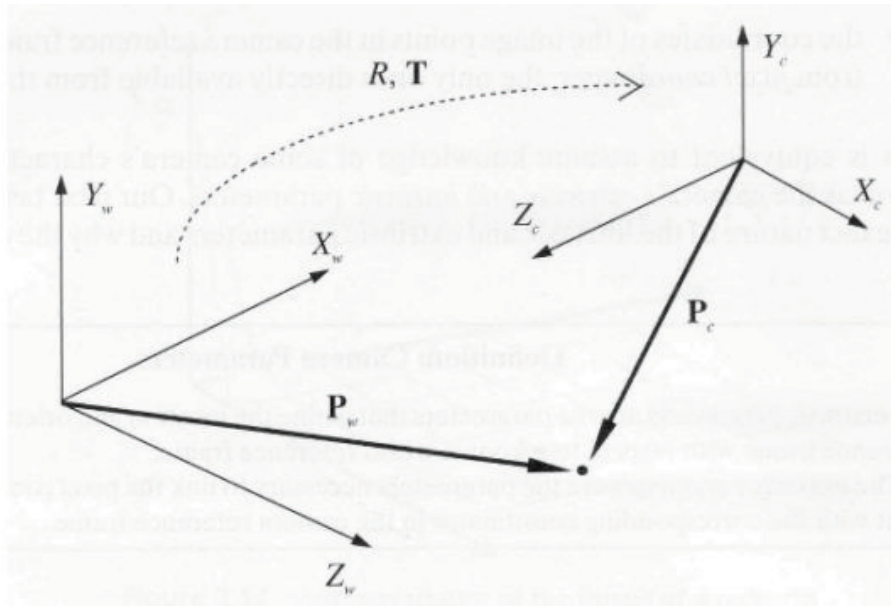
FIGURE B.2: Correspondence between world and camera coordinates

• the geometric distortion introduced by the optics.

The basic idea is the same: we have our camera model $x = MX$, where $X = (X, Y, Z, 1)$ is a 3D world point, $x = (x, y)$ is the corresponding point in pixel coordinates and $\mathbf{M}$ is the camera model matrix, which has to be solved. There are a set of correspondences $\{(X_i, x_i) | i = 1, ..., N\}$ where X_i is a world point and x_i its projection to pixel coordinates. It is possible to write:

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \frac{1}{z_i} M \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

that can be written out as

$$z_i \begin{pmatrix} x_1 \\ y_i \\ 1 \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

It gets the following equation

$$z_i x_i = m_{11} X_i + m_{12} Y_i + m_{13} Z_i + m_{14}$$
$$z_i y_i = m_{21} X_i + m_{22} Y_i + m_{23} Z_i + m_{24}$$
$$z_i = m_{31} X_i + m_{32} Y_i + m_{33} Z_i + m_{34}$$

By resolving the equation system, the factors that allows to project a point $\mathbf{P}$ in the 3D world in a correct point in 2D camera coordinates are then estimated.

Appendix C

Statistical tests for Software Analysis

Objective measurement of software is crucial to enable control of projects, products and process. *"You cannot control what you cannot measure"* [149]. Without measurements, it is not possible to have the desired control and, consequently, infer any conclusion about the level of performance implied by the choices made during the design of a process.

Measurement and measure are defined as: *"**Measurement** is a mapping from the empirical world to the formal, relational world. Consequently, a **measure** is the number or the symbol assigned to an entity by this mapping in order to characterize an attribute"* [150]. This section briefly introduces the statistical instruments useful to formally measure the performance achieved in software engineering. Considering the breadth of the subject, this section only discusses the statistical measurements used in this work of thesis.

Formulation of the Hypothesis

A *statistical hypothesis* is a scientific hypothesis that is testable on the basis of observing a process that is modelled via a set of random variables [151]. A statistical hypothesis is stated formally and the data collected during the experiment is used, possibly, to reject the hypothesis. Once that happens, it is possible to draw conclusions taking into account possible statistical risks.

During the planning phase of a statistical hypothesis test, two hypothesis have to be formulated:

1. **the null hypotesis,** H_0: that states that there is not a significant difference between two populations respect to a given aspect. For example, the null hypothesis for a new method M it that it gives on average the same count of red balls in several different groups of coloured balls respect to an old one. $H_0 : \mu_{M_{old}} = \mu_{M_{new}}$.

2. **the alternative hypothesis,** H_1: that is the accepted hypothesis if the null hypothesis is rejected. Referring to the example method above, $H_1 : \mu_{M_{old}} < \mu_{M_{new}}$.

The number of statistical tests that use this formulation of hypothesis to discuss the outcomes of an experiments is really high. They all aim at rejecting the null hypothesis in favor of the alternative one to draw conclusions. However, once the null hypothesis is rejected, it must be taken into account that testing hypothesis implies some risks. They are formally expressed as:

1. **Type-I-error**: occurs when a statistical test rejects the null hypothesis even if the difference between the populations is not particularly significant for the considered domain.

2. **Type-II-error**: occurs, conversely, when the null hypothesis is accepted even if there is actually a significant difference between the populations for the particular aspect that is analysed.

The definition of risks in hypothesis testing allows to define the **power** of a statistical test that is meant as the probability that the test reveals a significant difference when H_0 is rejected.

Before choosing a statistical test to analyse a particular aspect of the software design, it is extremely important to consider the above mentioned measures. In addition, the design of the test varies depending on the type of data to be collected and on the number of populations and factors that have to be compared.

In the book of Wholin et al. [152] a comprehensive and complete collection of most frequently used tests has been presented. It discusses how to choose the proper statistical test according to the number of factors and treatments in the testing design:

1. **one factor, two treatments**: when two different approaches have to be compared against each other.

2. **one factor, more than two treatments**: as the previous one, for all possible pairs of treatments, two treatments are compared against each other.

3. Two factors, two treatments: when increasing the number of factors to two (or more), the experiment gets more complex. The hypothesises become three, that is one for the effect from one of the factor, the second for the other one and the third for the interaction between the two factors.

Statistical tests can be **one-tailed** or **two-tailed**. In statistical significance testing, a one-tailed test and a two-tailed test are alternative ways of computing the statistical significance of a parameter inferred from a data set, in terms of a test statistic.
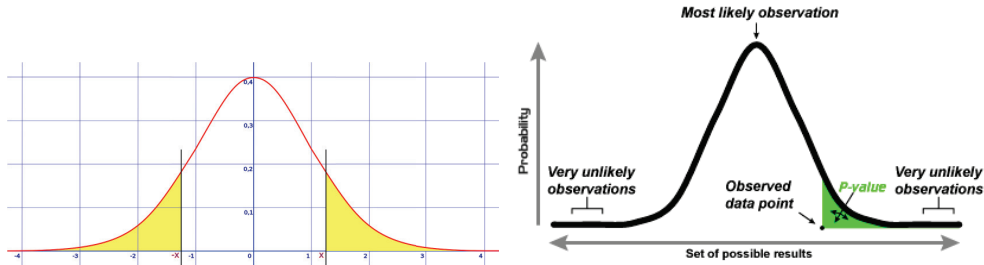
FIGURE C.1: Examples of the meaning of two-tailed and one-tailed statistical tests.

A two-tailed test is used if deviations of the estimated parameter in either direction from some benchmark value are considered theoretically possible; in contrast, a one-tailed test is used if only deviations in one direction are considered possible. Alternative names are one-sided and two-sided tests; the terminology "tail" is used because the extreme portions of distributions, where observations lead to rejection of the null hypothesis, are small and often "tail off" toward zero as in the normal distribution or "bell curve" (see figure C.1 [153]).

Considering the limited amount of factors and treatments analysed in this work of thesis, this section only presents technical details about the statistical tests used during the experimentations, that are the *Student's test* and *Mann-Whitney test*. Before introducing them, it is important to clarify the difference between parametric e non-parametric tests. The *parametric test* is based on a model that involves a specific distribution. For example, it can assume that the populations are normally distributed. Moreover, parametric tests require that data can be measured at least on an interval continuous scale. Conversely, *non-parametric test* do not make assumptions concerning the distribution of the data. When data cannot be measured on a interval scale, e.g., the scores of a questionnaire, non-parametric test are used rather than the parametric ones. In general, whenever there are doubts about the most appropriate test to perform, or they are non-normally distributed, a non-parametric test should be preferred to the parametric one. The *Shapiro-Wilk* test is in general used to assert the normal distribution of the samples before running

a parametric test.

Student's test

The *Student's test (also known as t-test)* is one of the most famous and used parametric tests. It is particularly applied for comparisons between two samples according to a single factor. The samples have to be independent and normally distributed. Given two independent samples $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y2, ..., y_m)$ and the null hypothesis $H_0 = \mu_x = \mu_y$, i.e., the expected mean values are the same in both treatments, the *t-statistic* is calculated as:

$$t_0 = \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \text{ where } S_p = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$$

and, S_x^2 and S_y^2 are the individual sample variance.

The two-sided test ($H_1 : \mu_x \neq \mu_y$): reject the null hypothesis H_0 if $|t_0| > t_{\alpha/2, n+m-2}$ where $t_{\alpha/2, n+m-2}$ is a known t-distribution critical value with significance threshold α (generally set to 0.05) and degree of freedom $n + m - 2$.

In its one-sided formulation ($H_1 : \mu_x > \mu_y$), the test reject H_0 if $t_0 > t_{\alpha, n+m-2}$.

In R scripting

Given a and b the samples to be compared, the Student's test is calculate by calling the following R functions, for two-tailed and one-tailed respectively:

```
shapiroTest_a = shapiro.test(a);
shapiroTest_b = shapiro.test(b);
if (shapiroTest_a < 0.01 & shapiroTest_b < 0.01){
  t.test(a, b, conf.value=0.05, alternative = "two.sided");
```

```
    t.test(a, b, alternative="less", conf.value=0.05);
}
```

where *shapiro.test* checks if the samples are normally distributed. The option *alternative* specifies that the t-test is of type one-tailed, $(H_1 : \mu_x > \mu_y)$ for *greater* and $(H_1 : \mu_x < \mu_y)$ for *less*. The output value in t.test is then compared to the critical value obtained by calling $qt(\alpha, f)$ where α is the significance threshold used and f the degrees of freedom $(n + m - 2)$.

Mann-Whitney test

The *Mann-Whiteny* test is a non-parametric statistical test alternative to the Student's test. Being non-parametric, it is always possible to use it instead of the t-test when the assumptions made by the t-test about the samples look uncertain. The test involves the calculation of a statistic, usually called U or W, whose distribution under the null hypothesis is known. In the case of small samples, the distribution is tabulated, but for sample sizes above 20, approximation using the normal distribution is fairly good. Some books tabulate statistics equivalent to U, such as the sum of ranks in one of the samples, rather than U itself.

Given two independent samples $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y2, ..., y_m)$ and the null hypothesis that the samples come from the same distribution, the *Mann-Whitney* test is calculated as:

$$U = N_A N_B + \frac{N - A(N_A + 1)}{2} - T \text{ and } U' = N_A N_B - U$$

where $N_A = min(n, m)$, $N_B = max(n, m)$ and T is the sum of the ranks of the smallest sample (the samples are ranked first).

The null hypothesis H_0 is reject if $min(U, U')$ is less than or equal to the known U-distribution critical value with significance threshold α (generally set to 0.05)

In R scripting

Given a and b the samples to be compared, the Mann-Whitney U-test is calculate by calling the following R functions, for two-tailed and one-tailed test respectively:

```
values = wilcox.test(a, b, alternative = "two.sided")
values = wilcox.test(a, b, alternative="less")
```

where the option *alternative="less"* or *alternative="greater"* specifies that the U-test is of type one-tailed.

Bibliography

[1] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3): 231 – 268, 2001. ISSN 1077-3142. doi: http://dx.doi.org/10.1006/cviu. 2000.0897. URL `http://www.sciencedirect.com/science/article/pii/S107731420090897X`.

[2] Roberto Brunelli and Daniele Falavigna. Person identification using multiple cues. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17 (10):955–966, 1995.

[3] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006.

[4] Dariu M Gavrila. The visual analysis of human movement: A survey. *Computer vision and image understanding*, 73(1):82–98, 1999.

[5] A.W.M. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah. Visual tracking: An experimental survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(7):1442–1468, July 2014. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.230.

[6] M.O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, pages 306–311, Oct 2010. doi: 10.1109/IIHMSP.2010.83.

[7] Song-Yi Han, Hyun-Ae Park, Dal-Ho Cho, Kang Ryoung Park, and Sangyoun Lee. Face recognition based on near-infrared light using mobile phone. In *Adaptive and Natural Computing Algorithms*, pages 440–448. Springer, 2007.

[8] Kai Xi, Tohari Ahmad, Fengling Han, and Jiankun Hu. A fingerprint based bio-cryptographic security protocol designed for client/server authentication in mobile computing environment. *Security and Communication Networks*, 4 (5):487–499, 2011.

[9] Ronald T Azuma et al. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.

[10] Stuart Eve. Augmenting phenomenology: using augmented reality to aid archaeological phenomenology in the landscape. *Journal of archaeological method and theory*, 19(4):582–600, 2012.

[11] Anthony Webster, Steven Feiner, Blair MacIntyre, William Massie, and Theodore Krueger. Augmented reality in architectural construction, inspection and renovation. In *Proc. ASCE Third Congress on Computing in Civil Engineering*, pages 913–919, 1996.

[12] Hannes Kaufmann and Dieter Schmalstieg. Mathematics and geometry education with collaborative augmented reality. *Computers & Graphics*, 27(3): 339–345, 2003.

[13] Adrian David Cheok, Fong Siew Wan, Xubo Yang, Wang Weihua, Lee Men Huang, Mark Billinghurst, and Hirokazu Kato. Game-city: A ubiquitous large area multi-interface mixed reality game space for wearable computers. In *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*, pages 156–157. IEEE, 2002.

[14] Andrea F Abate, Fabio Narducci, and Stefano Ricciardi. An augmented interface to audio-video components. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 254–257. ACM, 2012.

[15] Ihsan Rabbi, Sehat Ullah, and Siffat Ullah Khan. Augmented reality tracking techniques a systematic literature. *IOSR Journal of Computer Engineering (IOSRJCE)*, 2(2):23–29, 2012.

[16] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE Computer Society, 2008.

[17] Ihsan Rabbi and Sehat Ullah. A survey on augmented reality challenges and tracking. *Acta Graphica znanstveni časopis za tiskarstvo i grafičke komunikacije*, 24(1-2):29–46, 2013.

[18] Hirokazu Kato. Artoolkit. *http://www. hitl. washington. edu/artoolkit/*, 1999.

[19] Michael L Dertouzos and JF Reintjes. *Computer Aided Electronic Circuit Design*. Defense Technical Information Center, 1970.

[20] Leon O Chua. A computer-oriented sophomore course on nonlinear circuit analysis. *Education, IEEE Transactions on*, 12(3):202–208, 1969.

[21] Z. Fazarinc. Designer-oriented cad. *Circuit Theory, IEEE Transactions on*, 20 (6):673–682, Nov 1973. ISSN 0018-9324. doi: 10.1109/TCT.1973.1083774.

[22] Jianrong Tan. High quality-oriented product cooperate design in virtual environments. In *Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on*, pages II–II, April 2010. doi: 10.1109/ CSCWD.2010.5472014.

[23] S. Jayaram, Yong Wang, U. Jayaram, K. Lyons, and P. Hart. A virtual assembly design environment. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 172–179, Mar 1999. doi: 10.1109/VR.1999.756948.

[24] Lin Guimin and Sun Surong. Desktop virtual assembly system based on pc. In *Computer-Aided Industrial Design and Conceptual Design, 2006. CAIDCD '06. 7th International Conference on*, pages 1–5, Nov 2006. doi: 10.1109/ CAIDCD.2006.329413.

[25] Giovanni Garibotto, Pierpaolo Murrieri, Alessandro Capra, Stefano De Muro, Ugo Petillo, Francesco Flammini, Mariana Esposito, Cocetta Pragliola, Giuseppe Di Leo, Roald Lengu, et al. White paper on industrial applications of computer vision and pattern recognition. In *Image Analysis and Processing– ICIAP 2013*, pages 721–730. Springer, 2013.

[26] Bernd Schwald and Blandine De Laval. An augmented reality system for training and assistance to maintenance in the industrial context. 2003.

[27] SuWoong Lee, Jong gook Ko, Seokbin Kang, and Junsuk Lee. An immersive e-learning system providing virtual experience. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 249–250, Oct 2010. doi: 10.1109/ISMAR.2010.5643591.

[28] Yuan Liu. Virtual neurosurgical education for image-guided deep brain stimulation neurosurgery. In *Audio, Language and Image Processing (ICALIP), 2014 International Conference on*, pages 623–626, July 2014. doi: 10.1109/ ICALIP.2014.7009869.

[29] Yanlin Luo, Ping Guo, S. Hasegawa, and M. Sato. An interactive molecular visualization system for education in immersive multi-projection virtual environment. In *Multi-Agent Security and Survivability, 2004 IEEE First Symposium on*, pages 485–488, Dec 2004. doi: 10.1109/ICIG.2004.36.

[30] L. Gallo, G. De Pietro, and I. Marra. User-friendly inspection of medical image data volumes in virtual environments. In *Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on*, pages 749–754, March 2008. doi: 10.1109/CISIS.2008.33.

[31] F. Martinez-Reyes and I. Hern'ndez-Santana. The virtual maze: A game to promote social interaction between children. In *Intelligent Environments (IE), 2012 8th International Conference on*, pages 331–334, June 2012. doi: 10.1109/IE.2012.42.

[32] Yiyu Cai, N.K.H. Chia, D. Thalmann, N.K.N. Kee, Jianmin Zheng, and N.M. Thalmann. Design and development of a virtual dolphinarium for children with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(2):208–217, March 2013. ISSN 1534-4320. doi: 10.1109/TNSRE.2013. 2240700.

[33] M. Khan, S. Sulaiman, A.M. Said, and M. Tahir. Exploring the quantitative and qualitative measures for haptic systems. In *Information Technology (IT-Sim), 2010 International Symposium in*, volume 1, pages 31–36, June 2010. doi: 10.1109/ITSIM.2010.5561305.

[34] S. Katsura, N. Tsunashima, W. Yamanouchi, and Y. Yokokura. Preservation and reproduction of real-world haptic information. In *Power Electronics Conference (IPEC), 2010 International*, pages 2216–2221, June 2010. doi: 10.1109/IPEC.2010.5542005.

[35] M. Hosseini, F. Malric, and Nicolas D. Georganas. A haptic virtual environment for industrial training. In *Haptic Virtual Environments and Their Applications, IEEE International Workshop 2002 HAVE*, pages 25–30, 2002. doi: 10.1109/HAVE.2002.1106909.

[36] M. Radi and G. Reinhart. Industrial haptic robot guidance system for assembly processes. In *Haptic Audio visual Environments and Games, 2009. HAVE 2009. IEEE International Workshop on*, pages 69–74, Nov 2009. doi: 10. 1109/HAVE.2009.5356135.

[37] C.L. Clover, G.R. Luecke, J.J. Troy, and W.A. McNeely. Dynamic simulation of virtual mechanisms with haptic feedback using industrial robotics equipment. In *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, volume 1, pages 724–730 vol.1, Apr 1997. doi: 10.1109/ROBOT.1997.620121.

[38] A. Bolopion and S. Regnier. A review of haptic feedback teleoperation systems for micromanipulation and microassembly. *Automation Science and Engineering, IEEE Transactions on*, 10(3):496–502, July 2013. ISSN 1545-5955. doi: 10.1109/TASE.2013.2245122.

[39] K. Kahol. Ambient haptics in clinical environments. In *Haptic Audio visual Environments and Games, 2008. HAVE 2008. IEEE International Workshop on*, pages xi–xi, Oct 2008. doi: 10.1109/HAVE.2008.4685288.

[40] A.M. Okamura, C. Basdogan, S. Baillie, and W.S. Harwin. Haptics in medicine and clinical skill acquisition [special section intro.]. *Haptics, IEEE Transactions on*, 4(3):153–154, May 2011. ISSN 1939-1412. doi: 10.1109/TOH.2011. 47.

[41] M. Solanki and V. Raja. Haptic based augmented reality simulator for training clinical breast examination. In *Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on*, pages 265–269, Nov 2010. doi: 10.1109/IECBES.2010.5742241.

[42] Hyung-Soon Park, Jonghyun Kim, and D.L. Damiano. Development of a haptic elbow spasticity simulator (hess) for improving accuracy and reliability of clinical assessment of spasticity. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 20(3):361–370, May 2012. ISSN 1534-4320. doi: 10.1109/TNSRE.2012.2195330.

[43] Kangdon Lee. Augmented reality in education and training. *TechTrends*, 56 (2):13–21, 2012.

[44] M.E.C. Santos, A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato. Augmented reality learning experiences: Survey of prototype design and evaluation. *Learning Technologies, IEEE Transactions on*, 7(1):38–56, Jan 2014. ISSN 1939-1382. doi: 10.1109/TLT.2013.37.

[45] Robert Rosenthal, H Cooper, and LV Hedges. Parametric measures of effect size. *The handbook of research synthesis*, pages 231–244, 1994.

[46] Mau-Tsuen Yang and Wan-Che Liao. Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *Learning Technologies, IEEE Transactions on*, 7(2):107–117, April 2014. ISSN 1939-1382. doi: 10.1109/TLT.2014.2307297.

[47] Andrea F Abate, Fabio Narducci, and Stefano Ricciardi. An image based approach to hand occlusions in mixed reality environments. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, pages 319–328. Springer, 2014.

[48] Andrea F Abate, Fabio Narducci, and Stefano Ricciardi. Mixed reality environment for mission critical systems servicing and repair. In *Virtual, Augmented and Mixed Reality. Systems and Applications*, pages 201–210. Springer, 2013.

[49] H. KATO. Artoolkit. *http://www.hitl.washington.edu/artoolkit/*. URL `http://ci.nii.ac.jp/naid/10019957376/en/`.

[50] Wolfgang Wohlgemuth and Gunthard Triebfürst. Arvika: Augmented reality for development, production and service. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, DARE '00, pages 151–152, New

York, NY, USA, 2000. ACM. doi: 10.1145/354666.354688. URL `http://doi.acm.org/10.1145/354666.354688`.

[51] Kiyoshi Kiyokawa, Mark Billinghurst, Bruce Campbell, and Eric Woods. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 133. IEEE Computer Society, 2003.

[52] BIAN Zhiqiang, Hirotake Ishii, Hiroshi Shimoda, Hidekazu Yoshikawa, Yoshit-sugu Morishita, Yoshiki Kanehira, and Masanori Izumi. Development of a tracking method for augmented reality applied to npp maintenance work and its experimental evaluation. *IEICE TRANSACTIONS on Information and Systems*, 90(6):963–974, 2007.

[53] Erick Mendez, Denis Kalkofen, and Dieter Schmalstieg. Interactive context-driven visualization tools for augmented reality. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 209–218. IEEE Computer Society, 2006.

[54] Katharina Pentenrieder, Christian Bade, Fabian Doil, and Peter Meier. Augmented reality-based factory planning-an application tailored to industrial needs. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 31–42. IEEE, 2007.

[55] Gerhard Schall, Erick Mendez, Ernst Kruijff, Eduardo Veas, Sebastian Jung-hanns, Bernhard Reitinger, and Dieter Schmalstieg. Handheld augmented reality for underground infrastructure visualization. *Personal and ubiquitous computing*, 13(4):281–291, 2009.

[56] Francesca De Crescenzio, Massimiliano Fantini, Franco Persiani, Luigi Di Stefano, Pietro Azzari, and Samuele Salti. Augmented reality for aircraft maintenance training and operations support. *Computer Graphics and Applications, IEEE*, 31(1):96–101, 2011.

[57] Daniel Wagner, Tobias Langlotz, and Dieter Schmalstieg. Robust and unobtrusive marker tracking on mobile phones. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, pages 121–124. IEEE, 2008.

[58] Madjid Maidi, Marius Preda, et al. Markerless tracking for mobile augmented reality. In *Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on*, pages 301–306. IEEE, 2011.

[59] Andrew I Comport, Eric Marchand, Muriel Pressigout, and Francois Chaumette. Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):615–628, 2006.

[60] Pascal Fua and Vincent Lepetit. Vision based 3d tracking and pose estimation for mixed reality. *Emerging Technologies of Augmented Reality Interfaces and Design*, pages 43–63, 2005.

[61] Can Liu, Stéphane Huot, Jonathan Diehl, Wendy Mackay, and Michel Beaudouin-Lafon. Evaluating the benefits of real-time feedback in mobile augmented reality with hand-held devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2973–2976. ACM, 2012.

[62] DWF Van Krevelen and R Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9 (2):1, 2010.

[63] Tobias Langlotz, Holger Regenbrecht, Stefanie Zollmann, and Dieter Schmalstieg. Audio stickies: visually-guided spatial audio annotations on a mobile augmented reality platform. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, pages 545–554. ACM, 2013.

[64] Andreas Dünser, Raphaël Grasset, and Mark Billinghurst. *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand, 2008.

[65] Cédric Bach and Dominique L Scapin. Obstacles and perspectives for evaluating mixed reality systems usability. In *Acte du Workshop MIXER, IUI-CADUI*, volume 4. Citeseer, 2004.

[66] Brian F Goldiez, Ali M Ahmad, and Peter A Hancock. Effects of augmented reality display settings on human wayfinding performance. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(5): 839–845, 2007.

[67] Steven Henderson and Steven Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *Visualization and Computer Graphics, IEEE Transactions on*, 17(10):1355–1368, 2011.

[68] Si Jung Jun Kim. A user study trends in augmented reality and virtual reality research: A qualitative study with the past three years of the ismar and ieee vr conference papers. In *Ubiquitous Virtual Reality (ISUVR), 2012 International Symposium on*, pages 1–5. IEEE, 2012.

[69] Xiangyu Wang and Phillip S Dunston. Comparative effectiveness of mixed reality-based virtual environments in collaborative design. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(3): 284–296, 2011.

[70] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[71] John Craig and Victor Patterson. Introduction to the practice of telemedicine. *Journal of Telemedicine and Telecare*, 11(1):3–9, 2005.

[72] AC Norris. Telemedicine service providers and applications. *Essentials of Telemedicine and Telecare*, pages 61–75.

[73] O Ferrer-Roca, RD Diaz De Leon, FJ de Latorre, M Suárez-Delgado, L Di Persia, and M Cordo. Aviation medicine: challenges for telemedicine. *Journal of telemedicine and telecare*, 8(1):1–4, 2002.

[74] Francesco Beltrame, K Boddy, and Piergiulio Maryni. Adopting telemedicine services in the airline framework. *Information Technology in Biomedicine, IEEE Transactions on*, 5(2):171–174, 2001.

[75] Michael Weinlich, Nadine Nieuwkamp, Uwe Stueben, Ingo Marzi, and Felix Walcher. Telemedical assistance for in-flight emergencies on intercontinental commercial aircraft. *Journal of telemedicine and telecare*, 15(8):409–413, 2009.

[76] F Amenta, A Dauri, and N Rizzo. Telemedicine and medical care to ships without a doctor on board. *Journal of telemedicine and telecare*, 4(suppl 1): 44–45, 1998.

[77] Kenneth L Wilson, Jayfus T Doswell, Olatokunbo S Fashola, Wayne Debeatham, Nii Darko, Travelyan M Walker, Omar K Danner, Leslie R Matthews, and William L Weaver. Using augmented reality as a clinical support tool to assist combat medics in the treatment of tension pneumothoraces. *Military medicine*, 178(9):981–985, 2013.

[78] Marcello Bracale, Mario Cesarelli, and Paolo Bifulco. Telemedicine services for two islands in the Bay of Naples. *J Telemed Telecare*, 8(1):5–10, 2002.

[79] Alexandra Martiniuk, Joel Negin, Fred Hersch, Tenneth Dalipanda, Rooney Jagilli, Patrick Houasia, Lilijana Gorringe, and Annie Christie. Telemedicine in the Solomon Islands: 2006 to 2009. *J Telemed Telecare*, 17(5):251–6, 2011. doi: 10.1258/jtt.2011.100920.

[80] Barbara J Drew, Claire E Sommargren, Daniel M Schindler, Jessica Zegre, Kent Benedict, and Mitchell W Krucoff. Novel electrocardiogram configurations and transmission procedures in the prehospital setting: effect on ischemia and arrhythmia determination. *J Electrocardiol*, 39(4 Suppl):S157–60, Oct 2006. doi: 10.1016/j.jelectrocard.2006.05.033.

[81] M Oldenburg, X Baur, and C Schlaich. Assessment of three conventional automated external defibrillators in seafaring telemedicine. *Occupational medicine*, page kqr200, 2011.

[82] Heidi J Silver and Nancy S Wellman. Family caregiver training is needed to improve outcomes for older adults using home care technologies. *Journal of the American Dietetic Association*, 102(6):831–836, 2002.

[83] Heidi J Silver, Nancy S Wellman, Daisy Galindo-Ciocon, and Paulette Johnson. Family caregivers of older adults on home enteral nutrition have multiple unmet task-related training needs and low overall preparedness for caregiving. *J Am Diet Assoc*, 104(1):43–50, Jan 2004. doi: 10.1016/j.jada.2003.10.010.

[84] M Cristina MAZZOLENI, Paolo BIFULCO, Nicola FERRARA, and Franco RENGO. Telemedicine fetal phonocardiography surveillance: An italian satisfactory experience. In *Seamless Care, Safe Care: The Challenges of Interoperability and Patient Safety in Health Care: Proceedings of the EFMI Special Topic Conference, June 2-4, 2010, Reykjavik, Iceland*, volume 155, page 176. IOS Press, 2010.

[85] Manish N Shah, Ryan McDermott, Suzanne M Gillespie, Erin B Philbrick, and Dallas Nelson. Potential of telemedicine to provide acute medical care for adults in senior living communities. *Academic Emergency Medicine*, 20(2): 162–168, 2013.

[86] Manish N Shah, Dylan Morris, Courtney Jones, Suzanne M Gillespie, Dallas L Nelson, Kenneth M McConnochie, and Ann Dozier. A qualitative evaluation of a telemedicine-enhanced emergency care program for older adults. *Journal of the American Geriatrics Society*, 61(4):571–576, 2013.

[87] Kelly McCann, Anna Holdgate, Rima Mahammad, and Adam Waddington. Accuracy of ecg electrode placement by emergency department clinicians. *Emergency Medicine Australasia*, 19(5):442–448, 2007.

[88] Alain Rudiger, Jens P Hellermann, Raphael Mukherjee, Ferenc Follath, and Juraj Turina. Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings. *The American journal of emergency medicine*, 25(2):174–178, 2007.

[89] William Wenger and Paul Kligfield. Variability of precordial electrode placement during routine electrocardiography. *Journal of electrocardiology*, 29(3): 179–184, 1996.

[90] Richard A Harrigan, Theodore C Chan, and William J Brady. Electrocardiographic electrode misplacement, misconnection, and artifact. *The Journal of emergency medicine*, 43(6):1038–1044, 2012.

[91] Colin James Sutherland. An augmented reality haptic training simulator for spinal needle procedures. 2011.

[92] N Alberto Borghese and Iuri Frosio. Compact tracking of surgical instruments through structured markers. *Medical & biological engineering & computing*, 51 (7):823–833, 2013.

[93] Mark A Livingston, William F Garrett, Gentaro Hirota, Mary C Whitton, Etta D Pisano, Henry Fuchs, et al. Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 439–446. ACM, 1996.

[94] Henry Fuchs, Mark A Livingston, Ramesh Raskar, Kurtis Keller, Jessica R Crawford, Paul Rademacher, Samuel H Drake, Anthony A Meyer, et al. *Augmented reality visualization for laparoscopic surgery.* Springer, 1998.

[95] Nassir Navab, Joerg Traub, Tobias Sielhorst, Marco Feuerstein, and Christoph Bichlmeier. Action-and workflow-driven augmented reality for computer-aided medical procedures. *Computer Graphics and Applications, IEEE*, 27(5):10–14, 2007.

[96] Marco Feuerstein, Thomas Mussack, Sandro Michael Heining, and Nassir Navab. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *Medical Imaging, IEEE Transactions on*, 27(3):355–369, 2008.

[97] Sebastian Vogt, Ali Khamene, and Frank Sauer. Reality augmentation for medical procedures: System architecture, single camera marker tracking, and system evaluation. *International Journal of Computer Vision*, 70(2):179–190, 2006.

[98] Grigore Burdea. Keynote address: Virtual rehabilitation-benefits and challenges. In *1st International Workshop on Virtual Reality Rehabilitation (Mental Health, Neurological, Physical, Vocational) VRMHR*, pages 1–11. sn, 2002.

[99] Michael Villiger, Dominik Bohli, Daniel Kiper, Pawel Pyk, Jeremy Spillmann, Bruno Meilick, Armin Curt, Marie-Claude Hepp-Reymond, Sabina

Hotz-Boendermaker, and Kynan Eng. Virtual reality–augmented neurorehabilitation improves motor function and reduces neuropathic pain in patients with incomplete spinal cord injury. *Neurorehabilitation and neural repair*, 27 (8):675–683, 2013.

[100] Aaron Kotranza and Benjamin Lok. Virtual human+ tangible interface= mixed reality human an initial exploration with a virtual breast exam patient. In *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 99–106. IEEE, 2008.

[101] Ryutarou Ohbuchi, Mike Bajura, and Henry Fuchs. Case study: Observing a volume rendered fetus within a pregnant patient. In *Visualization: Proceedings of the IEEE Conference on Visualization*, volume 5. Citeseer, 1998.

[102] Tobias Sielhorst, Tobias Obst, Rainer Burgkart, Robert Riener, and Nassir Navab. An augmented reality delivery simulator for medical training. In *International Workshop on Augmented Environments for Medical Imaging-MICCAI Satellite Workshop*, volume 141, 2004.

[103] V Masero, FM Sanchez, and J Uson. A telemedicine system for enabling teaching activities. *Journal of telemedicine and telecare*, 6(suppl 2):86–88, 2000.

[104] Georgi Graschew, Stefan Rakowsky, Panagiota Balanou, and Peter M Schlag. Interactive telemedicine in the operating theatre of the future. *Journal of telemedicine and telecare*, 6(suppl 2):20–24, 2000.

[105] Mark Green and Lloyd White. The cave-let: a low-cost projective immersive display. *Journal of telemedicine and telecare*, 6(suppl 2):24–26, 2000.

[106] H Van Dijk and HJ Hermens. Distance training for the restoration of motor function. *Journal of telemedicine and telecare*, 10(2):63–71, 2004.

[107] Christen Lundsgaard and Donald D Van Slyke. Studies of lung volume. i relation between thorax size and lung volume in normal adults. *The Journal of experimental medicine*, 27(1):65–86, 1918.

[108] Frank N Wilson, Franklin D Johnston, Francis F Rosenbaum, and Paul S Barker. On einthoven's triangle, the theory of unipolar electrocardiographic leads, and the interpretation of the precordial electrocardiogram. *American heart journal*, 32(3):277–310, 1946.

[109] Burch G Goldberger E Graybiel A Hecht H Johnston F Lepeschkin E Myers G Wilson F, Kossmann C. RECOMMENDATIONS for standardization of electrocardiographic and vectorcardiographic leads. *Circulation*, 10(4):564–573, October 1954.

[110] Paolo Bifulco, Fabio Narducci, Raffaele Vertucci, Pasquale Ambruosi, Mario Cesarelli, and Maria Romano. Telemedicine supported by augmented reality: an interactive guide for untrained people in performing an ecg test. *Biomedical engineering online*, 13(1):153, 2014.

[111] Fabio Narducci and Stefano Ricciardi. Augmented reality aided maintenance for industrial applications. In *Eurographics Italian Chapter Conference*, pages 67–72, 2011.

[112] Narducci Fabio Abate, Andrea F and Stefano Ricciardi. Biometrics empowered ambient intelligence environment. In *Proceedings of Accademia Peloritana dei Pericolanti Journal*, in course of publication.

[113] Narducci Fabio Abate, Andrea F and Stefano Ricciardi. Photorealistic virtual exploration of an archaeological site. In *Proceedings of Accademia Peloritana dei Pericolanti Journal*, in course of publication.

[114] Zhigeng Pan, Adrian David Cheok, Hongwei Yang, Jiejie Zhu, and Jiaoying Shi. Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics*, 30(1):20–28, 2006.

[115] R Rajaganeshan, CL Ludlam, DP Francis, SV Parasramka, and R Sutton. Accuracy in ecg lead placement among technicians, nurses, general physicians and cardiologists. *International journal of clinical practice*, 62(1):65–70, 2008.

[116] H. Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, USA, October 1999.

[117] JP Royston. Algorithm as 181: The w test for normality. *Applied Statistics*, pages 176–180, 1982.

[118] Myles Hollander, Douglas A Wolfe, and Eric Chicken. *Nonparametric statistical methods*. John Wiley & Sons, 2013.

[119] Keith C Hendy, Kevin M Hamilton, and Lois N Landry. Measuring subjective workload: when is one scale better than many? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(4):579–601, 1993.

[120] Ky Waegel and Frederick P Brooks. Filling the gaps: Hybrid vision and inertial tracking. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 1–4. IEEE, 2013.

[121] Nappi M. Narducci F. Abate, A.F. and S. Ricciardi. Measuring the benefits of augmented reality in industrial environment. *IEEE Transactions on Visualization and Computer Graphics*, under review.

[122] PA Tresadern, Chris McCool, Norman Poh, Pavel Matejka, Abdenour Hadid, Christophe Levy, TF Cootes, and Sebastien Marcel. Mobile biometrics (mobio): Joint face and voice verification for a mobile platform. *IEEE pervasive computing*, 99, 2012.

[123] Anil K Jain and Stan Z Li. *Encyclopedia of Biometrics: I-Z.*, volume 1. Springer Science & Business Media, 2009.

[124] M. De Marsico, M. Nappi, and R. Daniel. Isis: Iris segmentation for identification systems. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2857–2860, Aug 2010. doi: 10.1109/ICPR.2010.700.

[125] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1148–1161, 1993.

[126] Richard P Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.

[127] Wageeh W Boles and Boualem Boashash. A human identification technique using images of the iris and wavelet transform. *IEEE transactions on signal processing*, 46(4):1185–1188, 1998.

[128] Shinyoung Lim, Kwanyong Lee, Okhwan Byeon, and Taiyun Kim. Efficient iris recognition through improvement of feature vector and classifier. *ETRI journal*, 23(2):61–70, 2001.

[129] Hugo Proença and Luís A Alexandre. Ubiris: A noisy iris image database. In *Image Analysis and Processing–ICIAP 2005*, pages 970–977. Springer, 2005.

[130] Kevin W Bowyer. The results of the nice. ii iris biometrics competition. *Pattern Recognition Letters*, 33(8):965–969, 2012.

[131] Dae Sik Jeong, Jae Won Hwang, Byung Jun Kang, Kang Ryoung Park, Chee Sun Won, Dong-Kwon Park, and Jaihie Kim. A new iris segmentation method for non-ideal iris images. *Image and vision computing*, 28(2):254–260, 2010.

[132] Kwang Yong Shin, Gi Pyo Nam, Dae Sik Jeong, Dal Ho Cho, Byung Jun Kang, Kang Ryoung Park, and Jaihie Kim. New iris recognition method for noisy iris images. *Pattern Recognition Letters*, 33(8):991–999, 2012.

[133] Dal Ho Cho, Kang Ryoung Park, and Dae Woong Rhee. Real-time iris localization for iris recognition in cellular phone. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on*, pages 254–259. IEEE, 2005.

[134] Dal-ho Cho, Kang Ryoung Park, Dae Woong Rhee, Yanggon Kim, and Jonghoon Yang. Pupil and iris localization for iris recognition in mobile phones. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2006. SNPD 2006. Seventh ACIS International Conference on*, pages 197–201. IEEE, 2006.

[135] Kang Ryoung Park, Hyun Park, Byung Jun Kang, Eui Chul Lee, and Dae Sik Jeong. A study on iris localization and recognition on mobile phones. *EURASIP Journal on Advances in Signal Processing*, 2008:20, 2008.

[136] Jin-Suk Kang. Mobile iris recognition systems: An emerging biometric technology. *Procedia Computer Science*, 1(1):475–484, 2010.

[137] Maria De Marsico, Chiara Galdi, Michele Nappi, and Daniel Riccio. Firme: face and iris recognition for mobile engagement. *Image and Vision Computing*, 32(12):1161–1172, 2014.

[138] Andrea F Abate, Michele Nappi, Fabio Narducci, and Stefano Ricciardi. Fast iris recognition on smartphone by means of spatial histograms. In *Biometric Authentication*, pages 66–74. Springer, 2014.

[139] Gabriel Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, 1991.

[140] Ciarán O Conaire, Noel E O'Connor, and Alan F Smeaton. An improved spatiogram similarity measure for robust object localisation. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–1069. IEEE, 2007.

[141] Silvio Barra, Andrea Casanova, Fabio Narducci, and Stefano Ricciardi. Ubiquitous iris recognition by means of mobile devices. *Pattern Recognition Letters*, 2014.

[142] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.

[143] Michal Dobes and Libor Machala. Upol iris image database, 2004.

[144] Yooyoung Lee, P Jonathon Phillips, and Ross J Micheals. An automated video-based system for iris recognition. In *Advances in Biometrics*, pages 1160–1169. Springer, 2009.

[145] Fengliang Xu, Xia Liu, and Kikuo Fujimura. Pedestrian detection and tracking with night vision. *Intelligent Transportation Systems, IEEE Transactions on*, 6(1):63–71, 2005.

[146] Carlo Tomasi and Takeo Kanade. *Detection and tracking of point features.* School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[147] Bernardin Keni and Stiefelhagen Rainer. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008.

[148] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge university press, 2003.

[149] Tom DeMarco. *Controlling software projects : management, measurement & estimation.* Englewood Cliffs, N. J. Yourdon Press, 1982. ISBN 0-13-171711-1. URL `http://opac.inria.fr/record=b1086414`. Index.

[150] Norman Fenton and James Bieman. *Software metrics: a rigorous and practical approach.* CRC Press, 2014.

[151] Maurice Kendall, Alan Stuart, Keith J Ord, and Steven Arnold. Kendall's advanced theory of statistics: Volume 2a–classical inference and and the linear model (kendall's library of statistics). *A Hodder Arnold Publication,*, 1999.

[152] C Wohlin, P Runeson, M Host, MC Ohlsson, B Regnell, and A Wesslen. Experimentation in software engineering: an introduction. 2000, 2000.

[153] Wikipedia. One- and two-tailed tests — wikipedia, the free encyclopedia, 2015. URL `http://en.wikipedia.org/w/index.php?title=One-_and_two-tailed_tests&oldid=646105912`. [Online; accessed 18-December-2014].