UNIVERSITE BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Informatique*

Par

Guillaume CORTES

Contribution to the Study of Projection-based Systems for Industrial Applications in Mixed Reality

Thèse à présenter et soutenir à Rennes, le 24/10/18
Unité de recherche : IRISA – UMR6074
Thèse N° :

Rapporteurs avant soutenance :

Marie-Odile Berger	Directrice de Recherche Inria à Inria Nancy
Martin Hachet	Directeur de Recherche Inria à Inria Bordeaux

Composition du Jury :

Rapporteurs :	Marie-Odile Berger	Directrice de Recherche Inria à Inria Nancy
	Martin Hachet	Directeur de Recherche Inria à Inria Bordeaux
Examinateurs :	Sabine Coquillart	Directrice de Recherche Inria à Inria Grenoble
	Guillaume Moreau	Professeur à l'Ecole Centrale de Nantes
Co-dir. de thèse :	Anatole Lécuyer	Directeur de Recherche Inria à Inria Rennes
Co-dir. de thèse :	Eric Marchand	Professeur à l'Université de Rennes 1

Acknowledgements

First of all, I would like to sincerely thank my two advisors, Anatole Lécuyer and Eric Marchand, for supervising this work and for their unrelenting support on whatever difficulty was encountered during this journey. This manuscript and the PhD would not exist without Jérôme Ardouin and Guillaume Brincin thus I would like to espacially thank them for creating this adventure.

I would also like to thank the members of the PhD committee for taking the time to read the manuscript and for traveling to Rennes in order to participate to the defense.

Thank you to the many great minds that I met at Inria and Realyz. Special thanks to my Inria office mates, Hakim and Yoren. Thank you for your good mood and the many laught we had.

To conclude, I thank my friends and family for their support. Special thanks to all the *Rennunion* group for these great lunch meals, afterworks and week-ends that we spent together. Of course, I especially thank my parents for coming to the defense from many miles away as well as for hosting one of the best defense buffet ever!

Finally, thank you Laurie for being there all along.

Contents

List of Acronyms v

Notations ix

1 Introduction 1

2 Related work 11
2.1 Visual displays for mixed reality . . . 11
2.1.1 Non-projection-based visual displays . . . 13
2.1.2 Projection-based displays . . . 19
2.2 Tracking systems for mixed reality . . . 24
2.2.1 Non-optical tracking systems . . . 25
2.2.2 Optical tracking systems . . . 29
2.2.3 Discussion on optical tracking systems for projection-based displays 35
2.3 Industrial applications of mixed reality . . . 37
2.3.1 Training applications . . . 39
2.3.2 Assistance applications . . . 40
2.3.3 Design applications . . . 42
2.3.4 Planning and validation applications . . . 43
2.4 Conclusion . . . 44

3 Pilot study: Analysis of user motion in a specific CAVE-based industrial application 47
3.1 Analysis of user motion in a CAVE-based industrial application . . . 48
3.1.1 Selection of the industrial application . . . 48
3.1.2 Participants . . . 51
3.1.3 Procedure . . . 51
3.1.4 Collected data . . . 51
3.2 Results . . . 52
3.2.1 Cyclop (Head) and Wand (Hand) 3D positions . . . 52
3.2.2 Cyclop and Wand 3D orientations . . . 54
3.2.3 Cyclop and Wand speeds . . . 55
3.3 Discussion . . . 55
3.4 Main outcomes and guidelines . . . 56
3.5 Conclusion . . . 57

4 Increasing optical tracking workspace for mixed reality applications 59

4.1 General approach: Increasing the optical tracking workspace of mixed reality applications . . . 60
4.2 Fundamentals of optical tracking . . . 62
4.2.1 Prespective camera model . . . 62
4.2.2 Epipolar geometry . . . 65
4.3 MonSterTrack: Increasing optical tracking workspace with hybrid stereo/monocular tracking . . . 68
4.3.1 Off-line system calibration . . . 69
4.3.2 On-line real-time stereo tracking . . . 71
4.3.3 Monocular tracking mode . . . 76
4.4 CoCaTrack: Increasing optical tracking workspace with controlled cameras . 77
4.4.1 Off-line controlled camera calibration . . . 78
4.4.2 Registration . . . 80
4.4.3 Controling camera displacements: visual servoing . . . 80
4.5 Proofs of concept . . . 82
4.6 Performance . . . 84
4.6.1 Main results of our global approach . . . 84
4.6.2 Comparison with Vicon's optical tracking . . . 87
4.7 Conclusion . . . 90

5 Mobile spatial augmented reality for 3D interaction with tangible objects 91
5.1 The MoSART approach: Mobile spatial augmented reality on tangible objects 93
5.2 Proof of concept . . . 94
5.2.1 Optical tracking . . . 96
5.2.2 Projection mapping . . . 97
5.2.3 Interaction tools . . . 99
5.2.4 Adding collaboration . . . 100
5.2.5 Characteristics and performances . . . 101
5.3 Use cases . . . 102
5.3.1 Virtual prototyping . . . 102
5.3.2 Medical visualization . . . 103
5.4 Discussion . . . 104
5.5 Conclusion . . . 106

6 Introducing user's virtual shadow in projection-based systems 107
6.1 Related work on virtual shadows . . . 108
6.2 Studying the use of virtual shadows in projection-based systems . . . 110
6.2.1 Objective . . . 110
6.2.2 Apparatus . . . 110
6.2.3 Participants . . . 112
6.2.4 Experimental task . . . 113
6.2.5 Experimental protocol . . . 113
6.3 Results . . . 115
6.3.1 Performance measurements . . . 115
6.3.2 User experience questionnaires . . . 116

6.4 Discussion 119
6.4.1 Virtual shadows and virtual embodiment 119
6.4.2 Virtual shadows and spatial perception 120
6.5 Conclusion 121

7 Conclusion **123**

Author's publications **129**

A Appendix : Résumé long en français **131**

List of Figures **147**

Bibliography **164**

List of Acronyms

AR Augmented Reality. 1–3, 7, 8, 11, 12, 14–16, 20, 21, 23, 24, 28–30, 32, 35, 36, 38–45, 75, 92, 104–106, 124, 127

AV Augmented Virtuality. 1

CAD Computer-Aided Design. 42, 44

CAVE Cave Automatic Virtual Environment. 3, 19, 42–45, 123, 125

CoCaTrack Controlled Camera Tracking. 8, 60, 77, 90, 123–125

DoF Degrees of Freedom. 5, 25, 37, 51, 92

EKF Extended Kalman Filter. 28

FoR Field-of-Regard. 5, 12, 15, 18, 23, 24

FoV Field-of-View. 12, 14, 15, 18, 23, 24, 34, 94, 101, 104

fps frames per second. 12

HMD Head-Mounted Displays. 3–5, 7, 13–16, 36, 39, 41–45, 121, 123, 124

HMPD Head-Mounted Projection-based Displays. 36

Hz Hertz. 12, 17, 26, 36, 37, 82

IMU Inertial Measurement Units. 27, 28

IPS Immersive Projection-based Systems. 7, 119

MEMS MicroElectroMechanical Systems. 27

MonSterTrack Monocular and Stereo Tracking. 8, 60, 78, 90, 123, 125

MoSART Mobile Spatial Augmented Reality on Tangible objects. 8, 92, 93, 98, 102, 104–106, 125, 126

MR Mixed Reality. 1–8, 11, 13, 24, 25, 34, 37–45, 48, 57, 60, 75, 90, 106, 123, 124, 127

OST Optical See-Through. 7, 14, 15, 21, 92, 104, 105

P3P Perspective from 3 Points. 76

PBS Projection-Based Systems. 2–8, 21, 23, 24, 35–37, 44, 45, 48, 56, 57, 60, 90, 92, 106, 108, 110, 121, 123–125, 127

PnP Perspective from n Points. 76

SAR Spatial Augmented Reality. 20, 22, 36, 37, 92, 99, 105, 124

SMB Small and Medium Businesses. 3, 48

SSD Surround-Screen Displays. 19, 23, 36, 37

SVD Single Value Decomposition. 67, 70, 71

UAV Unmanned Aerial Vehicle. 5, 90, 124

VE Virtual Environment. 13, 24

VR Virtual Reality. 1–3, 5, 7, 11, 13, 15–17, 23, 24, 29, 30, 32, 35, 36, 39, 40, 42–45, 48, 49, 52, 56, 60, 68, 71, 75, 82, 87, 123–125, 127

VST Video See-Through. 7, 14, 16, 43, 92, 105

Notations

In this document we use the following conventions for every equation and mathematical notation.

- $\mathcal{F}_w$ refers to an orthonormal frame.
- $\mathbf{a}$ refers to a column vector.
- $\mathbf{a}^\top$ refers to a row vector, written as the transpose of column vector $\mathbf{a}$.
- $[\mathbf{a}]_\times$ refers the skew-symmetric matrix of the 3-vector $\mathbf{a}$ defined by $[\mathbf{a}]_\times \mathbf{b} = \mathbf{a} \times \mathbf{b}$.
- $\mathbf{A}$ refers to a matrix.
- $\mathbf{A}^{-1}$ refers to the inverse of matrix $\mathbf{A}$.
- $\mathbf{A}^\top$ refers to the transpose of matrix $\mathbf{A}$.
- A_{ij} refers to the value of the i^{th} row and j^{th} column of matrix $\mathbf{A}$.
- $\bar{\mathbf{x}}$ refers to an image point coordinates expressed as $\bar{\mathbf{x}} = (x, y)^\top$.
- $\mathbf{x}$ refers to an image point homogeneous coordinates expressed as $\mathbf{x} = (x, y, 1)^\top$.
- $\bar{\mathbf{X}}$ refers to a 3D point coordinates expressed as $\bar{\mathbf{X}} = (X, Y, Z)^\top$.
- $\mathbf{X}$ refers to a 3D point homogeneous coordinates expressed as $\mathbf{X} = (X, Y, Z, 1)^\top$.
- ${}^w\mathbf{X}$ refers to point $\mathbf{X}$ which 3D coordinates are expressed in frame $\mathcal{F}_w$.
- ${}^i\mathbf{M}_j$ refers to the homogeneous transformation matrix from $\mathcal{F}_j$ to $\mathcal{F}_i$.
- ${}^i\mathbf{R}_j$ refers to the rotation matrix from frame $\mathcal{F}_j$ to frame $\mathcal{F}_i$.
- ${}^i\mathbf{t}_j$ refers to the translation vector from frame $\mathcal{F}_j$ to frame $\mathcal{F}_i$.

Introduction

1

This PhD thesis is entitled "**Contribution to the Study of Projection-based Systems for Industrial Applications in Mixed Reality**". This work has been carried out in an industrial context and aimed at improving the usage of mixed reality projection-based systems for industrial applications.

This thesis belongs to the field of Mixed Reality (MR). Mixed Reality was introduced by Milgram et al. [1995] as a concept that combines several immersive technologies. These technologies are placed along a virtual continuum (Figure 1.1) that takes the human from a real environment into a virtual one, better known today as Virtual Reality (VR).

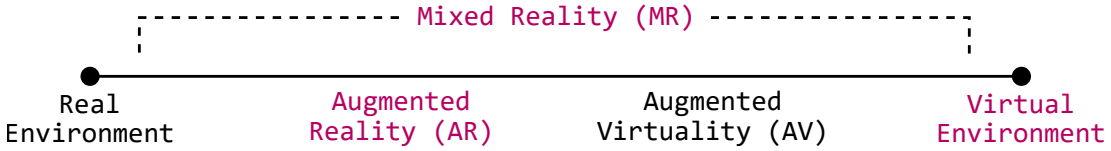


Figure 1.1 – The virtuality continuum proposed by Milgram et al. [1995].

Several definitions of VR have been proposed in the literature and, in this manuscript, we refer to the one proposed by Arnaldi et al. [2003]:

> "Virtual Reality is a technical and scientific area making use of computer science and behavioral interfaces in order to simulate 3D entities behavior in a virtual world that interact in real time among themselves and with the user in pseudo-natural immersion through sensory-motor channels."

Regarding Augmented Reality (AR), we refer to the definition from Azuma [1997], revised by Azuma et al. [2001] and supported by the recent book from Schmalstieg and Hollerer [2016]. This definition describes AR as:

> "[...] any system that has the following three characteristics:
>
> - combines real and virtual objects in a real environment;
> - runs interactively, and in real time; and
> - registers (aligns) real and virtual objects with each other."

Still displayed in this continuum, Augmented Virtuality (AV) consists in enhancing a virtual world with objects of the real world. For example Bruder et al. [2009] used AV to display the physical hands of the user over the virtual environment. Augmented

Virtuality will not be taken into account within this thesis. In the following we will only consider VR and AR systems and their applications.

Different MR systems have been developed for either AR or VR applications. We can classify these systems into three main categories: near-eye displays, screen displays and projection-based displays (as depicted in Figure 1.2).

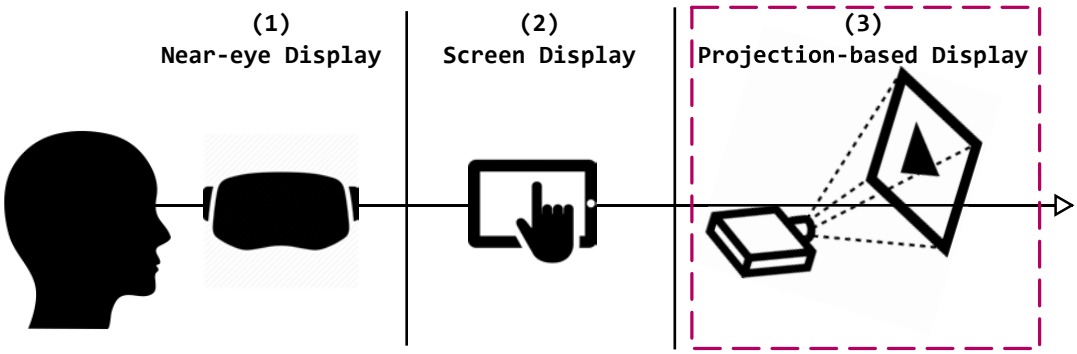


Figure 1.2 – Simplified classification of Mixed Reality systems: (1) near-eye displays, (2) screen displays and (3) projection-based displays. This PhD thesis is focused on projection-based displays.

These categories are detailed as follows:

- **Near-eye displays** are systems that "place the image directly in front of the user's eye using one or two small screens" [Bowman et al., 2004]. Near-eye displays, which are commonly head-mounted (or head-worn), have been popularized recently with VR systems such as the HTC/Vive or the Oculus rift and AR systems such as the Microsoft Hololens (see Figure 1.3-left).

- **Screen displays** use standard screens to provide MR environment. For VR applications, any 3D screen can be used as a display. Holobenches are more advanced technologies that are composed of two 3D screens that are placed in right angle as depicted in Figure 1.3-middle. Regarding AR, using a tablet or smartphone is the most popular way to enhance reality with virtual objects.

- **Projection-Based Systems (PBS)** use projectors to display virtual content. For VR applications, Cruz-Neira et al. [1993] introduced one of the first immersive PBS, the CAVE display. This display consists of 4 screens on which 3D content is front- or rear-projected (see Figure 1.3-right). Regarding AR, projection-based AR systems project virtual content directly over physical objects or flat surfaces.

Within the scope of this thesis we focus mainly on **Projection-Based Systems (PBS)** for industrial usage in MR.

This PhD thesis has been conducted within the frame of a CIFRE[1] industrial partnership with Realyz. Realyz is a french company that is specialized in the design of VR and AR Projection-Based Systems for the industry. Although the industry has been one

[1]Conventions Industrielles de Formation par la Recherche (Industrial agreements for training through research) - http://www.enseignementsup-recherche.gouv.fr/cid22130/les-cifre.html

Figure 1.3 – Examples of mixed reality displays: the Microsoft Hololens near-eye display (left), an Holobench screen display from Realyz (middle) and a CAVE projection-based system from Realyz inspired by Cruz-Neira et al. [1993] (right).

of the first actors to be attracted by MR technologies [Schmalstieg and Hollerer, 2016], many industrial facilities are not yet familiar with them. Indeed these technologies are recent and one of the goals of Realyz is to popularize the immersive technologies within the Small and Medium Businesses (SMB). In fact, even for SMB, Mixed Reality proposes an alternative and faster way to dynamically visualize, modify and validate the design of many projects [Mourtzis et al., 2014]. For example mixed reality applications can help visualizing the design of an engine or validating the arrangements and dispositions inside a factory as depicted in Figure 1.4. MR also makes it possible to train or assist operators on a specific task or on maintenance/repair procedures with lower expenditures [Nee et al., 2012]. Nevertheless, in order to attract the smallest industrial actors, cost-effective systems should be proposed. Thus, Realyz aimed at exploring different ways of improving the user experience and the usage of projection-based systems for industrial applications in order to make PBS a competitive alternative to near-eye head-mounted displays.

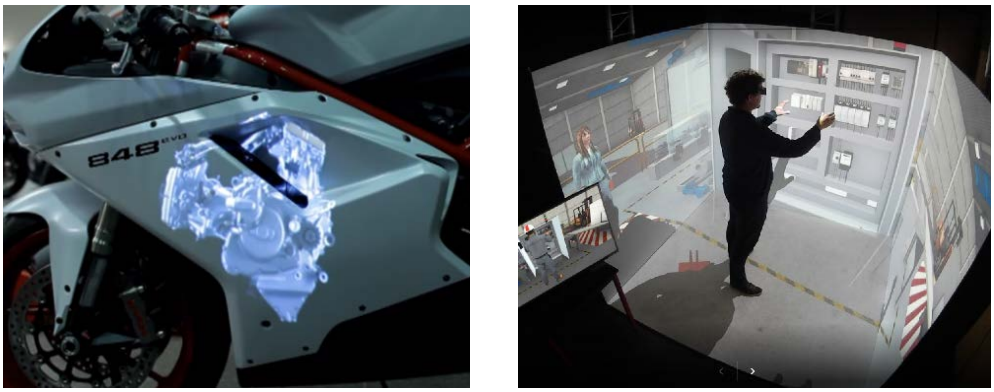


Figure 1.4 – Mixed Reality industrial use cases at Realyz: visualizing a motorcycle engine using AR (left) or visualizing and validating a factory layout project using VR (right).

Even though several industrial actors have already adopted near-eye Head-Mounted Displays (HMD) (e.g., the Hololens, the HTC/Vive) there is still an increasing demand for immersive Projection-Based Systems (PBS) such as CAVE displays. Indeed one main benefit of immersive PBS for industrial purposes is that, in addition to providing adequate immersion for the main user, projective displays generally also enable to visu-

alize the virtual environment from an external point of view. Such configuration makes it possible for several external persons to straightforwardly share the experience and collaborate with the main user (e.g., by directly pointing virtual objects). Moreover, using near-eye HMD raises ethical and sanitary issues. Such issues induce cumbersome processes like cleaning the device area that is in contact with the skin after each use. Thus, the industrial structures that are not familiar with MR technologies are commonly reluctant to the idea of sharing a HMD compared to sharing stereoscopic glasses.

Challenges of projection-based systems

Figure 1.5 depicts the standard framework of Mixed Reality (MR) applications using Projection-Based Systems (PBS). As illustrated in Figure 1.5, the tracking system captures the motion of user. The tracking data is then interpreted by the interaction techniques that send commands to update the application. Then the MR application modifies the interaction state and is displayed on the projective screen. Finally, the projective screen provides visual feedback to the user. Figure 1.5 also displays the three research axes of this PhD thesis that are detailed in the next section.

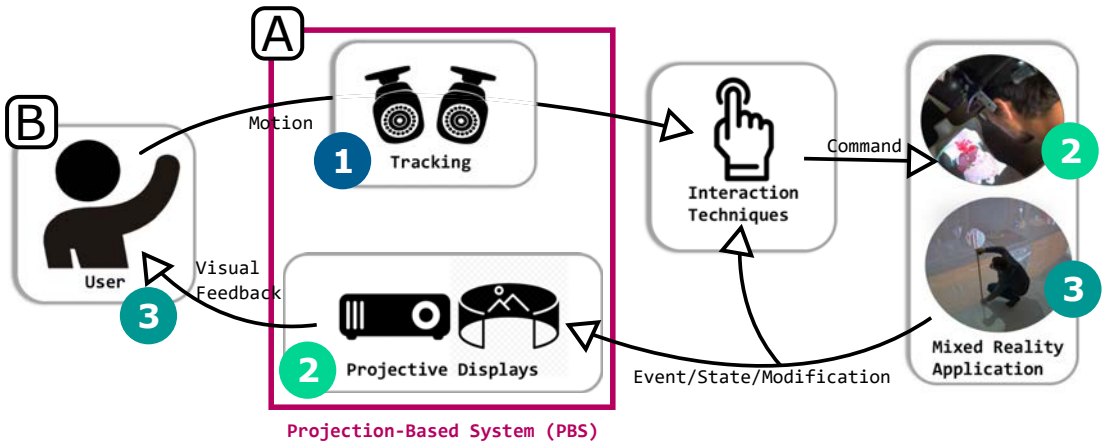


Figure 1.5 – Global framework of a mixed reality PBS that is considered in this work. The two main challenges of the PhD thesis are: (A) improving the technical components of PBS and (B) improving the user experience when using PBS. The three research axes are: (1) Improving the tracking systems used in PBS, (2) Proposing a novel paradigm for PBS and (3) Increasing the user perception and experience in immersive PBS.

This thesis will focus on two main challenges that still need to be addressed to make PBS an alternative to near-eye HMD: **(A) A techno-centered challenge: improving the technical components of PBS** and **(B) A user-centered challenge: improving the user experience when using PBS**.

(A) Improving the technical components of projection-based systems

The first challenge is a techno-centered challenge that aims at improving the technical components of Mixed Reality (MR) Projection-Based Systems (PBS). According to the

framework presented in Figure 1.5 we consider two technical components (depicted in red): the **tracking system** and the **projective display**.

Regarding **the tracking systems**, they have been widely used to track movement and objects in several fields such as surgery [Taylor et al., 1994], virtual reality [Pintaric and Kaufmann, 2007] or Unmanned Aerial Vehicle (UAV) localization [Mustafah et al., 2012]. To provide an optimal tracking system for MR several requirements should be addressed. As stated by Welch and Foxlin [2002] an optimal tracking system should be: tiny, self-contained, complete (6DoF), accurate, fast, immune to occlusion, robust, tenacious, wireless and cheap. Lots of work has been carried out to fulfill as many requirements as possible, but, as Welch and Foxlin [2002] said: there is still "no silver bullet, but a respectable arsenal" of tracking technologies. Hence, there is a need to **optimize and adapt the tracking systems and performances to the application requirements**, in our case, to industrial applications in mixed reality projection-based systems.

Regarding **the projective displays**, one of the main limitations of Projection-Based Systems (PBS) is that using projectors is generally cumbersome and that the system is generally static. When using projectors there is a compromise between the projection area, the spatial resolution of the images and the projection distance. In fact, high resolution projectors are generally bulky and projecting a large image (for immersive systems for example) generally requires a long distance between the projectors and the screens. Solutions to overcome such limitations, like using high-resolution short-throw pico-projector, are generally expensive and the technologies are not yet easily available. Moreover these limitations prevent the usage of these systems in a wearable way, like head-mounted PBS. Thus there is a need to study ways to compact such systems and **propose wearable and mobile PBS in MR**.

(B) Improving the user experience and interaction techniques when using projection-based systems

The second challenge is a user-centered challenge. When using Projection-Based Systems (PBS) the users are generally aware of their own body and of the real environment that surrounds them. Therefore, unlike Virtual Reality (VR) HMD, it is difficult to make the user embody a virtual character in PBS. Indeed avatars cannot be straightforwardly introduced in the virtual environment due to the physical presence of the user's body. Finally, immersive PBS are generally less immersive than HMD since they do not propose a 360° Field-of-Regard (FoR). Regarding the 3D interaction techniques, the industrial fields generally require the interaction to be adapted to the real task and to simulate it as faithfully as possible. Therefore there is still a need to propose **novel immersive techniques and interaction paradigms to improve the user experience when using PBS**.

Research axes

This PhD thesis focuses on studying and adapting projection-based system to industrial requirements when using mixed reality applications. We considered three main axes of research: **(1) Improving the tracking systems used in projection-based systems**, **(2) Proposing a novel paradigm for MR projection-based systems** and **(3) Increasing the user perception and experience in immersive projection-based systems**. These three axes of research led to three main contributions that are illustrated in Figure 1.5 and 1.6 and are detailed in the following.

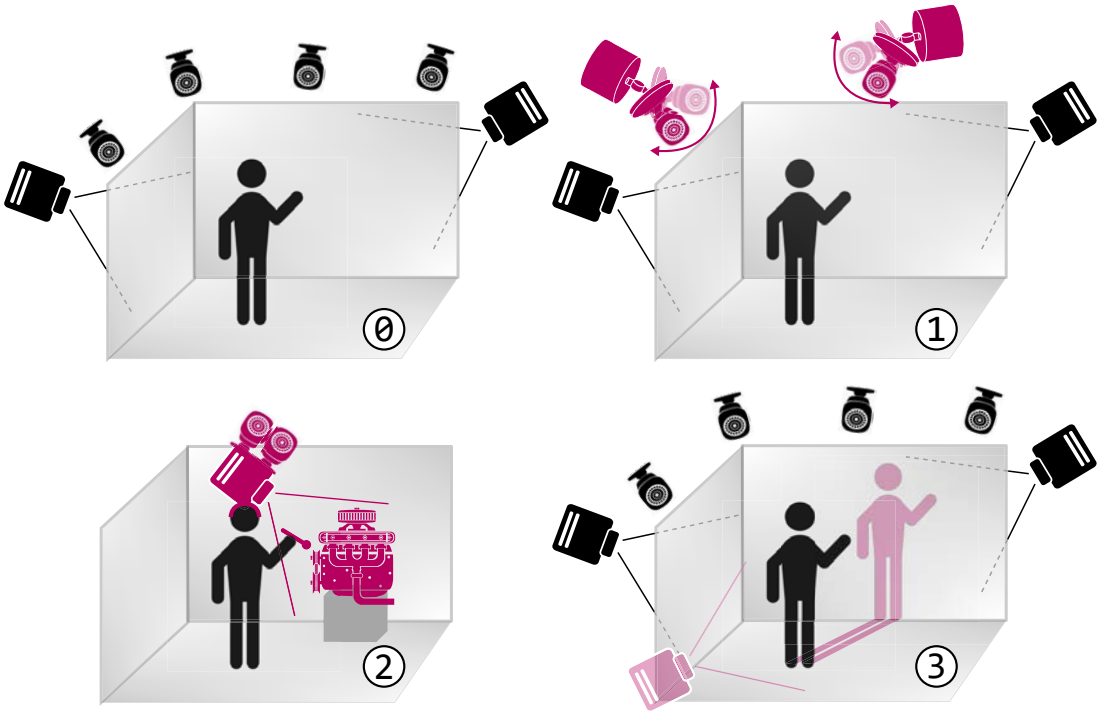


Figure 1.6 – The three contributions of this PhD thesis. Standard projection-based VR systems are illustrated in (0). Contribution (1) aimed at improving optical tracking systems to increase the workspace, notably by using controlled cameras. Contribution (2) aimed at proposing a novel paradigm for mobile spatial augmented reality. Contribution (3) aimed at improving the user perception and experience in projection-based systems by introducing virtual shadows.

(1) Improving the tracking systems used in projection-based systems: Toward an increased workspace

In this axis of research (illustrated in Figure 1.6-top-right) we do not claim to have found the "silver bullet" tracking system (as defined by Welch and Foxlin [2002]) but we chose to address some tracking requirements. Regarding PBS, having a tiny, self-contained and wireless tracking system might not be a priority since PBS are generally cumbersome and already involve having wired connections. Nevertheless for MR ap-

plications the tracking system has to be fast, accurate, complete, robust, and as much as possible, immune to occlusions. Also the tracking system should adapt itself to the range of interactions proposed by the application. Therefore, the tracking system needs to be tenacious and, if needed, it should be able to follow the objects as far as possible. Finally, when considering industrial applications, the tracking system should be as cost-effective as possible. Hence our first objective is to propose an optical tracking system that fulfills as many requirements as possible from the ones selected above. Among other things we propose an approach to **increase the workspace and reduce the occlusion issues of optical tracking systems in PBS for VR and AR applications**.

(2) Proposing novel paradigms for projection-based systems: Toward mobile spatial augmented reality

As mentioned before, industrial actors often require to be able to collaborate directly with the main user and Projection-Based Systems (PBS) can be a good candidate for such applications. Nevertheless PBS generally require to have large room to work at their optimal performance (e.g., large field of projection) and are generally static systems. These constraints generally prevent from using PBS for mobile applications for example. However several industrial applications may require to use a MR system over large areas (or large objects) and mobility can be a fundamental requirement for such use cases. These use cases are generally related to augmented reality applications that require to enhance sparse or large physical 3D objects with virtual content. To overcome this mobility limitation, near-eye displays (like Optical See-Through (OST) or Video See-Through (VST) HMD) can be used but these displays can limit straightforward collaboration compared to PBS. Therefore, in this contribution (illustrated in Figure 1.6-bottom-left) we propose to **design a novel approach for mobile spatial augmented reality on tangible objects**.

(3) Increasing the user perception and experience in immersive projection-based systems: Toward users virtual shadows

When using Immersive Projection-based Systems (IPS) the users are generally aware of the real environments and in particular they are aware of their own body. In some cases this awareness can be an asset (e.g., having a real reference, reducing motion sickness). However the awareness of the real environment can, in some cases, reduce the immersion and the user experience. Even though the interaction techniques designed for HMD are generally compatible with IPS, the compatibility is not straightforward regarding virtual embodiment. The full immersion provided by HMD enables to create virtual body ownership illusions and virtual embodiment thanks to virtual avatars [Kilteni et al., 2012]. But is it possible to embody someone else in an IPS even though the users are aware of their own body? In this contribution (illustrated in Figure 1.6-bottom-right) we propose to address this issue by **introducing a virtual shadow of the user. This virtual shadow represents the virtual projection of the user in the virtual environment**. This approach is expected to improve the user experience when using IPS.

Outline

This manuscript is composed of 6 chapters that are summarized as follows.

In **Chapter** 2 we present an overview of related work on the main components of Mixed Reality (MR) systems. First we present the different mixed reality displays together with their main drawbacks and benefits. Then MR tracking technologies are described including a close-up of optical tracking techniques. Finally we present the main industrial applications that can benefit from MR systems.

In **Chapter** 3 we present a pilot study that aims at defining guidelines for designing and deploying Projection-Based Systems (PBS) better adapted to the end-use industrial application. The study was conducted in "out-of-the-lab" conditions during a professional construction industry exhibition. We recorded users interactions as well as head and hand motions. 3D motion data was analyzed in order to extract the users' behavior which could be used afterwards to improve the configuration of the PBS system. The proposed methodology could also be used to analyze other PBS and application scenarios in order to extract the user behavior and optimize them accordingly.

In **Chapter** 4 we propose an approach which intends to maximize the workspace of optical tracking systems used in PBS and overcome some occlusion problems. As such, since adding sensors can be expensive and is not always possible due to the lack of space, we propose not to use additional cameras. Our approach is based on two complementary methods. As a first method, when the tracked target becomes no longer visible by at least 2 cameras (e.g., due to occlusions, or when moving outside the stereo workspace), we occasionally enable switching to a monocular tracking mode using 2D-3D registration algorithms (*MonSterTrack* method). Then we propose to use controlled cameras mounted on automated motors that follow the tracked objects through the workspace (*CoCaTrack* method). Thus, the stereo registration can be achieved as much as possible using the moving cameras.

In **Chapter** 5 we promote an alternative approach for head-mounted spatial augmented reality which enables mobile and direct 3D interactions with real tangible objects, in single or collaborative scenarios. Our novel approach, called *MoSART* (for Mobile Spatial Augmented Reality on Tangible objects) is based on an "all-in-one" headset gathering all the necessary AR equipment (projection and tracking systems) with a set of tangible objects and interaction tools. The tangible objects and tools are tracked thanks to an embedded optical tracking. The user can walk around, grasp and manipulate the tangible objects and tools that are augmented with projection mapping techniques. The experience can be shared with other users thanks to direct projection/interaction. In a nutshell, our approach is the first one which enables direct 3D interaction with tangible objects, mobility, multi-user experiences, in addition to a wider field-of-view and low latency in AR.

In **Chapter** 6 we introduce the use of virtual shadows in PBS. The virtual shadows make it possible to provide a virtual representation of the user which differs from their own physical body even though they are still able to see it. We carried out an experiment in order to assess the user's virtual embodiment and the user's perception

in presence of the virtual shadows. In particular, we study how the users appropriate different virtual shadows (male and female shadow) and how does the virtual shadow affects the user behavior when performing a 3D positioning task.

Finally, **Chapter** 7 concludes the manuscript and discusses short-term future work for each of our contributions as well as long-term perspectives regarding the futuristic mixed reality projection-based systems.

Related work

2

Contents

2.1 Visual displays for mixed reality **11**
2.1.1 Non-projection-based visual displays 13
2.1.2 Projection-based displays 19
2.2 Tracking systems for mixed reality **24**
2.2.1 Non-optical tracking systems 25
2.2.2 Optical tracking systems 29
2.2.3 Discussion on optical tracking systems for projection-based displays . 35
2.3 Industrial applications of mixed reality **37**
2.3.1 Training applications 39
2.3.2 Assistance applications 40
2.3.3 Design applications 42
2.3.4 Planning and validation applications 43
2.4 Conclusion **44**

This chapter presents an overview of previous work that relates to mixed reality systems. First an overview and classification of the existing Mixed Reality (MR) visual displays is proposed. Next the different tracking systems are presented. We discuss to what extend each technology is promising and we make a close-up on optical tracking. Finally we present different usages of mixed reality systems for industrial application together with concrete examples in different industrial fields of activity.

2.1 Visual displays for mixed reality

Mixed Reality (MR) originally proposes to stimulate the visual sense of the users by modifying their current visual perception of the reality. Later, MR systems also focused on stimulating other senses such as the haptic, auditory or olfactory ones. In the following we focus on the visual sense and present the visual displays that are used for either Virtual Reality (VR) or Augmented Reality (AR).

MR systems are evaluated according to the user experience when using the application. For example, having a system that provides a realistic environment while being comfortable and immersive is generally appreciated. Different characteristics of MR systems have to be considered to optimize the user experience. According to LaViola Jr. et al. [2017] some of these characteristics are the following ones :

- *Stereoscopy*: The systems can display either monocular or stereoscopic images. Monocular images are the standard images, displayed by standard TV screens. Stereoscopic (or binocular) displays take into account the human binocular vision and display different images for each eye to create a stereoscopic effect.

- *Field-of-View (FoV)*: Maximum number of degrees of visual angle that can be seen instantaneously on a display horiontally (HFoV) and vertically (VFoV). Human eye HFoV is of around 200°-220° and VFoV of around 120°.

- *Field-of-Regard (FoR)*: Amount of physical space surrounding the user in which images are displayed. *FoV* and *FoR* are not linked since some displays can, for example, provide large *FoR* and small *FoV*.

- *Spatial resolution*: Ratio between the number of pixels that are displayed and the size of the screen. It is considered as a measure of visual quality. Two display with the same pixel resolution may not have the same *spatial resolution*.

- *Screen geometry*: Geometry of the display. It can be rectangular, hemispherical, L-shaped, etc. Different geometries can lead to different visual quality due to the distortions of the images for example.

- *Refresh rate*: Speed with which a visual display refreshes the displayed images. It is generally expressed in Hertz (Hz). The refresh rate of the display is different from the frames per second (fps) of the application. The fps is the rate with which the images are rendered by the application. It can be faster than the refresh rate of the display.

Other characteristics can be considered, e.g., the number of depth cues, the ergonomics or the production cost.

A classification of AR visual displays has been proposed by Bimber and Raskar [2005] and Schmalstieg and Hollerer [2016] according to the distance between the display and the user. The different displays are classified as follows: **near eye displays**, **handheld displays**, **stationary displays** and **projection-based displays**.

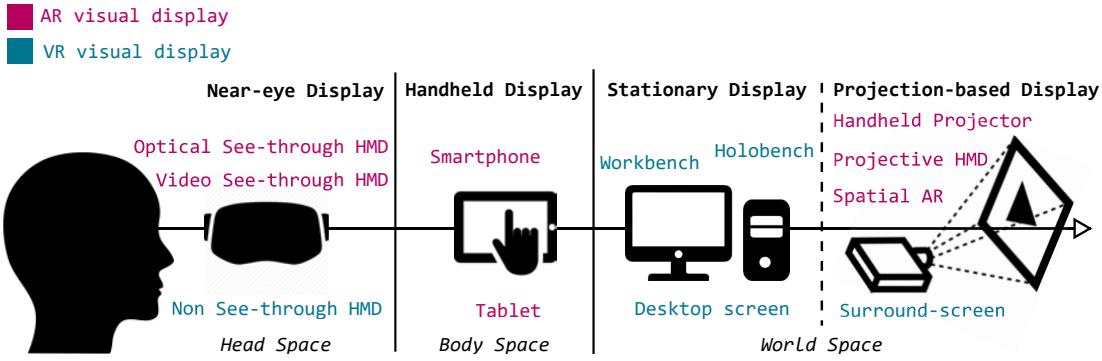


Figure 2.1 – Classification of MR visual displays according to the distance between the display and the user's eye, inspired from Bimber and Raskar [2005] and Schmalstieg and Hollerer [2016]

In the following we present an overview of MR visual displays by following the same classification as depicted in Figure 2.1. We adapted the classification by introducing the VR displays into the different categories.

2.1.1 Non-projection-based visual displays

2.1.1.1 Near-eye displays

Near-eye displays are devices that "place the image directly in front of the user's eye using one or two small screens" [Bowman et al., 2004]. Nowadays near-eye displays are generally head-mounted (or head-worn [LaViola Jr. et al., 2017]). Wearing a device on the head involves complex designs in order to make the device unobstrusive and comfortable [Rolland and Cakmakci, 2009]. Thereby, the first Head-Mounted Display (HMD), the "Sword of Damocles", invented by Sutherland [1968] (illustrated in Figure 2.4-left) was suspended from the ceiling. Since then, miniaturizing HMD and reducing their weight has been achieved. Different technologies are available to design near-eye displays: non see-though, optical see-though and video see-though. These technologies are presented in the following.

Non see-though HMD

Non see-through HMD are the most common near-eye displays used for VR applications. These displays are completely opaque and do not enable to see any piece of the real world. Therefore, non see-through HMD provide complete immersion in the Virtual Environment (VE). Nowadays some of the major companies have developed their own non see-through HMD like HTC, Oculus or Samsung (see Figure 2.2). These devices use either two small LCD (or OLED) retinal screens or a smartphone combined with optical lenses that redirect the images in either the right or left eye. Generally the devices that use smartphones are limited by the smartphone performances and it is common to have more latency and a lower refresh rate.

Figure 2.2 – Examples of non see-through HMD used for VR applications: the HTC Vive (left), the Oculus Rift (middle) and the Samsung Gear VR which relies on a smartphone (right).

Video see-through HMD

Video See-Through (VST) Head-Mounted Displays (HMD) are commonly used for AR applications by enhancing virtual content over regular video streams [Mohring et al., 2004] (see Figure 2.3-left). One straight forward way to design VST-HMD is by using the front camera of a smartphone to capture the real environment (see Figure 2.2-right). Then the virtual content is added over the video stream and displayed on the smartphone screen. The smartphone is then embedded on a case which is mounted on the user's head. As for non see-through HMD using smatphones generally involves a higher latency. An alternative to smartphones is the design of HMD with embedded cameras, generally two, that are responsible of capturing the real environment in front of the user. Nevertheless since the captured video stream generally has a lower resolution than the user's eye, VST technologies present a limitation in terms of content resolution.

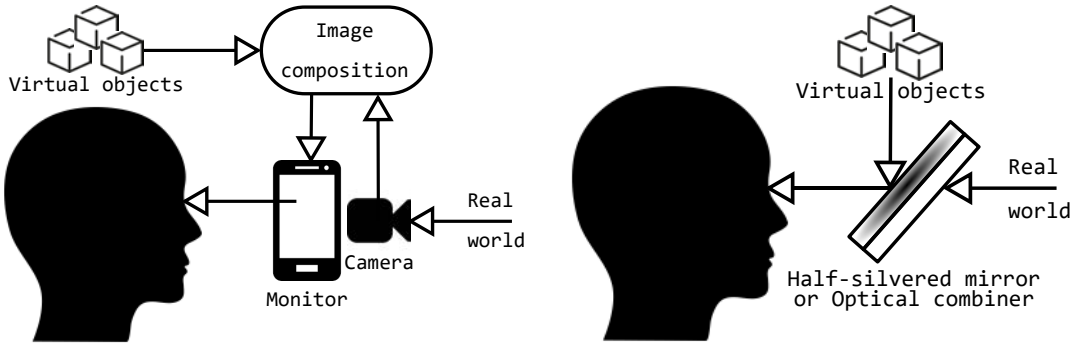


Figure 2.3 – Video see-though (left) and Optical see-through (right) technologies

Optical see-through HMD

Optical See-Through (OST) Head-Mounted Displays (HMD) enable to directly see the real environment though a half-silvered mirror (or optical combiners) on which virtual content is displayed (see Figure 2.3-right). The first HMD was in fact an OST-HMD [Sutherland, 1968] (see Figure 2.4-left). The main advantage of this technology is that it provides a direct view of the real environment. Nevertheless, nowadays, OST devices generally provide a restricted FoV for displaying virtual content, hardly more than $40° \times 20°$ (HFoV×VFoV). Work from Cheng et al. [2009] introduced a technology to provide a higher FoV by using a free-form prism. Later, Zheng et al. [2010] designed an off-axis optical system using polynomial surfaces for OST-HMDWork has also been done in order to miniaturize and lighten OST-HMD while increasing their technical performance [Kiyokawa et al., 2003; Maimone et al., 2014; Olwal et al., 2005; Sutherland, 1968]. Recently, OST-HMD, particularly the Microsoft Hololens (see Figure 2.4-right), have been democratized and are used in many AR industrial applications.

Benefits and drawbacks of near-eye displays

In general near-eye displays are portable and, apart from wearing a large but light enough display on the head, the user does not have many constraints. Some near-

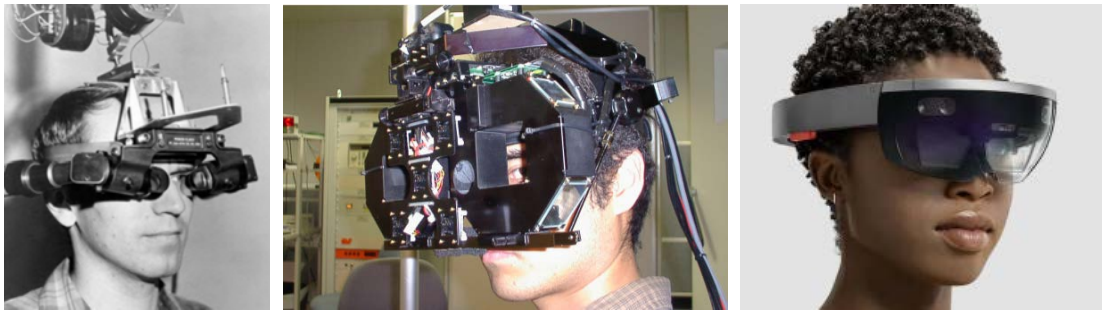


Figure 2.4 – Examples of OST-HMD: the Sutherland pioneer HMD [Sutherland, 1968] (left), the design of the ELMO-4 from Kiyokawa et al. [2003] (middle) and the Hololens from Microsoft (right).

eye displays still require to have a computer to render the image. This computer can be carried on a backpack to overcome the issues arouse by the wired connections (particularly when using non see-through HMD since the user is not aware of the position of the wires). Near-eye displays relying on smartphones can compute the virtual content on the embedded phone unit. Compared to other displays, there is no viewpoint problem when using near-eye displays and the images do not need to be rendered depending on who is looking at them. Indeed the rendering is made only for the main user. Regarding the immersion, Non see-through HMD can be a good candidate for VR applications since the virtual environment is displayed within a 360° FoR and generally no piece of the real world is visible.

However, the usage of near-eye displays is most of the time limited to one user and there is no straightforward way to collaborate with an external group of persons. Also, the lack of real landmark when using non see-though HMD can create more motion sickness that other visual displays. As of today, near-eye HMD can sometimes involve viewing discomfort due to either the vergence-accomodation conflict described by Kramida [2016] or the variable inter-ocular distance between users as noted by Speranza et al. [2006]. Finally the FoV of most near-eye displays is reduced which can lead to a biased perception of the environment [Knapp and Loomis, 2004]. Also, their spatial resolution is smaller than the one of surround-screen displays for example (see section 2.1.2).

2.1.1.2 Handheld displays

Handheld displays are one of the first democratized displays that provide Augmented Reality (AR) environments. Indeed handheld display are mainly tablets or smartphones that can be held on the hand. For AR purposes, the real environment is captured thanks to the back camera of the device and is displayed on its main screen where virtual content is added (see Figure 2.5). Using such approach generally limits the point of view of the user to the one of the camera of the device.

Recent work has been proposed to provide the users with a coherent viewpoint when using handheld devices [Baricevic et al., 2012; Hill et al., 2011]. Such approaches are more expensive but enable tracking the user head (e.g., thanks to the front camera of the device) and reconstructing the view of the back camera (see Figure 2.5).

Even though handheld displays are well-known for AR purposes, these devices gen-

erally do not propose stereoscopic viewing. The absence of stereoscopic effect involves that no immersion can be provided by these systems. Therefore handheld devices are not commonly used for VR applications.

Figure 2.5 – Handheld displays are mainly used for AR purposes. Devices have been proposed by Hill et al. [2011] (left) and Baricevic et al. [2012] (right) to display a coherent viewpoint when using handheld displays.

Benefits and drawbacks of handheld displays

Handheld devices propose a fast and efficient way to enhance the reality with accessible and cheap hardware like handheld tablets and smartphones. Nevertheless one main drawback of such technology is that it generally requires to hold the display with one or both hands which limits the interaction possibilities and the comfort when using the device. Holding the display on the hand facing toward the area of interest can lead to a fatigue in the arms after a rather short period [LaViola Jr. et al., 2017]. Many handheld displays are not equipped with stereoscopic rendering and therefore they provide less depth cue than stereoscopic displays. Moreover, like for Video See-Through (VST) Head-Mounted Displays (HMD), the real environment is captured thanks to the cameras of the handheld device which generally involves higher latency and lower quality than other displays. Finally handheld devices are commonly small what makes them very portable but does not make them immersive.

2.1.1.3 Stationary displays

Desktop screens

One of the easiest way to provide visual feedback for either AR or VR applications is by using desktop displays. As for handheld displays like tablets, desktop displays provide a straightforward AR device. By using a webcam, the real world can be captured and augmented in the desktop screen. Nevertheless contrary to handheld devices these systems are commonly stationary which restricts their workspace. As for tablets, when using AR, stereoscopy is generally not available on desktop screens. Indeed, reconstructing the stereoscopic view of the real environment is not straightforward [Baricevic et al., 2012].

For VR purposes, stereoscopic desktop screens can be used with stereo glasses to display the virtual environment. Either passive or active glasses can be used. Passive stereo glasses can either work on the spectral emission (*spectral multiplexing*) of the light or on its polarity (*polarization multiplexing*). When using spectral multiplexing the stereo glasses are equipped with colored filters and are commonly called anaglyphic glasses (Figure 2.6-left). Classical anaglyphic glasses are built with a cyan and a red filter so that one eye filters cyan and the other red. These devices are very inexpensive but present a color limitation. Using the polarity of the image is another technique that uses two opposite polarized filters (Figure 2.6-middle) to filter the images. Classical examples are the horizontal polarized and the vertical polarized filters that enable the images going through either the left or right eye. Active stereoscopic glasses (Figure 2.6-right) are generally considered as the best solution in terms of quality for providing stereoscopic effect [LaViola Jr. et al., 2017]. Active glasses work by sequentially opening and closing a shutter at the same refresh rate as the display (*temporal multiplexing*). Thus when the screen displays the image destined for the right eye, the left shutter closes and the right opens and so on. Using temporal multiplexing generally involves that the screen has to have a refresh rate twice bigger than the intended refresh rate. For example if a VR application is required to run at 60Hz then the display needs to have a 120Hz refresh rate.

Figure 2.6 – Examples of stereoscopic glasses: Anaglyphic stereo glasses that use spectral filters (left), polarized stereo glasses that use polarized filters (middle) and active stereo glasses that work with shutters (right).

Some desktop displays, called auostereoscopic displays, do not require to wear stereo glasses to have a stereoscopic effect. The survey from Holliman et al. [2011] provides an overview of existing autostereoscopic techniques and the one from Wetzstein et al. [2012] introduces light-field screens for stereoscopic rendering.

Workbench displays

The workbench displays "were designed to model and augment interaction that take place on desks, tables and workbenches" [Bowman et al., 2004]. Workbenches use projectors or LCD screens to display stereoscopic images over surfaces that are entirely reachable by arm length. The users need to wear active or passive stereoscopic glasses (see above) and their point of view is tracked.

The first workbench display, the Responsive Workbench, was introduced by Krüger et al. [1995]. This system was designed after analyzing the working environment of physicians, engineers, architects, etc. The display is made of a projector, a mirror and a glass surface as depicted in Figure 2.7-left. The display is placed horizontally to simulate an interactive table. Inspired from the Responsive Workbench, Czernuszenko et al. [1997] designed the Immersadesk. The Immersadesk is made of wood in order to

use a magnetic tracking systems and not interfere with a stain structure. Moreover, to ease its usage, the Immersadesk is composed of one screen that is no more horizontal but is almost vertical (see Figure 2.7-middle). Smaller versions of workbench displays have been designed to provide some portability. Infinite Z company has developed the zSpace which is a portable monitor-sized workbench.

A main variation of the workbench is the holobench which is commonly used in many industrial applications. The holobench is composed of two screen in right angle (one vertical and one horizontal (see Figure 2.7-right) and provides a holographic perception of the objects that are visualized or manipulated.

Figure 2.7 – Examples of workbench displays: the Responsive Workbench [Krüger et al., 1995] (left), the Immersadesk [Czernuszenko et al., 1997] (middle) and a Holobench from PSA Peugeot Citroën (right).

Benefits and drawbacks of stationary displays

Workbench displays provide higher spatial resolution than surround-screen displays because of their reduced size, which leads to a better visual quality. Furthermore, if they are equipped with an appropriate tracking and stereoscopic rendering, they can provide the same depth cue as other displays. Many workbench displays offer the possibility to rotate the screen and adapt it to different application, e.g., horizontal for surgical training and slightly tilted for 3D drawing.

However as for surround-screen displays there is a viewpoint problem when using workbench displays with more than one user. This issue can be solved using techniques similar to the one proposed by Agrawala et al. [1997] that involve displaying enough images for each eye of each user. Nowadays the technology only allows two user to have stereoscopic rendering with these displays. Another limitation of these displays is that the user motion is restricted. Indeed, the displays are not attached to head and are generally quite small and static. The size of the stationary displays involves that the FoV and FoR are limited and that immersion is not always achieved.

In terms of interaction, the stationary displays are generally well suited for selection techniques because the user's arm can reach all the surface without moving. Moreover these devices have been designed to fit with the common workstations of, e.g., physicians or architects and are adapted to object manipulation. However they are less suited for physical navigation techniques than other displays because of the lack of space the users can move in.

2.1.2 Projection-based displays

Projection-based displays rely on projectors that display images over real objects, flat surfaces or reflective screens (as depicted in Figure 2.8).

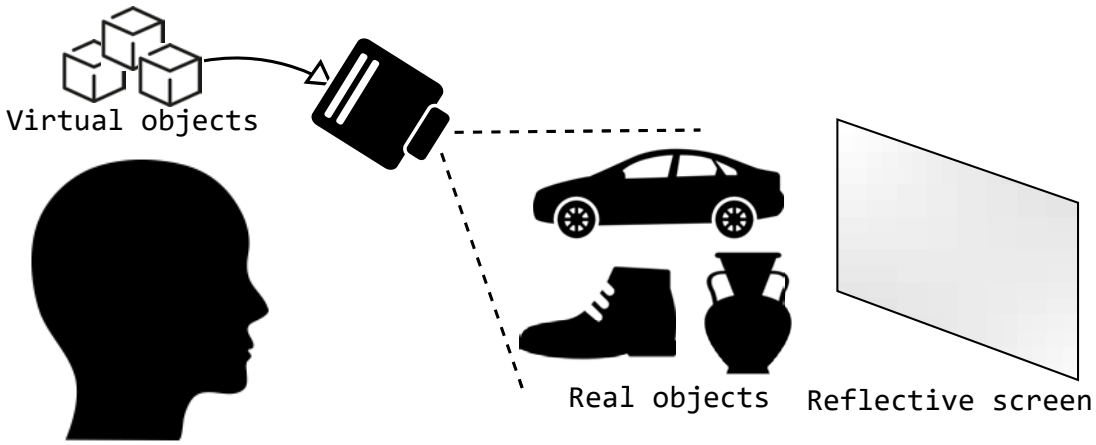


Figure 2.8 – Projection-based display technology: the virtual content is directly projected over the real environment or on reflective screens.

Surround-screen displays

Surround-Screen Displays (SSD) are defined as a "visual output device that has three or more large projection-based screens that surround the human participant" [Bowman et al., 2004]. They are generally stationary but since most of them are projection-based we chose to place them in the "projection-based" category in Figure 2.1.

The most common surround-screen displays are the ones made of several wall-sized screens. Projectors are used to display the virtual content on the screens. The images can be front-projected or rear-projected to prevent the real shadow of the user to be cast on the screens. One of the most famous surround-screen displays was developed by Cruz-Neira et al. [1993] and called the CAVE. The original CAVE had four screens (three walls and the floor) (see Figure 2.9). Four projectors were used to project immersive content over the screens. The users generally wear stereoscopic glasses to perceive a stereoscopic environment. Most of the current systems use active-shutter glasses for better rendering but polarized glasses or polarized lenses can also be used. With active shutter glasses the projectors have to display alternatively the right and left image and thus are required to have a high refresh rate.

Extensions to the original CAVE have been made by adding one or two screens (back screen and/or ceiling screen), allowing the user to be completely surrounded by the virtual environment. Indeed Febretti et al. [2013] designed the CAVE2 by using several LCD screens that can be disposed in different circular dispositions. Other variations in the structure have been designed such as the Computer-driven Upper Body Environment (CUBE) from the Entertainment Technology Center. The CUBE is a 360° FOR display composed of 4 screens that are suspended from the ceiling and turn around the user. The screens go from the head to the waist of the user.

A different approach for surround-screen displays consists in using spherical environments. Some examples of spherical environments are the Allosphere from Amatriain et al. [2009] of the University of California, the TORE display that has been developed by Antycip Simulation[1] (see Figure 2.9) or the Cybersphere from Fernandes et al. [2003]. The Cybersphere is made of a large, hollow, translucent sphere and the images are rear-projected on 5 segments of the outer sphere. It has been made to allow the user to navigate more naturally in the environment. Indeed, like in a hamster ball, when the user walks the sphere rotates and its rotation is measured in order to render the images accordingly. Spherical displays present a main limitation in terms of image quality due to the image distortion corrections that need to be applied in some cases.

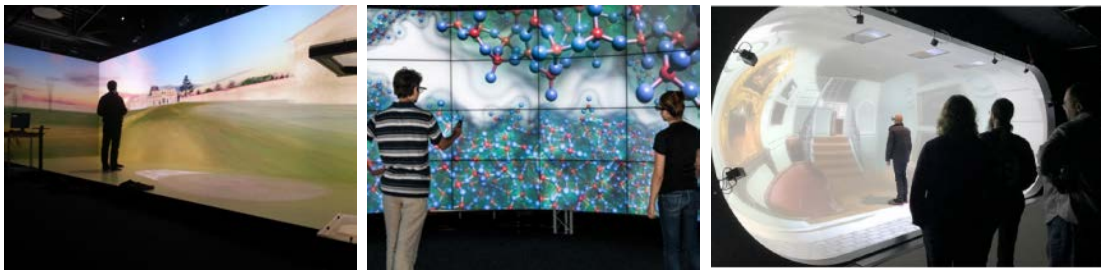


Figure 2.9 – Examples of surround-screen displays: the Immersia CAVE-like display from Inria Rennes-Bretagne Atlantique and IRISA (left), the Cybersphere [Fernandes et al., 2003] (middle) and the TORE Simulator from Antycip Simulation (right).

Spatial augmented reality

Projection-based AR systems are commonly associated with Spatial Augmented Reality (SAR) systems. Nevertheless SAR designate a more generic concept. In fact according to Raskar et al. [1998], with SAR "the users physical environment is augmented with images that are integrated directly in the users environment, not simply in their visual field" which do not necessary involve using projectors. Even so, SAR environments are generally achieved by using projection-based systems (see Figure 2.10 for examples of SAR) and we chose to merge both terms.

Spatial Augmented Reality (SAR) systems are generally used to project textures and augment stationary objects [Aliaga et al., 2012; Siegl et al., 2015]. Projecting on stationary objects with a stationary system gives good performances once everything is correctly calibrated. Nevertheless the use cases of such systems can be limited and few mobility or direct interactions can be considered. Thus more dynamic systems were designed to augment movable [Zhou et al., 2016] or deformable [Punpongsanon et al., 2015] objects. Work from Hochreiter et al. [2016] introduces multi-touch detection for interacting on augmented stationary objects directly with the fingers. Benko et al. [2012] proposed the Miragetable: a dynamic spatial AR system with projection on a curved surface (see Figure 2.10-left). These systems widen the possibilities of interaction since the real environment and the user motions are taken into account. However, since the projection is made on a stationary screen (or object) the usable workspace is rather limited. To overcome such limitation several spatial AR system were designed to project

[1] Antycip Simulation TORE display - http://www.antycipsimulation.com - Accessed: 2018-06-06

Figure 2.10 – Examples of spatial augmented reality displays: The Mirage Table from Benko et al. [2012] (left), the Lumipen from Okumura et al. [2012] (middle) and the Pmomo approach from Zhou et al. [2016] (right).

on movable 3D objects. The Lumipen, designed by Okumura et al. [2012], provides projection mapping for high-speed or high-frequency objects thanks to an high-speed vision sensor and a projector with an high-speed optical gaze controller. The Lumipen works well on simple 3D objects such as spheres and balls due to the insignificance of their rotation (Figure 2.10-middle). In more recent work, Sueishi et al. [2015] proposed an improvement of the Lumipen. Nevertheless their system is far too cumbersome and is still used on simple geometries. Such limitations do not provide an ideal environment for tangible interaction. Zhou et al. [2016] proposed the Pmomo: a projection mapping system on movable objects (Figure 2.10-right). The Pmomo handles more complex geometries with acceptable tracking performances. Even though the system is lighter than the previous approaches it is still stationary and is not designed to be portable or embedded. Moreover the current version of the system does not enable tracking several objects which can be inconvenient in many interaction scenarios. To compensate the limitations of a stationary projector, Benko et al. [2015] propose to combine it with Optical See-Through (OST) AR and provide more freedom to the user with a larger field of view induced by the projection. Nevertheless this approach is interesting whenever the user is in the workspace of the projector. Indeed outside of this workspace the field of view becomes limited again by the OST system. To overcome stationary displays a french company, Diota[2], proposes a SAR device that is able to move without being held in the hand or mounted on the head. This solution is based on robotic arms that move the projectors around the objects. Nevertheless such solution is not designed to be portable or to be used in small indoor environments.

Handheld projectors

A first approach to overcome stationary Projection-Based Systems (PBS) is to design handheld projectors. With handheld devices the projector needs to have knowledge of the geometry of the scene since it needs to be aware of its localization at each instant. Work from Raskar et al. [2006] introduces the iLamps (see Figure 2.11-left), geometrically aware projector. The approach is illustrated with a handheld projector and single-user applications. Handheld projectors have been studied in several posterior works. In 2007, Cao et al. [2007] introduced multi-user interactions with two projectors

[2]Diota Augmented Reality for Industry - http://www.diota.com - Accessed: 2017-09-09

that are tracked with feature-based tracking (see Figure 2.11-right). The users can interact by moving the projectors in the workspace with a visual feedback projected on a flat wall. Still, the interactions are limited to planar objects and no 3D objects are considered. Ni et al. [2011] introduced handheld projection in medical applications to improve doctor-patient communications. With such system the doctor is able to project anatomical information directly over the patient body. Nevertheless the authors pointed out that the proposed system was more usable when projecting on a wall. More recent work has been proposed based on same approach with the SideBySide system [Willis et al., 2011]. The SideBySide system tracks several projectors that project fiducial markers on a wall but the system is not adapted to tracking 3D tangible objects. Even though handheld SAR devices provide more mobility than stationary SAR systems they are not always adapted to direct interactions since the user's hands are not free.

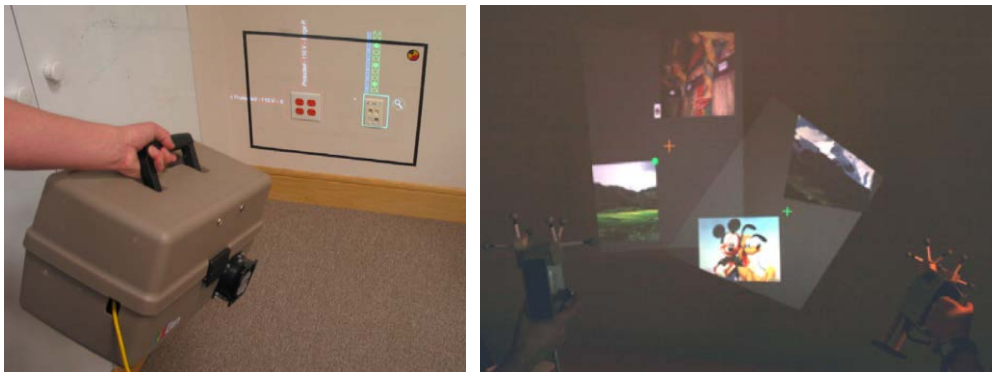


Figure 2.11 – Examples of handheld projector displays: The iLamps, one of the first handheld projector displays [Raskar et al., 2006] (left) and an application of multi-user interaction with two handheld projectors [Cao et al., 2007] (right).

Projection-based HMD

Since holding the projector in the hand is not always satisfying, work has been done to project from the head or the shoulder. Nevertheless mounting a projector on the head (or shoulder) can be more complicated due to the weight it induces. One of the first work going in that direction has been carried out by Karitsuka and Sato [2003]. They propose a shoulder-mounted projection system to augment a planar target. Then the users are able to interact with the augmented planar target by using their fingers. Bolas and Krum [2010] introduced head-mounted projection on reflective surfaces. Nevertheless they do not introduce interaction techniques for augmented reality and they only project informative content that cannot be modified. CastAR[3], a start-up company, implemented an industrial product based on head-mounted spatial augmented reality (see Figure 2.12). Their system projects 3D images over reflective surfaces that can have different predefined simple shapes and enables the users to interact with the environment. The prototype proposed by CastAR gets close to a virtual reality projective holobench system and they do not propose any augmentation of tangible 3D objects.

[3]CastAR Augmented Reality Glasses - http://en.wikipedia.org/wiki/CastAR - Accessed: 2017-04-11

Unfortunately, CastAR closed their doors in 2017 due to a lack of interest for this technology in the industry they were targeting. Work from Akşit et al. [2014] also proposes an approach to project on planar surfaces from an head-worn mixed reality system based on a laser pico-projector and a smartphone. Unlike CastAR the authors chose to focus on motion capture applications. Thus, the system has been prototyped to work in a larger and non-friendly infra-red environment. However the projection over 3D tangibles objects is still not considered and no tracking system is required other than the smartphone sensors. More recent work from Harrison et al. [2011] introduces a shoulder-mounted system implementing direct hand interaction techniques. Indeed mounting the projector on the shoulder also leaves the hands free to interact. They proposed a tactile interaction on simple surfaces and body parts. The projection over these surfaces is still planar and the geometry of tangible objects is not taken into account.

Figure 2.12 – Examples of head-mounted projector displays: head-mounted projector for planar reflective surfaces from Bolas and Krum [2010] (left) and the CastAR prototype of head-mounted projection glasses (right).

Benefits and drawbacks of projection-based displays

One of the main benefits of Surround-Screen Displays (SSD) is that, if needed, the user can naturally walk within the display. Also, since the user is not cut off the real world, real objects can be mixed with the virtual environment. However there are depth cues problems when mixing real and virtual object because the user is unable to put a real object behind a virtual one that is projected on the screens. SSD provide a large FoV, which is generally almost equal to the human FoV and a large enough FoR, even though it does not reach 360°. Moreover the users have a world landmark that can reduce the effects of motion sickness, e.g., by looking at their own body.

All Projection-Based Systems (PBS) (including both Virtual Reality (VR) SSD and Augmented Reality (AR) PBS) generally present the possibility for external persons to partially share the experience with the main user. Indeed, since the images are displayed in the environment (relatively far from the user's eye) external persons have a similar visual feedback as the one provided to the main user, even though this feedback generally lacks of stereoscopic effect. Nevertheless one main disadvantage of PBS is that these systems generally require large physical space to be built and are often expensive. Such bulk can prevent the usage of these system for mobility use cases. Handheld

projectors overcome these issue by providing mobility. But holding the display in the hand can limit the range of interactions and therefore the range of applications. Finally head-mounted PBS relieve the user from holding the device in the hand. Still, head-mounted projection-based displays present a main challenge in reducing the weight and optimizing the ergonomics of the device that is mounted (or worn) on the head.

To sum up, compared to other displays, Projection-Based Systems (PBS) generally provide a greater spatial resolution, a faster refresh rate and a wider FoV. However they are, in some cases, less efficient in terms of FoR and ergonomics. But they, sometimes, enable sharing the experience with external persons.

Conclusion

Mixed Reality (MR) systems are composed, among others, of a visual display that provides visual feedback to the user about the current state of the MR applications. Visual displays are also generally responsible of the user's immersion and contribute to increase the user experience and comfort. In this section we have discussed several MR visual displays that can be classified according to the distance between the screens and the user's eye as follows: near-eye displays, handheld displays, sattionary displays and projection-based displays. Even though several visual display technologies exist, each of them has its own benefits and drawbacks. According to the use cases and applications one or another visual display can be better adapted and provide an optimal experience.

Even though visual displays represent a main and indispensable component of a MR system, they are closely related to the tracking systems. Both systems (the displays and the tracking) provide a complete and usable MR system. In the following we propose an overview of research work that relate to the different tracking technologies and we make a close-up on the optical tracking technologies that are commonly used for VR and AR applications.

2.2 Tracking systems for mixed reality

Based on the definition of Ribo et al. [2001],

> "In Virtual Reality (VR) and Augmented Reality (AR), tracking denotes the process of tracing the scene coordinates of moving objects in real-time."

In their survey Welch and Foxlin [2002] presents several purposes of tracking systems:

- *View-control* : Tracking systems allow to control the position and orientation of the viewpoint to make the rendering coherent to a first person point of view.
- *Interactions* : Tracking systems help the user navigating in the Virtual Environment (VE) and selecting and manipulating virtual objects.
- *Instrument tracking* : Instruments can be tracked so that their virtual representation matches their real position and orientation (co-localization).
- *Avatar animation* : Perhaps one of the most common use of tracking systems ha been the animation of virtual characters through full body motion capture.

Nevertheless Welch and Foxlin [2002] also affirm that "there is no silver bullet", meaning that none of the tracking systems presented below fulfills all the tracking requirements, namely : tiny, self-contained, complete, accurate, fast, immune to occlusion, robust, tenacious, wireless and cheap. Several tracking technologies are briefly introduced in the next section. Optical tracking, which is the most common tracking technology used in MR applications, will be discussed largely in section 2.2.2.

2.2.1 Non-optical tracking systems

In the early 1990s, mechanical, acoustic and magnetic tracking technologies were proposed. Later, inertial tracking was introduced in mobile devices and is nowadays used combined with other technologies. In this section we present an overview of the mentioned tracking technologies.

Mechanical tracking devices

Mechanical systems track the end-effector of an articulated arm composed of several limbs. The joints between the different limbs can have up to 3 Degrees of Freedom (DoF) in rotation which is measured from rotary encoders or potentiometers. Then knowing the length of every limb of the articulated arm a kinematic chain can be used to determine the position and orientation of the end-effector. Mechanical systems are commonly used for haptic and robotics purposes like the commercialized Novint Falcon characterized by Martin and Hillier [2009]. Mechanical tracking systems can also be built as exoskeletons that enable tracking the users limb. As an example the Gypsy exoskeleton (Figure 2.13-left) enables tracking the upper body of a user. Similar techniques are used for hand tracking using mechanical hand exoskeletons like the HEXOTRAC [Sarakoglou et al., 2016] (Figure 2.13-right) that can also be used as haptic devices.

Figure 2.13 – Examples of mechanical tracking systems: Gypsy's upper body exoskeleton (left) and the HEXOTRAC hand mechanical exoskeleton [Sarakoglou et al., 2016] (right).

Mechanical trackers provide good precision and fast update rate. Nevertheless they require a cumbersome calibration process to determine the extend of every limb and

the angles of every joint. Moreover such systems generally limit the workspace of interaction to the range of action of the articulated arm. Thus they are commonly used in very restricted workspace. On top of that these systems generally require to have a large room to be installed and they can be quite obtrusive.

Acoustic tracking devices

Acoustic systems use sound waves to determine the position and orientation (pose) of an object. All known commercial acoustic ranging systems operate by timing the flight duration of a brief ultrasonic pulse but one of the first implemented method was the phase-coherent method. Sutherland [1968] built a continuous carrier-phase acoustic tracking system to supplement his mechanical system. This system used a continuous-wave source and determined range by measuring the phase shift between the transmitted signal and the signal detected at a microphone. One of the main limitations of the phase-coherent method is that it can only measure relative distance changes within a cycle [Meyer et al., 1992]. To measure absolute distance, one needs to know the starting distance and then keep track of the number of accumulated cycles.

The acoustic time-of-flight method operates by timing the flight duration of a brief ultrasonic pulse. Commercial devices such as the hybrid acousto-inertial system Intersense IS-900 [Foxlin et al., 1998] use this method. Pulsed time-of-flight acoustic systems can overcome most multipath reflection problems by waiting for the first pulse to arrive. The first pulse will arrive via the direct path unless the signal is blocked. This method works better for acoustic systems than for radio frequency or optical systems because the sound travels relatively slowly, allowing a significant time difference between the arrival of the direct path pulse and the first reflection.

According to Welch and Foxlin [2002] the speed of sound changes about 0.1 percent per degree Fahrenheit of temperature differential. This corresponds to about a one millimeter error per degree Fahrenheit at one meter. Thus the accuracy of the acoustic systems is affected by environmental conditions such as humidity, wind and temperature. Furthermore, their update rate is affected by reverberations (echoes). It may be necessary, depending on the system, to wait for up to 100ms (10Hz update rate) to allow echoes from the previous measurement to die before initiating a new one. Acoustic systems require a line of sight between the emitters and the receivers, but they are somewhat more tolerant to occlusions than optical trackers (which we discuss later) because sound can find its way through and around obstacles more easily.

Magnetic tracking devices

Magnetic tracking systems measure the local magnetic field vector and its absolute value. The sensors measure a quasi-static direct current fields or a changing magnetic field produced by an active source. The magnetic field vector indicate the orientation of the object relatively to the excitation. To measure the orientation and the position, three orthogonal triaxial coils are used at both the transmitter and receiver [Bishop et al., 2001]. A method was proposed by Paperno et al. [2001] to increase the speed of magnetic tracking and to simplify the computation algorithm. This method is based on a magnetic field that rotates continuously by using a pair of excitation coils that are in space and phase quadrature.

Well-known devices such as the Polhemus magnetic tracker [Raab et al., 1979] or the Sixense Razer Hydra use magnetic technology to track position and orientation of an object. Some research and commercial products are illustrated in Figure 2.14.

Figure 2.14 – Examples of magnetic tracking systems: the Sixense Razer Hydra (left), the Polhemus G4 wireless tracker (middle) and the Ascension flock of birds magnetic tracking system (right).

According to Welch and Foxlin [2002] and Bishop et al. [2001] one of the main limitations of magnetic systems is that ferromagnetic and conductive material in the environment can affect the magnetic field's shape. With alternative sensors, such as current sensors, eddy currents can be be induced in conducting materials by the magnetic field. These currents produce a magnetic field around the material which creates interference. To overcome these problems, every conducting material should be removed from the working space and thus, have structures made of wood or plastic for example. Another disadvantage of this magnetic systems is that the performance of the tracking decreases with the distance from the emitter, limiting the operating range between 1 and 3 meters.

However magnetic tracking systems do not require a line of sight since magnetic waves can go through objects. Also, they can track several users and they are cheap wireless devices.

Inertial tracking devices

Inertial tracking technologies were often used around 1950, embedded in ships, airplanes or submarines. However the technologies were too heavy to be attached to a person's body. Around 1990, with the advent of MicroElectroMechanical Systems (MEMS), inertial technologies for body tracking began to be used.

Inertial Measurement Units (IMU) are composed of accelerometers, gyroscopes and often a magnetometer. Nowadays the IMU are built on MEMS and are often composed of three accelerometers to compute the position according to the X, Y and Z axes and of three gyroscope to compute the orientation around the three axes (roll, pitch and yaw). The accelerometers are used to measure the acceleration of an object in order to compute its position using the Newton's second law of motion $\mathbf{F} = m\mathbf{a}$. The gyroscopes are used to measure the Coriolis force of a vibrating object and to obtain the rotational velocity of the object. Acceleration and rotational velocity are integrated to obtain the position and the orientation of the tracked objects. Figure 2.15 illustrates the process of computing the position and orientation of an object from the IMU measurement.

As referenced by Welch and Foxlin [2002], inertial systems are the closest thing to the "silver-bullet". Indeed inertial tracking systems are self-contained and thus do not require the use of emitters and do not need to have a line of sight. Moreover there are

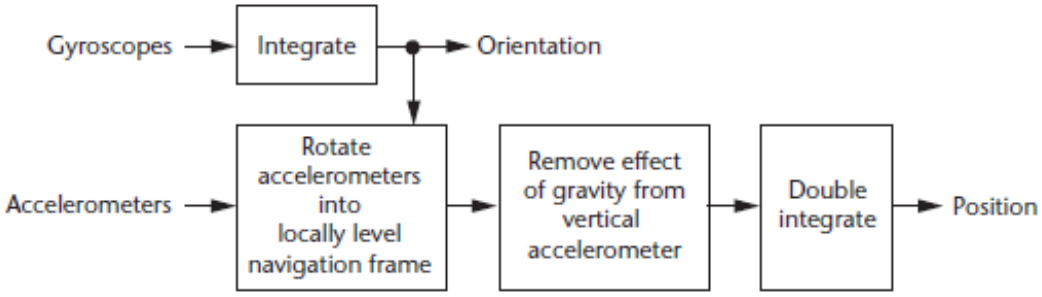


Figure 2.15 – Position and orientation integration in Inertial Measurement Units [Welch and Foxlin, 2002]

no wires and no physical limit on the tracked volume. They have very low latency and low jitter.

Intertial tracking systems present a limitation due to the drift they generally involve. This drift is caused by the error measurements that are propagated through the double integration process that is required to compute the position from acceleration measurements. Indeed, if one of the accelerometers has a bias error of just 1 milli-g, the reported position output would diverge from the true position with an acceleration of 0.0098 $m.s^{-2}$. After barely 30 seconds, the estimates would have drifted by 4.5 meters [Welch and Foxlin, 2002].

However the advantages of inertial technologies make them a good candidate for hybrid tracking systems. By combining inertial tracking with another technology the drift can be periodacally corrected [Bishop et al., 2001].

Hybrid tracking devices

Hybrid tracking devices combine different tracking technologies in one system. Inertial systems are good candidates to complement other technologies. Indeed they have many advantage but need to be corrected over time. Coupling them with another tracking technique can improve the overall tracker performance.

The Intersense IS-900 implements inertial tracking which is assisted by acoustic technologies as presented by Foxlin et al. [1998] (Figure 2.16-left).

Both You et al. [1999] and Foxlin et al. [2003] present inertial and optical tracking hybridization. These systems track slow movements with optical system and fast ones with the inertial unit. The fusion of the data from the two streams can be achieved with an Extended Kalman Filter (EKF) [Julier and Uhlmann, 2004]. Similar work has been carried out by Jiang et al. [2004] to provide tracking to outdoor Augmented Reality (AR) applications. The system (depicted in Figure 2.16-middle) is based on gyroscopes and line-based optical tracking that corrects gyroscopes' drift.

Inertial and magnetic tracking has been implemented for human motion tracking by Zhu and Zhou [2004]. Their system is built with sensors that are composed of tri-axis microelectromechanical accelerometers, rate gyroscopes, and magnetometers. Moreover a Kalman-based fusion algorithm is applied to obtain dynamic orientations and further positions of segments of the user's body. Also, HRL Laboratories developed an inertial and magnetic tracking system that is, in addition, provided with GPS data

Figure 2.16 – Examples of hybrid tracking devices: the IS-900 acoustic and inertial tracking system from Intersense (left), an inertial and optical tracking system from Jiang et al. [2004] (middle) and the HRL Laboratories tracking device with GPS, inertial and magnetic sensors (right).

(Figure 2.16-right).

2.2.2 Optical tracking systems

Optical tracking is based on the sensing of light waves to compute the position and orientation of visual features. Optical tracking systems are composed of one or multiple cameras and use computer vision algorithms. Optical tracking technologies are well known because of their usage in the animation and film making industry through motion capture [Moeslund and Granum, 2001]. Moreover optical tracking has also been widely used in different fields of applications such as robotics [Mustafah et al., 2012], medical applications [Taylor et al., 1994] or video surveillance [Cohen and Medioni, 1999]. Nowadays optical tracking is also commonly used in Virtual Reality (VR) and Augmented Reality (AR) [Pintaric and Kaufmann, 2007]. It enables to compute the pose (position and orientation) of the cameras, the 3D objects or the users.

Being the more common and promising tacking technique we go deeper into optical tracking technologies. In the following we present different components of these systems and the main workflow of optical tracking based on rigid body features.

2.2.2.1 Spatial arrangements

When using multiple optical sensors, one must consider whether to put the light sources on the moving target and the sensors in the environment, or vice versa. Figure 2.17 illustrates both configurations, that are called Inside-out and Outside-in.

Inside-Looking-Out

The Inside-Out spatial arrangement means that the camera is attached to the tracked object (the head of the user for example). The position and orientation are then given by the localization of static markers placed on the floor or on the ceiling like for the HiBall device proposed by Welch et al. [2001]. Figure 2.17 depicts an Inside-Out configuration compared to an Outside-In one. A more recent study from Hutson and Reiners [2011] also proposed an inside-out tracking with detection of fiducial markers

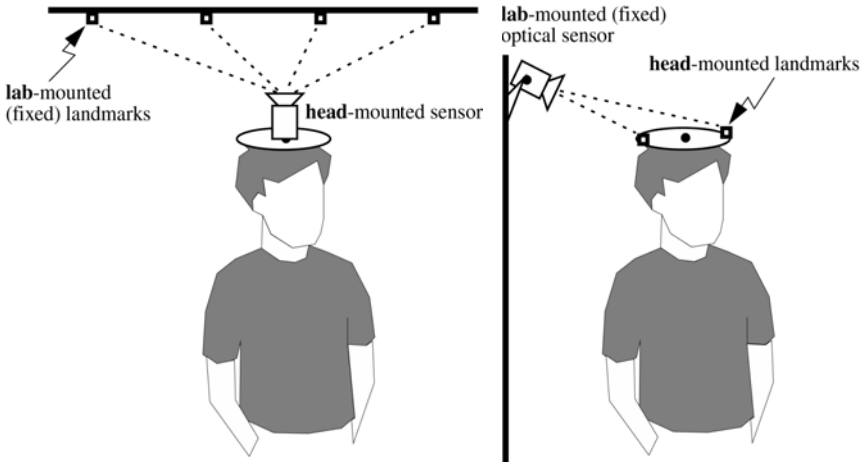


Figure 2.17 – Inside-Out (left) and Outside-In (right) optical tracking spatial arrangements [Welch et al., 2001]

in a CAVE environment. The fiducial markers are displayed on the screen behind the user and the camera (attached to the user's head) is facing backwards.

Outside-Looking-In

The Outside-In spatial arrangement consists in having static cameras positioned in the environment looking to markers that are positioned on the tracked objects. Most of the industrial actors, such as Vicon, OptiTrack or ARTracking use this configuration. Research work is also focused on outside-in configurations with devices like the iotracker [Pintaric and Kaufmann, 2007] because of the ergonomics of outside-in compared to the inside-out. Indeed it can be inconvenient and cumbersome to have a camera fixed on the user's hand to track its movements.

2.2.2.2 Visual features

Optical tracking technologies need to detect some inhterst or key points in the images. These interest points define the objects that are tracked. When using optical tracking technologies for VR application it is generally require to add visual markers over the tracked objects. Even though natural features can be used for optical tracking, they are generally not adapted to track independent objects which shape is unknown. However natural features are used to compute the position of the camera in the real world (the real scene). Thus natural features are not generally used in VR applications but are common in AR contexts [Martin et al., 2014] even though using markers can also be a good solution for many AR applications [Wagner et al., 2008]. Since almost every VR application requires to compute the position of at least one unknown object (e.g., the user's head) then markers are generally added over this unknown object.

In the following we present both natural-feature-based and marker-based tracking techniques.

Natural features

Natural features can be detected in the images without adding any marker. They are directly extracted from the video stream of the camera. Thus the captured images need to have a good resolution and quality. Within the images, interest or key points are detected. These points generally require to be salient in the images and to be static according to the objects of the scene. Several techniques enable extracting key points such as Harris detector [Harris and Stephens, 1988], SUSAN [Smith and Brady, 1997], SIFT [Lowe, 2004] or FAST [Rosten et al., 2010] (see Figure 2.18). Even though the extraction of natural features works well on dense and irregular textures it generally does not provide good results on homogeneous images. In such cases, edges features can be used as natural features to enable target detection.

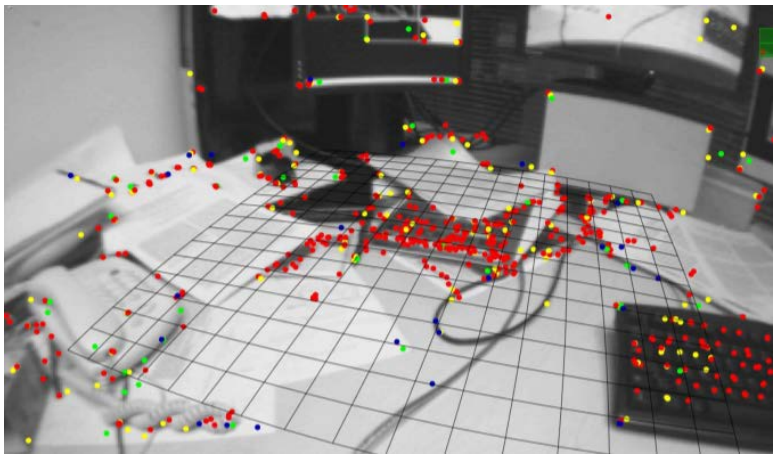


Figure 2.18 – Natural features used as interest points for the Parallel Tracking and Mapping (PTAM) approach [Klein and Murray, 2007].

Marker features

Optical tracking based on markers detection has ben studied in several previous work such as the one from Pintaric and Kaufmann [2007], Mathieu [2005], Welch et al. [2001] or Ribo et al. [2001].

Marker-based optical tracking computes the position and orientation of several predefined markers. The markers can either be active or passive markers. As opposed to passive markers, active markers generate an illumunation that can be captured by the optical sensors (e.g., cameras). Regarding the sensors, they can either work on the visible or on the infrared light bandwidth. In the case of infrared light the sensors are filtered so that they are only able to capture infrared light [Pintaric and Kaufmann, 2007]. Therefore active markers emit infrared light and passive markers are lighten with an infrared lighting rig [Ribo et al., 2001]. If the system works on visible light, active markers can be colored such as in the PSMove device from Sony. Common passive markers that work on visible light are the fiducial markers that can either be squared [Hutson and Reiners, 2011] or circular [Foxlin et al., 2003]. Squared fiducials enable to straight forwardly compute the pose of the object while circular fiducials have to be

used in group with a know disposition. Figure 2.19 shows examples of markers that can be used with optical tracking devices.

Figure 2.19 – Examples of markers used in optical tracking devices: passive markers used by Pintaric and Kaufmann [2007] (left), active colored markers used in the PSMove device from Sony (middle) and a squared fiducial marker [Hutson and Reiners, 2011] (right).

2.2.2.3 Rigid body optical tracking

Marker-based optical tracking technologies are well adapted to Virtual Reality (VR) purposes. These techniques provide accurate, robust and fast pose estimation for several targets. Among the marker-based techniques, using constellations is one of the most common techniques used for VR [Mathieu, 2005; Pintaric and Kaufmann, 2007; Ribo et al., 2001]. A ***constellation*** is defined as a set of markers that are rigidly attached together and that form a rigid structure that its call a ***rigid body***. Tracking rigid bodies can also be used in AR applications [Dorfmüller, 1999; Wang et al., 2008] even though it requires to add several markers over the real objects; markers that can be visible through the AR system.

Three steps are needed to perform accurate rigid body optical tracking (see Figure 2.20): (1) rigid body construction, (2) system calibration and (3) tracking.

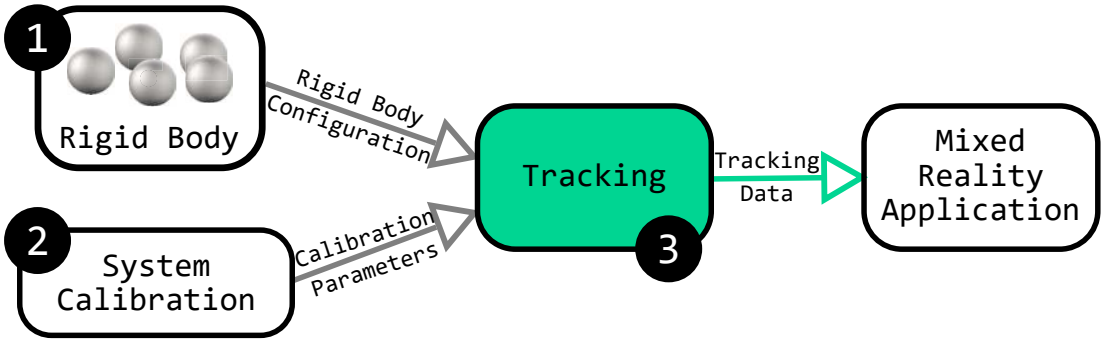


Figure 2.20 – Overall rigid body tracking workflow

(1) Rigid body construction

An essential step of using rigid bodies to track objects is to define and build a convenient rigid body that will be attached to the tracked object [Steinicke et al., 2007]. This process is made by the user beforehand. When using rigid constellations, the structure

of the different constellations has to be define in order to distinguish and identify the different tracked objects. Figure 2.21 shows the design of a contellation made by Pintaric and Kaufmann [2007]. This design allows to identify each constellation thanks to the distance between each marker. Every distance has to be unique and the different markers should be non co-planar. This non-coplanarity cannot be achieved with only 3 markers and it is recommended to use at least 4 markers to avoid having many solutions when estimating the pose of the objects. Indeed using only 3 markers provides 4 solutions to the pose estimation problem and ambiguities need to be handled [Hartley and Zisserman, 2003].

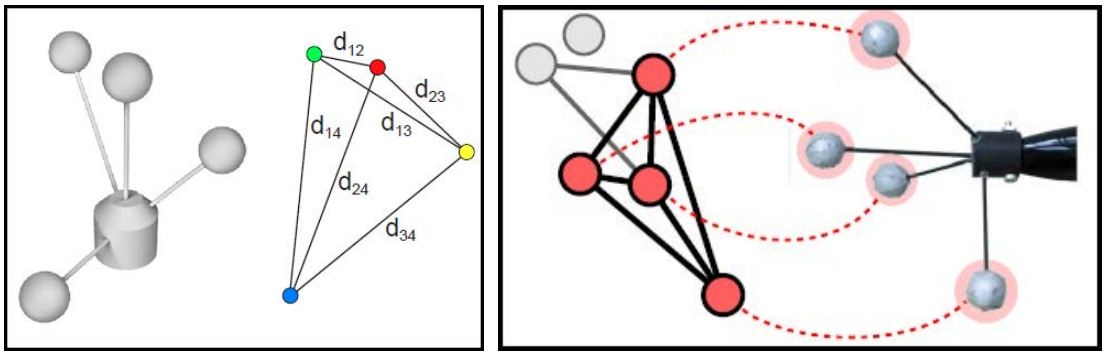


Figure 2.21 – Constellation structure (left) and Constellation 3D cloud (right) defined by Pintaric and Kaufmann [2007]

(2) Calibration

To fully calibrate an optical system, the intrinsic parameters of each camera and the extrinsic parameters of the system have to be computed. The intrinsic parameters define the geometry of the camera (e.g., focal length, distortion coefficients) and the extrinsic parameters define the respective transformation between the different cameras of the system.

The intrinsic parameters can be computed by capturing several views of a known grid (see Figure 2.22). Once these parameters are computed the external parameters can be found by collecting, from the different cameras, several points in the workspace. These points can be collected by using passive or active markers.

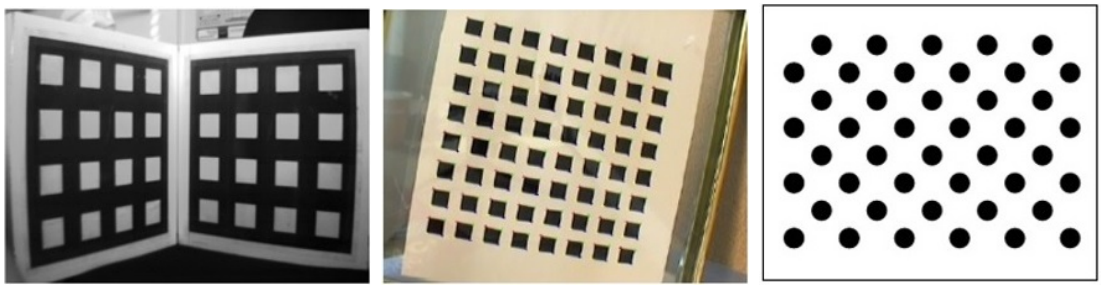


Figure 2.22 – Patterns used for internal camera calibration: a 3D calibration pattern [Faugeras and Toscani, 1986] (left), a planar chessboard pattern [Zhang, 2000](middle) and a circular calibration grid (right).

The different algorithms and the calibration processes used for either intrinsic and extrinsic calibration are detailed in section 4.3.1.

(3) Tracking

Once the calibration is done and the structure defined, the tracking can be performed. Within the tracking process several operations are consecutively executed as follow:

- *Feature detection*: Detecting the markers in the camera frames is the first step of optical tracking. The goal of this step is to obtain the image coordinates of the projection of the 3D markers in the different cameras frame.
- *Feature correction*: The correction step undistorts the projected points according to the intrinsic parameters of the cameras.
- *Feature correlation*: The correlation matches the projected points from one image with the projected points in another image. The matched points are the projection of the same 3D point in the different images.
- *Triangulation*: Once the correlation is done, the triangulation recovers the 3D points from their projection in several cameras. A 3D point cloud of reconstructed points is obtained.
- *Model matching*: This step associates the 3D points from the 3D point cloud to one or another rigid body. In their work, Kurihara et al. [2002] proposed a polyhedra search algorithm for real-time matching. Later work from Pintaric and Kaufmann [2007] introduced a step by step process for designing rigid bodies that aimed at optimizing the matching process.
- *Registration*: When the rigid bodies have been identified in the 3D point cloud, the registration consist in estimating the pose of the tracked objects (see section 4.3.2).

Once every step has been carried out, the pose of every tracked object is accessible and can be sent to any MR application.

2.2.2.4 Benefits and drawbacks of optical tracking systems

The main limitation of optical tracking systems is that they are sensitive to occlusion and require a clear line of sight to be used. If multiple cameras are used, the tracking is simplified by using stereoscopic techniques but it can go berserk if the features are only visible by one camera. This issue generally forces the users to stay into the overlapping Field-of-View (FoV) of at least 2 cameras which can reduce the working volume of the end-use application. Moreover, optical tracking systems require image processing and computer vision which can increase the computational time. This computational time can be particularly challenging when using natural features since the extraction of these features generally consumes a lot of resources.

However optical techniques are barely sensitive to environment conditions (although lighting conditions can affect the tracking performance). Moreover, in some cases, they

only require the placement of markers on the user or on objects which is a light and wireless technique. Marker-based tracking makes it possible to track accurately several targets and proposes robust and fast registration which is a main requirement for many real-time applications.

2.2.3 Discussion on optical tracking systems for projection-based displays

One the most common tracking technologies used in Virtual Reality (VR) Projection-Based Systems (PBS) is the optical one. As defined by Pintaric and Kaufmann [2007] this technology uses multiple cameras and active or passive markers that form a rigid body or constellation (as presented in the previous section). Industrial implementation of optical tracking systems have been made by VICON, Optitrack or ART Tracking. Even thought these systems present many advantages like their accuracy, speed or lightness (they only require wearing light markers) they still present a limitation in terms of working volume. Indeed these systems generally require to have a stereo configuration to be able to localize an object. Such configuration involves that the tracked targets need to be visible by at least two cameras which reduces the working volume to the overlapping of the cameras.

As discussed in section 2.2.2.1 inside-out tracking configuration can be a alternative to standard outside-in tracking. Inside-out tracking has been introduced in surround-screen displays with the JanusVF from Hutson and Reiners [2011]. The JanusVF tracking uses fiducial markers that are directly displayed on the screen that is behind the user (see Figure 2.23). Then the fiducial markers are captured with a camera that is worn on the user's head. However such tracking technique requires to continually have a screen behind the user to be able to display the markers. Moreover, since it is a inside-out tracking, the camera is worn on the user's head which generally reduces the comfort when using the VR application.

Figure 2.23 – Examples of tracking techniques used in surround screen displays environments: JanusVF Inside-out tracking [Hutson and Reiners, 2011] (left) and an outside-in infrared optical tracking from Realyz (right).

Regarding Augmented Reality (AR) Projection-Based Systems (PBS) several technologies have been used to track real objects that are augmented. As a first example,

Zhou et al. [2016] combined inertial tracking with optical tracking to localize 3D objects on which virtual information is projected. The optical tracking is used to correct the inertial measurements by localizing the 3D model of the object in the images.

A summary of constraints and requirements of the tracking system according to the PBS it addresses is proposed in Table 2.1 and explained hereafter.

Table 2.1 – Requirements of the optical tracking systems according to the projection-based displays. (●: Acceptable, ● ●: High, ● ● ●: Very High, *Small*: Up to $6m^3$, *Large*: Up to $100m^3$)

Requirement	Surround-screen	SAR	Handheld projector	HMPD
Working volume	*Large*	*Small*	*Large*	*Large*
Accuracy	● ●	● ● ●	● ● ●	● ● ●
Robustness	● ● ●	● ●	● ●	● ●
Min. Speed (Hz)	120	60	60	120
DoF	6	6	6	6

Working volume

Many SAR systems are static which involves that their workspace can be rather limited. Indeed the tracking systems only requires to track around the projection area. When moving the projector (e.g., handheld projector or head-mounted projection displays) the tracking system should cover a larger workspace. Regarding Virtual Reality (VR) Surround-Screen Displays (SSD), depending on the display, the working volume can be very large and, since the user should be able to move around this volume, the tracking system should cover a large workspace.

Accuracy

Compared to Virtual Reality HMD that do not enable seeing the real environment, the PBS require to have a higher accuracy. Indeed the users will be more sensitive to accuracy errors since they have a real reference. For example when projecting a texture over a 3D object, the tracking error will propagate to the location of the texture according to the object. In a similar way, when users are immersed in SSD they can use their hand as reference to evaluate if the object they manipulate is correctly co-localized with their hand. Nevertheless we believe that users will in general be less sensitive to accuracy errors when using VR content than AR content.

Robustness

In terms of robustness, when considering VR, every tracking drop-out has an influence on the rendering of the virtual environment. Since the users are generally fully immersed, an inconsistency in the rendering can generate nasty sensations that can lead to motion sickness. Also, if the tracking fails the interactions that are being carried out have no more meaning. AR systems present the same requirements regarding the interaction. However, when using AR the users are generally not fully immersed in a virtual environment and despite the tracking failing they will generally be less affected

in terms of feelings and sensations. In any case, having a tacking that is not robust leads to a discomfort and decreases the user experience and appreciation.

Speed

From the study of Wells and Venturino [1990], voluntary head movements may have accelerations up to thousand degrees per second squared. Thus Velger [1998] and Merhav and Velger [1991] have recommended to compute the head position and orientation at a rate of 120 to 240 Hz. When considering SSD or head-mounted projector the users head movements are taken into account and the speed of the tracking system should be of 120Hz, as recommended. However, since SAR and handheld projector do not always require to track the users head, we estimate that a slower tracking (60Hz) can be acceptable for these applications. It is noteworthy that the movement of the head is also related to the user's role-playing and to the interaction techniques used.

Degrees of freedom

Most of the Mixed Reality (MR) applications enable the users to move freely in the real environment. Therefore the tracking system should be able to provide 6 Degrees of Freedom (DoF) positioning data (3 for the position and 3 for the translation) whatever the MR system, even Projection-Based Systems (PBS).

Conclusion

Several techniques have been proposed to localize an object in a 3D space. Acoustic, magnetic, mechanical, inertial and optical tracking systems are present in the literature. The tracking systems need to fulfill many requirements and there is still no technology that can outperform the others. Depending on the final use case, the tracking technique can be chosen to better suit the application.

Regarding Mixed Reality (MR), the most common tracking technology is the optical one, that uses visual sensors (commonly cameras) to recover the position and orientation of visual markers. The high update rate and accuracy of such systems coupled with their ergonomics make them one of the most promising tracking technologies whenever there is a clear line of sight between the sensors and the target.

However, optical tracking techniques still present a limitations in terms of outdoor usage. In fact the usability of optical techniques can be affected by direct sunlight exposure. Moreover, to cover a large workspace optical tracking systems commonly use numerous sensors which requires to have enough room and can considerably increase the price of the overall tracking system. Optimizing the number of cameras and increasing the workspace represents a main challenge for optical tracking devices.

2.3 Industrial applications of mixed reality

Industrial pipelines and processes generally involve large expenditures and the consumption of many resources and energy. Moreover these processes become more complex with the uprising of versatile products that are adressed to mass consumption

[Ong et al., 2008]. According to the definiton from Chung [2003], "Simulation modeling and analysis is the process of creating and experimenting with a computerised mathematical model of a physical system". Simulating processes and operations can be of great value in several industrial domains since it enables designing, experimenting and validating products, operations and systems [Mourtzis et al., 2014].

The manufacturing industries (e.g., ***aerospace***, ***automotive***) can be a good example to illustrate both the need for simulation and the main challenges of product design and process optimization. Thus, this section focuses mainly on the manufacturing industry but several applications can be transposed to other industrial domains such as: ***military***, ***medical***, ***construction***, ***architecture*** or ***retail*** (Figure 2.24). According to Chryssolouris [2013] manufacturing is defined as "the transformation of materials and information into goods for the satisfaction of human needs". With the recent advances of information technology, digital manufacturing has been considered to reduce product development times and cost. Digital manufacturing also addresses the need for customization, increased quality and fast mass distribution [Chryssolouris et al., 2009].

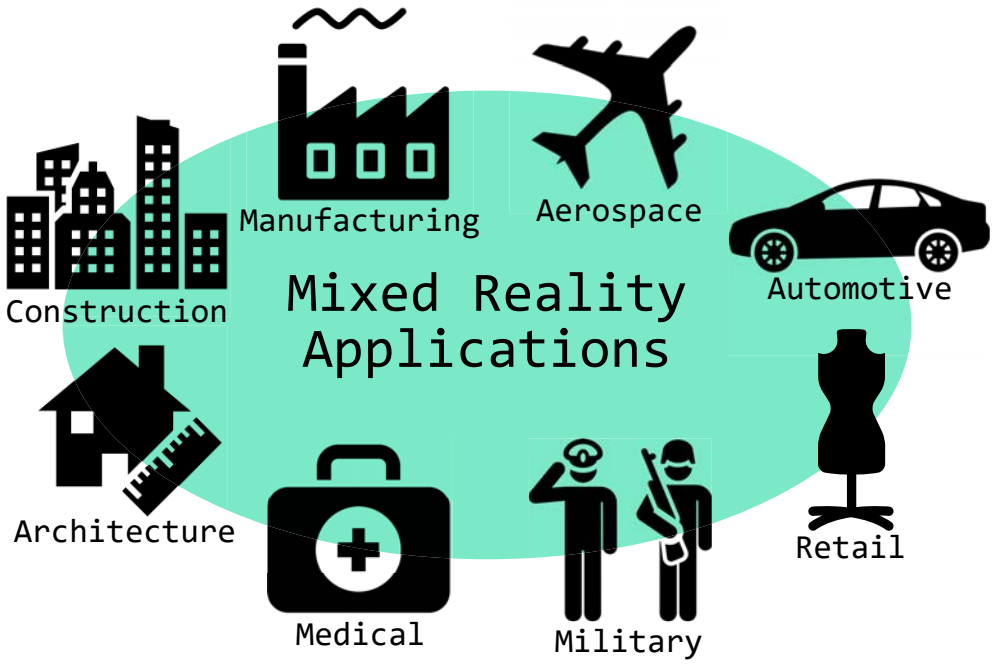


Figure 2.24 – Industrial fields use Mixed Reality applications: e.g., architecture, construction, manufacturing, aerospace, automotive, retail, military, medical.

Mixed Reality (MR) techniques and applications seem to be a good solution to help visualize, validate and assist the conception of manufacturing processes before they are carried out. According to the early study from Lu et al. [1999], in the automotive industry MR technologies could help reduce the production time from 2 years to 8 months. The work carried out at DaimlerChrysler [Baratoff and Regenbrecht, 2004] is commonly cited as the reference in terms of AR applications in the automotive industry that could benefit to the product realization processes [Nee et al., 2012]. The survey from Ong et al. [2008] presents several Mixed Reality solutions and demonstrations in both manufacturing and other fields such as medical, military, maintenance and

entertainment. A more specific survey from Mujber et al. [2004] gives en overview of the different virtual reality applications in manufacturing process. They also detail the added value of using VR and its benefits in manufacturing applications.

Inspired by Mourtzis et al. [2014] and Seth et al. [2011] we propose to classify the MR industrial applications in four different categories:

- training applications
- assistance applications
- design applications
- planning and validation applications

In the following we present previous work and systems that have been proposed for the different industrial applications.

2.3.1 Training applications

Training applications in Mixed Reality present several advantages compared to standard training session in real environments. According to Lourdeaux [2001] they enable:

- Executing task without risks.
- Making mistakes without having an impact on security.
- Reconfiguring the environment (terrain, meteorology).
- Modeling inaccessible training field (space, enemy field, frequented railways).
- Simulating scenarios that cannot be simulated in reality (technical incidents).
- Being free of time constraints and other necessities.
- Using a limited space (the MR system volume).
- Using the same system for different training applications.

An alternative to MR training could be the 2D training. Nevertheless an early study from Boud et al. [1999] proved that training operators with VR or AR systems improves their performances. Indeed, when performing assembly task, the task completion time was smaller for users that were trained using MR technologies compared to the ones trained with 2D training.

Due to its previously mentioned potential, MR training has been adopted in several industrial field either for maintenance or operation training. In the automotive industry PSA Citröen introduced a projection-based VR driving simulator (see Figure 2.25-left). Other simulators have been developed for driving purposes such as the Simu Cabine PL truck driving simulator from Thales (Figure 2.25-middle). Even though VR is commonly more used for simulating driving scenarios, work from Regenbrecht et al. [2005] proposed an AR-based driver safety training application. The applications is based on two video-see through HMD that are used in a real car by both the trainer

and the driver. Results showed that the driver reacted in a very similar way compared to real scenes. Simulators have also been introduced in other industrial domains such as aerospace and military. Regarding aerospace, simulators are generally flight simulators. For space flights Osterlund and Lawrence [2012] developed a training simulator to estimate ergonomic risks and evaluate spacecraft flight systems. In the military industry simulators are used to train infantry, navy, or air-force. VirTra proposed a projection-based VR application for police training (Figure 2.25-right). The officers are immersed in real life conditions and the instructors are able to modify the environment and the avatars behavior depending on the situation.

Figure 2.25 – Mixed Reality training applications for the industry: the PSA Citröen projection-based VR driving simulator (left), the Thales truck driving simulator (middle) and the VirTra simulator for police officer training in real life conditions (right).

Apart from real conditions simulators, MR training applications can also be used to train operators to maintenance and/or operation tasks. Li et al. [2003] designed a low-cost VR desktop system for maintenance training and illustrated its usage with a training on the maintenance of a centrifugal pump system. The results showed the potential of VR training for reducing the cost of maintenance tasks. Several VR training solutions have been proposed for medical [Albani and Lee, 2007; Basdogan et al., 2004; Gosselin et al., 2010], military [Shu et al., 2010], maintenance [Buriol et al., 2009; Chang et al., 2010; Schwald and De Laval, 2003], and assembly training [Brough et al., 2007].

Regarding AR several training applications have been developed for the aeronautic and aerospace industries. In 2004, Macchiarella and Vincenzi [2004] proposed an AR application based on a video-see through desktop display as a learning paradigm for maintenance training. The approach was then transposed to mobile AR [Haritos and Macchiarella, 2005] for aerospace maintenance training. They evaluated the learning effects on long term memory. Results showed that the operators did not recall more information but they recalled a greater percentage of what they learned. Following these results Rios et al. [2011] developed a laptop based AR application for aeronautic maintenance and training. They demonstrate that, with simple AR training devices, the performance of the workers when performing complex tasks is increased. Similar work has been carried out in the military [Brown et al., 2006] and medical industry [Yeo et al., 2011].

2.3.2 Assistance applications

Assisting operators with guidance tools is commonly done in several manual tasks such as assembling or disassembling objects. One straight forward way to provide assistance

is by using classical textbooks that can be used as a construction guide. In order to continuously increase the workers production new techniques have been explored to guide the operators through maintenance and assembly tasks. For this purpose, MR technologies have been introduced in several industrial fields.

Within the Boeing company, Mizell [2001] aimed at using AR for guiding operators in an electrical wire bundle assembly task with HMD technologies. Even though they did not found any productivity improvements with their approach they supported the idea that AR could improve the productivity of the workers. Recent studies from Büttner et al. [2016] and Funk et al. [2016] confirmed the results found by Mizell [2001]. They compared the effects of displaying the guiding information either on a paper, an AR HMD or with in-situ projection. Both results showed that using a paper or in-situ projection provides better guiding information in terms of task completion time and average errors compared to HMD. Büttner et al. [2016] also carried out subjective tests on the participants and the results endorsed the fact that being guided by in-situ projection or paper was more helpful, joyful and easy to use than HMD techniques. The weak results of the HMD is explained by Büttner et al. [2016] by their lack of robustness under bright light and their restricted field of view.

Regarding MR assistance applications that do not use HMD, Echtler et al. [2004] introduced the Intelligent Welding Gun (Figure 2.26-left) that helped users shooting studs with high precision on prototype vehicles. The Intelligent Welding Gun was conceive with an instrument-based approach with a rendering display attached to the gun to give feedback on the performed task. An ART infrared tracker was used to track the gun and the components. Since 2009 Diota[4], a french company, conceives and distributes projection-based AR systems to different industrial fields. Their products are based on projectors that display 2D information over industrial part to assist the operators in assembly and maintenance operation (Figure 2.26-middle).

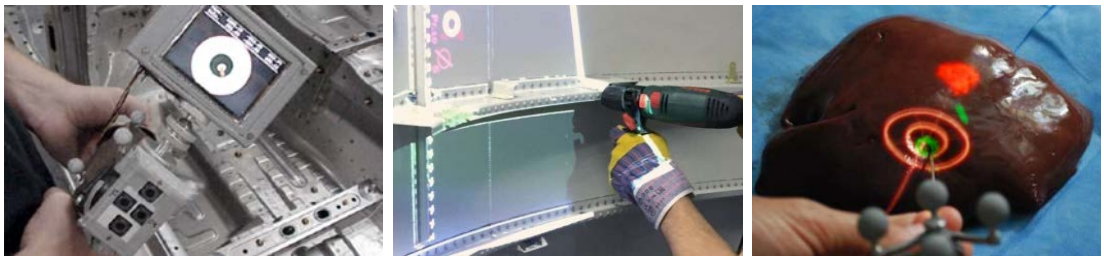


Figure 2.26 – Example of AR industrial applications for maintenance and task assistance: the Intelligent Welding Gun [Echtler et al., 2004] (left), Diota's projection-based AR assistance for assembly operations (middle) and projection-based AR for medical assistance on liver surgery [Gavaghan et al., 2011] (right).

Medical operations can be critical and involve heavy responsibilities on the operator. Therefore computer-aided techniques have been developed to assist surgeons. As an example, Gavaghan et al. [2011] designed a portable handheld-navigated projection device. The device is based on a laser projector that projects information over the surface of the liver for assisting surgery operations (Figure 2.26-right).

[4]Diota Augmented Reality for Industry - http://www.diota.com - Accessed: 2017-09-09

2.3.3 Design applications

Design processes can step in at different stage of a product life-time cycle [Mourtzis et al., 2014]. Mixed reality application have been widely used in product design. In particular VR has proven to provide adequate interactions to designers during the product design phase [Nee et al., 2012]. Early VR systems have been developped for design purposes [Butterworth et al., 1992; Chapin et al., 1994; Dani and Gadh, 1997]. The 3DM was designed in 1992 to adapt Computer-Aided Design (CAD) and drawing programs to VR Head-Mounted Displays (HMD). The DesignSpace [Chapin et al., 1994] was introduced in 1994 to enable conceptual design and assembly planning using voice and gesture in a virtual environment. The COVIRDS system [Dani and Gadh, 1997] (first implemented with a desktop display) brings together the CAD modeling, the user interface design and VR technologies.

Based on these works, several MR design applications have been proposed. The Spacedesign [Fiorentino et al., 2002] introduces a complete design process using both AR and VR. It proposes tools for creating and editing 3D curves and surfaces. For early conceptual design, Israel et al. [2009] and Stark et al. [2010] introduced 3D sketching techniques (Figure 2.27-left) and carried out a study with experts and users on the efficiency of 3D sketching compared to 2D paper sketching. Regarding 3D sketching, later work from De Araùjo et al. [2012] proposes bi-manual interaction techniques for sketching and designing objects over planar surfaces such as workbenches. Nevertheless workbench systems propose limited immersion and are adapted to design objects at a reduced scale. For designing large objects at scale one the users can be immersed in large projection-based systems. Therefore Hughes et al. [2013] proposed CaveCAD (Figure 2.27), a VR architectural design application for immersive environments displayed on a CAVE. They implemented interaction techniques and several features that enable modifying the geometry of the objects. Their preliminary study suggested that an arm fatigue appears for manipulating the control device in mid air for a long period of time compared to desktop displays. Similarly to the CaveCAD, the SculptUp system [Ponto et al., 2013] proposes an alternative way of designing objects in VR CAVE systems.

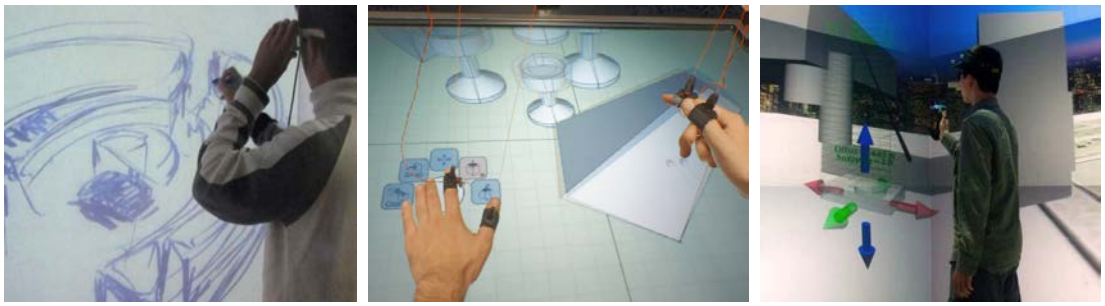


Figure 2.27 – Design applications of Mixed Reality in the industry: Early conceptual 3D sketching [Israel et al., 2009] (left), the Mockup Builder for modeling on workbenches [De Araùjo et al., 2012] (middle) and architectural design and modeling in a CAVE environment [Hughes et al., 2013] (right).

Regarding AR design applications, Xin et al. [2008] introduced Napkin Sketch, a tablet-based system for supporting artistic sketching in 3D. Nevertheless the users

seem to have adopted the technique in a similar way as paper and pencil rather than 3D modeling techniques. Tablet-based AR limits the interaction to a one-hand interaction since the tablet is generally hold in the other hand. Therefore an alternative to these techniques that free both hands can be projection-based AR(other than hand-held projectors). Saakes and Stappers [2009] proposed SKIN: a tangible design tool for sketching materials over products. By projecting computer generated texture they enable designers to better appreciate the 3D shape of the objects. They made a proof-of-concept on a ceramic design use case and suggested its introduction to industrial practical uses.

2.3.4 Planning and validation applications

Among the different MR systems, VR systems have been widely used for factory layout planning. Iqbal and Hashmi [2001] were one of the first to use virtual environments for factory layout planning and to propose alternative layouts solutions. Similar work from Calderon et al. [2003] proposed an on-line layout planning that can help the users in exploring alternative planning solutions. Nevertheless these approaches do not propose to insert the virtual environment into immersive displays. Work from Menck et al. [2012] and Menck et al. [2013] introduced collaborative virtual environments for factory layout planning tasks. They proposed an approach enabling to simultaneously visualize, investigate and analyze factory plans.

Regarding assembly planning, an early comparative study was carried out to evaluate the potential of using virtual environments for supporting assembly planning [Ye et al., 1999]. The study aimed at comparing three environments: traditional engineering, non-immersive VR, and CAVE VR. The results showed that participants performed better in VR environments for tasks related to assembly planning. Following these results several applications in CAVE environments have been designed for layout planning and validation. Regarding AR, Doil et al. [2003] used AR systems to plan the layout of factories by displaying virtual content over the actual factory floor thanks to tracked fiducial markers (Figure 2.28-left). For planning complex manufacturing systems, Dangelmaier et al. [2005] proposed a system that uses both AR an VR technologies. The CAVE VR systems enables several people of a project team to study and validate the planned manufacturing system (Figure 2.28-left). In the other hand the AR system enables editing the layout and model other plans similarly to the approach proposed by Doil et al. [2003]. Work from Medeiros et al. [2013] introduced a 3D interaction tablet-based tool within a CAVE system. The tool was evaluated on both a gas factory and a photo-voltaic solar plant for investigating and validating the final layout (Figure 2.28-right).

Regarding validation, De Sa and Zachmann [1999] introduced VR tools for reviewing and verifying assembly and maintenance task using Head-Mounted Displays (HMD) in the automotive industry. They carried out a user survey which results promote the use of VR for virtual prototyping in the automotive industry. Still in the automotive industry, the virtual center from PSA Peugeot Citröen is equipped with a CAVE display where a project team can interactively validate the design and plan the assembly of cars [Arnaldi et al., 2006]. Regarding AR validation, Caruso and Re [2010] developed an AR design review system. The system is based on a Video See-Through (VST) HMD

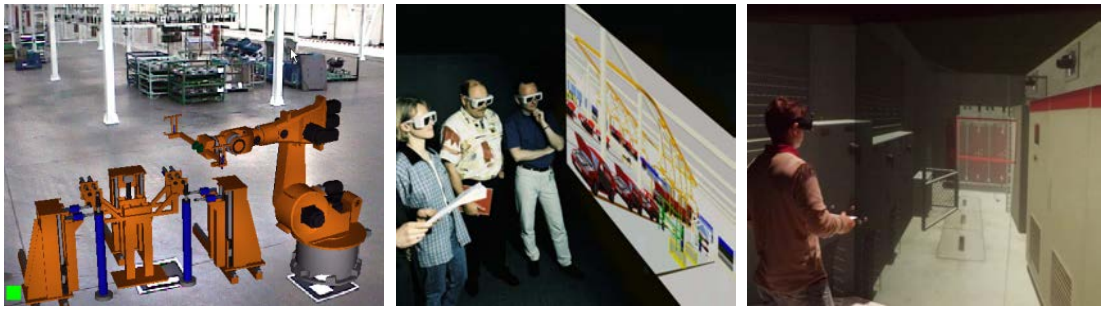


Figure 2.28 – Planning applications of Mixed Reality in the industry: an AR planning applicaiton for factory layout [Doil et al., 2003] (left), a project team validating the planned manufacturing system [Dangelmaier et al., 2005] (middle) and the validation of a solar plant in a CAVE environment [Medeiros et al., 2013] (right).

that helps visualizing and interacting in AR with a virtual object during the virtual prototyping review process.

Conclusion

The need for simulating most of the industrial operations has promoted the use of Mixed Reality (MR) technologies in many industrial applications. Immersive environments such as Virtual Reality (VR) are a good candidate to propose simulations in training context. Simulating training operations in MR enables to put the trainees in "almost" real conditions without putting them or other persons in danger. Moreover studies have proven that training in VR environment increases the performances of the operators compared to standard training sessions. VR environments are also appropriate to designing objects or processes thanks to imported CAD models and 3D drawing interaction paradigms. In terms of planning and validations VR Projection-Based Systems (PBS), like the CAVE system, present an advantage since they enable a complete project team to interactively plan and validate products and operations.

Regarding AR even though it is used for several operations, it is mainly used for assisting the operators in their task. Even though HMD technologies have proven to be less efficient than standard textbooks for guiding operators, in-situ AR projection is at least as efficient. Moreover in-situ projection is visible by other operators which can give better performances regarding multi-user tasks. Since in-situ projection has not overcome textbook performances, a challenge remains on improving the use of projection-based AR systems.

2.4 Conclusion

Mixed Reality (MR) systems encompass several components, namely: the visual display, the tracking and the application. In this chapter we have made an overview of MR systems and their usage in industrial applications.

Regarding the MR visual displays, four main categories are referenced: near-eye displays, handheld displays, stationary displays and projection-based systems. The

near-eye systems display virtual content directly in front of the user's eyes and are generally head-mounted. They generally present a restricted field-of-view but are wearable and mobile systems that can provide full immersion. Handheld systems are held on one or both hands and are generally video-based. They do not provide immersion but are a portable and straightforward way to porpose MR applications. Stationary displays generally require to wear stereoscopic glasses. These systems are not portable and hardly movable but they propose adapted environments for interacting at arm length. Finally Projection-Based Systems (PBS) are large environments that can also be stationary and can enable several persons to visualize the environment. Even though PBS propose more immersion than handheld and stationary displays, they do not overcome near-eye displays.

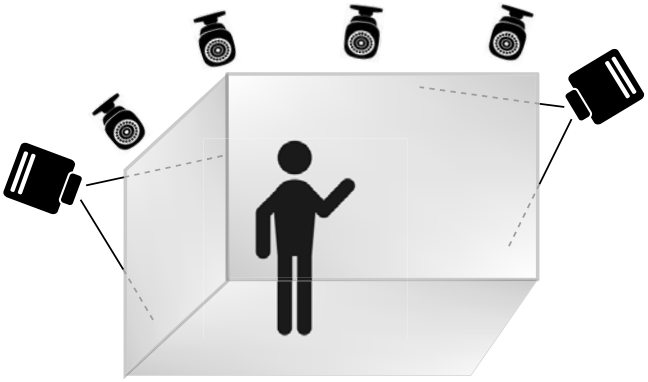
The different MR systems require a tacking to at least display consistent stereoscopic rendering. Several tracking techniques have been reported: mechanical, acoustic, magnetic, inertial and optical. Nevertheless "there is no silver bullet" [Welch and Foxlin, 2002] and tracking technologies can be coupled to give different performances. Nevertheless a challenge still needs to be addressed in order to design an optimal tracking system or, at least, to increase the performance of the existing ones. The optical tracking is commonly used for both Virtual Reality (VR) and Augmented Reality (AR) applications. Optical tracking requires to have visual feature that are identified by the sensors. Generally the optical tracking systems used in VR require to add physical markers over the objects while most of the AR applications use natural features. Nevertheless optical tracking techniques still present some limitations in terms of robustness to lighting and workspace covering. Optical tracking systems generally require for the target to be visible by at least two cameras.

Both visual displays and tracking systems define a MR system that can be used for many industrial applications. Indeed the need for simulating most of the industrial operations has legitimized the use of MR for simulation purposes. Several applications have been developed to increase overall productivity and ease complex and expensive industrial task. For instance training, assistance, design and validation applications have developed in several industrial domains such as aerospace, automotive, construction, architecture, military and retail. Regarding the performance of the MR systems several studies have proven that the use of near-eye displays such as HMD does not have a positive impact on the operators productivity compared to in-situ projection for assembly or maintenance task for example. Moreover several projection-based VR systems, like the CAVE, are used to validate and review projects in team since they enable several users to visualize the virtual environment.

Even though we are not yet be able to propose a "silver bullet" tracking nor an optimal visual display, several systems have been developed to fulfill, in an acceptable way, the needs of industrial actors. A variety of visual displays and tracking systems can offer a large selection of MR systems for different industrial applications. Nevertheless the studies have still not proven the efficiency of using MR in the industry. Therefore work is still required to adapt these systems and propose alternatives ways of using MR for industrial applications.

Pilot study: Analysis of user motion in a specific CAVE-based industrial application

3



Contents

3.1 Analysis of user motion in a CAVE-based industrial application 48
3.1.1 Selection of the industrial application 48
3.1.2 Participants 51
3.1.3 Procedure 51
3.1.4 Collected data 51
3.2 Results 52
3.2.1 Cyclop (Head) and Wand (Hand) 3D positions 52
3.2.2 Cyclop and Wand 3D orientations 54
3.2.3 Cyclop and Wand speeds 55
3.3 Discussion 55
3.4 Main outcomes and guidelines 56
3.5 Conclusion 57

Every VR system presents a different set of properties and characteristics in terms of space occupation and user experience. The tracking systems, that provides the application with information about user's head and hand 3D motions, are mostly designed to cover the entire workspace provided by the VR system. As for today the deployment of these consumers systems is often made based on previous experience and intuition.

As such, the design of Projection-Based Systems (PBS) is not always well adapted to the end-use application, partly due to the lack of usage analysis. Even though usage data is strongly dependent on the application, it could help defining guidelines for designing and deploying PBS systems better adapted to their industrial context of use. Such guidelines or rules could assist engineers when designing cost-efficient and application-driven PBS systems.

In this chapter we propose a pilot study that follows this path. We focused on a specific **planning/validation application** of the **construction industry** and a standard **CAVE-like projection-based system**. We recorded the end-user's behavior (head and hand 3D motions) in "out-of-the-lab" conditions of experimentation. The experiment was carried out during a construction industry exhibition in which data from 58 participants was recorded. The acquired information was used to analyze the user's motion behavior when immersed in the MR application.

In the remainder of this chapter we first present our analysis and its main components such as the application, the projection-based setup, or the recording procedure. Second, we discuss the results and the resulting design guidelines. The chapter ends with a general conclusion.

3.1 Analysis of user motion in a CAVE-based industrial application

In this thesis we aim at exploring ways to improve the usage of Projection-Based Systems (PBS) for mixed reality industrial application. In order to approach PBS and identify the main leads and perspectives for improvement, we first propose a pilot study to characterize the usage of such systems in an industrial context. Therefore, we carried out a study in "out-of-the-lab" conditions during an industrial exhibition. A PBS was presented during a french construction exhibition that aimed at popularizing Virtual Reality (VR) in the construction industry. Fifty-eight naive participants from the construction industry were able to test the system. Their 3D motions (head and hand) were recorded in order to analyze the trends that can be found in the usage of a specific CAVE-based VR environment. In this section we present the different components of the study.

3.1.1 Selection of the industrial application

In the context of construction industry, we chose to present the users with a **planning/validation construction application**. The application was mainly addressed to Small and Medium Businesses (SMB) in order to introduce them to Virtual Reality (VR) technologies. The concerned SMB were mainly kitchen and bathroom designers.

Such companies can take advantage of virtual reality to visualize and dynamically design their project together with their own clients that are generally private individuals. Using VR in this context generally enables Realyz clients to save time during the design phase of the project due to a fastest validation with their clients.

In the following we give details about the applications, the projection-based system and the 3D interaction that we chose for the study.

Virtual reality application

The participants were immersed in a construction application (Figure 3.1). The proposed virtual environment was a 3D model of a $200m^2$ house (constructed in a $900m^2$ field) in which most of the objects could be selected, manipulated and changed. The house could be redesigned and rearranged at will.

The house was built on 2 floors. The ground floor presented an open kitchen and a living room. From the living room a corridor led to a large bathroom and a bedroom. In the living room, a bay window led to a large garden with a swimming pool and stairs led to the first floor. The first floor was composed of a bedroom and bathroom located at the end of a corridor. It also had a playroom with toys and playful objects. Figure 3.1-left illustrates the application with a top view of the 3D model of the ground floor. Figure 3.1-right illustrates a density map that corresponds to the accumulated trajectories of the participants. We can notice that the participants spent most of the time in the kitchen and bathroom.

Figure 3.1 – 3D scene of our VR application dedicated to the construction field: a top-view of the ground floor (2-floor house, garden and parking lot), the kitchen is depicted in red, the bathroom in green (left) and the density map corresponding to the accumulated trajectories of the 58 naive users who participated in our study (red-blue-green code, red=high density) (right).

Projection-based system

The participants were immersed in a CAVE-like display composed of four front-projected screens providing a volume of 2.5×2.2×2 meters (Figure 3.2-left). An Optitrack[1] tracking system made of four Prime 13 cameras was used to provide the application with user's 3D motion information. Figure 3.3 illustrates the positioning of the optical tracking system on the VR setup and the corresponding worskspace covered.

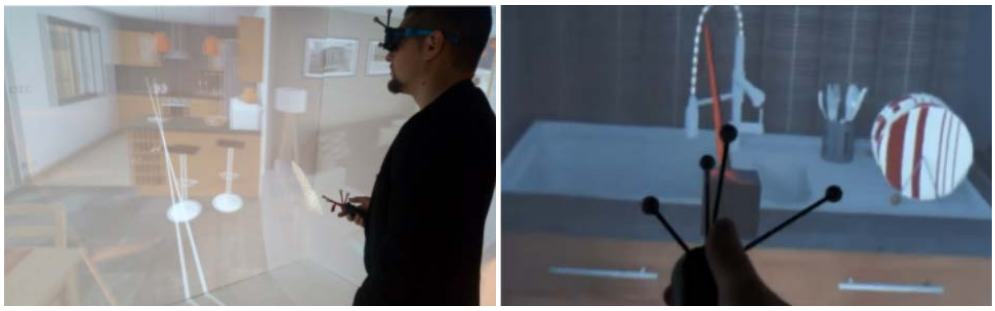


Figure 3.2 – Our pilot study VR application in use: a user interacting in the CAVE display (left) and a first-person view during the selection an object made with a virtual ray controlled by the Wand device (right).

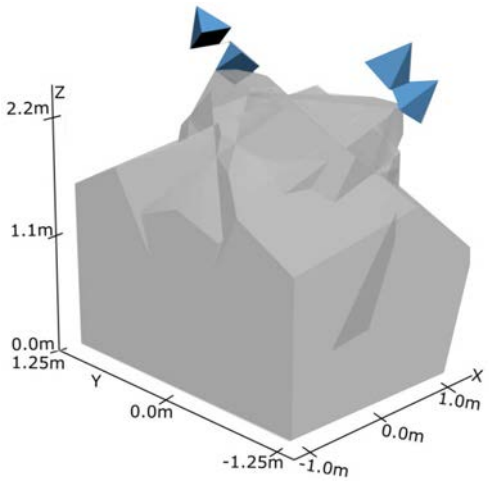


Figure 3.3 – Workspace covered by the tracking system (in grey) used in our pilot study. The blue cones are the 4 tracking cameras.

3D user interactions

The 3D interaction techniques implemented in our construction application are standard implementations of the literature [Mine et al., 1995]:

- The **3D navigation** in the application is a hand-directed navigation technique that makes use of a wand-based pointing technique [Mine et al., 1995] to define the navigation direction. Then, the wand's joystick enables the users to move

[1]Optitrack - http://optitrack.com - Accessed: 2016-10-08

forward and backward. When pushing the joystick toward the right or left a rotation around the vertical axis (Z) is performed. Maximum navigation speed is set to $2.0m.s^{-1}$ for translation and $30°.s^{-1}$ for the rotation. A linear ratio is used, thus when the users push the joystick to 50% of its maximum they directly navigate at 50% of the maximum speed.

- The **3D selection** technique is a standard ray-casting technique as implemented by Mine et al. [1995]. The users can select the object pointed by the ray by pressing a button (Figure 3.2-right). The **3D manipulation** of a selected object is a free manipulation. The users move the wand controller they are holding to move the object through the environment with a direct 6DoF mapping where the pivot is located in the wand.

3.1.2 Participants

A total of 58 participants were recorded, with 47 males (81.0%) and 11 females (19.0%) leading to a mean height of 1.74m. In terms of age there were: 8 under 20 (13.8%), 28 participants having between 20 and 40 years old (48.3%), and 22 over 40 (37.9%).

Most of the users did not had previous experience with VR. 41 participants were beginners (70.7%), 10 intermediates (17.2%) and 7 experts (12.1%). Finally most of the users, 54, were right-handed (93.1%) and only 4 (6.9%) were left-handed .

A high number of participants were from the construction industry. Most of them were kitchen or bathroom designers that show interest in VR to visualize and modify the arrangement and design of their construction project before validating it.

3.1.3 Procedure

The experiment was conducted during the famous ARTIBAT[2] construction industry exhibition that took place in Rennes in October 2016. The participants had to sign a consent form informing that they could stop the data acquisition whenever they wanted or could ask that their data was erased. They were also informed of the anonymity of the collected data and of the research purpose of the recording.

Since the data acquisition was made in end-use conditions during an exhibition, the data recording was triggered manually by the experimenter. Indeed, participants were accompanied of a salesperson that was in charge of introducing them to the CAVE-like system and the application. The salesperson assisted the user during the first seconds of usage by equipping them with the glasses and the wand. Then the participant was left alone in the VR system to manipulate and move freely. Data was recorded from the moment the participant was alone and stopped when the participant wanted to leave the CAVE. The initial position of the users and the types of interaction were not previously defined and were set free to user's will.

3.1.4 Collected data

For every participant we collected the following data:

[2]ARTIBAT Salon de la Construction - http://www.artibat.com - Accessed: 2016-09-20

- The **position and orientation of the user's head (cyclop)** were recorded respectively in meters (3D vector) and radians (quaternion). In the following the Cyclop refers to the point in the middle of the user's eyes.
- The **position and orientation of the user's hand (wand)** were recorded respectively in meters (3D vector) and radians (quaternion). In the following the Wand refers to the 3D pointing device held in the user's hand and used to interact with the application.
- The **interaction state** that corresponds to the interaction the user is currently performing. It can be: *Navigation* if the user is navigating through the environment, *Manipulation* if the user has selected an object, or *No Interaction* if neither *Navigation* nor *Manipulation* are active.

Tracking data was provided by the OptiTrack tracking software, which is computed from four cameras placed at the top of the VR system. Around 3 hours of raw data was acquired for both the Cyclop and the Wand at a rate of 10Hz.

3.2 Results

3D and 2D graphs were computed according to the acquired data and illustrate the different parameters of the usage of the VR system. The positions of either the Cyclop or Wand illustrate the volume where the users moved their head and hand when using the application. As a first result, a density map is presented in Figure 3.1-left to visualize the areas of the application where the participants spent most of the time. Regarding the interactions, the users spent 58.0% of the time in the *No Interaction* state, 39.3% in the *Navigation* state and 2.7% in the *Manipulation* state. In the following we present the analysis of the acquired data.

3.2.1 Cyclop (Head) and Wand (Hand) 3D positions

In order to visualize the space distribution of both Cyclop and Wand, each position was drawn as a 3D point in a representation of the 2.5m×2.2m×2m VR display. Figure 3.4 illustrates the spatial distribution of the Cyclop and Wand positions. An ellipsoid was computed to cover 84% of the data [Payton et al., 2003]. The point cloud was projected along each axis to visualize the distribution according to every plane of the system. The ellipses drawn on each plane cover 84% of the 2D projected data. Such elliptic shapes were computed with a Principal Component Analysis (PCA) algorithm which extracts the 2 or 3 principal axes of the data distribution for respectively 2D and 3D data.

Then the 3D positions of both Cyclop and Wand were separated according to the interaction state. Figure 3.5-top illustrates the planar distribution of the data according to each 3D axis as function of the different interactions. The data distribution of the Cyclop and Wand positions are approximated by 84% confidence ellipses on each plane.

Finally the data was also separated, according to the user's past experience with VR, in two groups: *Beginners* and *Intermediates and Experts*. The results for both

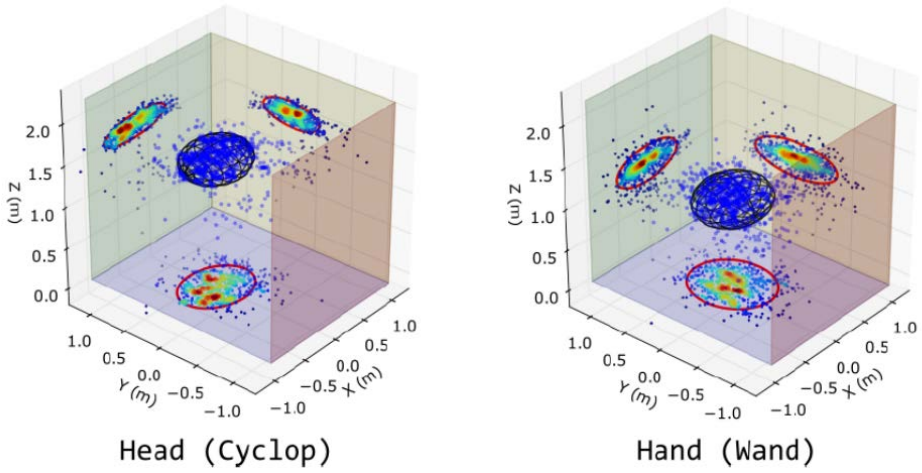


Figure 3.4 – 3D positions of the Cyclop (left) and the Wand (right) in a 2.5m×2.2m×2m CAVE-like display. The 3D point cloud is projected on each plane to visualize the 2D distribution of the points according to each direction. The 2D and 3D point clouds are estimated with elliptic shapes that include 84% of the data.

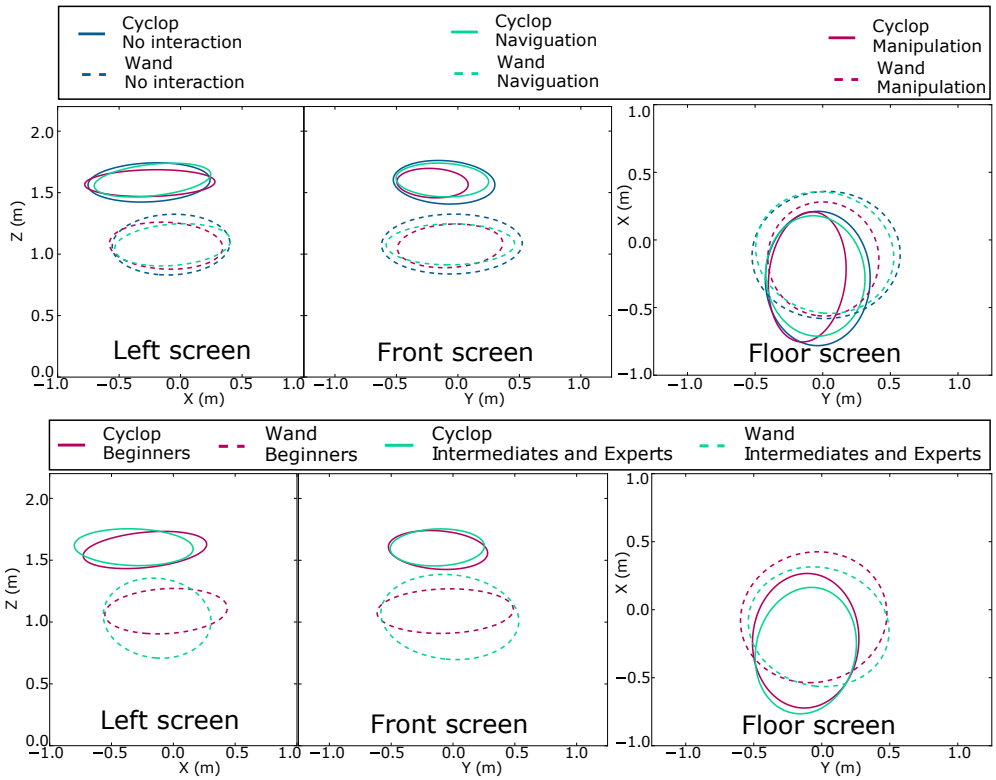


Figure 3.5 – Influence of interaction state and user's past experience with VR on head and hand 3D positions. Projection of the data on the left, front and floor screens according to the interaction state (top) or the user's past experience with VR (bottom). Each dataset projection is approximated with a 84% confidence ellipse.

groups are presented in Figure 3.5-bottom with the planar distribution of the data according to each 3D axis.

3.2.2 Cyclop and Wand 3D orientations

The orientations of the Cyclop and the Wand represent, respectively, the direction the user is looking at or pointing the wand at. Such data can be visualized by depicting the point of intersection between the ray cast from either the Cyclop or the Wand and the different screens of the VR display. Figure 3.6 presents the results by approaching the data with a 84% confidence ellipse shared by the four screens. Similarly to the 3D position analysis, the orientations were also separated according to interaction state (Figure 3.6-top) or user's past experience with VR (Figure 3.6-bottom).

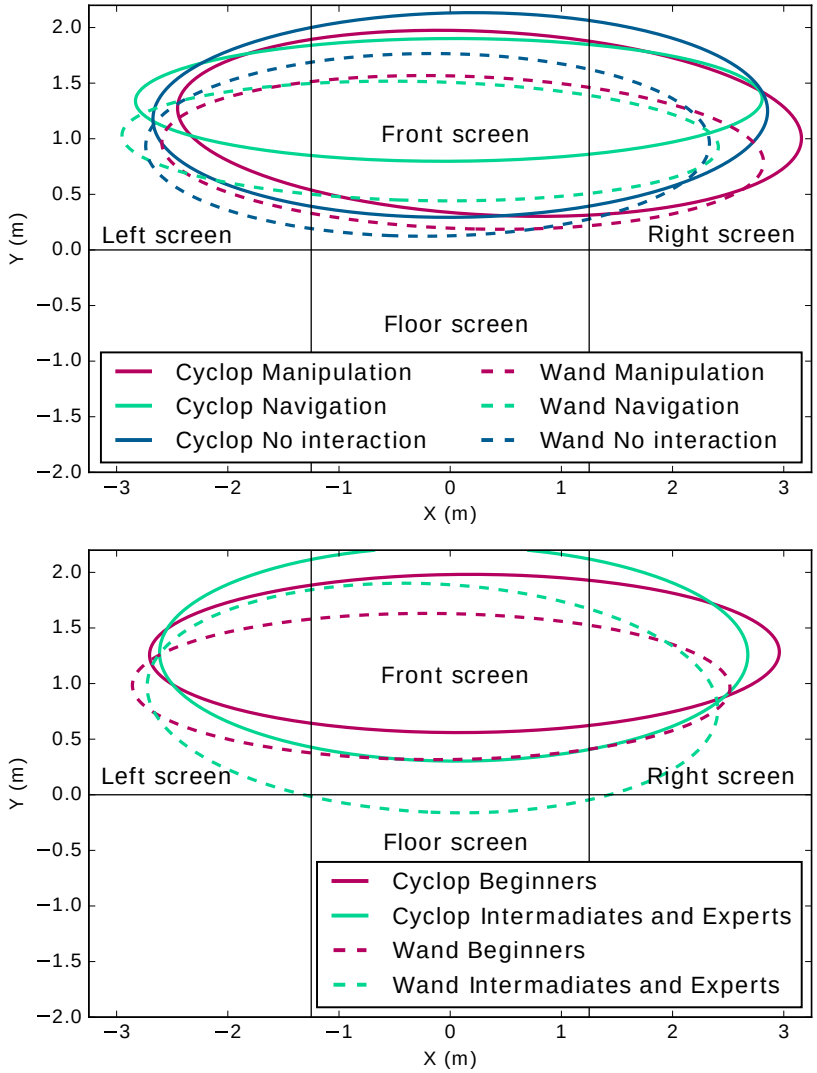


Figure 3.6 – Influence of the interaction state and the user's past experience with VR on 3D head and hand orientations. Distribution of the intersection between the 4 screens and the Cyclop's sight or Wand's pointing direction according to each interaction state (top) or the user's past experience with VR (bottom). Each dataset is approximated with a 84% confidence ellipse.

3.2.3 Cyclop and Wand speeds

The Cyclop and Wand speeds were computed from the raw position data. Speeds were computed for each interaction state. Figure 3.7-top illustrates the percentage of sample moving at every speed for both Cyclop and Wand during each interaction. The 95% confidence speed of each curve is drawn as a vertical line in Figure 3.7. Same procedure was used to depict the head and hand velocities according to user's past experience with VR (see Figure 3.7-bottom).

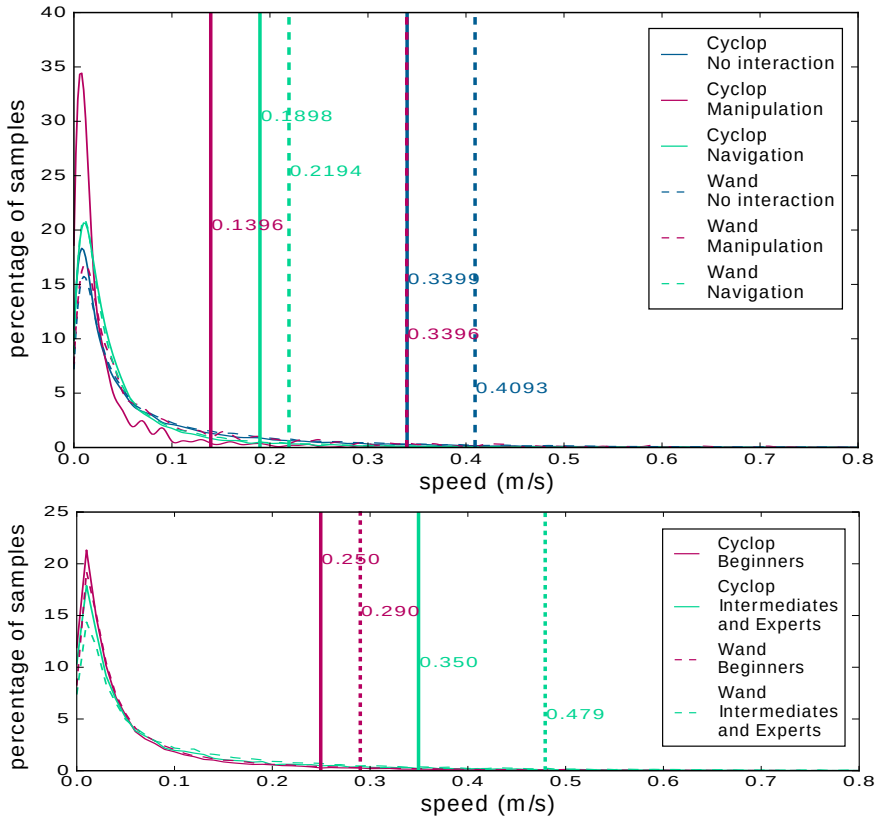


Figure 3.7 – Speeds of the Cyclop and the Wand depending on: the interaction state (top) or the user's past experience with VR (bottom). The vertical lines represent the 95% confidence speed values for each condition.

3.3 Discussion

The recorded data is homogeneous and shows that the participants were well engaged in the application in "out-of-the-lab" conditions. Figure 3.1 shows that the users did not try to cross walls and did not principally go outside. Since most participants were kitchen and bathroom designers, they spent most of their time in the kitchen and bathroom. Moreover the hand speed profiles are consistent with the speeds found for aiming tasks in virtual environments [Liu et al., 2009].

An unexpected result was that the users only used a minimal amount of the available tracking volume of the VR system. Indeed, from the above data an approximated motion volume was computed for the Cyclop and Wand. By computing the volume of the 84% confidence ellipsoids the actual Cyclop's motion volume is of around $0.15m^3$ and the Wand's one is of around $0.25m^3$. By simply adding both volumes one can consider an effective volume of $0.40m^3$. Since the VR display provides $11m^3$, the effective volume represents between 5 and 10% of the total volume. When considering an ellipsoid covering the 95% of the used volume the effective volume still corresponds to 10% of the global volume.

Moreover the users, no matter their past experience with VR, were mainly stationary in the setup. Indeed, Figure 3.7-bottom depicts slow motions (from $0.13m.s^{-1}$ to $0.47m.s^{-1}$). Nevertheless experienced users tend to move faster (40% faster for the head and 65% faster for the hand) and to move their hand in a larger range along the Z-axis (Figure 3.5-bottom). According to Figure 3.4 and 3.5 the user's hand and head rarely left the center of the tracking volume which almost corresponds to the center of the VR display.

Interaction state has also an impact on the head and hand motions. As illustrated in Figure 3.5, head range of motion is restricted when the user is manipulating an object. Such result is confirmed by Figure 3.7-top, head is almost stationary when manipulating. When navigating both head and hand motions are restricted and their speeds are close one from another (Figure 3.7-top). This can be due to the fact that the hand directs the navigation and the users may need to look at the direction they are moving to.

The orientations of the head and hand (Figure 3.6-bottom and 3.6-top) also provide unexpected results since almost no user looked or pointed the wand toward the floor screen. We can even say that they spent most of the time looking and interacting with the front screen. The orientations also confirm that expert users are more comfortable with hand motion.

3.4 Main outcomes and guidelines

The results of the analysis highlighted trends in the usage of projection-based systems. In general the participants tend to be static within the display and to move slowly. An open question remains on how to encourage the users to take advantage of the overall system. Also, a difference can be perceived regarding the usage of the systems when comparing beginners and experienced users. Some users, who were mainly inexperienced with VR, pointed out that sometimes "*it is difficult to apprehend the virtual environment*". Such a difference can suggest that using the system may not always be intuitive or comfortable. Therefore and open question remains on how to adapt the system and the application to increase the user experience and comfort.

From the analysis we identified some leads that could help adapting the design of Projection-Based Systems (PBS) to specific industrial applications:

- The tracking system could be adapted to provide optimal performances in the areas of interest while providing enough liberty on the usage of the overall PBS.

- The overall system could be revised to encourage the users to move through all the display volume and take advantage of the entire environment and immersion.
- The application could be adapted to increase the user experience and perception in PBS leading to a better understanding of the virtual world and its physical constraints.

To sum up, from this pilot study and for the purpose of this thesis, we can identify two main challenges that could be addressed to adapt PBS to the end-use industrial application. First a technical challenge could aim at adapting the technical components (tracking and display) of PBS in order to enable the users to take advantage of the overall range of interactions and workspace provided. Then, a user-centered challenge could aim at improving the user experience and propose more realistic real-life professional situations when using industrial applications in PBS.

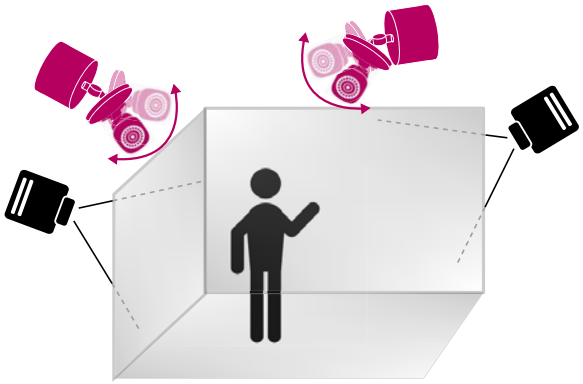
3.5 Conclusion

We carried out an analysis of head and hand 3D motions in a CAVE-like VR system within a predefined construction VR application. Our results show that, in "out-of-the-lab" conditions, 90% of the volume proposed by the VR system was not used. Moreover the users remained mainly stationary and their motion velocities were small (less than $0.5m.s^{-1}$). The user's past experience with VR influenced to some extent its behavior with wider hand motion for expert users. The interaction state has an impact on the head and hand motions since the head is more stationary when manipulating. It is noteworthy that the orientations of head and hand are centered in the middle of the front screen and spread across the horizontal axis.

As a pilot study, these results promote the idea of adapting different components of a MR system to the application and usage. In fact the tracking system could be adapted to the actual usage of the MR system and propose cost-effective solutions that provide a larger workspace and interaction volume. Therefore it will facilitate the conception of the final MR system and ease the usage of the application.

Regarding the user behavior, this study could suggest doing some improvement on how the users could have a better perception of the virtual environment. Such improvement could lead to a better user experience that enables user to make the most of the MR applications.

4 Increasing optical tracking workspace for mixed reality applications



Contents

4.1 General approach: Increasing the optical tracking workspace of mixed reality applications . . . 60
4.2 Fundamentals of optical tracking . . . 62
4.2.1 Prespective camera model . . . 62
4.2.2 Epipolar geometry . . . 65
4.3 MonSterTrack: Increasing optical tracking workspace with hybrid stereo/monocular tracking . . . 68
4.3.1 Off-line system calibration . . . 69
4.3.2 On-line real-time stereo tracking . . . 71
4.3.3 Monocular tracking mode . . . 76
4.4 CoCaTrack: Increasing optical tracking workspace with controlled cameras . . . 77
4.4.1 Off-line controlled camera calibration . . . 78
4.4.2 Registration . . . 80
4.4.3 Controling camera displacements: visual servoing . . . 80
4.5 Proofs of concept . . . 82
4.6 Performance . . . 84
4.6.1 Main results of our global approach . . . 84
4.6.2 Comparison with Vicon's optical tracking . . . 87
4.7 Conclusion . . . 90

The optical tracking techniques are probably the most commonly used in MR applications. They perform with infrared (IR) light visible by the sensors. Several industrial actors' optical tracking systems, such as Vicon, NaturalPoint or ARTracking, use a similar technique performing with high accuracy and frequency (metrology instrument). These methods require a stereo configuration (multiple cameras) to provide tracking data. Thus such systems present an inherent limitation regarding the workspace they cover.

In this chapter we present an approach which intends to maximize the tracking volume (or workspace) of optical tracking systems used in Projection-Based Systems (PBS). Our approach could also overcome some occlusion problems by relaxing the current constraints on camera positioning and multi-view requirements. As such, since adding cameras can be expensive and is not always possible (due to the lack of space) we propose not to use additional ones.

To achieve this goal, the contributions of this chapter are:

- *MonSterTrack* (Monocular and Stereo Tracking): When the tracked target is no longer visible by at least two cameras, we occasionally enable switching to a monocular tracking mode using 2D-3D registration algorithms (Figure 4.1-top-right). The localization accuracy remains precise enough although more noisy.
- *CoCaTrack* (Controlled Camera Tracking): A control scheme that allows the camera to remain centered on the target (Figure 4.1-bottom-left). It allows to track the target through a larger volume. Thus, the multi-view registration can be achieved as much as possible. This is carried out by a visual servoing process with cameras mounted on pan-tilt heads.

Numerous previous works in computer vision exist on designing and improving these two techniques independently. In this work, we adopt a different perspective and consider their introduction in the field of Virtual Reality (VR) and 3D interaction. We propose to reconsider these tracking techniques as alternate means to extend the 3D workspace, enabling to relax the current constraints on camera positioning and stereo requirements.

In the remainder of this chapter we first introduce our global approach for maximizing VR optical tracking workspace. Second we detail the geometry of two-view optical systems. Third, we detail our first method, i.e., our hybrid scheme based on stereo and monocular registrations. Fourth, we detail our second method, i.e., our VR tracking based on controlled cameras. Fifth, we present two illustrative prototypes of VR setups that were developed based on our approach. Sixth, we give the performance and results obtained with our systems. The chapter ends with a discussion and a general conclusion.

4.1 General approach: Increasing the optical tracking workspace of mixed reality applications

We propose an approach that intends to maximize the workspace of optical tracking systems using a small number of cameras. This approach is composed of two methods, that can be combined.

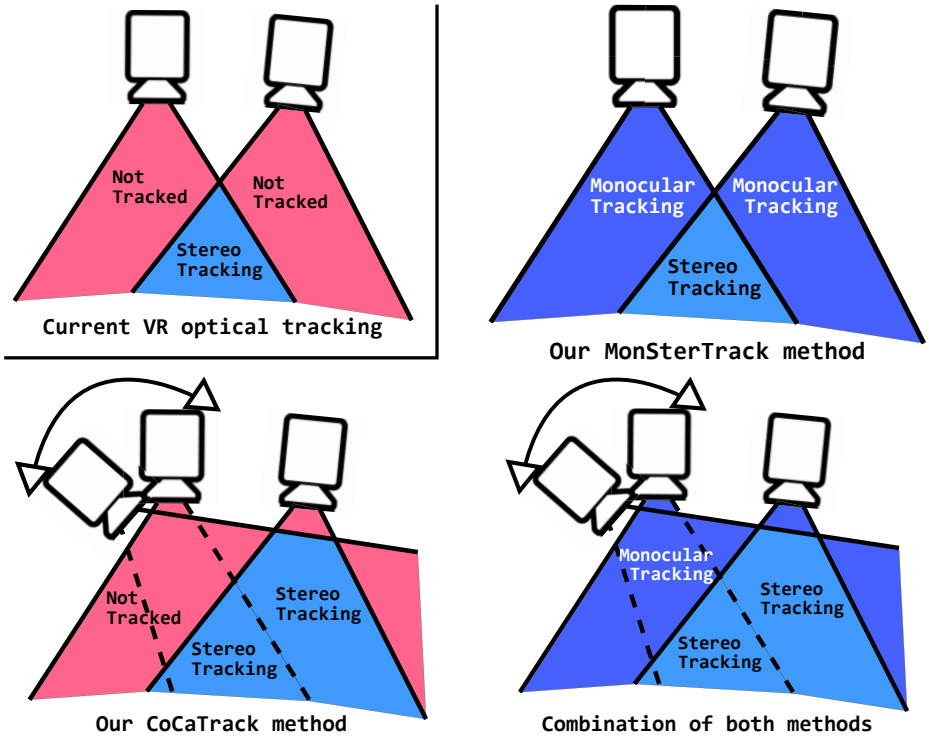


Figure 4.1 – Compared to current optical tracking systems (top-left) our approach increases the workspace with two methods: (top-right) enabling a monocular tracking mode when stereo registration is no longer available (MonSterTrack method), or (bottom-left) using controlled cameras able to follow a target (CoCaTrack method - illustrated here with one controlled camera). Both methods can be combined (bottom-right) to provide an even larger tracking workspace.

We propose a first method called ***MonSterTrack*** that occasionally switches to a monocular tracking mode if and only if the stereo mode is no longer available. We expect to increase the workspace by providing tracking data for the entire monocular space covered by each camera. Besides, some occlusion problems should be solved when the occlusion is not present in all the views. We anticipate to have slightly lower performances with monocular registrations. However we consider that the monocular registration should be used at the boundaries of the workspace, whereas stereo-based tracking should be kept for the most critical (inner) area of the workspace.

As a second and complementary method, we propose ***CoCaTrack***, a method that allows the cameras to move during the tracking. Using controlled cameras enables to follow the tracked marker through the workspace and make it visible as long as possible by as many cameras as possible. Such method should considerably increase the stereo workspace of a two-camera based tracking systems. We anticipate to have a bias on 3D reconstruction when using the controlled cameras due to calibration errors. Nevertheless these errors can be mitigated using a thorough calibration step.

Figure 4.2 illustrates the global architecture of our approach. Whenever a controlled camera is available it is used to follow the markers thanks to visual servoing. As such the stereo workspace should increase. The visual servoing loop is independent of the

tracking, so that the camera movements do not impact the tracking latency. Then, when stereo is no longer available the system switches to a monocular tracking. Calibration steps are required for the tracking to be robust and accurate. They will be presented in the following sections together with the stereo registration algorithms.

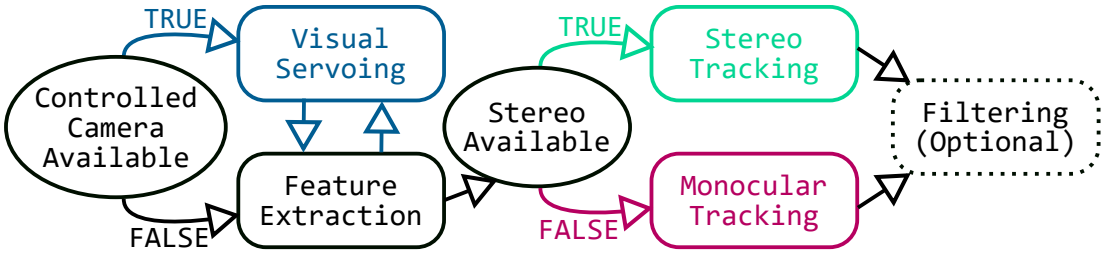


Figure 4.2 – Global architecture of our approach for maximizing the tracking workspace with controlled cameras and hybrid stereo/monocular tracking. This pipeline is performed when the marker is visible by at least one camera. The visual servoing runs in a parallel loop when controlled cameras are available.

4.2 Fundamentals of optical tracking

In order to introduce the contributions of this chapter we first describe the fundamentals of optical tracking. The fundamentals of optical tracking are detailed in the following by first presenting the camera geometry and then the epipolar geometry.

The camera geometry is used to define the mapping between a 3D space point and a 2D image point. In fact the 2D image point will be the projection of the 3D point in the focal plane of the camera according to a perspective projection model.

Then, the epipolar geometry represents the relations that exists between two views. It is independent of the scene structure and only depends on the cameras' internal parameters and relative pose. In the following we first present a standard camera model, then we introduce the principle of epipolar geometry used in two-view systems.

4.2.1 Prespective camera model

The perspective camera model is based on the human vision and and on the theory of perspective which was discovered in 1435. One of the most basic and common camera model is the pinhole camera model created from the camera obscura invention. This model uses central projection to determine the image of an object (see Figure 4.3).

4.2.1.1 The pinhole model

When using the pinhole camera model we assume that the image is formed on the plane $Z = f$ where f is the distance from the center $\mathbf{C}$ of the camera to the focal plane of the camera. Then the projection of a point $\mathbf{X} = (X, Y, Z, 1)^\top$ on the focal plane is given by the point $\mathbf{x} = (x, y, 1)^\top$ with:

$$\begin{cases} x = fX/Z \\ y = fY/Z \end{cases} \tag{4.1}$$

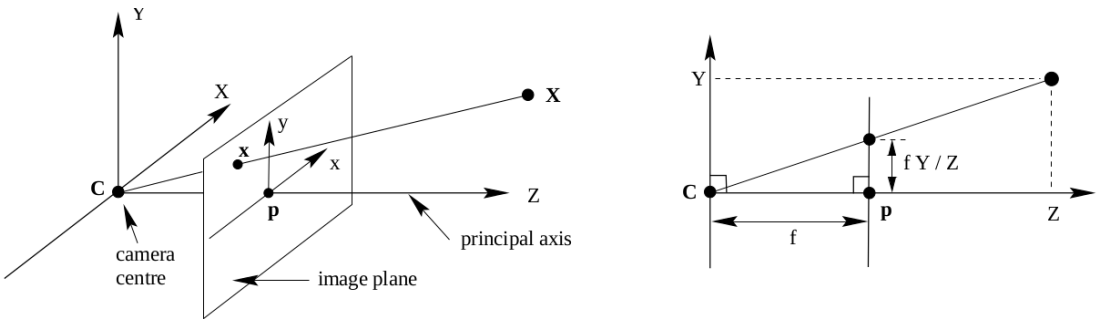


Figure 4.3 – Central projection on a focal plane for a camera whose center is **C** [Hartley and Zisserman, 2003].

4.2.1.2 Principal point offset and pixels size

The principal point of the focal plane is the point $\mathbf{p} = (0, 0, 1)^\top$ in the camera frame (expressed in the normalized metric space). However this point is not always placed at the origin of the focal image frame. Moreover the pixels may have different vertical and horizontal physical sizes on the sensor (l_x and l_y). Assuming that the principal point's pixel coordinates in the image frame are $\mathbf{p} = (u_0, v_0, 1)^\top$, the projected point $\mathbf{u} = (u, v, 1)^\top$ expressed in pixel in the image frame has the following expression:

$$\begin{cases} u = p_x x + u_0 \\ v = p_y y + v_0 \end{cases} \tag{4.2}$$

where $p_x = f/l_x$ and $p_y = f/l_y$. Equation (4.2) can be expressed as:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} p_x & 0 & u_0 \\ 0 & p_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{K}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{4.3}$$

where matrix **K** is called *camera calibration matrix* and its entries are the intrinsic parameters that can be computed thanks to a calibration process [Brown, 1971; Zhang, 2000] (see section 4.3.1.1).

4.2.1.3 Lens distortion

Lens distortion are part of the intrinsic parameters of the camera that can be estimated through calibration. Lens distortions are visible on the image because of a deformation that can take different shapes. The most important distortion is the radial distortion [Zhang, 2000] that impacts the image as shown in Figure 4.4. According to Wei and De Ma [1994] modeling the distortions could be limited to the radial distortions since taking into account more elaborated distortions can cause numerical instabilities.

If no distortion model is considered equation (4.3) is valid. Nevetheless when considering radial distortions, with coefficient k_1 and k_2, the undistorded point $\mathbf{x}' = (x', y', 1)^\top$ will have the following coordinates [Zhang, 2000]:

$$\begin{cases} x' = x + x(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) \\ y' = y + y(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) \end{cases} \tag{4.4}$$

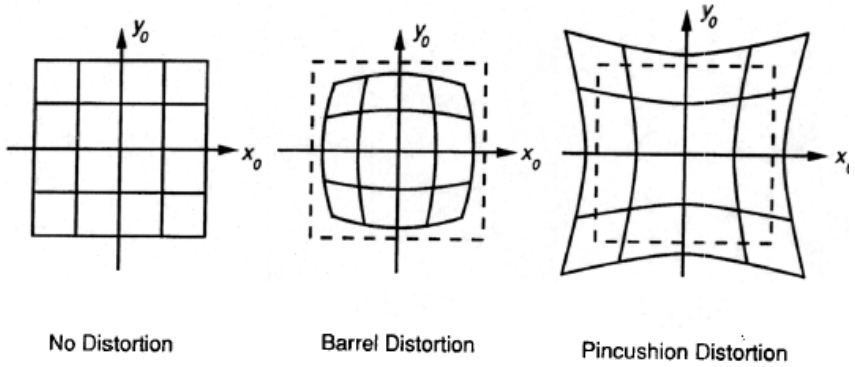


Figure 4.4 – Impact of different radial distortions on an image.

Other distortions can be present in the images as reported by Weng et al. [1992]. Decentering distortions combine radial and tangential deformations. These distortions occur because the lens is not always parallel to the image plane which creates trapezoid deformations. Correcting decentering distortions is performed in two steps. First a radial correction is applied to the original point $\mathbf{x}$ to obtain $\mathbf{x}'$ with equation (4.4). Then tangential corrections are applied to $\mathbf{x}'$ to determine the undistorded point $\mathbf{x}'' = (x'', y'', 1)^\top$ as follows [Weng et al., 1992]:

$$\begin{cases} x'' = x' + p_1(3x^2 + y^2) + 2p_2xy \\ y'' = y' + 2p_1xy + p_2(x^2 + 3y^2) \end{cases} \tag{4.5}$$

where p_1 and p_2 are the distortion coefficients of the tangential deformations.

Important note: In the following we consider that the cameras have been calibrated off-line [Zhang, 2000] and we assume that the images of the cameras have been corrected by applying radial en tangential corrections. Therefore we consider rectified cameras and the projection $\mathbf{x}$ of a 3D point $\mathbf{X}$ is always expressed in the normalized metric space (in meters).

4.2.1.4 Transformation between 3D frames

Considering two frames $\mathcal{F}_a$ and $\mathcal{F}_b$, a 3D point $\bar{\mathbf{X}} = (X, Y, Z)$ in non-homogeneous coordinates can either be expressed in $\mathcal{F}_a$ or $\mathcal{F}_b$ coordinate system as, respectively, ${}^a\bar{\mathbf{X}}$ and ${}^b\bar{\mathbf{X}}$. The transformation from $\mathcal{F}_a$ to $\mathcal{F}_b$ is given by a rotation matrix ${}^b\mathbf{R}_a$ and a translation vector ${}^b\mathbf{t}_a$ and we have:

$${}^b\bar{\mathbf{X}} = {}^b\mathbf{R}_a\,{}^a\bar{\mathbf{X}} + {}^b\mathbf{t}_a \tag{4.6}$$

Equation (4.6) can be rewritten in a matrix form in homogeneous coordinates as:

$${}^b\mathbf{X} = \underbrace{\begin{pmatrix} {}^b\mathbf{R}_a & {}^b\mathbf{t}_a \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}}_{{}^b\mathbf{M}_a} {}^a\mathbf{X} \tag{4.7}$$

where ${}^{b}\mathbf{M}_{a}$ represents the transformation matrix from $\mathcal{F}_a$ to $\mathcal{F}_b$.

By inverting equation (4.6) we have:

$$ {}^{a}\bar{\mathbf{X}} = {}^{b}\mathbf{R}_{a}^{\top}{}^{b}\bar{\mathbf{X}} - {}^{b}\mathbf{R}_{a}^{\top}{}^{b}\mathbf{t}_{a} \tag{4.8} $$

and

$$ {}^{a}\mathbf{M}_{b} = \begin{pmatrix} {}^{b}\mathbf{R}_{a}^{\top} & -{}^{b}\mathbf{R}_{a}^{\top}{}^{b}\mathbf{t}_{a} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix} \tag{4.9} $$

As an example, in order to use equation (4.1) the original 3D point $\mathbf{X}$ needs to be expressed in the camera frame $\mathcal{F}_c$. However this point is generally known in the object frame $\mathcal{F}_o$. Thus point ${}^{o}\mathbf{X}$ needs to be transformed to ${}^{c}\mathbf{X}$ through the transformation matrix ${}^{c}\mathbf{M}_{o}$ as follows:

$$ \mathbf{x} = \mathbf{\Pi}{}^{c}\mathbf{X} = \mathbf{\Pi}{}^{c}\mathbf{M}_{o}{}^{o}\mathbf{X} \tag{4.10} $$

where $\mathbf{\Pi}$ is the projection matrix deduced from equation (4.1).

4.2.2 Epipolar geometry

In this section we consider two cameras c_1 and c_2. The projection of a point $\mathbf{X}$ in c_1 and c_2 are respectively $\mathbf{x}_1$ and $\mathbf{x}_2$ and the center of the cameras are respectively $\mathbf{C}_1$ and $\mathbf{C}_2$. Given a point in an image, the epipolar geometry enables finding its correspondence in the image of the second camera. The standard situation is depicted in Figure 4.5 [Hartley and Zisserman, 2003].

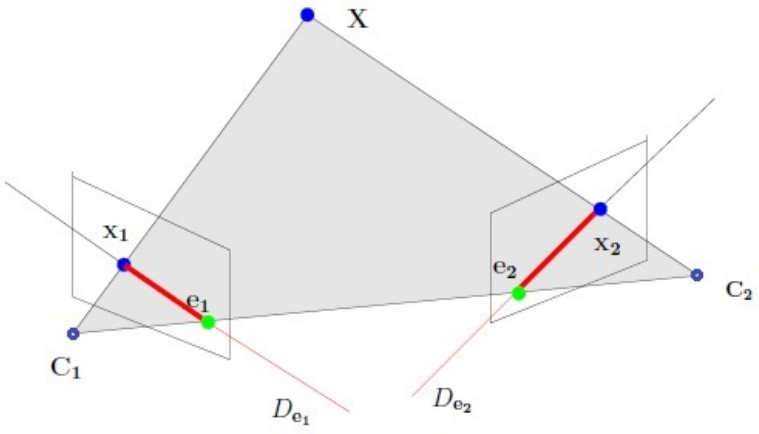


Figure 4.5 – Standard two-view situation

As depicted in Figure 4.5, if a 3D point $\mathbf{X}$ is projected on camera on the 2D point $\mathbf{x}_1$, then the corresponding 2D point $\mathbf{x}_2$ should be lying on a line called *epipolar line*. This line goes through the *epipole* ($\mathbf{e}_2$) which is the projection of the center of the first image in the second image (see section 4.3.2.2 for additional details).

4.2.2.1 Essential matrix

As we can see in Figure 4.5 the points $\mathbf{X}$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are coplanar. This coplanarity involves that:

$$\mathbf{C}_1\mathbf{x}_1.(\mathbf{C}_1\mathbf{C}_2 \times \mathbf{C}_2\mathbf{x}_2) = 0 \tag{4.11}$$

Developping equation (4.11) leads to the following relation [Ma et al., 2012] :

$$\mathbf{x}_1^\top [{}^{c_1}\mathbf{t}_{c_2}]_\times {}^{c_1}\mathbf{R}_{c_2}\mathbf{x}_2 = \mathbf{x}_1^\top {}^{c_1}\mathbf{E}_{c_2}\mathbf{x}_2 = 0 \tag{4.12}$$

where $[{}^{c_1}\mathbf{t}_{c_2}]_\times$ is the skew-symmetric matrix [Hartley and Zisserman, 2003] defined by the translation between camera c_1 and c_2 and ${}^{c_1}\mathbf{R}_{c_2}$ is the rotation associated. By definition the matrix ${}^{c_1}\mathbf{E}_{c_2}$ is called *essential matrix* and is given by:

$${}^{c_1}\mathbf{E}_{c_2} = [{}^{c_1}\mathbf{t}_{c_2}]_\times {}^{c_1}\mathbf{R}_{c_2} \tag{4.13}$$

By taking the transpose of equation (4.12) we have $\mathbf{x}_2^\top {}^{c_1}\mathbf{E}_{c_2}^\top \mathbf{x}_1 = 0$ where ${}^{c_1}\mathbf{E}_{c_2}^\top = {}^{c_2}\mathbf{E}_{c_1} = [{}^{c_2}\mathbf{t}_{c_1}]_\times {}^{c_2}\mathbf{R}_{c_1}$.

Note that the points use to define the essential matrix are taken in the respective camera frame (normalized coordinates) meaning that the calibration matrices $\mathbf{K}_1$ and $\mathbf{K}_2$, from c_1 and c_2 are known. If this is not the case the essential matrix is called *fundamental matrix*. It is generally written as ${}^{c_1}\mathbf{F}_{c_2}$ and verifies the epipolar constraint in a digitized space (pixel coordinates) as follows:

$$\mathbf{x}_1^\top {}^{c_1}\mathbf{E}_{c_2}\mathbf{x}_2 = \mathbf{u}_1^\top \mathbf{K}_{c_1}^{-\top} {}^{c_1}\mathbf{E}_{c_2}\mathbf{K}_{c_2}^{-1}\mathbf{u}_2 = \mathbf{u}_1^\top {}^{c_1}\mathbf{F}_{c_2}\mathbf{u}_2 = 0 \tag{4.14}$$

This epipolar constraint specifies the equation of the epipolar line where the correspondent point should lie. Indeed equation (4.12) constraints the point $\mathbf{x}_2$ to be on the line directed by the vector $\mathbf{x}_1^\top {}^{c_1}\mathbf{E}_{c_2}$.

4.2.2.2 Essential matrix estimation

As defined in the previous section, the essential matrix can be simply computed if the translation and rotation between the two views is known. However this is not always the case and in many application the essential matrix needs to be estimated in order to retrieve the relative position between the two cameras. Many algorithms in the literature propose linear and non-linear solutions for estimating the essential matrix. One of the first implemented algorithms was the 8-point algorithm [Longuet-Higgins, 1987]. This algorithm as been improved by Hartley et al. [1997] with normalization techniques but is still outperformed by the non linear approaches from Luong et al. [1993], Zhang [1998] or by the "Gold-Standard" approach that can be found in the text book from Hartley and Zisserman [2003].

8-point algorithm

The 8-point algorithm as been first referenced by Longuet-Higgins [1987] and estimates the essential matrix from several points. Considering a set of i corresponding points

$(\mathbf{x}_{1_i}, \mathbf{x}_{2_i})$, the epipolar constraint is satisfied for every i: $\mathbf{x}_{1_i}^{\top\, c_1}\mathbf{E}_{c_2}\mathbf{x}_{2_i} = 0$. This set of epipolar constraint equations can be rewritten as

$$\mathbf{A}\mathbf{e} = 0 \tag{4.15}$$

where $\mathbf{e}$ contains the element of the essential matrix

$$\mathbf{e} = (E_{11}\ E_{12}\ ...\ E_{32}\ E_{33})^{\top} \tag{4.16}$$

and $\mathbf{A}$ is a $n \times 9$ matrix depending on the observations

$$\mathbf{A} = \begin{pmatrix} & & & & & & & & \\ & & & & & & & & \\ x_{2_i}x_{1_i} & x_{2_i}y_{1_i} & x_{2_i} & y_{2_i}x_{1_i} & y_{2_i}y_{1_i} & y_{2_i} & x_{1_i} & y_{1_i} & 1 \\ & & & & & & & & \\ & & & & & & & & \end{pmatrix} \tag{4.17}$$

A solution to equation (4.15) can be found using the least-squares method and assuming that $||\mathbf{e}|| = 1$. However the essential matrix is of rank 2 meaning that it only has 2 non null singular values. Thus, the smallest single value of the Single Value Decomposition (SVD) of ${}^{c_1}\mathbf{E}_{c_2}$ should be set to 0.

Normalized 8-point algorithm

Hartley et al. [1997] proposed an improvement on the condition of the matrix $\mathbf{A}^{\top}\mathbf{A}$ which is used to find the least-squares solution of equation (4.15). They claim that these conditions should lead to improvements on the stability of the results. The normalization is made by transforming and scaling the input data before writing the equations to solve.

The normalization step suggests that the coordinates in each image are translated so that the centroid of the set of points is at the origin. Then the coordinates are scaled so that the average distance of a point $\mathbf{x}$ from the origin is $\sqrt{2}$, meaning that the average point has the coordinates $(1, 1, 1)^{\top}$

Non-linear approaches

Non linear approaches of the essential matrix estimation are referenced in many papers and proceed by minimizing either the distance from the 2D point to its corresponding epipolar line [Luong et al., 1993] or the re-projection error [Hartley et al., 1997].

A physically meaningful quantity to minimize should be something measured in the image plane because the extracted information (2D points) comes from an image. Such quantity can be the distance from a point $\mathbf{x}_i$ to its corresponding epipolar line $\mathbf{l}_i = (l_{i_1}, l_{i_2}, l_{i_3})$ as define by Zhang [1998] :

$$d(\mathbf{x}_i, l_i) = \frac{\mathbf{x}_i^{\top}\mathbf{l}_i}{\sqrt{l_{i_1}^2 + l_{i_2}^2}} \tag{4.18}$$

The criterion to minimize is then defined by the following formula for N points :

$$C = \sum_{i=1}^{N} d(\mathbf{x}_{1_i}, {}^{c_1}\mathbf{E}_{c_2}\mathbf{x}_{2_i})^2 + d(\mathbf{x}_{2_i}, {}^{c_1}\mathbf{E}_{c_2}^{\top}\mathbf{x}_{1_i})^2 \tag{4.19}$$

This problem can be solved using an iterative linear method or a non linear minimization in parameter space. Interested readers are invited to look at the work from Zhang [1998] for more details.

4.3 MonSterTrack: Increasing optical tracking workspace with hybrid stereo/monocular tracking

In this chapter we propose an approach for increasing the optical tracking workspace that is based on two methods. As a first method we introduce MonSterTrack ("MONocular and STEReo TRACKing"), a method that enables extending the workspace of optical tracking systems by occasionally switching to a monocular tracking mode when the stereo tracking is no longer available. The tracking takes two different paths depending on the condition "Stereo Available" (Figure 4.2). If stereo is available then the localization is performed using 3D-3D registration else 2D-3D registration is performed. As in many tracking devices, the system requires an off-line calibration process to determine the internal parameters of the cameras and their relative positions.

Monocular tracking is generally constrained to use an external motion capture system to precalibrate the markers [Tjaden et al., 2015] and to define a reference frame [Vogt et al., 2002]. Indeed information about the markers structure is required to perform the pose estimation. By using a stereo mode we are relieved of using external systems. The stereo mode enables reconstructing the targets' points and defining their structure. Moreover the reference frame $\mathcal{F}_w$ is set with a specific target that defines its X and Z-axes.

In the following we consider an optical system composed of two cameras (see Figure 4.6). Nevertheless everything can be transposed to multi-view systems [Hartley and Zisserman, 2003] composed of N sensors.

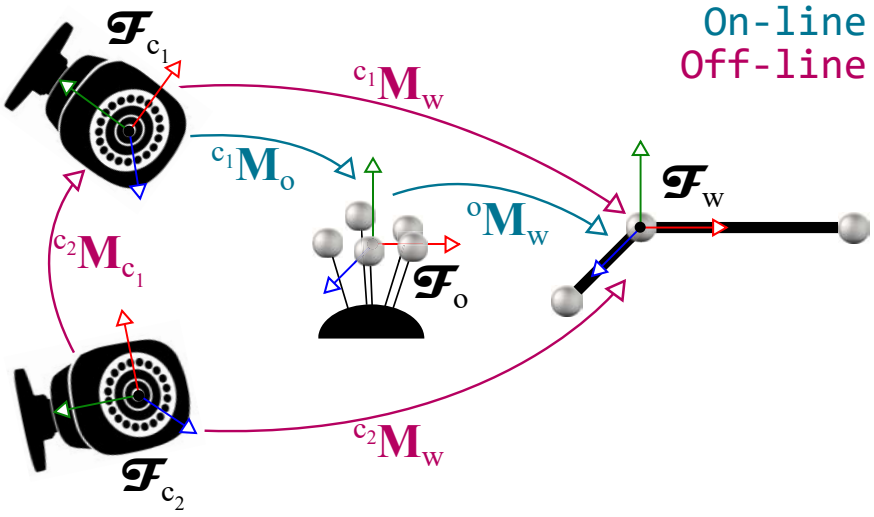


Figure 4.6 – Two-view optical tracking aims at recovering the pose ${}^{o}\mathbf{M}_w$ of an object ($\mathcal{F}_o$) in the reference frame $\mathcal{F}_w$. In VR $\mathcal{F}_w$ links the tracking coordinate system to the visual display coordinate system and therefore link the virtual and real environments.

4.3.1 Off-line system calibration

The calibration process aims at recovering the internal and external parameters of the system. The calibration is performed in two steps: internal camera calibration and calibration of the relative transformations between the cameras.

4.3.1.1 Internal calibration

The goal of internal calibration is to estimate the *internal parameters* (also called *intrinsic parameters*) of each camera (principal point, focal length and distortion coefficients). To perform the internal calibration, several points are needed and their coordinates in the object frame $\mathcal{F}_o$ need to be known. In practice we use a calibration chessboard that has several point which position are known within the pattern. Then the positions of the corners of each black square are extracted in the image. Moreover several views of the chessboard are captured (as depicted in Figure 4.7) to provide a more reliable estimation.

Figure 4.7 – During the internal camera calibration process, several views of a chessboard are captured and 2D features are extracted to compute the internal parameters of the cameras.

Once the different views are captured (hereafter indexed j), the following equation is solved:

$$(\widehat{\mathbf{q}}_j, \widehat{\mathbf{K}}) = \arg\min_{\mathbf{q},\mathbf{K}} \sum_{j=1}^{M} \sum_{i=1}^{N} d(\mathbf{x}_i, \mathbf{K}\mathbf{\Pi}\,{}^{c}\mathbf{M}_{o_j}\,{}^{o}\mathbf{X}_i)^2 \tag{4.20}$$

with $\mathbf{q}_j = ({}^{c}\mathbf{t}_{o_j}, \theta_j \mathbf{u}_j)^\top$ a minimal representation of ${}^{c}\mathbf{M}_{o_j}$ where θ_j is the angle and $\mathbf{u}_j$ the axis of rotation ${}^{c}\mathbf{R}_{o_j}$. $d(\,\cdot\,,\,\cdot\,)$ represents the Euclidean distance between two points in homogeneous coordinates. $N \times M$ represent the overall number of points captured in all the chessboard views.

A linear solution can be computed as the one proposed by Faugeras and Toscani [1986]. Then, the solution can be improved with a closed-form solution, like the one

from Zhang [2000]. However the system is non-linear and non-linear approaches are very popular in the literature [Brown, 1971]. They are even called "Gold-Standard methods" in the text book from Hartley and Zisserman [2003]. The non-linear solution of equation (4.20) can be found by using an iterative minimization method such as Gauss-Newton or Levenberg-Manquartd approaches. These non-linear approaches are detailed in section 4.3.2.4 and the Jacobian is given by Marchand and Chaumette [2002].

4.3.1.2 Calibration of the relative transformation between two cameras

The second step of the calibration process determines the relative transformation between each pair of cameras c_1 and c_2 (*external* or *extrinsic calibration*). This transformation, ${}^{c_1}\mathbf{M}_{c_2}$, is determined by both a rotation matrix ${}^{c_1}\mathbf{R}_{c_2}$ and a translation vector ${}^{c_1}\mathbf{t}_{c_2}$ (see section 4.2.1.4).

According to equation (4.13) both ${}^{c_1}\mathbf{R}_{c_2}$ and ${}^{c_1}\mathbf{t}_{c_2}$ are present in ${}^{c_1}\mathbf{E}_{c_2}$. Therefore, by estimating and decomposing ${}^{c_1}\mathbf{E}_{c_2}$ the relative pose between the cameras can be determined.

Estimation of the essential matrix. The essential matrix is estimated with a normalized 8-point algorithm with RANSAC refinement [Hartley and Zisserman, 2003]. As mentioned in section 4.2.2.2 at least 8 points visible from both views are required. In practice, more points are used. In our case we calibrated the external parameters of the system by moving a single marker across the workspace visible by both cameras (see Figure 4.8-left). By doing so, we acquire hundreds of images from both cameras where one point is present. This is equivalent to acquiring one image from both cameras where hundreds of points are present. The estimation of the matrix is carried out on all the images once enough images have been captured.

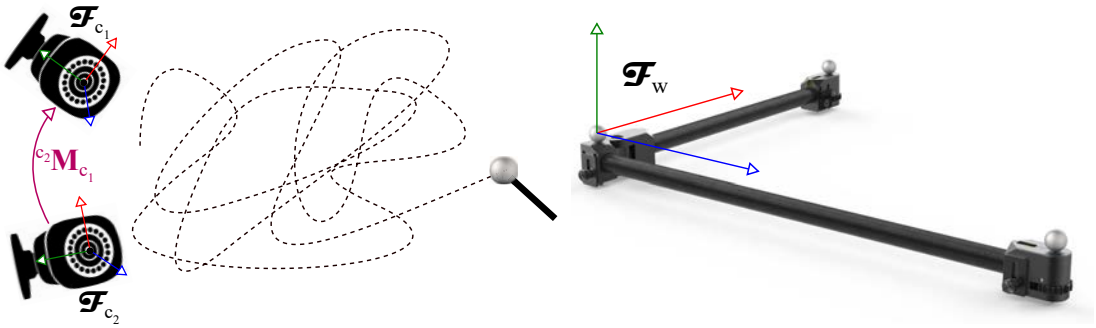


Figure 4.8 – Calibration of the relative transformation between two cameras. The transformation ${}^{c_2}\mathbf{M}_{c_2}$ relating both cameras is estimated by moving a single marker over the working volume (left) and the world reference frame $\mathcal{F}_w$ is given with a particular constellation (right).

Decomposition of the essential matrix. To decompose ${}^{c_1}\mathbf{E}_{c_2}$, a Single Value Decomposition (SVD) is carried out to have ${}^{c_1}\mathbf{E}_{c_2} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ where $\mathbf{\Sigma}$ is a diagonal matrix. Then ${}^{c_1}\mathbf{E}_{c_2}$ is rewritten ${}^{c_1}\mathbf{E}_{c_2} = \mathbf{SR}$ where $\mathbf{R}$ is a rotation matrix and $\mathbf{S}$ a skew-symmetric matrix. This is made possible by introducing two new matrices [Hart-

ley and Zisserman, 2003]:

$$\mathbf{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.21}$$

when using $\mathbf{W}$, $\mathbf{Z}$ and the SVD of ${}^{c_1}\mathbf{E}_{c_2}$, four solutions $(\mathbf{R}_i,\mathbf{t}_i)$ are found for ${}^{c_1}\mathbf{R}_{c_2}$ and ${}^{c_1}\mathbf{t}_{c_2}$. Among the four solutions, two will not be valid because the matrix $\mathbf{R}_i$ will not be a rotation matrix ($det(\mathbf{R}_i) = -1$). The two remaining solutions are then given by (see [Hartley and Zisserman, 2003]) :

$$\mathbf{S}_1 = -\mathbf{U}\mathbf{Z}\mathbf{U}^\top,\ \mathbf{R}_1 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top \quad \text{and} \quad \mathbf{S}_2 = \mathbf{U}\mathbf{Z}\mathbf{U}^\top,\ \mathbf{R}_2 = \mathbf{U}\mathbf{W}\mathbf{V}^\top \tag{4.22}$$

From these equations, the final solution for ${}^{c_1}\mathbf{t}_{c_2}$ can be found in the third row of $\mathbf{U}$ and can be scaled with a scalar factor without modifying the null-space condition. Finally, in order to select the valid rotation between $\mathbf{R}_1$ and $\mathbf{R}_2$, one condition is added: the visible point has to lie in front of both cameras. Only one rotation satisfies this condition.

Positioning the world reference frame. Finally, many VR systems require to have a reference world which is placed according to the VR display. This reference frame enables linking the tracking coordinate system to the virtual environment coordinate system. In this work we define the reference frame $\mathcal{F}_w$ by using a specific marker that defines the X and Z-axis of the frame (see Figure 4.8-right). Therefore every parameter of the system is calibrated and the pose ${}^{w}\mathbf{M}_{c_i}$ of each camera in the reference frame is available.

4.3.2 On-line real-time stereo tracking

The on-line real-time stereo tracking performs the localization of a target in the reference frame whenever the target is visible by both cameras. It first requires a 2D feature extraction to determine the position of the markers in the different images. Then the 2D features are correlated and triangulated to perform a 3D-3D registration.

4.3.2.1 Feature extraction

The feature extraction determines the position of the bright markers of the target on the different camera images. The images are captured in black and white and the sensors are equipped with infrared filters. The markers emit or are lighten with infrared light. Therefore the markers appear really bright in the images while the environment appears generally dark (see Figure 4.9-left). To extract the markers position in the image, first a binary threshold is applied to the overall image (see Figure 4.9-right). Then a recursive algorithm is used to find the different sets of connected bright pixels before computing the barycenter of each independent set. These barycenters define the centers of the light blobs. Once the markers positions are retrieved in the image, they corrected by taking into account the internal parameters of the cameras. The resulting positions are then undistorted by applying radial and tangential corrections (see section 4.2.1.3). Two distortion models were implemented to correct either standard or wide angle lenses.

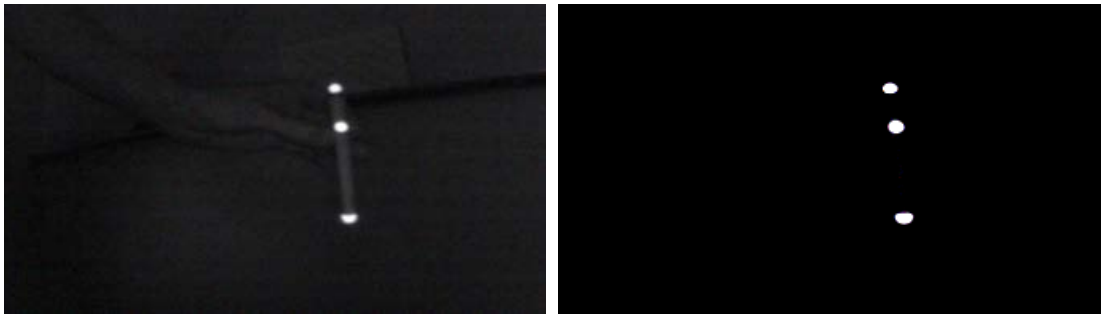


Figure 4.9 – The cameras capture the infrared light of the environment which appear dark (left) and threshold is applied to extract the objects that emit or reflect infrared light (right).

4.3.2.2 Feature correlation

The feature correlation associated the points from one image with their corresponding points in the other images. Matching the points from one image with the points in the other images is possible by using the epipolar constraint (see section 4.2.2) that states that two corresponding image points $\mathbf{x}_1$ and $\mathbf{x}_2$ (respectively from c_1 and c_2) in an epipolar configuration should fulfill:

$$\mathbf{x}_1^{\top {}^{c_1}}\mathbf{E}_{c_2}\mathbf{x}_2 = 0 \tag{4.23}$$

where ${}^{c_1}\mathbf{E}_{c_2}$ is the essential matrix computed as in section 4.2.2.2. Equation (4.12) constraints the point $\mathbf{x}_2$ to lie on the line directed by the vector $\mathbf{L}_1 = \mathbf{x}_1^{\top {}^{c_1}}\mathbf{E}_{c_2}$. Thus $\mathbf{L}_1$ is the epipolar line of $\mathbf{x}_1$ on the second image. Figure 4.10 depicts an epipolar configuration where the epipolar lines for each point are drawn over the camera planes.

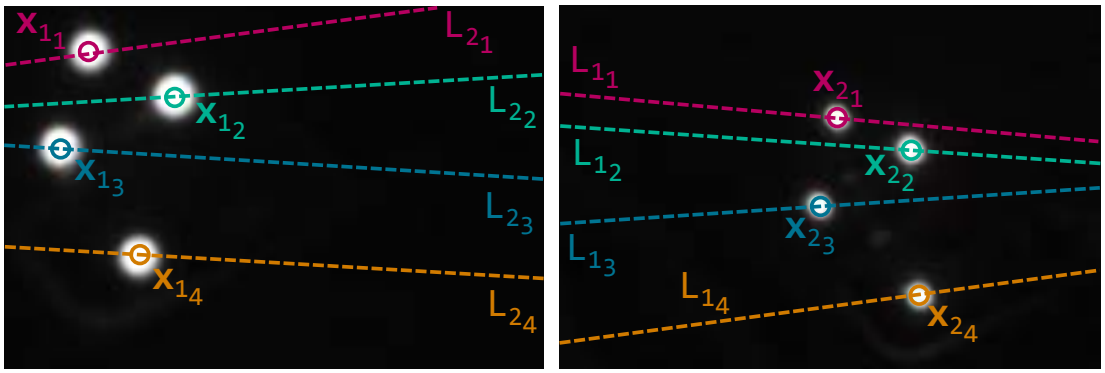


Figure 4.10 – Epipolar lines drawn for each blob visible in each camera frame. For each camera c_i and index j, the epipolar line corresponding to $\mathbf{x}_{i_j}$ is $\mathbf{L}_{i_j}$.

4.3.2.3 Triangulation

The problem of triangulation can be formulated as: *Given the projection* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *of a world point* $\mathbf{X}$ *respectively in camera* c_1 *and camera* c_2*, compute the 3D location of the world point* $\mathbf{X}$. This problems assumes that the transformation matrix ${}^{c_1}\mathbf{M}_{c_2}$ between the cameras is known. At first this problem seems straightforward to solve since the ray cast from the projected points should intersect in space.

Lets assume that we have two point $\mathbf{x}_1 = (x_1, y_1, 1)$ and $\mathbf{x}_2 = (x_2, y_2, 1)$. We want to find the coordinated ${}^{c_1}\mathbf{X}$ of point the 3D $\mathbf{X}$ in the camera frame $\mathcal{F}_{c_1}$. According to Longuet-Higgins [1987] we have:

$$Z_{c_1} = \frac{(\mathbf{r}_1 - x_2\mathbf{r}_3).{}^{c_1}\mathbf{t}_{c_2}}{(\mathbf{r}_1 - x_2\mathbf{r}_3).\mathbf{x}_1} \quad \text{and} \quad {}^{c_1}\mathbf{X} = \begin{pmatrix} x_1 Z_{c_1} \\ y_1 Z_{c_1} \\ Z_{c_1} \end{pmatrix} \tag{4.24}$$

where $\mathbf{r}_i$ is the i^{th} row of ${}^{c_1}\mathbf{R}_{c_2}$. However, there are errors in the measurements and the epipolar constraint is generally not be perfectly satisfied. Therefore the cast rays will generally not meet in space. These uncertainties can be due to essential matrix estimation errors, camera calibration errors or the noise that is present in the images.

To solve these accuracy issues, several triangulation algorithms have been tested and assessed in the literature survey from [Hartley and Sturm, 1997]. In our work we principally use the DLT algorithm [Hartley and Zisserman, 2003] as follows:

Considering the projection $\mathbf{x}_1$ and $\mathbf{x}_2$ of the 3D points $\mathbf{X}$ in camera c_1 and c_2, $\mathbf{x}_1$ and $\mathbf{x}_2$ can be written as:

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{\Pi}_1 {}^{c_1}\mathbf{M}_{c_1} {}^{c_1}\mathbf{X} = \mathbf{P}_1 {}^{c_1}\mathbf{X} \\ \text{and} \quad \mathbf{x}_2 &= \mathbf{\Pi}_2 {}^{c_2}\mathbf{M}_{c_1} {}^{c_1}\mathbf{X} = \mathbf{P}_2 {}^{c_1}\mathbf{X} \end{aligned} \tag{4.25}$$

where ${}^{c_1}\mathbf{X}$ denotes the coordinates of the 3D point $\mathbf{X}$ expressed in $\mathcal{F}_{c_1}$. From the previous equations we have:

$$\mathbf{x}_i \times (\mathbf{P}_i {}^{c_1}\mathbf{X}) = \mathbf{0}. \tag{4.26}$$

Therefore, by developing equation (4.26) and by denoting $\mathbf{p}_i^{j\top}$ the j^{th} row of $\mathbf{P}_i$

$$\begin{cases} x_i(\mathbf{p}_i^{3\top} {}^{c_1}\mathbf{X}) - (\mathbf{p}_i^{1\top} {}^{c_1}\mathbf{X}) = 0 \\ y_i(\mathbf{p}_i^{3\top} {}^{c_1}\mathbf{X}) - (\mathbf{p}_i^{2\top} {}^{c_1}\mathbf{X}) = 0 \\ x_i(\mathbf{p}_i^{2\top} {}^{c_1}\mathbf{X}) - y_i(\mathbf{p}_i^{1\top} {}^{c_1}\mathbf{X}) = 0 \end{cases} \tag{4.27}$$

where the two first equations are linearly independent. System equation (4.27) can be rewritten as:

$$\mathbf{A} {}^{c_1}\mathbf{X} = \mathbf{0} \quad \text{where} \quad \mathbf{A} = \begin{pmatrix} x_1\mathbf{p}_1^{3\top} - \mathbf{p}_1^{1\top} \\ y_1\mathbf{p}_1^{3\top} - \mathbf{p}_1^{2\top} \\ x_2\mathbf{p}_2^{3\top} - \mathbf{p}_2^{1\top} \\ y_2\mathbf{p}_2^{3\top} - \mathbf{p}_2^{2\top} \end{pmatrix} \tag{4.28}$$

The reconstructed 3D point ${}^{c_1}\mathbf{X}$ expressed in camera c_1 frame is obtained from equation (4.28) as the unit singular vector corresponding to the smallest singular value of $\mathbf{A}$.

4.3.2.4 Registration

The final step of real-time stereo tracking recovers the pose (position and orientation) of the target in the reference frame [Besl and McKay, 1992; Fitzgibbon, 2003; Marchand et al., 2016]. If the target is visible from several views then the registration matches a 3D point cloud to a 3D model.

When using a rigid constellation the position ${}^{o}\mathbf{X}_i$ of each marker i is known within the object frame $\mathcal{F}_o$. The 3D-3D registration estimates the transformation ${}^{c_1}\mathbf{M}_o$ that defines the pose of the tracked object in camera c_1 frame (see Figure 4.6). The estimation is achieved by minimizing the error between the 3D reconstructed points ${}^{c_1}\mathbf{X}_i$ (expressed in the camera frame) and their corresponding 3D points ${}^{o}\mathbf{X}_i$ (expressed in the object frame) transferred in the camera frame through ${}^{c_1}\mathbf{M}_o$. By denoting $\mathbf{q} = ({}^{c_1}\mathbf{t}_o, \theta\mathbf{u})^\top$ a minimal representation of ${}^{c_1}\mathbf{M}_o$ where θ is the angle and $\mathbf{u}$ the axis of ${}^{c_1}\mathbf{R}_o$, the problem is reformulated:

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}} \sum_{i=1}^{N} ({}^{c_1}\mathbf{X}_i - {}^{c_1}\mathbf{M}_o {}^{o}\mathbf{X}_i)^2. \tag{4.29}$$

The problem is solved by initializing the pose ${}^{c_1}\mathbf{M}_o$ with a linear solution [Arun et al., 1987]. Then the solution is refined with an iterative non-linear minimization such as the Gauss-Newton method which minimizes the cost given by:

$$\begin{array}{ll} & ||\mathbf{e}(\mathbf{q})|| = \mathbf{e}(\mathbf{q})^\top \mathbf{e}(\mathbf{q}) \\ \text{with} & \mathbf{e}(\mathbf{q}) = \mathbf{x}(\mathbf{q}) - \mathbf{x} \\ \text{where} & \mathbf{x}(\mathbf{q}) = (..., ({}^{c_1}\mathbf{R}_o | {}^{c_1}\mathbf{t}_o) {}^{o}\mathbf{X}_i, ...)^\top \\ \text{and} & \mathbf{x} = (..., {}^{c}\mathbf{X}_i, ...)^\top \end{array} \tag{4.30}$$

A first order Taylor expansion of $\mathbf{e}$ is given by:

$$\mathbf{e}(\mathbf{q} + \delta\mathbf{q}) \approx \mathbf{e}(\mathbf{q}) + \mathbf{J}(\mathbf{q})\delta\mathbf{q} \tag{4.31}$$

where $\mathbf{J}(\mathbf{q})$ is the Jacobian of $\mathbf{e}$ in $\mathbf{q}$. The minimization problem can be solved with an iterative least-squares approach and:

$$\delta\mathbf{q} = -\mathbf{J}(\mathbf{q})^{+}\mathbf{e}(\mathbf{q}) \tag{4.32}$$

where $\mathbf{J}^{+}$ is the pseudo inverse of the $3N \times 6$ Jacobian $\mathbf{J}$ given by:

$$\mathbf{J} = \begin{pmatrix} & \vdots & \\ -\mathbf{I}_{3\times3} & & [{}^{c_1}\mathbf{M}_o {}^{o}\mathbf{X}_i]_\times \\ & \vdots & \end{pmatrix} \tag{4.33}$$

Since the method is iterative, at each iteration $\mathbf{q}$ is updated as:

$$\mathbf{q}_{i+1} = \exp^{\delta\mathbf{q}} \mathbf{q}_i \tag{4.34}$$

where $\exp^{\delta\mathbf{q}}$ denotes the exponential map of $\delta\mathbf{q}$ [Ma et al., 2012].

A complete derivation of the problem, including the Jacobian derivation is given by Chaumette and Hutchinson [2006]. This method requires a good initialization of ${}^{c_1}\mathbf{M}_o$ in order to converge to the global minimum and can be initialized with linear solutions. The same optimization process can be used for both the essential matrix estimation and the internal calibration (see sections 4.2.2.2 and 4.3.1.1).

4.3.2.5 Transformation to reference frame

Once ${}^{c_1}\mathbf{M}_o$ is estimated, either with stereo or monocular registration, the pose ${}^{w}\mathbf{M}_o$ of the constellation in the world frame can be recovered with ${}^{w}\mathbf{M}_o = {}^{w}\mathbf{M}_{c_1}\,{}^{c_1}\mathbf{M}_o$ (Figure 4.6). Matrix ${}^{w}\mathbf{M}_{c_1}$ defines the pose of camera c_1 in the reference frame $\mathcal{F}_w$ and will vary with the controlled cameras. An additional calibration process will then be required (see section 4.4.1). Once the position of every object is registered in the world frame $\mathcal{F}_w$, the tracking data can be sent to the MR application. Note that when using AR systems, $\mathcal{F}_w$ is generally the same as $\mathcal{F}_{c_1}$. This is not the case for VR applications since $\mathcal{F}_w$ enables linking the tracking coordinate system to the virtual environment coordinate system.

4.3.2.6 Filtering

An optional low-pass filtering process is performed at the end of the registration. It may be useful for some applications in order to smooth the output of the registration process. Filtering helps reducing noise and preventing some drop outs but may add little latency. As an example we have implemented a predictive kalman filter [Bar-Shalom and Li, 1993] with a constant velocity state and position measurements model.

With a constant velocity and measured position model, the state vector at iteration k is given by:

$$\mathbf{x}_k = \begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix} \quad \text{with} \quad \begin{aligned} x_k &= x_{k-1} + (\Delta t)\dot{x}_{k-1} + \gamma \\ \dot{x}_k &= \dot{x}_{k-1} + \zeta \end{aligned} \tag{4.35}$$

where γ and ζ denote noise terms. Equation (4.35) can be rewritten as:

$$\mathbf{x}_k = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{k-1} \\ \dot{x}_{k-1} \end{pmatrix} + \begin{pmatrix} \gamma \\ \zeta \end{pmatrix} = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{w} \tag{4.36}$$

Regarding the measure, it is given at instant k by:

$$y_k = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix} + \eta = \mathbf{H}\mathbf{x}_k + \eta \tag{4.37}$$

where η denotes a noise term.

Note that every noise term is calculated thanks to the jitter evaluation that has been carried out on the system. This evaluation is presented in section 4.6.1.3.

The state vector is initialized at iteration 0 and 1 as follows:

$$\mathbf{x}_0 = \begin{pmatrix} y_0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{x}_1 = \begin{pmatrix} y_1 \\ (y_1 - y_0)/\Delta t \end{pmatrix} \tag{4.38}$$

This filtering model is applied independently to each parameter of the pose (3 for the translation and 3 for the rotation). However, the parameters of the pose are not independent since the rotation has an influence on the translation vector. Therefore filtering independently each parameter of the pose can not be theoretically done. A filtering process over the overall pose should then be considered, like the one used by Rambach et al. [2016]. But in practice, only a small difference will be perceived.

4.3.3 Monocular tracking mode

The monocular tracking mode is used to localize the target if it is visible by only one camera, camera c. Thus, the tracking is still carried out when the target is out of the field of view of almost all the cameras. Since the target is visible by only one camera, the feature correlation and triangulation steps can not be processed. Thus the localization is directly performed from the 2D extracted features and is called 2D-3D registration. During the 2D-3D registration, the transformation ${}^{c}\mathbf{M}_{o}$ between the camera frame and the object frame is estimated by using the 2D projected points $\mathbf{x}_i$ and their corresponding 3D target points ${}^{o}\mathbf{X}_i$.

The estimation of ${}^{c}\mathbf{M}_{o}$ can theoretically be represented by six independent parameters thus three points should be enough to have a solution. This problem is called Perspective from 3 Points (P3P). However with three points there are four possible solutions. It is then necessary to have four points [Hartley and Zisserman, 2003] (in our case we construct our rigid bodies with at least 4 markers). Quan and Lan [1999] extended the P3P problem to the P4P, P5P and eventually Perspective from n Points (PnP) problem. The solution of the P3P can be found in two steps:

1. Estimate the unknown depth ${}^{c}Z_i$ of each point in the camera frame. This is made by using the triangle $\mathbf{C}{}^{o}\mathbf{X}_i{}^{o}\mathbf{X}_j$, where $\mathbf{C}$ is the camera center. Since the real distance between ${}^{o}\mathbf{X}_i$ and ${}^{o}\mathbf{X}_j$ is known, as well as the directions $\mathbf{C}{}^{o}\mathbf{X}_i$ and $\mathbf{C}{}^{o}\mathbf{X}_j$, according to Quan and Lan [1999], the points depths can be estimated by solving a 4^{th} order polynomial equation.

2. Once the 3D points coordinates are known in the camera frame the rigid transformation that links the camera frame ($\mathcal{F}_c$) to the object frame ($\mathcal{F}_o$) is estimated. ${}^{c}\mathbf{M}_{o}$ is estimated by solving a least-squares equation with SVD [Ameller et al., 2000]. A close form solution proposed by Horn [1987] allows to retrieve the rotation represented by quaternions.

A more recent study from Kneip et al. [2011] proposes to directly compute the rigid transformation without the intermediate derivation of the points in the camera frame. This is achieved by adding an intermediate camera and world frame. This close-form solution is then faster to compute. Thus, by adding an outlier rejection process such as RANSAC, the fast P3P method [Kneip et al., 2011] seems to be a good candidate for pose estimation. RANSAC can also be used to solve the P3P in a iterative way as proposed by Fischler and Bolles [1981].

However pose accuracy increases with the number of points used in the estimation. Thus several studies have been made on PnP. For solving the PnP, one method can be to do like the P3P, by estimating the coordinates of the points in the camera frame and then achieve a 3D-3D registration. Other methods have been developed to perform the estimation in one step. DLT is one of the common linear methods performing un one step. Although this method is not very accurate and PnP is not a linear problem, a linear solution can be considered. The estimation is made for the 12 entries (4×3 matrix) of ${}^{c}\mathbf{M}_{o}$ by solving an equation $\mathbf{Ah} = 0$ where $\mathbf{A}$ contains the observations and $\mathbf{h}$ depends on the rotation and translation parameters [Sutherland, 1974]. The solution

of the system is given by the eigenvector of $\mathbf{A}$ corresponding to the smallest eigenvalue. Being over-parametrized, this solution is sensitive to noise.

An alternative method that takes into account the non-linearity of the problem has been proposed by Dementhon and Davis [1995], the POSIT method. With this method they propose to iteratively go back from the scaled orthographic projection model to the perspective one since the pose estimation is linear under the scaled orthographic projection model.

In our work we used, according to Hartley and Zisserman [2003] and Marchand et al. [2016], the "Gold-Standard" approach which is based on a non-linear estimation of the transformation ${}^{c}\mathbf{M}_{o}$. In this case the estimation problem can be written:

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}} \sum_{i=1}^{N} d(\mathbf{x}_i, \mathbf{\Pi}\, {}^{c}\mathbf{M}_{o}\, {}^{o}\mathbf{X}_i)^2 \tag{4.39}$$

where $\mathbf{\Pi}$ is the projection matrix, $\mathbf{q}$ is a minimal representation of ${}^{c}\mathbf{M}_{o}$ and $d(\mathbf{x}, \mathbf{x}')$ is the Euclidian distance between two points $\mathbf{x}$ and $\mathbf{x}'$. The problem is solved by initializing the pose ${}^{c}\mathbf{M}_{o}$ with a linear solution based on an EPnP approach [Lepetit et al., 2009] and refining with a non-linear Gauss-Newton estimation [Marchand et al., 2016] like the one presented in equation (4.30) but with:

$$\mathbf{x}(\mathbf{q}) = (..., \mathbf{\Pi}({}^{c_1}\mathbf{R}_{o} | {}^{c_1}\mathbf{t}_{o})\, {}^{o}\mathbf{X}_i, ...)^{\top} \quad \text{and} \quad \mathbf{x} = (..., \mathbf{x}_i, ...)^{\top} \tag{4.40}$$

In this case the $2N \times 6$ Jacobian $\mathbf{J}$ is given by, e.g., [Marchand and Chaumette, 2002]:

$$\mathbf{J} = \begin{pmatrix} & & & \vdots & & \\ -\frac{1}{Z_i} & 0 & \frac{x_i}{Z_i} & x_i y_i & -(1+x_i^2) & y_i \\ 0 & -\frac{1}{Z_i} & \frac{y_i}{Z_i} & 1+y_i^2 & -x_i y_i & -x_i \\ & & & \vdots & & \end{pmatrix} \tag{4.41}$$

Equation (4.39) provides several solutions when only three 3D points are considered. Thus we built the targets (constellations) (Figure 4.13) with at least four non-coplanar points so that the 2D-3D registration gives a unique solution. The registration algorithms presented above assume that the matching between the $\mathbf{x}_i$ and the ${}^{o}\mathbf{X}_i$ is known. In our implementation of the approach, the matching is carried out using brute force and minimizing equation (4.39) for all the combinations of points. Once ${}^{c}\mathbf{M}_{o}$ is estimated, the transformation to the reference frame and the filtering steps can be performed as for stereo tracking. Results of monocular tracking compared to stereo tracking are presented in section 4.6.

4.4 CoCaTrack: Increasing optical tracking workspace with controlled cameras

To further increase the workspace we propose a second method called Controlled Camera Tracking (CoCaTrack), that consists in moving the cameras. In that way, the cameras are able to track the movement of the target (constellation) and keep it in

their field of view. Even if only one camera can move, the stereo workspace increases. Moreover, the monocular workspace gained with MonSterTrack becomes even larger.

Controlled cameras have been used in video surveillance for a while to track a target. Our method uses similar techniques for visual servoing but, in our case, an additional difficulty is introduced. Indeed our method must perform 3D reconstruction and compute the 3D pose of the target in a reference frame.

A visual servoing process controls the camera so that the target projection is close to the image center. The automation is made through robots on which the cameras are fixed on. Using a camera mounted on a robot requires an off-line calibration process to determine the position of the camera frame, $\mathcal{F}_c$, in the robot's end-effector frame, $\mathcal{F}_e$, which is required to recover the position of the camera in the reference frame, $\mathcal{F}_w$, and perform pose estimation. In the following we consider a pair of camera c_1 and c_2 where c_1 is a controlled camera mounted on an actuator. Nevertheless the approach can be transposed as is to multiple controlled cameras.

4.4.1 Off-line controlled camera calibration

The controlled camera calibration process recovers the pose ${}^{e}\mathbf{M}_{c_1}$ of the camera in the end-effector frame of the robot [Tsai and Lenz, 1989] which is constant. In practice, when fixing a camera to a robot, matrix ${}^{e}\mathbf{M}_{c_1}$ needs to be known to compute the transformation ${}^{c_2}\mathbf{M}_{c_1(t)}$ that relates both cameras at each instant t. In fact when using controlled cameras, the calibration of the relative transformation between the cameras (section 4.3.1.2) needs to be performed on-line. The computation of ${}^{e}\mathbf{M}_{c}$ is carried out off-line.

Statement of the problem. When considering one controlled camera (in our case camera c_1) ${}^{c_2}\mathbf{M}_{c_1(t)}$ is computed as:

$$ {}^{c_2}\mathbf{M}_{c_1(t)} = {}^{c_2}\mathbf{M}_{w}\,{}^{w}\mathbf{M}_{c_1(0)}\,{}^{c_1(0)}\mathbf{M}_{e(0)}\,{}^{e(0)}\mathbf{M}_{e(t)}\,{}^{e(t)}\mathbf{M}_{c_1(t)} \tag{4.42} $$

where for each t

$$ {}^{c_1(t)}\mathbf{M}_{e(t)} = {}^{c_1(0)}\mathbf{M}_{e(0)} = {}^{c_1}\mathbf{M}_{e}. \tag{4.43} $$

${}^{c_2}\mathbf{M}_{w}$ is known by the previously made extrinsic calibration. Same goes for ${}^{w}\mathbf{M}_{c_1(0)}$ since the extrinsic calibration is made at $t = 0$. Matrix ${}^{e(0)}\mathbf{M}_{e(t)}$ which represents the transformation of the end-effector frame at instant t in the end-effector frame at instant $t = 0$ varies but is known by odometry measurements. Thus the only matrix that remains unknown in equation (4.42) is ${}^{e}\mathbf{M}_{c_1}$. Figure 4.11 illustrates the different frames that take part in equation (4.42) at instant $t = 0$ and justifies equation (4.43).

Estimation of ${}^{e}\mathbf{M}_{c_1}$. To obtain the unknown matrix ${}^{e}\mathbf{M}_{c_1}$ an additional calibration process is introduced. this calibration is performed off-line since matrix ${}^{e}\mathbf{M}_{c_1}$ is constant. To estimate ${}^{e}\mathbf{M}_{c_1}$ we used a stationary calibration chessboard and estimated its monocular pose for different positions of the robot's end-effector frame (In practice we used 4 positions). Figure 4.12 illustrates the calibration setup for 2 positions of the the robot's end-effector frame $e(t_0)$ and $e(t_1)$ at two distinct instants t_0 and t_1 that lead to 2 positions of camera c_1, respectively $c_1(t_0)$ and $c_1(t_1)$.

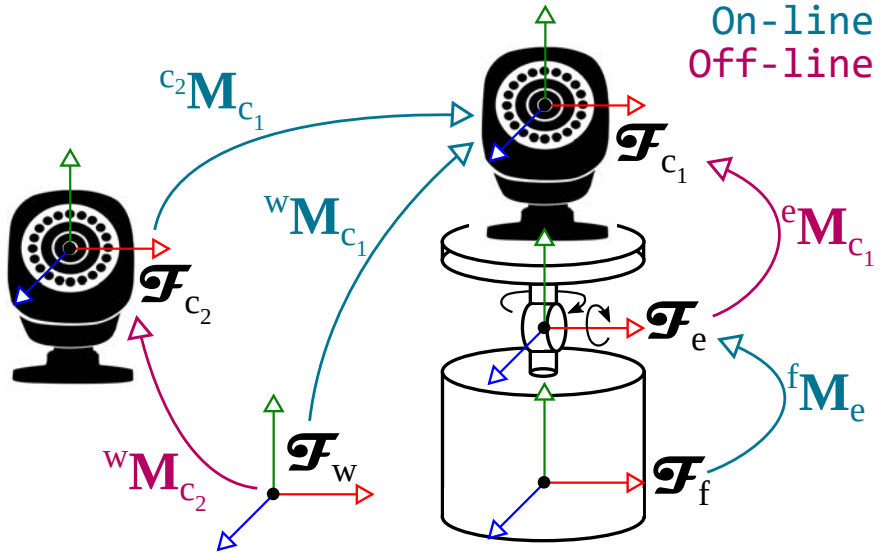


Figure 4.11 – Cameras configuration with two cameras and one pan-tilt head. The transformation matrices $\mathbf{M}$ are either estimated on-line or off-line.

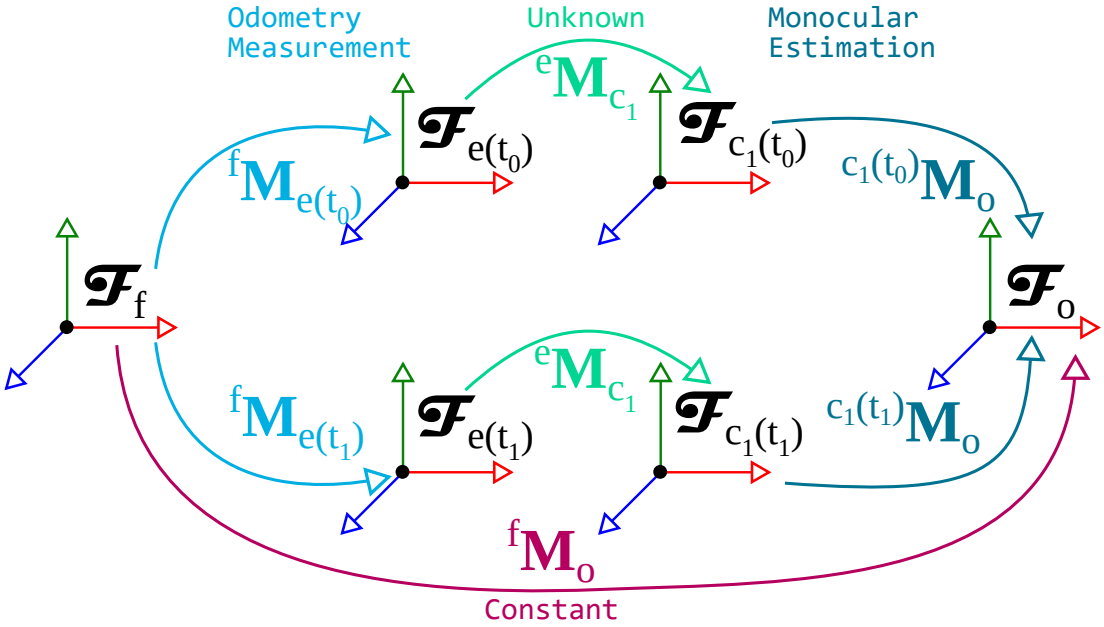


Figure 4.12 – Frame configuration for controlled camera calibration with 2 camera positions (in practice 4 positions were used). The only unknown matrix is $^{e}\mathbf{M}_{c_1}$ and its estimated thanks to the other matrices that are known.

Since the object frame, $\mathcal{F}_o$, and the robot reference frame, $\mathcal{F}_f$, are fixed $^{f}\mathbf{M}_{o}$ is constant and given by (see Figure 4.12):

$$^{f}\mathbf{M}_{o} = {}^{f}\mathbf{M}_{e(t_0)}\,{}^{e}\mathbf{M}_{c_1}\,{}^{c_1(t_0)}\mathbf{M}_{o} = {}^{f}\mathbf{M}_{e(t_1)}\,{}^{e}\mathbf{M}_{c_1}\,{}^{c_1(t_1)}\mathbf{M}_{o} \tag{4.44}$$

where for each instant t_i the transformation $^{f}\mathbf{M}_{e(t_i)}$ is given by the robot configuration (odometry measurements) and the transformation $^{c_1(t_i)}\mathbf{M}_{o}$ can be estimated through

monocular registration (see section 4.3.3). Therefore equation (4.44) needs to be solved to obtain ${}^{e}\mathbf{M}_{c_1}$.

From equation (4.44) the rotation and translation parts of the transformations can be separated to obtain two solvable equations [Tsai and Lenz, 1989]. For the rotation part, the following system has to be solved:

$$\mathbf{A}\,{}^{e}\mathbf{R}_{c_1} = {}^{e}\mathbf{R}_{c_1}\mathbf{B} \tag{4.45}$$

where $\mathbf{A}$ and $\mathbf{B}$ are rotation matrices computed from the measurements. For the translation, the system is the following:

$$\mathbf{A}\,{}^{e}\mathbf{t}_{c_1} = {}^{e}\mathbf{R}_{c_1}\mathbf{b} \tag{4.46}$$

where $\mathbf{A}$ and $\mathbf{b}$ are a matrix and a column vector computed from the measurements.

Equation (4.46) can be solved for ${}^{e}\mathbf{t}_{c_1}$ with a least-squares linear method once the solution ${}^{e}\mathbf{R}_{c_1}$ of equation (4.45) is found [Shiu and Ahmad, 1989]. Finding ${}^{e}\mathbf{R}_{c_1}$ involves converting equation (4.45) to a linear least-squares system by using a different representation of the rotation. For a rotation $\mathbf{R}$ of angle θ and unit axis $\mathbf{u}$, the vector $\mathbf{p}_R = 2sin(\theta/2)\mathbf{u}$ is defined and equation (4.45) can be rewritten as:

$$Skew(\mathbf{p}_A + \mathbf{p}_B)\mathbf{x} = \mathbf{p}_B - \mathbf{p}_A. \tag{4.47}$$

However $Skew(\mathbf{p}_A + \mathbf{p}_B)$ has rank 2 so at least 3 positions are required to solve the system. Finally the angle θ and the unit axis $\mathbf{u}$ can be extracted from $\mathbf{x}$ to recover ${}^{e}\mathbf{R}_{c_1}$. Once ${}^{e}\mathbf{R}_{c_1}$ is estimated, equation (4.46) can be solved [Tsai and Lenz, 1989] to recover ${}^{e}\mathbf{t}_{c_1}$.

Now that ${}^{e}\mathbf{M}_{c_1}$ is estimated, matrix ${}^{c_2}\mathbf{M}_{c_1(t)}$ can be computed by using equation (4.42) at each instant t even if camera c_1 moves. Then, since ${}^{c_2}\mathbf{M}_{c_1(t)}$ is known, all the steps of the tracking can be performed.

4.4.2 Registration

After computing ${}^{c_2}\mathbf{M}_{c_1(t)}$ on-line, the registration of the tracked targets can be carried out with either stereo or monocular registration algorithms:

- If stereo tacking is available, the stereo registration is performed as in section 4.3.2.4, by solving equation (4.29).
- Otherwise the monocular registration is performed as in section 4.3.3, by solving equation (4.39).

4.4.3 Controling camera displacements: visual servoing

To achieve the control of the cameras, we consider a visual servoing scheme [Chaumette and Hutchinson, 2006]. The goal of visual servoing is to control of the dynamic of a system by using visual information provided by one camera. The goal is to regulate an error defined in the image space to zero. This error, to be minimized, is based on visual features that correspond to geometric features. Here we consider the projection of the

center of gravity of the constellation $\mathbf{x} = (x, y)^\top$ that we want to see in the center of the image $\mathbf{x}^* = (0, 0)^\top$ (coordinates are expressed in normalized coordinates by taking into account the camera calibration parameters).

Considering the actual pose of the camera $\mathbf{r}$ the problem can therefore be written as an optimization process:

$$\widehat{\mathbf{r}} = \arg\min_{\mathbf{r}}((\mathbf{x}(\mathbf{r}) - \mathbf{x}^*)^\top(\mathbf{x}(\mathbf{r}) - \mathbf{x}^*)) \tag{4.48}$$

where $\widehat{\mathbf{r}}$ is the pose reached after the optimization process (servoing process). This visual servoing task is achieved by iteratively applying a velocity to the camera. This requires the knowledge of the interaction matrix $\mathbf{L_x}$ of $\mathbf{x(r)}$ that links the variation, $\dot{\mathbf{x}}(\mathbf{r})$, to the camera velocity and which is defined as:

$$\dot{\mathbf{x}}(\mathbf{r}) = \mathbf{L_x}\mathbf{v} \tag{4.49}$$

where $\mathbf{v}$ is the camera velocity (expressed in the camera frame). In the specific case of a pan-tilt camera that is considered in this chapter, $\mathbf{L_x}$ is given by[1]:

$$\mathbf{L_x} = \begin{pmatrix} xy & -(1+x^2) \\ 1+y^2 & -xy \end{pmatrix}. \tag{4.50}$$

This equation leads to the expression of the velocity that needs to be applied to the robot. The control law is classically given by:

$$\mathbf{v} = -\lambda \mathbf{L}_\mathbf{x}^+(\mathbf{x}(\mathbf{r}) - \mathbf{x}^*) \tag{4.51}$$

where λ is a positive scalar and $\mathbf{L}_\mathbf{x}^+$ is the pseudo-inverse of the interaction matrix. To compute, as usual, the velocity in the joint space of the robot, the control law is given by:

$$\dot{\mathbf{q}} = -\lambda \mathbf{J}_\mathbf{x}^+(\mathbf{x}(\mathbf{r}) - \mathbf{x}^*) \tag{4.52}$$

where $\dot{\mathbf{q}}$ is the robot joint velocity and where Jacobian $\mathbf{J_x}$ is given by Chaumette and Hutchinson [2006]; Marchand et al. [2005]:

$$\mathbf{J_x} = \mathbf{L_x}{}^c\mathbf{V}_e{}^e\mathbf{J}(\mathbf{q}) \tag{4.53}$$

${}^e\mathbf{J}(\mathbf{q})$ is the classical robot Jacobian expressed in the end effector frame (this Jacobian depends of the considered system). ${}^c\mathbf{V}_e$ is the spatial motion transform matrix [Chaumette and Hutchinson, 2006] from the camera frame to the end effector frame (computed using ${}^c\mathbf{M}_e$, see section 4.4.1). It is a constant matrix as soon as the camera is rigidly attached to the end effector.

The positive scalar λ is adapted as a function of the distance of the visual feature, $\mathbf{x}$, from the center of the image. We define a dead zone in the image where the velocity of the robot is set to zero whenever the visual feature is located in this zone. The dead zone is defined by an ellipse of center $\mathbf{x}^* = (0, 0)^\top$ with a long axis of length $w/8$ and a

[1] Note that the interaction matrix presented in equation (4.50) is defined for a pan-tilt system but the proposed method can scale to any kind of camera motions (up to 6 degrees of freedom).

short axis of length $h/8$, where w and h are, respectively, the width and height of the image. We choose a tanh function to adapt λ as follows:

$$\lambda(\mathbf{x}) = \alpha(\tanh(d(\mathbf{x}) + \beta) + 1) \tag{4.54}$$

where $d(\mathbf{x})$ defines the distance between $\mathbf{x}$ and the ellipse. The parameter α regulates the slope coefficient and β the slope position in the image.

Note that we considered that only one constellation was followed. If there are several constellations, one is free to define $\mathbf{x}$ as the barycenter of the points of both constellations or as the barycenter of a priority constellation.

4.5 Proofs of concept

We have designed two prototypes to test and illustrate our approach based on two different Virtual Reality (VR) setups: an holobench display and a wall-sized display.

Prototype 1: Holobench with MonSterTrack

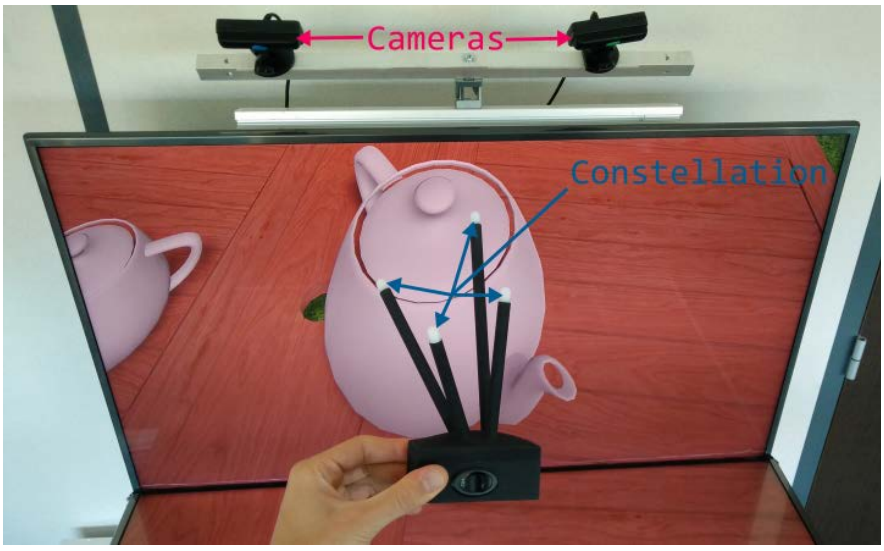


Figure 4.13 – First prototype based on an holobench display. Such a $75 tracking system hardware has been engineered and deployed at Realyz. Both cameras are Sony PSEye. An active constellation was built on purpose using a rigid structure and infrared emitting LEDs.

Our first prototype was designed within an holobench display (Figure 4.13). In such configuration the number of cameras and their positions are constrained. In order to make it suitable for SMBs, a low production cost was targeted, ending up with a minimal number of cameras ($n = 2$). The MonSterTrack method was implemented and enable covering a larger workspace and, in particular, the boundaries of the manipulation volume.

The tracking system was composed of two Sony PSEye cameras (providing 320x240 images) at a 150Hz refresh rate (Figure 4.13). The cameras were modified with short

focal length lenses (2.1mm) providing a final field-of-view of 87° by 70° each. An infrared band-pass filter was added to each lens. The constellations were designed on purpose with at least 4 non-coplanar active infrared LEDs and built on a 3D printed CAD rigid structure (Figure 4.13). Since active infrared markers were used, the cameras did not have to wear infrared LED rings. In order to make the light diffusion isotropic, a diffuser was added to each LED marker. Our setting has been deployed successfully at Realyz.

Prototype 2: Wall-sized display with MonSterTrack and CoCaTrack

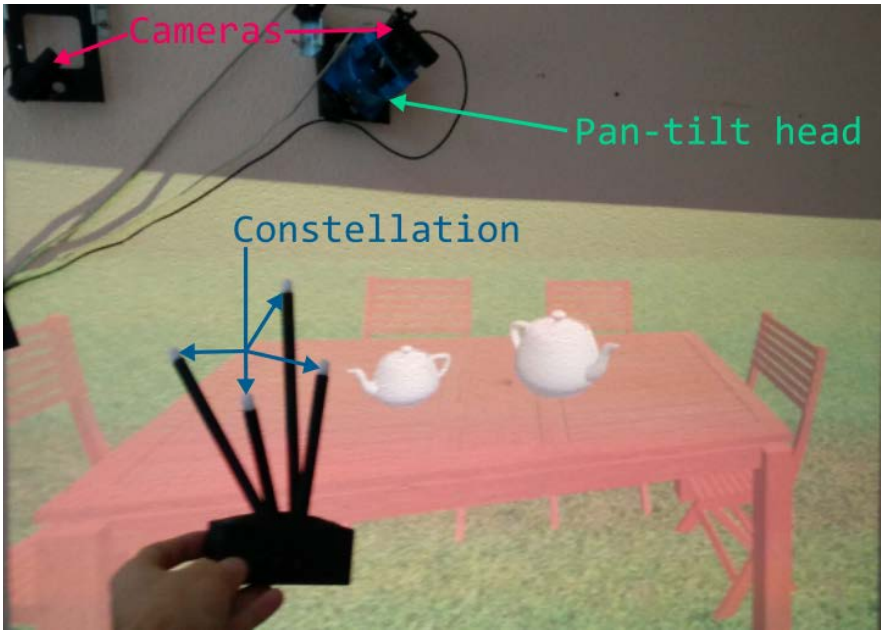


Figure 4.14 – Second prototype based on a wall-sized display. One of the two Sony PSEye cameras is embedded on a TracLabs Biclops pan-tilt head. A VR application is projected on the wall using stereo projection.

Our second prototype was designed within a wall-sized display (Figure 4.14). This prototype takes advantage from the combination of our two methods MonSterTrack and CoCaTrack. It relies again on two cameras, but one of them is a controlled camera that can follow the target across the VR workspace.

The system is still composed of two Sony PSEye camera. One camera is mounted on a TracLabs Biclops pan-tilt head (Figure 4.14). The pan-tilt head is controlled via a RS-232 connector with 115200 baud. The manufacturer framework is used to control the pan and tilt axes in velocity. The Biclops has two mechanical stops per axis allowing a range of rotation from $-170°$ to $+170°$ for the pan axis and from $-60°$ to $+60°$ for the tilt axis. The biclops is able to provide a rotation angle with a resolution of $0.03°$. The Traclabs Biclops pan-tilt head is very robust but relatively expensive. Cheaper pan-tilt actuators could be found in the market. They may not be made of aluminum and may not be usable in outdoor applications but they are cost-effective and they fulfill the requirements of many VR applications.

4.6 Performance

4.6.1 Main results of our global approach

The results presented in this section were obtained with our second prototype. The tests were run without filtering so that we could extract the exact jitter of the localization and its variation when switching tracking modes or when using controlled cameras.

4.6.1.1 Workspace gain

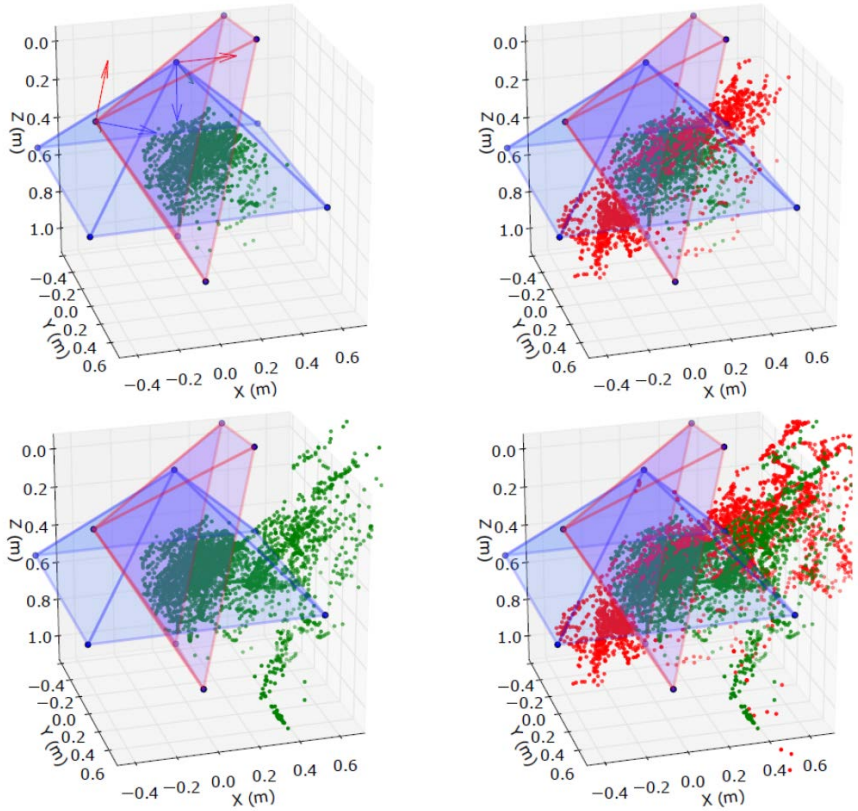


Figure 4.15 – Workspace gain compared to state of the art stereo optical tracking (top-left): our MonSterTrack method (top-right), our CoCaTrack method (bottom-left), and both methods at the same time (bottom-right). Points drawn in green were computed with the stereo tracking and the ones in red with the monocular one. The two pyramids illustrate the fields of view of the two cameras used by the system (blue: a controlled camera, red: a stationary camera).

We compared the optical tracking workspace of our approach with a state-of-the-art stereo tracking. To visualize the tracking workspace of the different solutions we computed tracking data through the entire workspace. First, we did it for state-of-the-art stereo tracking with two cameras (Figure 4.15-top-left). Then we computed tracking data using MonSterTrack. Thus the monocular tracking was active when the stereo was

not available. Several poses were computed with monocular tracking at both sides of the stereo space. In our case (Figure 4.15-top-right) there were almost as many stereo registrations as monocular registrations so we estimated a workspace gain of around 100%. A third test was to activate the controlled camera mounted on the pan-tilt actuator (blue cone in Figure 4.15) and compute stereo tracking data as depicted in Figure 4.15-bottom-left. Finally we merged our two methods and, by using the controlled camera and the monocular tracking, we obtained a far wider tracking workspace (Figure 4.15-bottom-right).

4.6.1.2 Calibration bias

The stereo mode was considered as ground truth and we compared the pose computed with the monocular mode and the controlled cameras to the pose given by the stereo one. In order to do that a constellation was placed at a stationary position (30cm of the cameras) in space and its stereo pose was logged. Then one camera was occluded in order to make the system switch to monocular mode and the monocular pose of the same constellation was logged. Figure 4.16-left illustrates the bias of the monocular mode which is of around 0.3cm. This bias can be mainly introduced by the internal calibration of the camera used to compute the tracking data (see section 4.3.1.1).

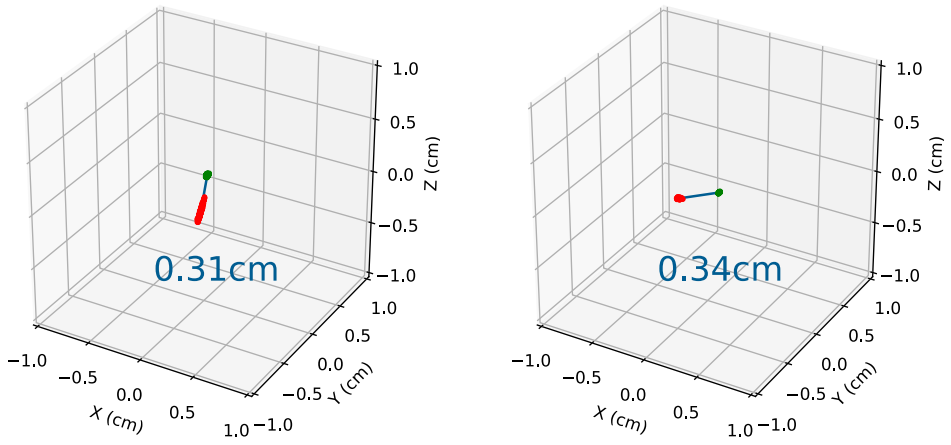


Figure 4.16 – Bias introduced by the monocular tracking mode (left) and by the calibration of the controlled camera (right). The green point cloud represents the ground truth and the red one represents the measurements with monocular tracking or controlled camera tracking. The blue distance represents the distance between the barycenters of the red and green point clouds.

An analogous test was made to quantify the controlled camera calibration bias. Indeed the controlled camera calibration has an impact on the position of the camera according to the reference frame and calibration errors may spread to the final tracking data. To estimate this bias a constellation was placed at a stationary position. This position was chosen so that the projection of the constellation in the controlled camera was on the right border of the camera frame. Thus, by activating the pan-tilt actuator the camera rotated toward the constellation. We computed the stereo pose for the initial position of the camera and for the final one. Figure 4.16-right illustrates the

bias introduced by the camera motion. This error is of around 0.3-0.4cm, due to the accuracy of the controlled camera calibration (see section 4.4.1).

4.6.1.3 Jitter

Jitter was measured using a protocol similar to one used by Pintaric and Kaufmann [2007]. A constellation was left at a stationary position and its pose was recorded during 600 measurements without filtering process. The constellation was placed at around 30cm of the cameras. Figure 4.17 illustrates the spatial distribution of the reconstructed constellation's position. The mean squared distance of the points from their mean-normalized center equals 0.08mm for stereo and 0.33mm for monocular. The 95% confidence radius of the distribution lies at 0.15mm for stereo and 0.76mm for monocular. Figure 4.18 illustrates the jitter in degrees of each rotation parameter of the computed poses.

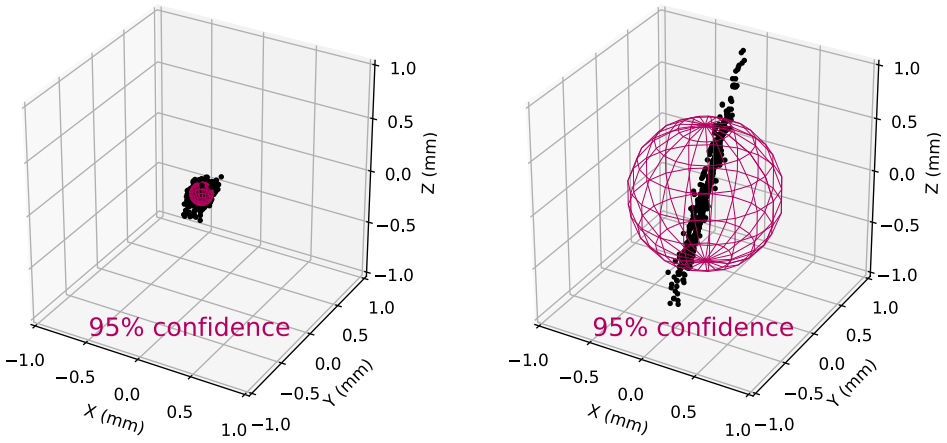


Figure 4.17 – Jitter of the localization computed with: stereo (left) and monocular (right) modes.

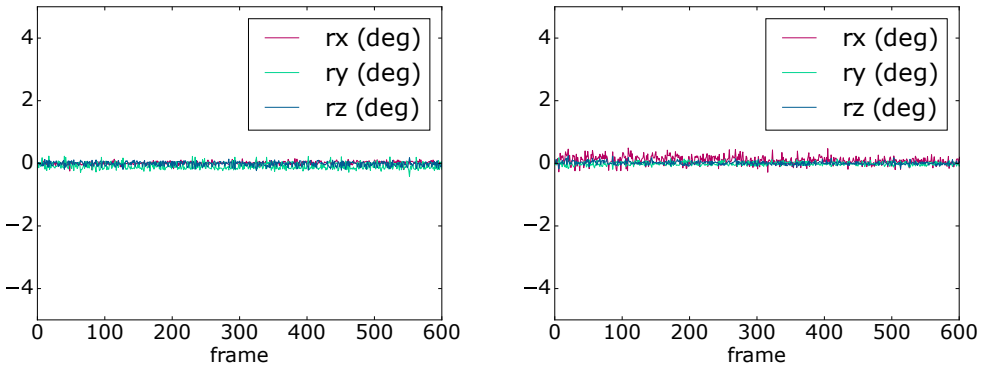


Figure 4.18 – Jitter of the rotation (in degree) computed with: stereo (left) and monocular (right) modes.

4.6.1.4 Transitions between monocular and stereo tracking

We analyzed the transition from one tracking mode to another by recording data for a random movement of the constellation translation of the constellation through the workspace. As depicted in Figure 4.19-left, the transition has a slight influence on the localization. Indeed, when changing the tracking mode, the localization can vary of up to 0.5cm. Figure 4.19-left also confirms the results on the jitter: monocular tracking is slightly more noisy than the stereo one. However Figure 4.19-right illustrates how a Kalman filtering process can be used to efficiently smooth the tracking data.

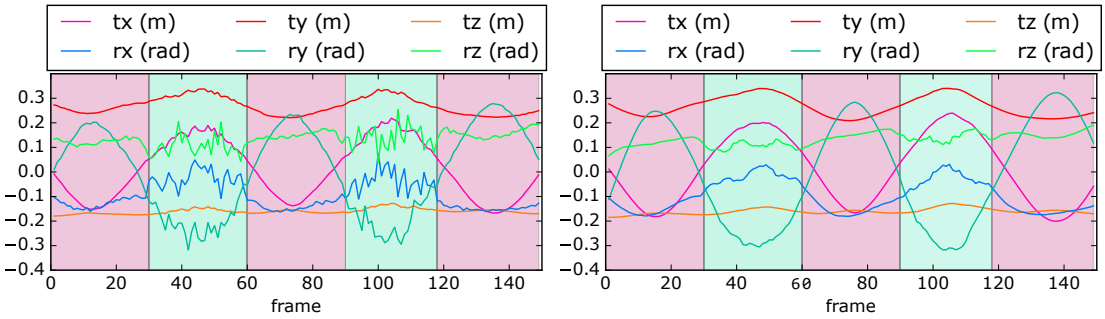


Figure 4.19 – Transitions from a stereo to a monocular tracking mode, and vice versa (green zone for monocular and red one for stereo) without filtering (left) and with a Kalman filter (right). A slight variation of the pose can be perceived in the data when changing the tracking mode. The filtering process absorbs this variation.

4.6.1.5 Latency

The VR application runs at 60Hz. Positions and orientations are provided to the application every 17ms. With a 150Hz refresh rate on the cameras, the internal latency of our current tracking software implementation is around 10ms. On the holobench setup (Figure 4.14), the end-to-end latency was measured around 50ms including rendering and display latency.

4.6.1.6 Summary

Our approach is based on two complementary methods which were implemented and tested on two prototypes. Using CoCaTrack together with MonSterTrack enabled to considerably increase the VR tracking workspace up to 100% and more. The implemented systems perform with ∼10ms internal latency, 50ms end-to-end latency on a specific VR use case, ±0.5cm accuracy and a jitter of 0.02cm for stereo and of 0.1cm for monocular.

4.6.2 Comparison with Vicon's optical tracking

Our tracking system was also installed in a room equipped with a Vicon optical tracking system composed of 8 Bonita 10 (1024×1024 at 250Hz) and 4 Vero v1.3 (1280×1024 at 250Hz) cameras. A Parrot AR.Drone 2.0 was tracked. Four non-coplanar active infrared LEDs were rigidly attached to the UAV (Figure 4.20). As active markers were

used, the cameras did not have to be equipped with infrared LED rings. A diffuser was added to each LED to provide isotropic light diffusion.

With such installation we were able to compare the performance of our approach with the Vicon's performance. Nevertheless we did not try to overtake Vicon. Our goal was to provide a large tracking volume at low cost. Indeed the Vicon installation costs around 150 000€ while our prototype costs around 3 000€. In the following a qualitative comparison between the two systems is introduced.

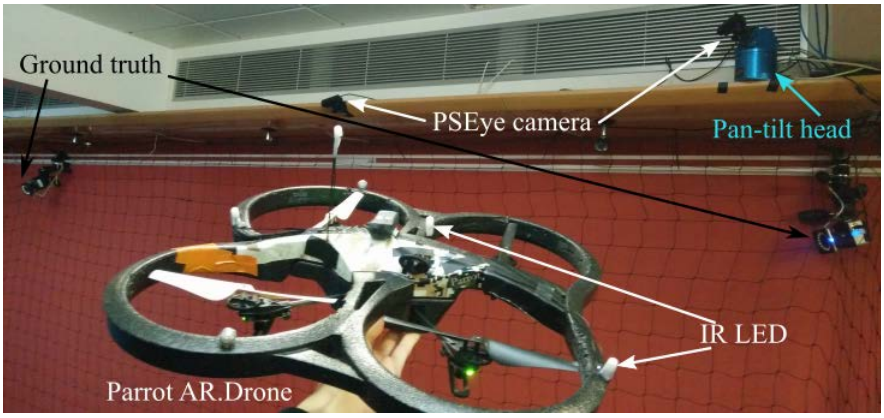


Figure 4.20 – Our optical tracking approach for UAV localization and for comparison with Vicon's optical tracking system.

Pose estimation

The pose of a flying UAV was estimated, at each instant, with our system and with the Vicon tracker. Figure 4.21-left illustrates the variations for the three components of the translation vector ${}^{w}\mathbf{t}_{o}$ while Figure 4.21-right illustrates the three components of the Rodrigues representation of the rotation matrix ${}^{w}\mathbf{R}_{o}$. The grey zones define the moments when our tracking was performed using MonSterTrack. At these instants the UAV was visible by only one of the cameras in our system. Since the calibration was made separately and the systems were not synchronized the error between both measurements is not of interest and a qualitative comparison of the pose is proposed.

Jitter

Jitter was measured by leaving the UAV at a stationary position and recording its pose during 7000 measurements without filtering process. The UAV was placed at around 2.5m of the cameras. Figure 4.22 illustrates the spatial distribution of the reconstructed positions. With Vicon measurements the 95% confidence radius of the distribution lies at 0.18mm. For our stereo tracking it lies at 0.86mm. The measurements with the monocular tracking are more noisy since the 95% confidence radius lies at 10.2mm. Nevertheless this noise is reduced when getting closer to the image frame of the camera. Some tests were carried out at 60cm of the camera and the 95% confidence radius lied at 1.4mm. These uncertainties are mainly oriented along the depth axis of the camera frame and are affected by the spatial resolution in the image.

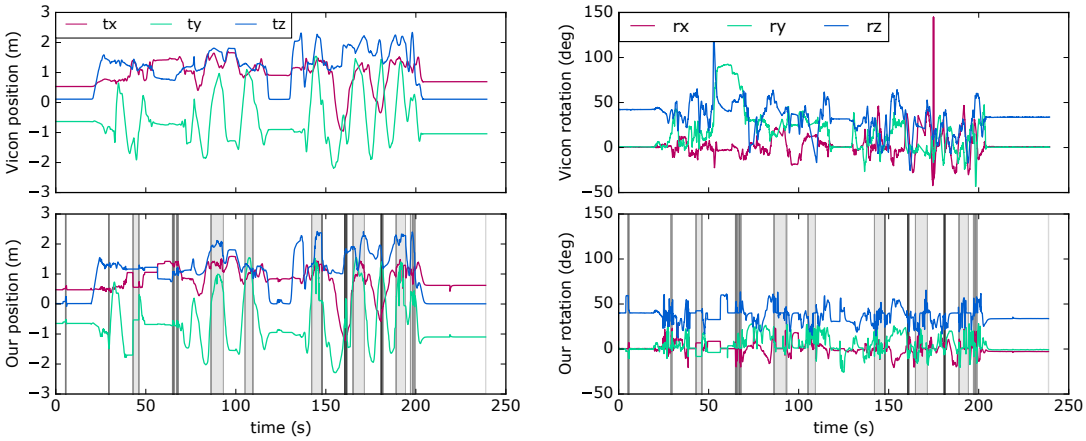


Figure 4.21 – Pose comparison with Vicon's optical tracking: Comparison of position components (left) and rotation components (right).

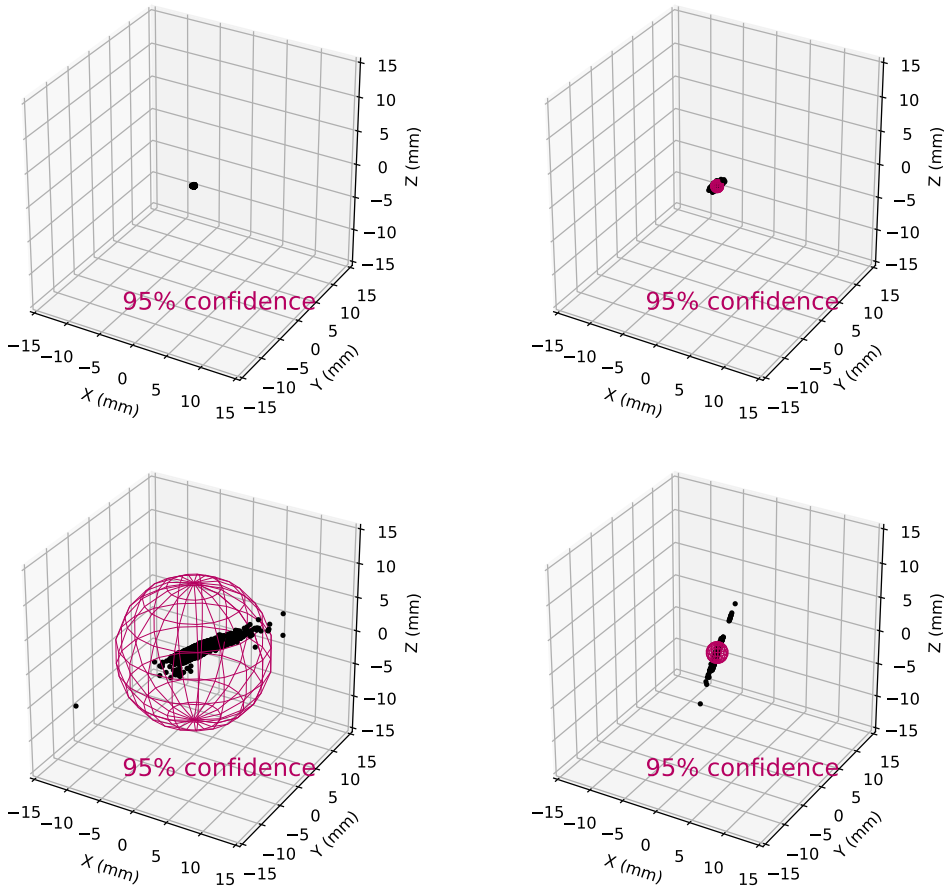


Figure 4.22 – Positional jitter of Vicon's optical tracking (top-left) compared to: our stereo tracking (top-right), our monocular tracking (bottom-left) and our close-range monocular tracking (bottom-right).

Summary

Considering the Vicon as ground truth, our system shows good performances for such a low-cost device. Nevertheless the calibration of both systems was made independently and they were not synchronized. Thus a quantitative comparison was unfortunately not possible. Since the Vicon tracker was composed of 12 high-resolution cameras (1MP) to cover all of the required volume it could be interesting to compare our performances with the ones of a Vicon tracker composed of at most 3 cameras.

4.7 Conclusion

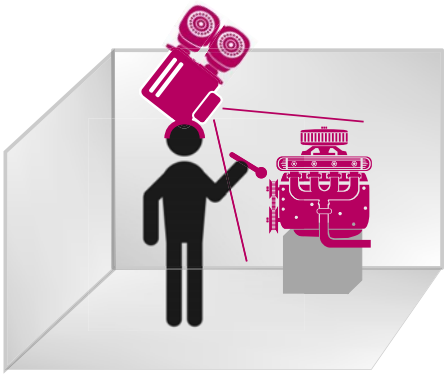
We proposed an approach that maximizes the workspace of optical tracking systems when using Projection-Based Systems (PBS). Our approach is based on two complementary methods.

As a first method, we proposed MonSterTrack, a method that enables to rely on a monocular registration when the stereo registration is no longer available. MonSterTrack can help providing a larger tracking workspace even with a small number of cameras. In fact when the targets are not visible by at least two cameras, the tracking is performed with monocular registration.

As a second method, we introduced CoCaTrack, a method that relies on controlled cameras. The controlled cameras are able to follow the target constellations through all the available workspace thanks to visual servoing. Therefore, CoCaTrack can bring more liberty when positioning the cameras in the Mixed Reality (MR) PBS. Also it enables performing stereo tracking in a larger workspace and can be combined with MonSterTrack to maximize the workspace.

We have designed two proofs of concept targeting different VR setups: a holobench and a wall-sized projection-based display. Our prototypes are based on two consumer-graded cameras and a pan-tilt head and implements both MonSterTrack and CoCaTrack. Our system has also been compared to a VICON tracking system in an Unmanned Aerial Vehicle (UAV) localization context. With both MonSterTrack and CoCaTrack the optical tracking workspace of PBS can be considerably increased, with up to 100% gain. The tracking is uninterrupted when the users move out of the stereo space. Some occlusion problems can be mitigated since a part of the constellation can be occluded from one camera but visible by another. Taken together our results suggest that our approach is affordable for Small and Medium Businesses with an interest in PBS for industrial applications.

Mobile spatial augmented reality for 3D interaction with tangible objects 5



Contents

5.1 The MoSART approach: Mobile spatial augmented reality on tangible objects . . . 93
5.2 Proof of concept . . . 94
5.2.1 Optical tracking . . . 96
5.2.2 Projection mapping . . . 97
5.2.3 Interaction tools . . . 99
5.2.4 Adding collaboration . . . 100
5.2.5 Characteristics and performances . . . 101
5.3 Use cases . . . 102
5.3.1 Virtual prototyping . . . 102
5.3.2 Medical visualization . . . 103
5.4 Discussion . . . 104
5.5 Conclusion . . . 106

Many industrial actors have adopted Augmented Reality (AR) Projection-Based Systems (PBS) [Bimber and Raskar, 2005] due to the possibility they offer in terms of direct collaboration. Indeed, compared to other AR technologies, such as Optical See-Through (OST) AR or Video See-Through (VST) AR, Spatial Augmented Reality (SAR) projects virtual content that can be directly shared with external persons. Also, SAR systems generally provide a larger field-of-view with a reduced latency. However SAR systems are mostly static (e.g., due to the use of a projector) which often restricts their usage when mobility is required.

In this chapter, we promote an alternative approach for head-mounted spatial augmented reality which enables mobile and direct 3D interactions with real tangible 3D objects, in single or collaborative scenarios. Our novel approach, called MoSART (for Mobile Spatial Augmented Reality on Tangible objects) is based on an "all-in-one" headset gathering all the necessary AR equipment (projection and tracking systems) together with a set of tangible objects and interaction tools (Figure 5.1). The tangible objects and tools are tracked thanks to an embedded feature-based optical tracking providing 6-DoF positioning data with low latency and high accuracy. The user can walk around, grasp and manipulate the tangible objects and tools augmented thanks to projection mapping techniques. Collaborative experiences can be shared with other users thanks to direct projection/interaction. In a nutshell, our approach is the first one which enables direct 3D interaction on tangible objects, mobility, multi-user experiences, in addition to a wider field-of-view and low latency in AR.

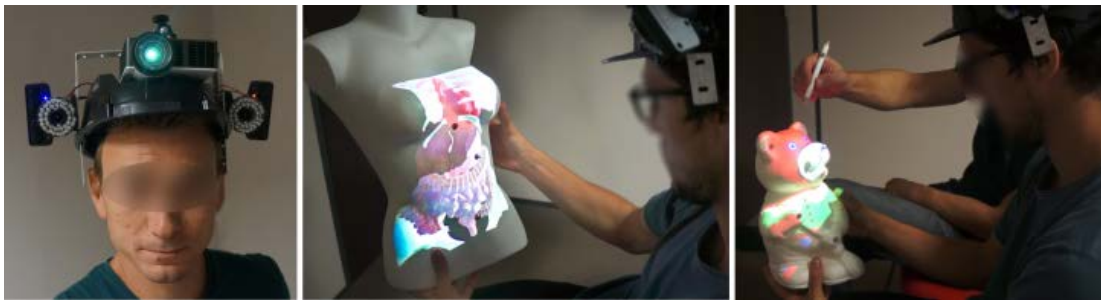


Figure 5.1 – Our MoSART approach enables Mobile Spatial Augmented Reality on Tangible objects. MoSART relies on an "all-in-one" Head-Mounted-Display (left). The users can walk around and manipulate tangible objects superimposed with the projected images (center). Tangible tools can also be used to interact with the virtual content (right).

To summarize, the main contributions of this chapter are:

- *MoSART*, a novel approach for spatial augmented reality that can simultaneously enable: mobile SAR on movable objects, 3D interactions with tangible objects, single and/or collaborative scenarios and experience sharing in AR applications.
- An operational prototype of the MoSART concept based on: 1) a novel all-in-one headset gathering head-mounted projection and optical tracking, and 2) a set of tangible objects and interaction tools.
- Several use cases that illustrate the potential of MoSART in different application contexts such as virtual prototyping and medical visualization.

In the remainder of this chapter we first describe the MoSART concept for mobile spatial augmented reality on tangible objects. Second, we present a proof of concept with the design of a MoSART prototype, and we assess its main characteristics and performances. Third, we propose two use cases of MoSART for virtual prototyping and medical visualization purposes. The chapter ends with a discussion and a general conclusion.

5.1 The MoSART approach: Mobile spatial augmented reality on tangible objects

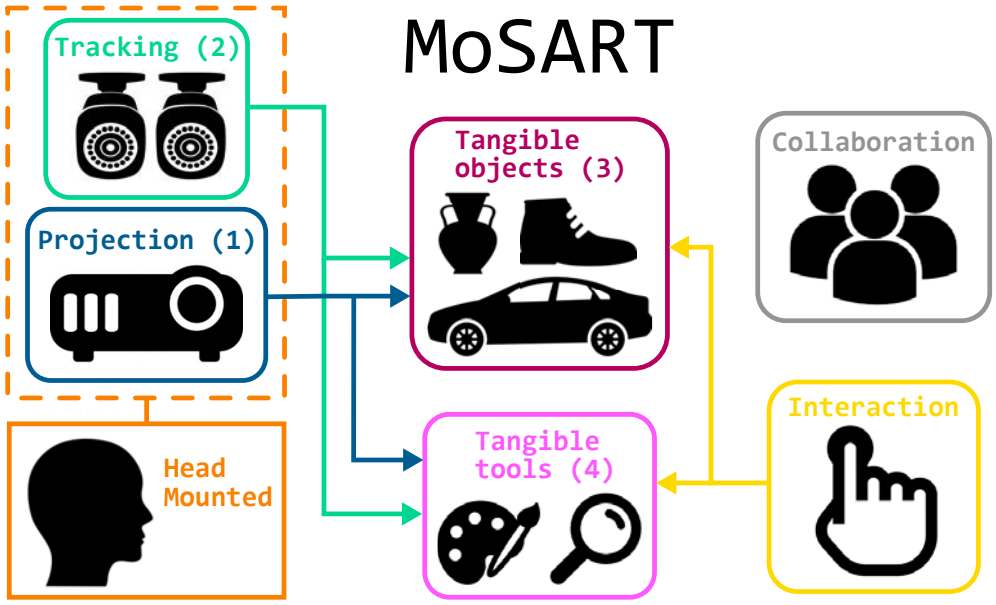


Figure 5.2 – Main components of the MoSART approach. MoSART involves head-mounted projection (1) and tracking (2). Direct 3D interactions are made possible with the tangible objects (3) and tools (4). Collaboration and multi-user scenarios can be addressed with or without additional headset(s).

The MoSART approach is a novel approach for mobile spatial augmented reality on tangible objects. MoSART enables mobile interactions with tangible objects by means of head-mounted projection and tracking. MoSART also allows to straightforwardly and directly manipulate 3D tangible objects and interact with them using dedicated interaction tools. It also makes it possible to share the experience with other users in collaborative scenarios.

The main components of the MoSART system are: (1) a head-mounted projection, (2) a head-mounted tracking, (3) tangible object(s), (4) several interaction tools. These main components are illustrated in Figure 5.2 and explained hereafter:

1. **Projection:** Head-mounted projection is used by MoSART to display the virtual content in the field-of-view and workspace of the user in a direct and unobtrusive way allowing to augment the objects located in the user's field of view. This also

implies that projection mapping techniques are required to match the 3D surface of the tangible object with the virtual content.

2. **Tracking:** Head-mounted tracking is used to follow the tangible objects and enable their manipulation in the workspace/FoV of MoSART. It enables users to walk and move around the objects, manipulate (rotate/translate) them at will. This naturally implies that the projector must be intrinsically tracked by the system.

3. **Tangible objects:** The use of 3D tangible objects is at the core of the MoSART approach. Thus the approach requires having both a physical model and a 3D virtual model of the object the users are interacting with.

4. **Interaction tools:** Tangible tools can also be incorporated straightforwardly within MoSART. Such tangible tools can benefit from the projection and tracking features of the system. This means that the tool surface can be used to project virtual content, and that the tools need to remain inside the projection/tracking volume. This also implies that dedicated 3D interaction techniques and metaphors need to be designed for every tool.

- **Head-Mounted:** To free the hands and provide an entire mobility to the user, all the projection and tracking features are mounted on the head.

- **Direct interaction:** Direct 3D interaction is a main advantage of MoSART thanks to the use of tangible objects. With MoSART the users can grasp tangible objects, and then manipulate (rotate/translate) them at will, within the field of view of the projector.

- **Collaboration:** Collaboration and multi-user interactions are a main advantage of MoSART. Two complementary collaborative modes are made possible. First, if there is only one user equipped with a MoSART headset (single-device configuration), MoSART allows other user(s) to share the direct projection controlled by the main user. The other users can also manipulate the tangible object(s) and/or some interaction tool. Second, if other headsets are available (multiple-devices configuration), the different projectors can be used to increase the projection area having for instance one user projecting on one side of the tangible object, and another user projecting on another side.

A prototype of the MoSART concept is introduced in the following section, and implementation details are provided regarding each MoSART component.

5.2 Proof of concept

We have designed a proof of concept of the MoSART approach. Our prototype includes a headset (Figure 5.3) and a specific set of tangible objects (Figure 5.4) and tangible tools (Figure 5.7), coming with dedicated 3D interaction techniques.

Our headset is composed of one short throw pico-projector (Optoma ML750ST) and two infrared cameras (PSEye). The cameras are rigidly attached on both sides of

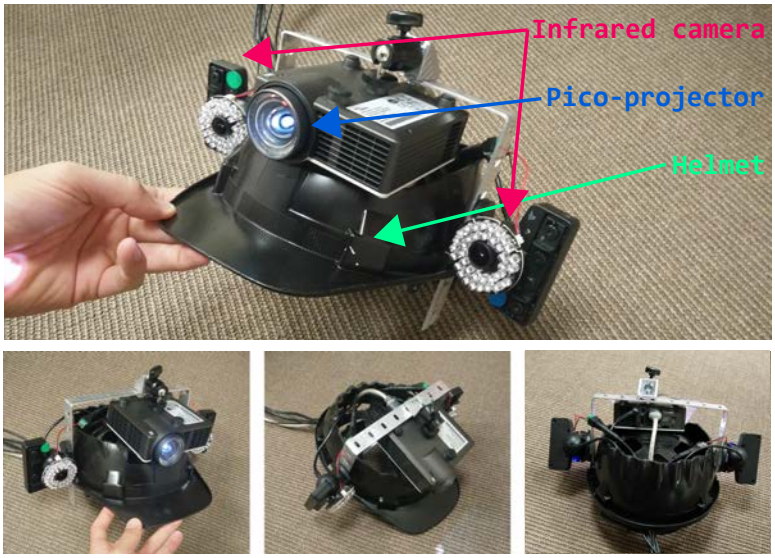


Figure 5.3 – Prototype of MoSART headset. The headset is composed of a pico-projector (for projection mapping) and two infrared cameras (for optical tracking).

Figure 5.4 – Example of tangible objects augmented with MoSART. The objects are white or covered with white painting. Reflective markers are positioned over the objects to facilitate their localization.

the projector. The whole projection/tracking system is mounted on the head and it is positioned so that the projection remains centered in the user's vision.

The projector is used to perform projection mapping (see section 5.2.2) on the tangible objects that are tracked with the optical tracking system (see Figure 5.3). The cameras are used to provide 6-DOF tracking data of the tangible objects thanks to a feature-based stereo optical tracking based on the one presented in chapter 4. An off-line initial calibration step is required to estimate the position and orientation of the projector with respect to the cameras. Such a configuration (projector and tracking

system attached) allows the system to be moved around the scene. The system does not need to track the projector localization anymore since the relative cameras/projector transformation is constant.

The tangible objects are ideally (but not necessary) white or covered with a reflective white paint coating, allowing to provide better results in terms of image color and contrast when projecting over the object. Several reflective markers (commonly 4 or 5) are positioned at the surface of every tangible object (see Figure 5.4), and are used to track and localize it using the optical tracking system.

5.2.1 Optical tracking

The tracking system mounted on the helmet is used to localize the tangible objects and interaction tools. The objective is to compute the position and orientation of the objects according to the projector. The system computes tracking data from the video streams provided by the two infrared cameras and it relies on feature-based stereo optical tracking. Feature-based optical tracking provides generally better performances than model-based tracking techniques in terms of accuracy and jitter. Localizing a rigid structure of markers (constellation) can be done generally faster than localizing a model. Moreover tracking several objects can be straightforwardly achieved by using different constellations for different objects. Also, using markers makes the tracking independent of the geometry of the objects, only the markers' disposition matters. Nevertheless it requires to add physical markers all over the tangible objects. To be able to localize a constellation it requires to have at least 3 markers (typically 4 or 5) and the distances between them have to be all different [Pintaric and Kaufmann, 2007]. Such constellation configuration reduces the ambiguities when computing the 3D registration to recover the pose of the objects (see section 4.3.3).

The tracking process is performed with both off-line and on-line steps. Optical tracking systems usually require an offline calibration process. Such calibration estimates the relative position ${}^{c_1}\mathbf{M}_{c_2}$ between camera c_1 and camera c_2 (see Figure 5.6) in order to be able to correlate the visual features in each view and to recover the 3D position of each reflective marker (see section 4.3.1.2). It also estimates the camera internal parameters and distortion coefficients (see section 4.3.1.1). Once the tracking system is calibrated four main online steps are performed to provide 6-DOF localization of the object as presented in section 4.3.2:

1. The **features extraction** determines the position of the bright markers in the images acquired by the two cameras.
2. The **features correlation** is performed thanks to the off-line calibration: the points from one image are associated with their corresponding points in the other images.
3. The **triangulation** process recovers the 3D points coordinates. The computation of the 3D coordinates is derived from the projections of the 3D point in the two image planes knowing the calibration parameters of the system.
4. The **3D registration** estimates the transformation ${}^{c_1}\mathbf{M}_o$ that defines the pose of the object in one of the camera frames (in our case c_1).

The tracking of several tangible objects can be performed. Each object pose is sent to update a virtual scene. Thus, this virtual scene matches the real environments and the projected image can be rendered (see Section 5.2.2.3).

The tracking system of the MoSART prototype was built based on the tracking system presented in chapter 4. It is composed of two Sony PsEye cameras providing 320×240 images at 150Hz. Infrared rings and infrared filters have been added to the cameras to capture the infrared light reflected by the reflective markers in order to ease the features extraction process.

5.2.2 Projection mapping

The projection mapping consists in mapping the virtual 2D image of a tangible object to the physical model on the same objects. To achieve this goal the application needs to have full knowledge of the object's shape and of the projection model. The projection model determines how a 3D point is projected into the image frame (3D-2D projection). In this case the projection model of the projector needs to be known to perform an "inverse-projection" (2D-3D projection). An off-line calibration process is used to determine such projection model. Regarding the object's shape, an off-line process is performed to scan and reconstruct the tangible object and incorporate its model in the application. Once both side of the system are known a virtual 3D scene can be generated to exactly fit the real scene. Figure 5.5 summarizes the projection mapping pipeline. The different steps of this pipeline, namely: ***virtual object creation***, ***projector calibration*** and ***virtual scene generation***, are detailed in the following.

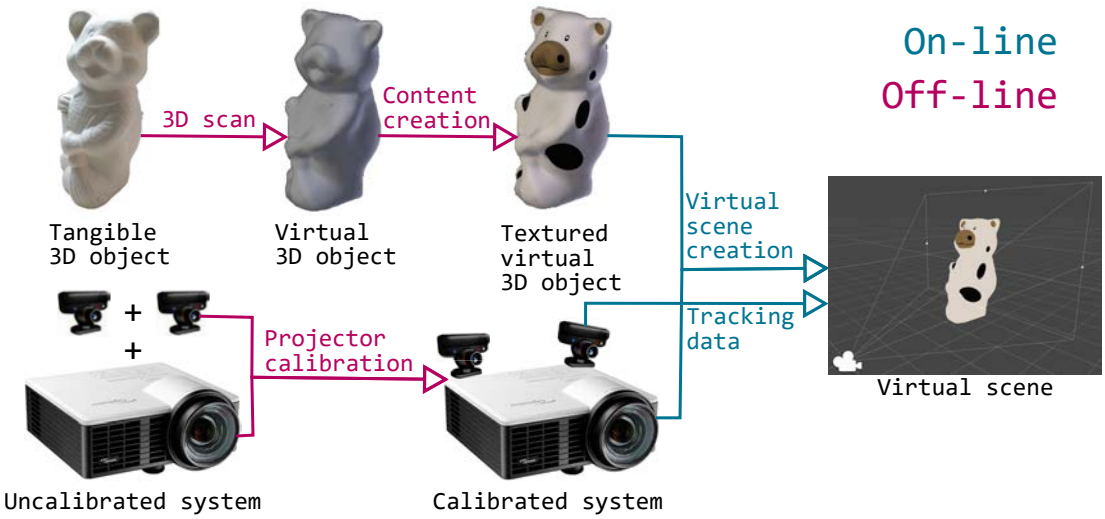


Figure 5.5 – Projection mapping pipeline. The object part (top) is responsible of the virtual object creation. The projection part (bottom) is responsible of the projector calibration. A virtual scene is created on-line to match the real environment.

5.2.2.1 Virtual object creation

A first requirement for projection-based augmented reality on tangible objects is to have access to the 3D model of the objects. This model is obtained with a 3D scanning technique. A structured light (infrared) depth sensor provides a depth map of the environment where the object is located and by moving the sensor around the object several maps are captured. An Iterative Closest Point (ICP) algorithm is used to match each dense map to the global model. Then the depth maps are fused to build a model of the tangible object (e.g., [Newcombe et al., 2011]). The 3D model of the object has its own coordinate system that we call virtual object frame ($\mathcal{F}_{vo}$). The physical objects has no predefined frame but its frame, $\mathcal{F}_o$, is defined by the tracking system. For a matching between the virtual scene and the real scene frame $\mathcal{F}_{vo}$ and frame $\mathcal{F}_o$ need to be the same. Thus another ICP algorithm is used to match at least four points of the virtual model to the same four points in the physical model. Once this matching process is carried out the transformation between $\mathcal{F}_{vo}$ and $\mathcal{F}_o$ is known and remains constant.

5.2.2.2 Projector calibration

The projector calibration is one of the most sensitive steps of projection mapping. Indeed the projection model and the projector pose need to be accurately known to ensure an acceptable mapping. This calibration is carried out once for each system so it needs to be as accurate as possible.

Figure 5.6 illustrates the different frames ($\mathcal{F}$) and transformations ($\mathbf{M}$) that take part in the MoSART calibration process.

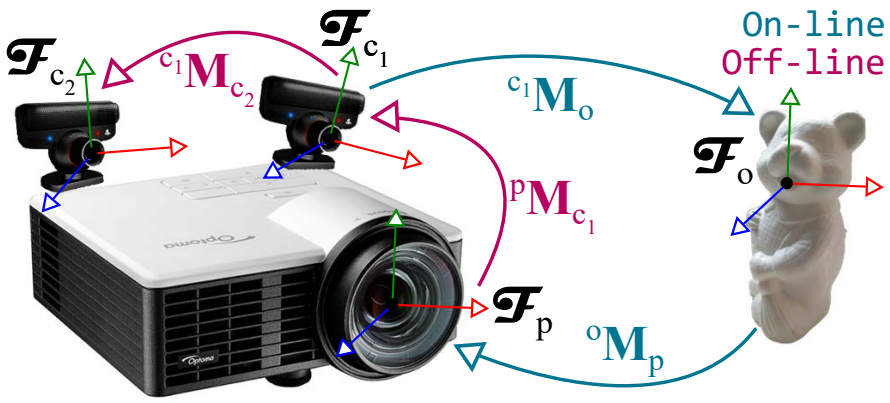


Figure 5.6 – Frame configuration for the MoSART calibration process. The estimation of $^{c_1}\mathbf{M}_{c_2}$ is explained in section 4.3.1.2, $^{c_1}\mathbf{M}_o$ is given by the tracking system and $^{o}\mathbf{M}_p$ is computed with monocular tracking similar to the one presented in section 4.3.3. $^{c_1}\mathbf{M}_p$ is computed during the projector pose estimation.

Projection model estimation

The projection model of a projector is very similar to a camera model and it consists in a projection matrix and distortion parameters. The projection matrix and the distortions

are estimated thanks to a calibration process [Yang et al., 2016]. The projection matrix determine the projection model and how a 3D point is projected into the image. The distortions enable to determine the corrections that need to be applied to the image to perfectly fit the tangible objects of the real scene. Camera-projector calibration algorithms are adapted to this case. The calibration involves a projector and a camera that are stationary relatively to each other. A 9x6 black and white chessboard is used for the calibration and several positions of this chessboard are capture by the camera. For each position of the chessboard 4 corners are selected by the user in the projector frame and an homography is computed between the projector frame and the camera frame. This homography is used to find the position of all the remaining corners of the chessboard in the projector frame. With these positions the projection matrix and distortions parameters of the projector can be estimated. The same method is used to calibrate the camera (Section 4.3.1.1 provides additional details on camera calibration).

Projector pose estimation

Once several views of the chessboard have been captured, the pose ${}^{c_1}\mathbf{M}_p$ of the projector in the camera frame can be computed from the measurements. Indeed the pose ${}^{o}\mathbf{M}_{c_1}$ of the camera and the pose ${}^{o}\mathbf{M}_p$ of the projector according to each chessboard position can be estimated with a PnP algorithm on planar objects [Lepetit et al., 2009; Marchand et al., 2016]. Then the relative pose between the projector and the camera can be computed with :

$$ {}^{c_1}\mathbf{M}_p = {}^{c_1}\mathbf{M}_o \, {}^{o}\mathbf{M}_p. \tag{5.1} $$

5.2.2.3 Virtual scene generation

The virtual scene is generated as a reconstruction of the real scene. This reconstruction is possible thanks to the tracking data and the 3D shape of the objects. The tracking data enables to position the 3D models of the tangible objects in the virtual scene. The projector is simulated as a virtual camera that has the exact same projection model and position. Thus the relative transformation between the 3D models and the virtual camera match as close as possible the real transformation between the projector and the tangible objects. The virtual scene is rendered in the virtual camera and the projector projects the rendered image over the real scene.

If the different steps of Figure 5.5 are correctly performed then the projection perfectly matches the real scene.

5.2.3 Interaction tools

Tangible tools enabling interaction techniques in Spatial Augmented Reality (SAR) scenarios have been proposed by Marner et al. [2009] and by Marner and Thomas [2010]. For our MoSART prototype we have specifically designed two tangible interaction tools. The first tool is the ***interactive Panel*** (Figure 5.7-left). It is a squared white board used to display information. The second tool is the ***interactive Stylus*** (Figure 5.7-right) which looks like a pen.

Several 3D interaction techniques have been designed to exploit these two tools within MoSART:

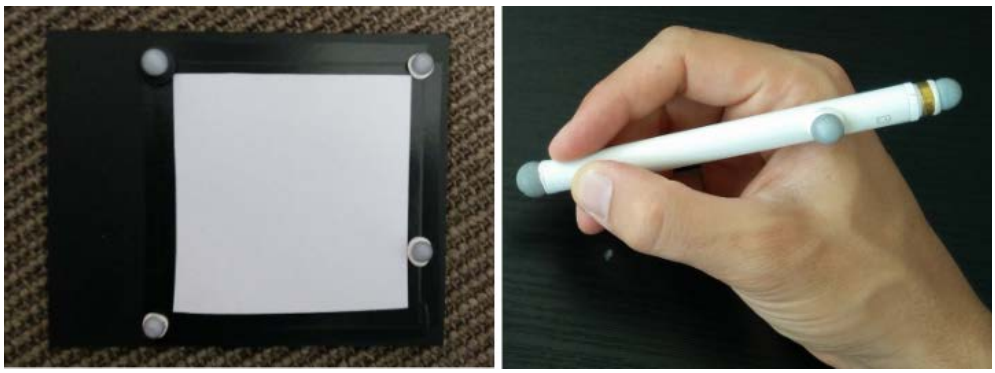


Figure 5.7 – Tangible interaction tools of MoSART: the interactive Panel (left) and the interactive Stylus (right).

1. The **interactive Panel** is primarily used as a control screen (Figure 5.8-left). It can be straightforwardly used to dynamically display 2D menus with various items. It can also be used as a specific tool, such as: a magnifying glass [Brown et al., 2003] or an "X-ray" visualizer.

2. The **interactive Stylus** is primarily used as a 3D pointer. The stylus serves as a selection tool in order to activate options and select items by touching them on the control panel (Figure 5.8-right). But it can also be used as a specific tool as well, such as: a painting tool, a light torch or a laser beam.

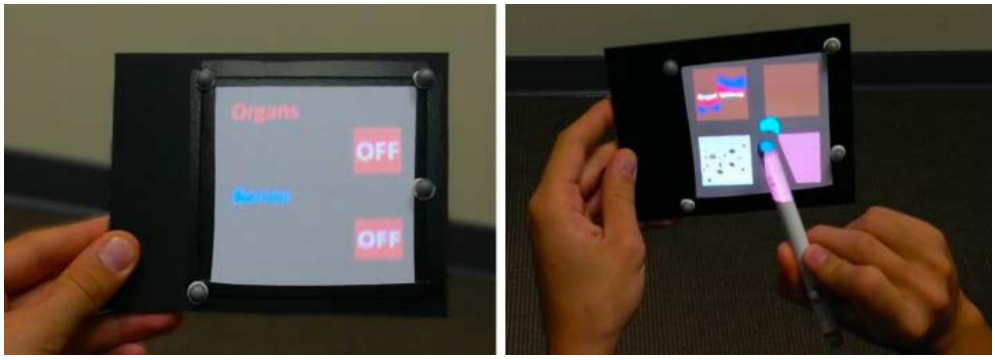


Figure 5.8 – Tangible interaction tools in use: here the Panel (left) is used to display the contextual items of a 2D menu that can be selected by pointing with the Stylus (right).

The user interactions are taken into account in the virtual scene generation so to modify the content projected over the tangible objects and tools. Other usages of these tools are depicted in the use cases presented later in section 5.3. Of course, other tangible tools and interaction techniques could be added to MoSART in the future.

5.2.4 Adding collaboration

An main advantage of MoSART is that it can be used in presence of multiple users. Indeed when only one user is equipped with a MoSART headset any external person

can still watch the augmented tangible object and exchange orally with the main user. The other persons can also manipulate the tangible object and/or the interaction tools, although being constrained to remain in the workspace of the main user corresponding to the field-of-view of the head-mounted projector. Such collaboration mode could be useful in the context of education/training scenarios in which an external user shares information with the main user.

An extension with multiple devices, which we have not implemented yet, is discussed in section 5.4. This mode could enable several users to be equipped with MoSART headsets (see a photomontage illustration in Figure 5.13).

5.2.5 Characteristics and performances

The main characteristics of our MoSART prototype are summarized in Table 5.1. The system performances have been computed using an MSI GE72 2QE laptop (CPU core I7 2.70GHz, 8Go RAM, SSD, GPU Nvidia GTX965M).

Table 5.1 – Main characteristics of the MoSART prototype

Characteristic	Value
Weight	1kg
FOV (H×V)	61° × 38°
Tracking Accuracy	1.0mm
Jitter	±0.08mm
End-to-end Latency	60ms
Resolution	1280×800
Contrast	20,000:1
Brightness	800 Lumens

The overall weight of the headset is around 1 kilogram, corresponding to: 472g for the projector, 170g for each camera, and around 200g for the helmet. The prototype currently runs on a laptop computer which can be worn in a backpack. Future work would be needed to further miniaturize the components and embed the computation and the battery directly in the headset.

The projector provides a short-throw ratio of 0.8:1.0 equivalent to an approximate effective Field-of-View (FoV) of $61° \times 38°$ with an image resolution of 1280×800 pixels. The projector provides images with a maximal brightness of 800 Lumens and a contrast of 20,000:1 which ends up with better performances when using the device at an arm distance (between 0.3 and 0.7 meter) and acceptable ones when projecting on large objects that are further away (e.g., a car mock-up at a scale close to 1).

The overall latency of the system depends on the tracking and projection display performances. Regarding the tracking system, the performances were computed similarly to the ones presented in section 4.6 by considering only stereo tracking. The implemented system performs with ∼10ms internal latency, 60ms end-to-end latency, ±0.1cm accuracy and a jitter of 0.08mm.

5.3 Use cases

The MoSART approach offers numerous possibilities in terms of interaction and visualization for single and/or collaborative situations. In this section, we present two different use cases designed and tested with our prototype, for: **(1) virtual prototyping** and **(2) medical visualization** purposes.

5.3.1 Virtual prototyping

MoSART enables to augment physical mock-ups with an infinite number of virtual textures. The users become able to interact with the mock-up directly, editing and visualizing the textured variants of the same object.

In our scenario, the user intends to choose the most suitable visuals and dressing of a teddy bear (see Figure 5.9). This use case could of course be transposed to other kinds of 3D objects, such as for the automotive or clothes industry. The user can switch between different textures that are applied to the tangible object. The selection of textures is made using a 2D menu displayed over the interactive Panel. A previsualization of each available texture is displayed on the Panel (Figure 5.8-right). The selection is achieved by pointing at the Panel's right location with the interactive Stylus.

Figure 5.9 – Several textures can be applied to an object for virtual prototyping purpose. The original teddy bear tangible object (left) is augmented with various textures selected on the interactive Panel.

Second, the user becomes able to edit and change the texture by applying virtual painting over the tangible mock-up. Our interaction tools are also used for this purpose. The interactive Panel is used to display several painting options such as the different available colors (Figure 5.10-left). The interactive Stylus acts like a paintbrush enabling the user to select a desired color, but also a brush size or a brush shape. Then, the user can directly paint the tangible mock-up with the Stylus as if he/she was painting a statue (see Figure 5.10-right).

Then, Figure 5.1-right illustrates how two users can collaborate during the painting task. One user is wearing a MoSART headset and holding the tangible object. The other user is painting the model according to the main user's instructions.

Figure 5.10 – Virtual painting. The user can select a color on the Panel (left) and paint the tangible object with the Stylus (right).

5.3.2 Medical visualization

Our second use case is a medical visualization scenario allowing to interact with a tangible body shape. To illustrate this use case, a women chest mannequin is used as a tangible object. The user can visualize different inner components (e.g., bones or organs) positioned with respect to the tangible human body. On Figure 5.11, the left image illustrates the visualization of the chest bones and the right image illustrates the visualization of both bones and organs of the human chest.

The interaction tools can first be used to change the visualization state of the application to either: display the bones, the organs, the digestive system or the whole. To do so, the interactive Panel displays a menu with two-state buttons (see Figure 5.8-left) that the user can toggle with the interactive Stylus used as a pointer. Then, in another interaction mode, the Panel and Stylus can be used to further explore the virtual inner components of the human body. The user can use the Panel as a magnifying glass (see Figure 5.12-left) to be positioned in front of an area of interest (such as for observing some small hidden organs of the chest). The Stylus can also serve as a flashlight (see Figure 5.12-right) to illuminate the organs and have a better perception of their geometry and material.

This visualization use case could inspire similar setups for education or training purposes, in single or collaborative conditions, without being limited to the medical field. Besides, by placing reflective markers over the body of a real person, MoSART could actually be used with a real body and a real patient [Ni et al., 2011]. This could be interesting for educational purposes, but also before or during a surgical operation. However, a technical challenge would consist here in accurately tracking the deformable body in real-time.

Figure 5.11 – Medical visualization on a tangible chest with MoSART. The user is able to visualize the bones (left) with or without the organs (right).

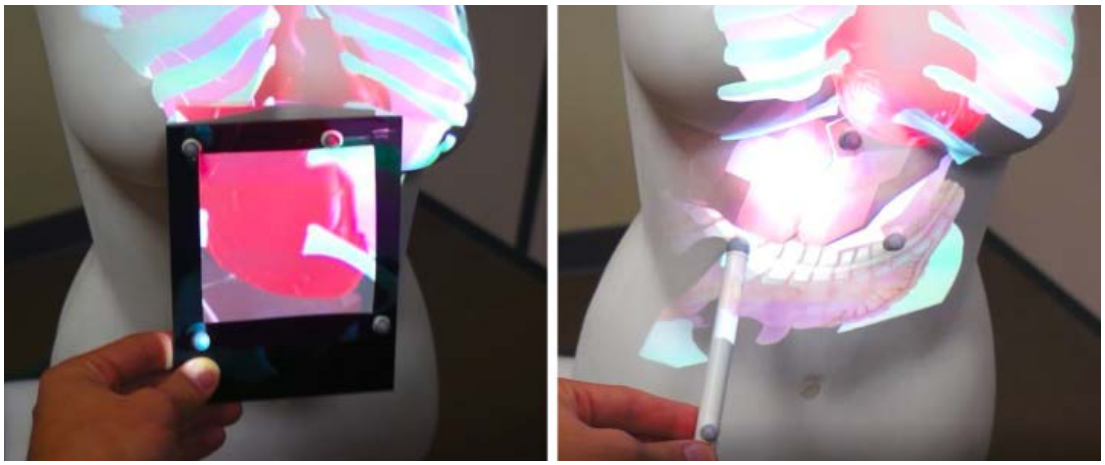


Figure 5.12 – Exploration of 3D medical models. The Panel can be used as a magnifying glass (left) to visualize details or hidden organs. The Stylus can simulate a flashlight (right) to better perceive depths and illuminate some parts of the virtual models.

5.4 Discussion

MoSART provides mobile spatial augmented reality on tangible objects and has been tested within several use cases. During the informal tests MoSART was found very promising in terms of 3D interactions, both for direct manipulation of the physical mock-ups, as well as by means of our dedicated interaction tools.

Considering the current limitation of a majority of AR systems regarding Field-of-View (FoV), it is noteworthy that MoSART can considerably increase the FoV and the interaction workspace, especially compared to OST-AR (e.g., Microsoft Hololens). The weight of the first MoSART prototype still remains above the usual weight of commercial OST-AR headsets (less than 600 grams for the Hololens for instance). But

the design of our prototype has not been fully optimized and miniaturized yet, and we can anticipate a reduction of the total size and weight in the following versions. Future studies could be carried out in order to compare the MoSART approach to other existing head-mounted AR systems (e.g., OST-AR or VST-AR).

Regarding the technological evolution of MoSART, we envision several paths for future works. The calibration process could first be fully automatized using a similar technique as the one proposed by Moreno and Taubin [2012] which could also improve the 2D-3D matching performance of our system. Then, as mentioned in section 5.2.4, a multi-user collaboration could be implemented to support several MoSART systems at the same time (see photomontage of Figure 5.13). In this case a master computer can handle the rendering of the virtual scene in the different virtual cameras corresponding to the different MoSART projectors enabling to generate the virtual scene only once and to avoid inconsistencies when projecting. Such implementation may required to have a blending of the multiple images and the jitter of the distinct devices should still be taken into account. Moreover, if the users are facing each other there is a risk of potential blindness due to the projection light. All these potential issues could thus be investigated.

Figure 5.13 – Concept of a collaborative setup with two MoSART systems (photomontage). The users could look at different sides of the object for an even wider projection space.

MoSART has only been tested with rather small tangible objects that can be held in the hand. Thus, it could also be interesting to test the MoSART concept with medium (e.g., chairs, tables) and large objects (e.g., cars, rooms). Using MoSART on larger objects might notably require an improvement of the projection performance to keep bright and contrasted projections. Also, a user study could be carried out to evaluate the performances of MoSART in terms of user experience and comfort in comparison with stationary SAR and OST-AR systems.

5.5 Conclusion

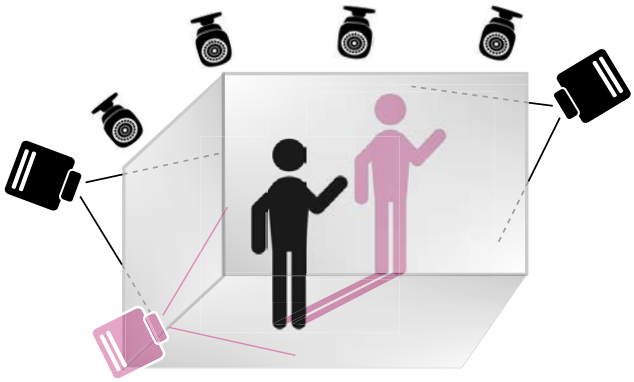
We have introduced a novel paradigm for Mixed Reality (MR) Projection-Based Systems (PBS). Our approach, called MoSART, proposes mobile spatial augmented reality with tangible objects. MoSART enables mobile interaction with tangible objects thanks to head-mounted tracking and projection. The tracking enables localizing both the tangible objects and tools while the projector is used to superimpose virtual content over them. With MoSART, the users are able to move around and to directly manipulate the tangible objects. They can also interact with the objects and modify their behavior by using dedicated interaction tools. In addition, due to the projection system, MoSART makes it possible for external users to share the experience with the main user for collaborative scenarios (e.g., training or assistance).

A proof of concept has been designed based on an "all-in-one" headset providing head-mounted projection and feature-based optical tracking. For this purpose, a pico-projector and two consumer-graded cameras were rigidly attached to a helmet. Localization algorithms were implemented based on infrared optical tracking and the objects were equipped with infrared markers. Projection mapping algorithms were implemented to make it possible for the projector to project consistent content over 3D tangible objects. The approach enables augmenting 3D objects that have complex geometries and structures. Two tangible interaction tools, the Stylus and the Panel, were designed for the users to interact with the tangible objects.

Our prototype shows good performances compare to current Augmented Reality (AR) systems. The system was illustrated on two industrial use cases: virtual prototyping and medical visualization. Taken together, our results suggest that the MoSART approach enables a straightforward, mobile, and direct interaction with tangible objects, for a wide range of projection-based augmented reality applications in single or collaborative conditions.

Introducing user's virtual shadow in projection-based systems

6



Contents

6.1 Related work on virtual shadows . . . 108
6.2 Studying the use of virtual shadows in projection-based systems . . . 110
6.2.1 Objective . . . 110
6.2.2 Apparatus . . . 110
6.2.3 Participants . . . 112
6.2.4 Experimental task . . . 113
6.2.5 Experimental protocol . . . 113
6.3 Results . . . 115
6.3.1 Performance measurements . . . 115
6.3.2 User experience questionnaires . . . 116
6.4 Discussion . . . 119
6.4.1 Virtual shadows and virtual embodiment . . . 119
6.4.2 Virtual shadows and spatial perception . . . 120
6.5 Conclusion . . . 121

When using Projection-Based Systems (PBS) the users are generally aware of the real environments and in particular they are aware of their own body. In some cases this awareness can be an asset (e.g., having a real reference, reducing motion sickness). However the awareness of the real environment can, sometimes, reduce the immersion and the user experience which can lead to a less accurate and comfortable perception of the virtual environment.

In this chapter we propose to project a respresentation of the user in the virtual environment by using a virtual and dynamic shadow in PBS. In fact, shadows are paramount in our everyday life as they provide information about depth, proximity, or shape of our environment [Puerta, 1989]. One particular shadow is the one cast by our own body which in combination with a virtual avatar can reinforce the user's virtual experience [Slater et al., 1995] and theoretically it can also enhance spatial perception. However using avatars is generally not possible in PBS. Therefore, is it possible to embody someone else in a CAVE environment even though the users are aware of their own body? Does the virtual shadow influence the users interaction behavior? To answer these questions we carried out an experiment in order to assess the user's virtual embodiment and the user's 3D performance in presence of a virtual shadow. The virtual shadows enabled to provide a virtual representation of the user which differed from their own physical body even though they were still able to see it. In particular, we studied how the users appropriate different virtual shadows (male and female shadow) and how does the virtual shadow affects the user behavior when performing a 3D positioning task. The results showed that the shadow can have an influence on the user's behavior while interacting, and that participants seemed to prefer virtual shadows which were closer to their own body.

To summarize, the contributions of this chapter are:

- An approach for introducing virtual and dynamic shadows as the representation of the user in the virtual environment of projection-based systems.
- A study to assess the influence of the users' virtual shadows on users' virtual embodiment and 3D performance when using projection-based systems.

In the remainder of this chapter we first make an overview of previous work that has been done on adding virtual shadows in virtual environment. Second we describe the experiment that aims at studying the influence of virtual shadows on the user comfort and presence and on the environment understanding. Third we present the results of the experiment. The chapter ends with a discussion and a general conclusion.

6.1 Related work on virtual shadows

Increasing the embodiment of the users in virtual reality applications has been studied in a number of different works. Although the majority of works have been focusing on the user's avatar [Kilteni et al., 2012], several works have also addressed the use of virtual shadows to reinforce the effect. For example, [Slater et al., 1995] carried out a study on the influence of the presence of shadows in HMD virtual environment. They did not find any influence of the shadows on depth perception but they found that

adding a static shadow increases the user presence and that adding a dynamic shadow increases it even more. This study focused on the object shadows and no user dynamic shadow was considered. In later work they introduced the dynamic shadow of the user in HMD to confirm that the realism of the scene has an impact on the user behavior [Slater et al., 2009]. Even though users virtual shadows have not been widely studied, avatars are commonly used in virtual reality HMDs. There even were studies on the influence of the morphology of the virtual avatar on the user behavior and embodiment in HMD [Peck et al., 2018]. Nevertheless they did not carry out any study on immersive projection systems.

Indeed using avatars in IPS or screen displays can be harder since the users are aware of their own body. When using such systems, casting shadows can still be a solution to enhance the user experience. Even though, in 1995, Slater et al. did not find any influence of the shadows on depth cue, later studies [Hubona et al., 1999] found out that when using screen displays adding objects shadows increases the accuracy during positioning tasks. However the study focused on the object shadows and no user shadow was considered. Moreover the study was carried out on screen displays that do not provide any immersion. Such results were confirmed in later studies carried out on augmented reality displays [Diaz et al., 2017; Sugano et al., 2003]. Adding shadows to the virtual objects integrated in the real world increases the objects presence and provides depth cues that increase the spatial perception. [Hu et al., 2000] also confirmed the results in VR HMD. Regarding the user shadow, altering the shadow behavior compared to the user's body behavior can modify the user perception of the environment. [Ban et al., 2015] carried out a study where the shadow was more or less independent from the user. The shadow was then able to move differently from the user movement. The users were then confused and were not always able to tell if they or their shadow were moving. Moreover their movement were altered by the shadow movements. Such study was carried out by projecting shadows on a wall but no virtual reality environment nor interaction were considered.

[Steinicke et al., 2005] were the first ones to introduce user shadows on IPS. They added the presence of users' virtual shadows and reflections on a responsive workbench. The real reflection of the user (captured with a camera) was added on a metallic surface and a virtual shadow of the user's hand was cast on the same surface. The authors claim that it increases the realism of virtual objects but no study was made to evaluate how this approach increases users perception and improves objects interaction. More recent work from [Yu et al., 2012] proposed to increase the realism of virtual environments in CAVE displays. They confirmed that the user presence was increased when having shadows and reflections that corresponded to the users' body movements. The body movements were forced by the application and the task consisted in naturally walking in the display to avoid a collision with a virtual character. Nevertheless no complex 3D interaction task was proposed to study the influence on depth perception. Later work from [Kwon et al., 2015] enhanced wall-sized VR application by adding objects shadows on the real floor in front of the display. They carried out a study were users had to touch the virtual objects with a direct touch metaphor. Their results suggested that the shadow cue is even more important than the stereoscopic cue. Regarding the shadow of the user, no study was considered.

6.2 Studying the use of virtual shadows in projection-based systems

6.2.1 Objective

We aimed at studying the use of virtual shadows in Projection-Based Systems (PBS). To do so we focused on the influence of a virtual shadow on the presence, the embodiment and the precision of the participants when performing a 3D interaction in a CAVE-like display. We designed a 3D positioning task in which participants had to place a physical ball over virtual targets placed on virtual planar surfaces, such as tables or walls. The goal of the positioning task was to place the ball as close as possible to the target areas without going through them. During the experiment the participants were presented with three virtual shadow conditions: A **male shadow (M)**, a **female shadow (F)** and **no shadow (N)**. We hypothesized that the possibility to see their virtual shadow can influence both the users perception on their shadow and the way users perceive and interact in the virtual environment, thus, two main research questions were addressed:

- **Q1**: Is it possible to embody someone else in an IPS even though the users are aware of their own body?
- **Q2**: Does a user's virtual shadow increases the user's spatial perception of the virtual environment in an IPS?

6.2.2 Apparatus

Participants were immersed in the virtual environment using a 9.6×2.9×3.1 meters CAVE display. The CAVE system is built with 4 screens: one on the floor, one on the front and one on each side. The projection on the screens is made using Barco F90-4K13 laser projectors. Every screen but the floor is back-projected. The tracking data is provided by an infrared optical tracking system from Optitrack[1]. The optical tracking is composed of 12 cameras (4 Optitrack Prime 13W and 8 Optitrack Prime 13).

The dynamic virtual shadow is created from a virtual 3D model of a humanoid. For the purpose of the experiment we chose to have a male and a female 3D model (see Figure 6.1). Such models have been chosen to fit the human proportions and to relieve the experiment from a cumbersome calibration process. Thus the matching between both the virtual and the real body are simply made by scaling the model so that its height corresponds to the participant's height.

Regarding the dynamic part of the shadow, an inverse kinematics (IK) algorithm was used to provide movement to the shadow and to make its position correspond to the user's one. The Final IK[2] Unity asset was used to optimize the position of the rigged models to fit the actual position of the user. The Final IK algorithms were able to converge to an optimized solution by providing the position of both feet, both hands

[1]Optitrack tracking hardware http://optitrack.com/hardware/
[2]Final IK asset http://www.root-motion.com/final-ik.html

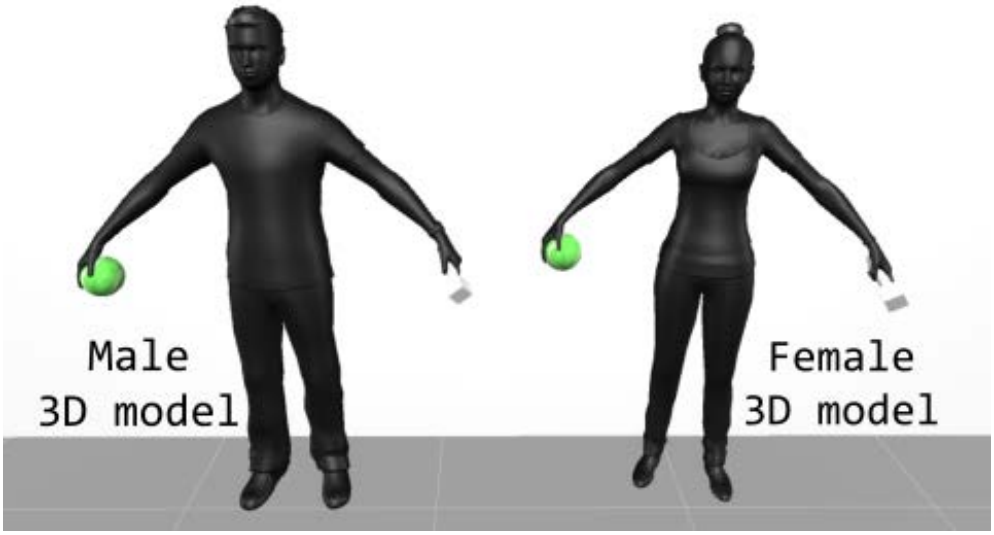


Figure 6.1 – Two 3D models were used to cast a virtual shadow in the virtual environments: a male 3D model (left) and a female 3D model (right).

and the head of the user. Therefore, the participants were equipped with tracking devices on the feet, hands and head (see Figure 6.2-left). Finally due to experimental testing we noticed that providing the orientation of the pelvis to the IK algorithm gave better visual performances. Thus, the participants were also equipped with an additional tracking marker in the waist to track the pelvis. To simplify the experiment and the recorded data, a physical ball was placed on the participants right hand as an extension of their arm and body. A controller was hold on their left hand (see Figure 6.2-right). The 3D model that matches the user position is introduced in the virtual environment but no rendering of the model's mesh was done, only the cast shadow was displayed.

Since the ball was not perfectly rigidly attached to its tracking constellation and that the participants were able to grab it as they wanted to, a calibration step was introduced during the experiment. The calibration step aimed at computing the distance between the physical ball and the virtual ball (used to cast the ball shadow). The participants were asked to place 3 times the physical ball over a physical table whose height was perfectly known (see Figure 6.3). Then the offset between the physical and virtual balls was computed and used to correct the recorded data. The calibration step was performed at the beginning and at the end of each experimental condition.

Regarding the lighting of the scene we chose to use 3 directional lights to provide the participants with 3 virtual shadows of themselves. Two lights were oriented of around 45° according to the wall normal vector and one light was almost collinear with the wall normal. Such lighting configurations provided the scene with two shadows cast half on the floor and half on the wall and one shadow that is almost only cast on the wall (see Figure 6.4). The VR application was developed using Unity 5.6.1f and ran at an average frame rate of 60 fps.

Figure 6.2 – Six infrared constellations (red) were placed over the participants (left) to provide tracking data to perform the inverse kinematics optimization. A ball was placed in their right hand, as an extension of their arm, and a controller in their left one (right).

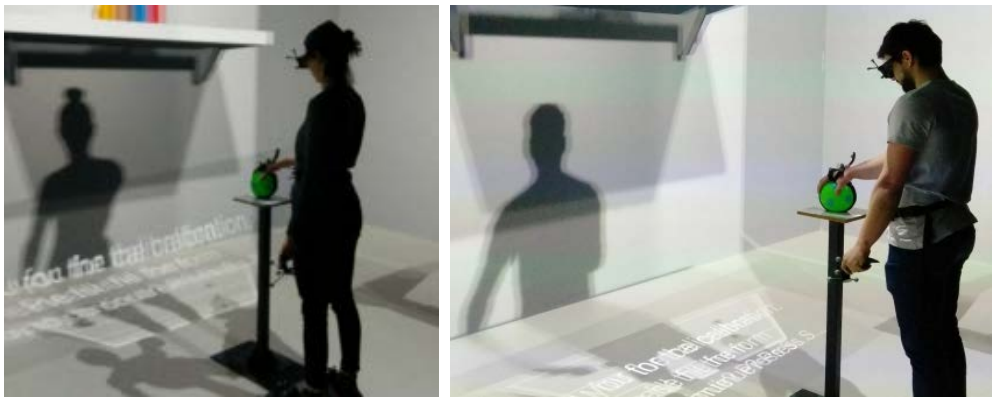


Figure 6.3 – The calibration step enables to compute the offset between the physical ball and the virtual ball at the beginning and end of each virtual shadow condition (N, F or M).

6.2.3 Participants

27 participants from the university campus took part in the experiment (Age: min = 23, max = 55, and avg = 31 ± 8), recruited both among general students and staff. For the purpose of the experiment we chose to recruit 14 women and 13 men since the participant were confronted to a male and a female shadow. Participants were recruited asking for minimal previous experience in VR: 18 subjects had none to very limited previous experience with virtual reality, 7 had some previous experience, and only 2 were familiar with VR. None of the participants knew about the experiment being tested, or that they would be presented with virtual shadows. All participants were right-handed since the ball was placed in their right hand for the positioning task.

6.2.4 Experimental task

Participants were asked to position a physical ball (7cm radius, see Figure 6.2-right) over circular green target areas (2.5cm radius). The target areas were positioned over 2 virtual tables and a virtual wall (Figure 6.4). Only one circular target was displayed at a time in order to help the users know which target they had to focus on. Every run of the experiment started with a target on the left table, then a target on the wall and then a target on the right table and so on. Once the participants were satisfied with the placement of the ball on the target, they had to validate the positioning of the ball by pressing the trigger of the controller placed on their left hand (see Figure 6.2-right). In order to reduce the required time to perform the task, a timer was added. The timer was depicted by changing the color of the current target area. The target areas appeared green, they went orange after 3 seconds and red after 6. Participants were asked to try and validate the positioning of the ball before the target became red. Nevertheless they were told to be as accurate as possible even if they had to spend more time for each target.

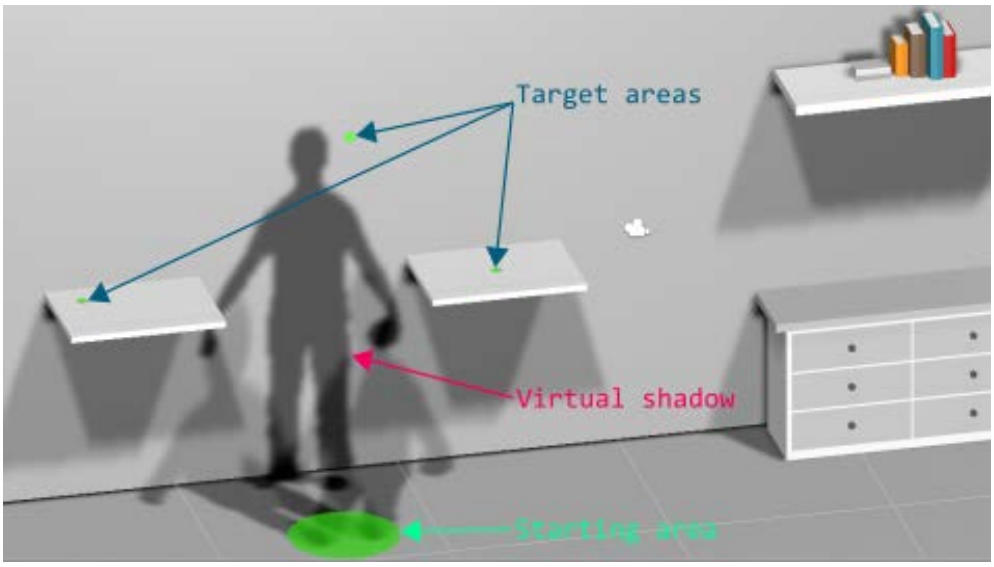


Figure 6.4 – Virtual environment of the experiment. The participants were asked to place the physical ball over the green target areas that were displayed on both tables and on the wall. Three different point lights generated three different virtual shadows at the same time.

In order to decrease learning effects, participants had to face a different number of target configurations. Three target positions were defined for each table and for the wall. Moreover the tables were positioned at a variable height (90cm or 110 cm).

6.2.5 Experimental protocol

An informed consent form was signed by each participant before starting the experiment. The form stated the participants' right to withdraw and presented the experiment and the main goal of the research. In addition, it also asked their consent

regarding image and video copyright. In order to minimize the priming of participants, little details were provided regarding the purpose of the shadow. Mainly, participants were told that the experiment aimed at assessing people precision when performing 3D positioning tasks. They were also told that the virtual shadow conditions of the scene could vary but no additional details about the shadows were given.

The experiment was divided into 3 blocks. During each block the participants were presented with one virtual shadow condition which was either **No shadow (N)**, **Male shadow (M)** or **Female shadow (F)** (see Figure 6.5). Each participant performed entirely the positioning task for each one of the 3 conditions.

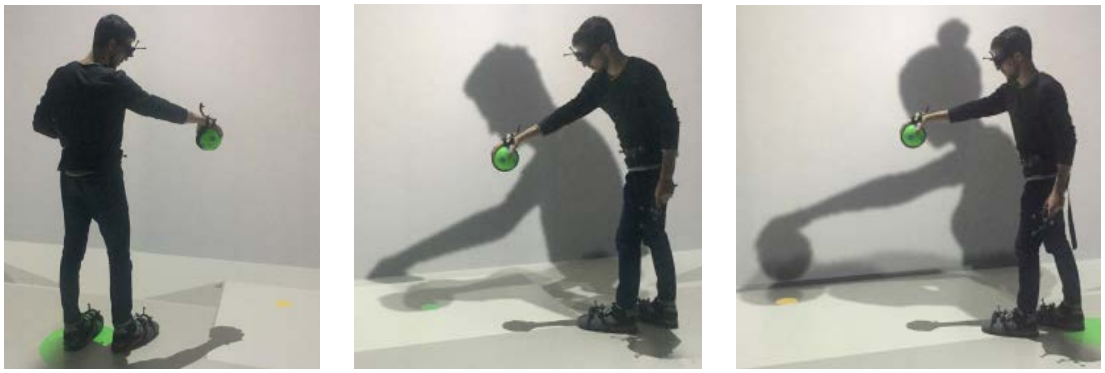


Figure 6.5 – The participants performed the positioning task with 3 different virtual shadow conditions: None (N) (left), Male (M) (middle), Female (F) (right). The real shadow of the user is visible on the floor but does not match the natural behavior of a shadow in the virtual environment and is not taken into consideration.

Considering all possible target combinations, three targets (wall, left table, right table), two heights and three positions, participants had to perform the placement task 18 times for each condition. Moreover, as three repetitions were considered, each block resulted in 54 trials. Finally, a training period of half a run (9 targets) was present at the beginning of each block. To counterbalance the influence of the running order of the conditions on the participants behavior, the participants were divided into 6 groups (M/F/N, M/N/F, F/M/N, F/N/M, N/M/F and N/F/M). Each group was composed of at least 2 male and 2 female participants.

After each block participants were asked to fill in a subjective questionnaire (see Table 6.1) in order to evaluate their subjective appreciation of the experiment and collect their feedback. The questionnaires began with 5 question to evaluate the presence of the user following the suggestions of Usoh et al. [Usoh et al., 2000]. Then, there were questions regarding the ownership and some questions about the task and the user comfort. Finally the participants were free to comment their strategy to perform the task and to detail their feelings and comments in presence or absence of a shadow.

While performing the positioning task the participants were immersed in the CAVE. After each condition the tracking constellations were removed and the users had to fill in the questionnaire on an independent laptop. Then they were reequipped with the tracking constellation and reintroduced in the CAVE. The whole experiment including the questionnaires lasted around 45 min in total (15 min per virtual shadow condition).

In addition to the subjective questionnaires assessments the following information was recorded for each positioning task:

Table 6.1 – Summary of the subjective questionnaire. (P: Presence, SA: Shadow Appreciation, TA: Task Appreciation, O: Ownership, A: Agency)

ID	Question
P_1	I had a sense of "being there" in the virtual house living room space
P_2	There were times during the experience when the living room space was the reality for me
P_3	The living room space seems to me to be more like images that I saw or somewhere that I visited
P_4	I had a stronger sense of being elsewhere or being in the living room space
P_5	During the experience I often thought that I was really standing in the living room space
SA_1	When positioning the ball on the tables I felt that the virtual shadow was useful
SA_2	When positioning the ball on the wall I felt that the virtual shadow was useful
SA_3	I felt that the virtual shadow was a good indicator of the proximity of the ball with the tables
SA_4	I felt that the virtual shadow was a good indicator of the proximity of the ball with the wall
TA_1	I felt that I was accurate on positioning the ball
TA_2	I felt that the positioning of the ball was rather easy
O_1	I felt as if the virtual shadow was my own shadow
O_2	I felt as if the virtual shadow was from someone else's
A_1	I felt as if the virtual shadow moved just like I wanted
A_2	I expected the virtual shadow to react in the same way as my own body
A_3	I felt like I controlled the virtual shadow as if it was my own shadow

- The depth error (Y axis for the tables and Z axis for the wall) between the ball and the target area when the position is validated by the participant. The error is positive when the ball is positioned over the surface and negative when its position inside the surface.
- The time the participant took to perform one positioning task. As the trials were performed sequentially, the task completion time matches the time between two validations.

6.3 Results

During the analysis we explored the effect of the participants' gender on the results. If the gender did not significantly influenced the results, data was pooled. Regarding the ANOVA analysis, effect sizes are expressed using the partial eta squared (η_p^2). The general rules of thumb given by [Miles and Shevlin, 2001] state that the qualifiers "small", "medium" and "large" correspond to cases where $\eta_p^2 > 0.01$, $\eta_p^2 > 0.06$ and $\eta_p^2 > 0.14$ respectively. Only significant effects are discussed. We first discuss performance measurements and then the subjective appreciations of participants.

6.3.1 Performance measurements

The main indicators of task performance were the final depth position (see Figure 6.6) and the task completion time. We first analyzed the effect of the Shadow type and the Task on the depth error using a two-way ANOVA analysis considering participants as a random factor. The ANOVA analysis showed a main significant effect for Shadow [$F_{2,52} = 8.99$, $p < 0.001$, $\eta_p^2 = 0.25$] and Task [$F_{2,52} = 39.67$, $p < 0.001$, $\eta_p^2 = 0.60$], no interaction effect was found [$F_{4,104} = 0.26$, $p = 0.9$]. Tukey post-hoc tests showed that for the male shadow condition participants were more conservative while performing the

task $M = 3.2cm$; $SD = 3cm$ compared to the female shadow $M = 1.4cm$; $SD = 3.3cm$ and without shadow $M = 0.9cm$; $SD = 3.7cm$. In general, in the condition without shadows participants were more prone to go through the target surface. On the contrary when a virtual shadow was present the user tend to stop their movement before going through

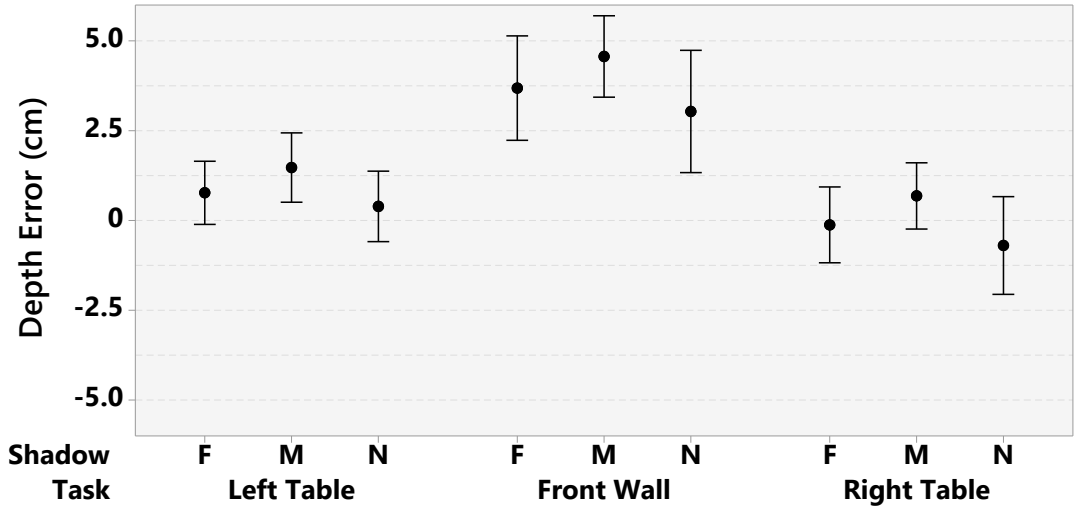


Figure 6.6 – Mean interval plot (CI 95%) for the depth positioning error, grouped by the Virtual shadow condition and the Task.

Regarding the task completion time (see Figure 6.7), the ANOVA analysis showed a main effect on the Task [$F_{2,52} = 68.24$, $p < 0.001$, $\eta_p^2 = 0.72$] and an interaction effect [$F_{4,104} = 4.13$, $p < 0.01$, $\eta_p^2 = 0.13$], there was no effect on Shadow [$F_{2,52} = 0.72$, $p = 0.49$]. Tukey post-hoc tests showed that participants required significantly more time to perform the task in the Left table condition $M = 2.65s$; $SD = 0.51s$ compared to the Right table condition $M = 2.28s$; $SD = 0.49s$ and the Wall condition $M = 2.32s$; $SD = 0.48s$. This result can be explained by the fact that all participants were right handed and required more time to access the left table. Post-hoc tests were not conclusive for the interaction effect.

6.3.2 User experience questionnaires

The different questions of the subjective questionnaires have been classed into several categories: Shadow Appreciation (SA), Task Appreciation (TA), Agency (A), Ownership (O) and Presence (P). Table 6.1 gathers all the questions for the different categories. In the following, the statistical analysis of each questionnaire category is presented.

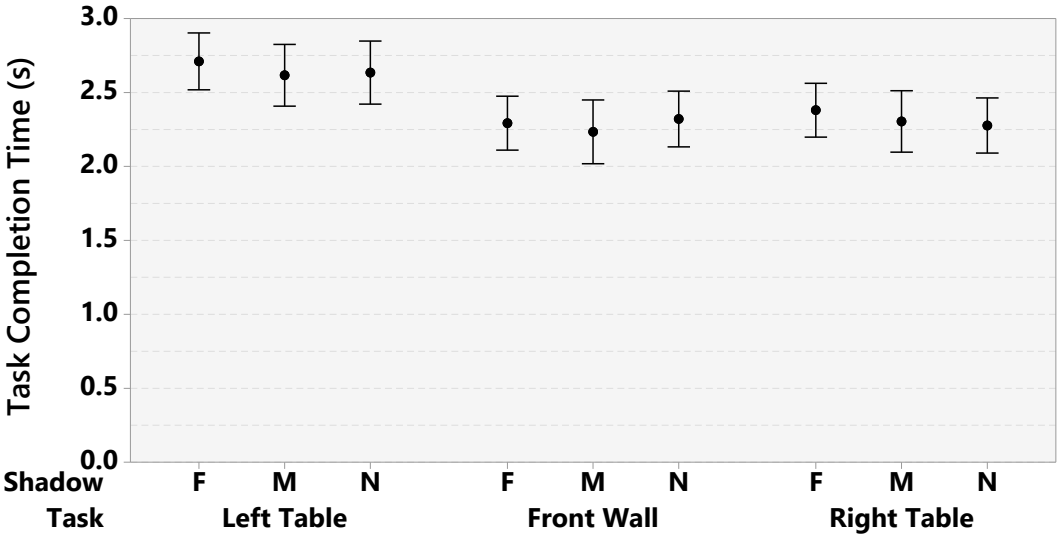


Figure 6.7 – Mean interval plot (CI 95%) for the task completion time, grouped by the Virtual shadow condition and the Task.

Shadow appreciation

In general participants considered that the shadow helped them to perform the task (SA_1: $M = 5$; $IQR = 2$, SA_2: $M = 5$; $IQR = 2$) and also that it was a good indicator of the proximity of the ball with respect to the targets (SA_3: $M = 5$; $IQR = 2$, SA_4: $M = 5.5$; $IQR = 2$) (Figure 6.8). Wilcoxon signed rank test showed that the male shadow was perceived to provide significantly better assistance when performing the table task (SA_3, $p < 0.05$). A similar trend was observed for the wall task, but results were not significant (SA_4, $p = 0.09$)

Agency

In overall, participants felt that the virtual shadow moved (A_1: $M = 6$; $IQR = 2$), reacted (A_2: $M = 6$; $IQR = 2$) and that they could control it (A_3: $M = 6$; $IQR = 1$) as if it was their own shadow (Figure 6.8). Wilcoxon signed rank tests did not show any significant differences between the male and female shadows.

Task appreciation

Friedman rank sum test was used to analyze how participants perceived their accuracy while performing the task (TA_1) and the perceived difficulty (TA_2) considering each level of Shadow (Figure 6.9). The Friedman analysis of TA_1 showed that the virtual shadow condition had a significant effect [$\chi^2(2) = 10.89$; $p < 0.01$]. Pairwise Wilcoxon tests showed that the condition without shadows was perceived as less accurate (both $p < 0.05$). Similarly, the analysis of TA_2 also showed a significant effect on Shadow [$\chi^2(2) = 12.11$; $p < 0.01$]. Again, pairwise Wilcoxon tests showed that the condition without shadows was perceived to be more difficult (both $p < 0.05$).

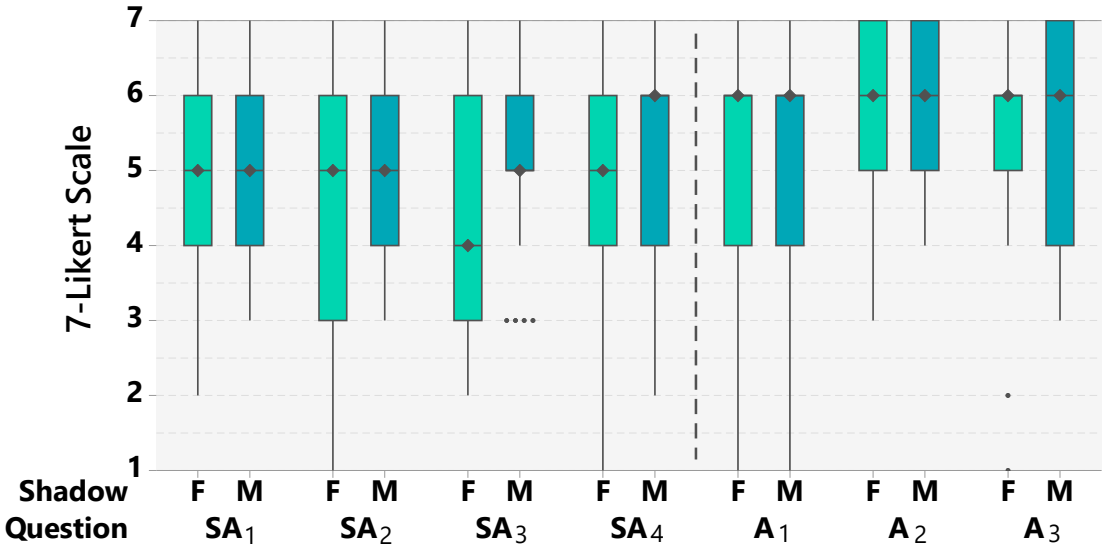


Figure 6.8 – Boxplot summarizing ratings for the shadow appreciation (SA) and agency (A) questionnaires. In general, participants appreciated the fact of having a shadow and felt a strong sense of agency.

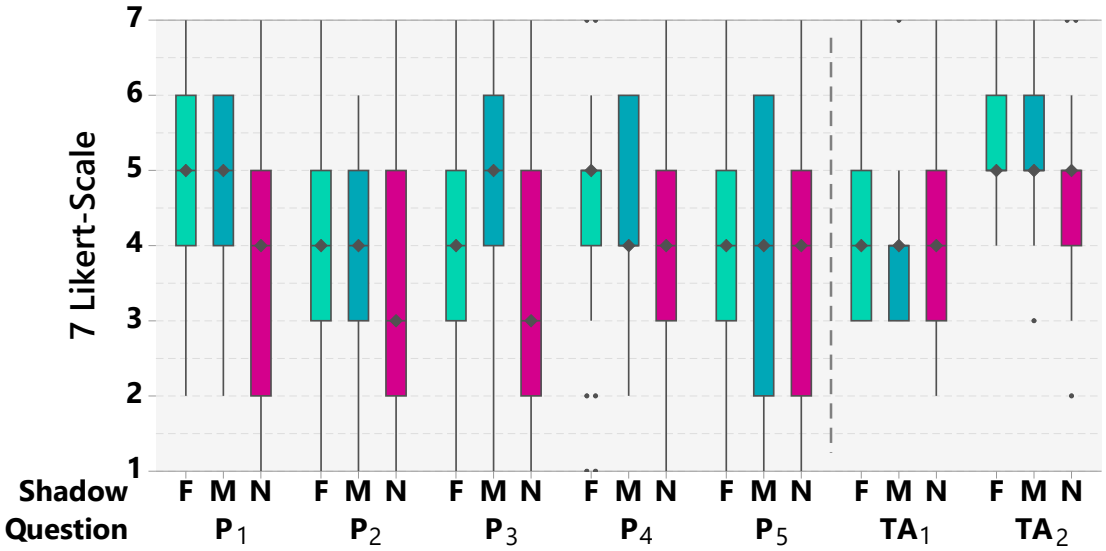


Figure 6.9 – Boxplot summarizing the ratings for the presence (P) and task appreciation (TA) questionnaires.

Presence

In general, participants experienced a moderate sense of presence. Although participants seemed to rate lower some of the questions for the condition without shadow (see Figure 6.9), the analysis of the presence questionnaire results did not show any significant effect on the Virtual Shadow condition.

Ownership

In the results related to the feeling of ownership (O_1 and O_2), we observed an interaction

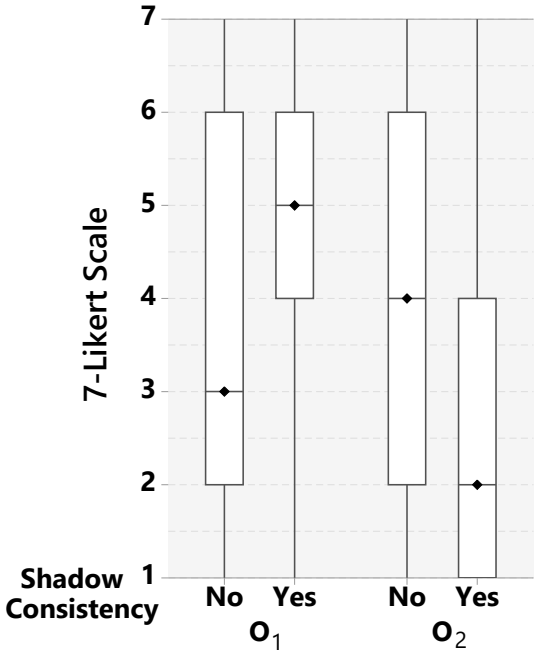


Figure 6.10 – Boxplot of the ownership ratings regarding the shadow consistency in terms of the gender. Ownership ratings were significantly higher when the shadow gender was consistent with the participant's gender.

6.4 Discussion

Taken together the results should help answering **Q1** and **Q2** to give a lead on, respectively, if virtual shadows can provide a sense of virtual embodiment in IPS and if the virtual shadows increase the spatial perception of the users. In the following we discuss the influence of the virtual shadow on both issues.

6.4.1 Virtual shadows and virtual embodiment

One of the major research questions in this chapter was: Is it possible to embody someone else in an Immersive Projection-based Systems (IPS)? The questions from categories A and O are taken into account to discuss the virtual embodiment of the users in presence of the virtual shadows. In terms of agency, subjective ratings showed that participants had a strong feeling of agency towards their virtual shadow. Participants

had the feeling that the virtual shadow was moving in a natural way and that it correspond to their shadow position. Some users did not even noticed the shadow at first since it is natural for them to have one: "*I did not payed attention to the shadow: it was natural I guess.*". This effect was not dependent on the morphology of the shadow. Nevertheless, in an IPS users are always aware of their own body and this limitation is reflected on what kind of body the users are able to appropriate. Indeed the ownership measurements depict that, when the morphology of the virtual shadow (or the gender) is not consistent with the user's morphology, the user tends to feel that the shadow is from someone else. However, as long as the virtual shadow morphology is close enough to the users' one they feel that the virtual shadow is their own. During the experiment some users reacted to the fact that the shadow was of the opposite gender or that the shadow did not had the same hair style for example: "*Is the bun hair style made on purpose ?*", "*Oh. I'm a male now.*", "*I clearly have a female outline*", "*I've got muscles !*", "*Do I wear a bun ?*". Finally, the results in the literature have shown that virtual shadows can increase the sense of presence. Our results show that there is a trend to rate lower the presence question in absence of the shadow which is consistent with the previous work. Some users even commented the enhanced realism of the scene in presence of the virtual shadow: "*The room is not realistic because of the absence of the shadows.*", "*I felt like the experiment was less realistic without the shadow.*", "*I felt it is more realistic with shadows than no shadow at all.*". To sum up, adding dynamic virtual shadows in IPS, such as CAVE displays, can enable the user to embody a virtual shadow. Nevertheless, in order to achieve a higher degree of ownership, the virtual shadow should be close enough to the users body since they are always aware of it.

6.4.2 Virtual shadows and spatial perception

On the other hand, does the presence of a virtual shadow increase the spatial perception of the users? The performance results, the SA and TA assessments are taken into account to answer it. If we consider SA ratings, participants felt that the virtual shadow could be a good indicator of proximity from the targets and that the shadow was a good assistant to position the ball: "*I mainly used the shadow to position the ball over the targets.*", "*The shadow has been useful even if it was not mine.*", "*The task is more complicated without the presence of a shadow, the distances were harder to estimate.*". The participants had an overall feeling of better perceiving and understanding the virtual environment physical limitations. Moreover, although performance measurements did not show any significant results in terms of task completion time, participants found it easier to perform the task with the presence of a virtual shadow. Some of them even commented it in the questionnaires: "*The shadow was helpful for the ball placements.*", "*It is helpful to have the shadow to place the ball, particularly on the wall.*". Finally, the analysis of the depth error showed that participants had a more conservative behavior when placing the ball on the targets in the presence of the virtual shadow. Indeed, the presence of the virtual shadow can warn the users that they are approaching a rigid object and that they may not be able to go through it, as if the object was physically there.

An interesting result was the fact that the participants were less accurate when the

target was placed in the wall. The most plausible explanation is that it is harder to estimate the actual position of the target when it is placed on the wall, but the actual reasons remain unknown. For the positioning task we chose to add a real ball in the participants right hand as an extension of their arm. Thus when the virtual shadow of the participant was removed (condition *No Shadow*), the shadow of the ball was also removed. Therefore we did not propose a condition with only the virtual shadow of the ball. According to the previous work the presence of the shadow of the objects adds a depth cue and a study with the shadow of the ball should lead to results that correspond to the previous studies. In summary, these results show that the presence of virtual shadows provides an increased awareness of the spatial relations between the users and the virtual environment (less inter-penetrations) and are positively perceived by the users.

6.5 Conclusion

In this chapter we proposed to introduce users' dynamic virtual shadows in Projection-Based Systems (PBS). Our approach aimed at increasing the user embodiment and spatial perception in such systems. Indeed since the user's are able to see their own body in PBS, embodying an avatar is not as straightforward as in Head-Mounted Displays (HMD). In order to provide the users with a virtual representation of themselves we propose to introduce their virtual shadow in PBS. The shadow was directly mapped to the users motion through inverse kinematics algorithms, making them fell like the shadow was their own.

We carried out a user study to evaluate the influence of the user's virtual shadow on the user's behavior and, in particular, on the virtual embodiment and the spatial perception. During the study, the participants were immersed in a CAVE display and they were asked to position a real ball over several virtual planar surfaces (a wall and two tables). The participants were presented with three virtual shadow conditions: a male virtual shadow, a female virtual shadow and no virtual shadow. The experiment showed that the participants had a better spatial perception since they were less prone to go through the virtual objects whenever a virtual shadow was present. Moreover they appropriate the virtual shadow whenever its morphology was close enough to theirs. According to the subjective questionnaires, the participants felt more comfortable when using the application and they generally felt that the experience was more realistic with their virtual shadow.

In a nutshell, the results of the experiment promote the use of dynamic virtual shadows in PBS and lead the way for further studies on "virtual shadow ownership" toward a better appreciation of the virtual environment in PBS.

Conclusion

7

In this manuscript, entitled "**Contribution to the Study of Projection-based Systems for Industrial Applications in Mixed Reality**", we focused on improving Projection-Based Systems (PBS) and their usage in Mixed Reality (MR) for industrial purposes. Mixed reality proposes an alternative and faster way to dynamically visualize, modify and validate the design of many industrial projects. It also provides new dimensions for training and assisting operators when performing, e.g., maintenance tasks. In the future PBS could constitute an alternative to near-eye displays for industrial applications since they present some advantages compared to Head-Mounted Displays (HMD) such as direct collaboration, communication and in-situ projection. Nevertheless PBS generally require more space, are more expensive, less immersive and are stationary compare to many HMD. To address these limitations, we identified three main axes of research. The first axis aimed at **improving the optical tracking by increasing its workspace**. The second axis focused on **proposing a novel paradigm for mobile spatial augmented reality**. Finally the third axis focused on **increasing the user perception and experience when using projection-based systems**.

Chapter 3 introduced a **pilot study of the use of projection-based systems in an industrial context**. The study was carried out on a planning/validation application aimed for the construction industry. The users were immersed in a CAVE display and were presented with a virtual house. They were able to move freely and interact with the environment freely. We recorded their behavior during the overall experiment. The results highlighted trends in the usage of projection-based displays and promoted the use of projection-based systems for specific applications and usages. From this study we identified that the tracking system could be adapted to industrial usages of projection-based systems enabling the users to take advantage of the overall range of interactions and workspace provided. Also the applications could be revised to improve the user experience and propose more realistic real-life professional situations.

Chapter 4 focused on improving the optical tracking systems and proposed an approach to increase the optical tracking workspace for Virtual Reality (VR) Projection-Based Systems (PBS). Our approach is based on two methods. The first method, called *MonSterTrack* for Monocular Stereo Tracking, enables **switching between stereo and monocular tracking modes** when a tracked target is visible by only one camera. The second method, called *CoCaTrack* for Controlled Camera Tracking, is based on **controlled cameras that are able to follow a target across the workspace** and keep it in their field of view. Both methods can be combined to provide an even

larger workspace. Our approach has been tested on different VR systems: a holobench and a wall-sized projection-based display. It has also been tested on an Unmanned Aerial Vehicle (UAV) application and compared to the VICON optical tracking system. The resulting systems showed acceptable performances for MR applications while providing a large workspace and reducing some occlusion problems. The comparison with VICON was promising since our system is close enough to the VICON performance while using only two cameras.

Chapter 5 focused on proposing a novel paradigm for mobile Spatial Augmented Reality (SAR). We introduced *MoSART*, an **"all-in-one" heaset for mobile spatial augmented reality on tangible objects**. The MoSART proof-of-concept has been designed with both the projection and tracking systems mounted on an helmet. Interaction tools have been designed such as an interactive Stylus and interactive Panel. The projector displays virtual content over the interaction tools and tangible 3D objects that are tracked. Then the interaction tools allow the users to interact with the objects and modify the content that is projected over them. Two MoSART use cases have been proposed to validate the concept: virtual prototyping and medical visualization. The users were able to dynamically edit the texture of a tangible object or to visualize medical content for training purposes. MoSART enables straightforward, mobile and direct interaction with tangible 3D objects. A wide range of AR application that require mobility could benefit from MoSART in single or collaborative conditions.

Finally, Chapter 6 focused on introducing **users' virtual shadows in Projection-Based Systems (PBS)**. Our approach aimed at increasing the user embodiment and spatial perception in such systems. Indeed since the user's are able to see their own body, embodying an avatar is not as straightforward as in Head-Mounted Displays (HMD). We carried out a study to evaluate the influence of users' dynamic virtual shadows on the user's behavior. The shadow was directly mapped to the users motion, making the users fell like the shadow was their own. During the study the users were asked to position a real ball over virtual planar surfaces. They were presented with three virtual shadow conditions: a male virtual shadow, a female virtual shadow and no virtual shadow. The results have proven that the users have a better understanding of the virtual environment and that they better respected its physical limitations when the virtual shadow was present. Moreover they appropriate the virtual shadow whenever the shadow morphology was close enough to their own. Adding virtual shadows also increased their comfort and immersion when using the application.

From the different contributions of this manuscript, several paths of improvement based on our approaches could be investigated in short/middle-term.

Toward an increased tracking workspace for PBS

- **Adapting controlled cameras**: When using our controlled cameras tracking method (CoCaTrack), a calibration bias is introduced if the cameras are not stationary relatively to each other. Therefore, mounting several cameras on the

same motor could be of interest since the performance of stereo tracking may remain the same whatever the cameras movements. Also, we used only one constellation for the visual servoing process. But if several constellations are used, it is possible to follow the barycenter of all the constellations or the barycenter of a priority constellation. Future work could then compare these techniques for handling multiple constellations with CoCaTrack.

- **Adapting hardware components**: The hardware components (e.g., cameras, constellation) could be revised to provide better performance for the overall tracking system. Following the results of Vogt et al. [2002], the structure of the constellations could be studied to provide better performance in monocular mode (MonSterTrack). We used wide-angle lenses to maximize the workspace when no controlled camera was available. But wide-angle lenses induce a loss in resolution that can degrade the feature extraction and increase jitter. Thus, our approach could be tested with standard lenses. Higher quality sensors (e.g., high-resolution cameras) and/or hardware synchronization could also help reducing jitter and increasing tracking stability and accuracy, but at a higher cost.
- **Performance test and user studies**: The performances of both MonSterTrack and CoCaTrack showed limited jitter and acceptable accuracy for Virtual Reality (VR) applications. However a concrete user study could be carried out to evaluate the influence of both methods on the user experience and comfort when using a VR Projection-Based Systems (PBS) compared to the pilot study we carried out in the CAVE display.

Toward mobile spatial augmented reality

- **Adding stereoscopic projection**: In some cases, the MoSART system could benefit from stereo projection. By using a 3D projector and shutter glasses, 3D content could also be projected over tangible objects. It could provide depth perception such as the one provided by optical see-through AR devices. However, a stereoscopic rendering generally induces the additional need of glasses and might prevent some collaborative scenarios.
- **Handling focus issue**: The focus of the projector can be an issue with our current prototype of MoSART. Indeed since the tangible objects can be manipulated directly, the projection may be done at closer or further distances than the one on focus. This issue can be solved by using either a laser projector or auto-focus algorithms. Nevertheless, auto-focus algorithms could add some latency to the overall system.
- **Providing full portability**: The tracking and projection mapping computations of the MoSART prototype are currently done on an external computer. This computer could be embedded on a backpack together with a battery that could power the projector. The entire system could also be ultimately miniaturized and put inside the headset.
- **Handling occlusions**: The use of tangible tools with our MoSART approach can generate partial occlusion problems since the tools can sometimes be located

between the projector and the tangible object. The direct manipulation of the objects with the hands can also be a cause of partial occlusions. This issue could be dealt with by detecting occlusions with a depth sensor and then removing the projection over occluding parts. Work from Zhou et al. [2016] already proposes a solution to this problem as long as the occluding objects are not too close to the manipulated object.

- **Overcoming resolution issues**: When projecting over small surfaces with MoSART (e.g., the interactive panel) the resolution of the image can be rather limited since only a small portion of the projector will be used. Thus displaying detailed information and interacting with small virtual objects over these surfaces can be difficult. A solution to overcome this limitation could be to use a real interactive tablet.

Toward user's virtual shadows in PBS

- **Exploring more realistic/unrealistic shadows**: When studying the influence of the virtual shadows we considered arbitrary human 3D models to generate the virtual shadow of the users. Nevertheless some users may have not been morphologically identified to neither the female nor the male shadow. It might then be interesting to test our approach with the scanned 3D model of the users in the virtual environment. On the other hand, the 3D models used were not excessively different from what a human can expect from a shadow. Thus the morphology of the shadow was generally not disturbing. A user study with a remarkably different shadow (see Figure 7.1) could lead to different results in terms of ownership.

Figure 7.1 – Virtual embodiment through virtual shadows in a entertainment VR application – The virtual shadow of a cowboy (Inspired from the famous *Lucky Luke* comic book) is displayed in a far west virtual scene in a CAVE display.

- **Studying the lighting conditions**: As mentioned in chapter 6 the lighting of the virtual environment was chosen to provide a noticeable virtual shadow

whatever the user's position. Nevertheless we did not carry out any study to evaluate the influence of the lighting conditions on the virtual shadow perception. Such study could help creating more natural and realistic virtual shadow configurations.

- **Removing the real shadow**: In this manuscript we proposed an approach to add virtual shadows of the user in immersive Projection-Based Systems (PBS). We noticed that during the experiment the users were less prone to notice their real shadow that was projected on the floor screen of the PBS. This real shadow presents a limitation when PBS are front projected since it can break the immersion and occlude some virtual objects. Therefore an open question remains on how to remove this shadow or at least reduce its influence on user perception.

Regarding long-term perspectives we believe that projection-based mixed reality could present a affordable and competitive alternative to near-eye displays. As of today, Mixed Reality (MR) has taken the path of mounting the overall system on the user's head (e.g., Microsoft Hololens). The mobility brought by such systems make them attractive and future studies will probably focus on improving and adapting these technologies. Therefore we believe that the studies on projection-based systems should also follow this path and focus on providing mobility. For providing **mobility** we believe that the system should be **light** and **tiny**, in case of mounting the system on the head. Also the system should be as **unobstrusive** as possible and should not require wired connections in order to not bother the user experience and provide more liberty when moving. Finally the system should be fast to equip and remove.

Other requirements can be considered to design the futuristic PBS. Apart from mobility, the futuristic PBS should also be **independent** and should not rely on other systems. To be independent we believe that the system should first be **self-contained** and should embed all the necessary equipment (e.g., tracking system, projection system, rendering unit and power supply). Also the system should be **cross-application** and adapt itself to the different targeted applications. It should be able to provide both Augmented Reality (AR) and Virtual Reality (VR) environments without using external systems.

Finally, regarding the application and the user experience the system should be **immersive** and provide the user with real-life conditions in industrial environments. Also, it should be **interactive** in order to propose a large range of interaction scenarios and enough possibilities for designing adapted interaction techniques for each industrial task. Last but not least, the system should propose **direct collaboration** with external persons and should enable designing collaborative scenarios for the fulfillment of complex multi-user tasks in industrial conditions.

To sum up, the futuristic mixed reality projection-based system could be an all-in-one mobile system with embedded tracking, projection and rendering. This system should be capable of providing both VR and AR projective environments with adaptive interactions. Therefore, such a system could enable mobile and interactive experiences within immersive environments in single and collaborative scenarios.

Author's publications

Journals

- **G. Cortes**, E. Marchand, G. Brincin, and A. Lécuyer. "MoSART: Mobile Spatial Augmented Reality for 3D Interaction with Tangible Objects". In *Frontiers in Robotics and AI - Virtual Environments*, 2018.

International conferences

- **G. Cortes**, E. Marchand, J. Ardouin, and A. Lécuyer. "Increasing Optical Tracking Workspace of VR Applications using Controlled Cameras". In *Proc. of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 22–25, 2017.
- **G. Cortes**, E. Marchand, J. Ardouin, and A. Lécuyer. "An Optical Tracking System based on Hybrid Stereo/Single-View Registration and Controlled Cameras". In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, p. 6. 2017.
- **G. Cortes**, F. Argelaguet, E. Marchand, and A. Lécuyer. "Virtual Shadows for Real Humans in a CAVE: Influence on Virtual Embodiment and 3D Interaction". *ACM Symposium on Applied Perception (SAP)*, 2018.

National conferences

- **G. Cortes**, E. Marchand, J. Ardouin, and A. Lécuyer. "Increasing Optical Tracking Workspace of VR Applications using Controlled Cameras". In *Annual Conference of the AFRV*, 2017.

Appendix : Résumé long en français A

Introduction

Dans ce manuscrit de thèse, intitulé “Contribution à l’étude des systèmes de projection pour des applications industrielles en réalité mixte” nous présentons des travaux de recherche visant à améliorer les systèmes de projection et leur utilisation pour des applications industrielles en réalité mixte.

La Réalité Mixte (MR) a été introduite par Milgram et al. [1995] comme un concept qui combine plusieurs technologies et qui amène l’utilisateur du monde réel dans le monde virtuel, plus connu aujourd’hui comme Réalité Virtuelle (VR), en passant par la Réalité Augmentée (RA). La réalité virtuelle (VR) a été définie par Arnaldi et al. [2003] comme étant “un domaine scientifique et technique exploitant l’informatique et des interfaces comportementales en vue de simuler dans un monde virtuel le comportement d’entités 3D, qui sont en interaction en temps réel entre elles et avec un ou des utilisateurs en immersion pseudo-naturelle par l’intermédiaire de canaux sensori-moteurs”.En ce qui concerne la réalité augmentée (RA) nous considérons la définition proposée par Azuma et al. [2001] qui décrit la RA comme étant tout système qui dispose des trois caractéristiques suivantes: combine des objets réels et virtuels dans un environnement réel; est exécuté de manière interactive en temps réel et aligne les objets réels et virtuels entre eux.

Cette thèse a été conduite en partenariat CIFRE avec l’entreprise Realyz, spécialisée dans la conception de systèmes de projection pour des applications industrielles en VR et RA. Ces technologies étant relativement récentes, un des objectifs de Realyz est de les démocratiser au sein des petites et moyennes entreprises (PME). En effet pour les PME, la réalité mixte propose de nouvelles méthodes permettant de visualiser, modifier et valider dynamiquement et rapidement la conception de nombreux projets [Mourtzis et al., 2014]. Cependant afin d’attirer les plus petits acteurs industriels, la réduction des coûts des systèmes tout en conservant la qualité attendue est essentielle.

Défis des systèmes de projection et axes de recherche

Le figure A.1 illustre le fonctionnement standard d’un application de réalité mixte basée sur un système de projection. Ces applications disposent d’un système le localisation dont les données sont interprétées par des techniques d’interactions qui envoient des commandes et mettent à jour l’application. Ensuite l’application modifie l’état des interactions et est affichée sur un écran de projection.

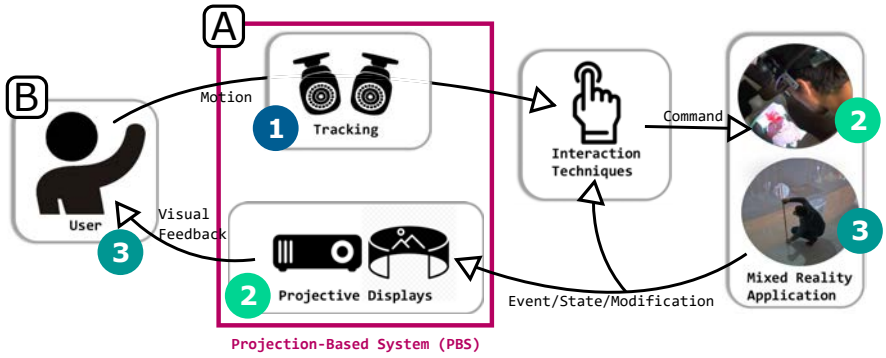


Figure A.1 – Fonctionnement général d'un système de projection pour la réalité mixte. Les deux principaux défis de la thèse sont: (A) améliorer les composants techniques des systèmes de projection et (B) améliorer l'expérience utilisateur dans les systèmes de projection. Les trois axes de recherche sont: (1) améliorer le système de localisation des systèmes de projection, (2) proposer un nouveau paradigme pour les systèmes de projection and (3) améliorer la perception utilisateur et son expérience dans les systèmes de projection.

Cette thèse adresse deux principaux défis qui peuvent améliorer l'utilisation des systèmes de projection: **(A) améliorer les composants techniques des systèmes de projection** et **(B) améliorer l'expérience des utilisateurs dans les systèmes de projection**.

Le premier défi vise donc à améliorer le système de localisation et l'interface visuelle de projection. En effet, Welch and Foxlin [2002] ont défini le système de localisation optimal comme étant: petit, indépendant, complet, rapide, précis, robuste, tenace, sans-fil, et pas cher. Cependant il n'existe pas aujourd'hui de système qui regroupe l'ensemble de ces caractéristiques et des besoins existent afin **d'optimiser et adapter le système de localisation aux exigences de l'application**. En ce qui concerne les interfaces visuelles de projection, ce sont des systèmes lourd et encombrants et qui sont généralement utilisés de manière statique. Il est donc nécessaire d'explorer les techniques permettant de rendre ces systèmes plus compacts et de **proposer une solution portable et mobile pour des systèmes de projection en réalité mixte**.

Le deuxième défi concerne l'expérience utilisateur. Lorsque les utilisateurs s'immergent dans un système de projection, ils ont généralement conscience de leur propre corps et peuvent souvent voir l'environnent qui les entoure. Ce phénomène rend l'incarnation virtuelle et l'appropriation d'un corps virtuel plus difficile et nécessite de **concevoir de nouvelles techniques immersives et des paradigmes d'interaction pour améliorer l'expérience utilisateur**.

Objectifs de la thèse et contributions

Nous considérons trois principaux axes de recherche: **(1) Améliorer les systèmes de localisation**, **(2) Proposer un nouveau paradigme pour les systèmes de projection en réalité mixte** et **(3) améliorer la perception et l'expérience utilisateur dans les systèmes de projection**. Ces trois axes de recherche ont donné lieu à trois contributions illustrées en Figure A.1 et A.2 et détaillées ci-dessous.

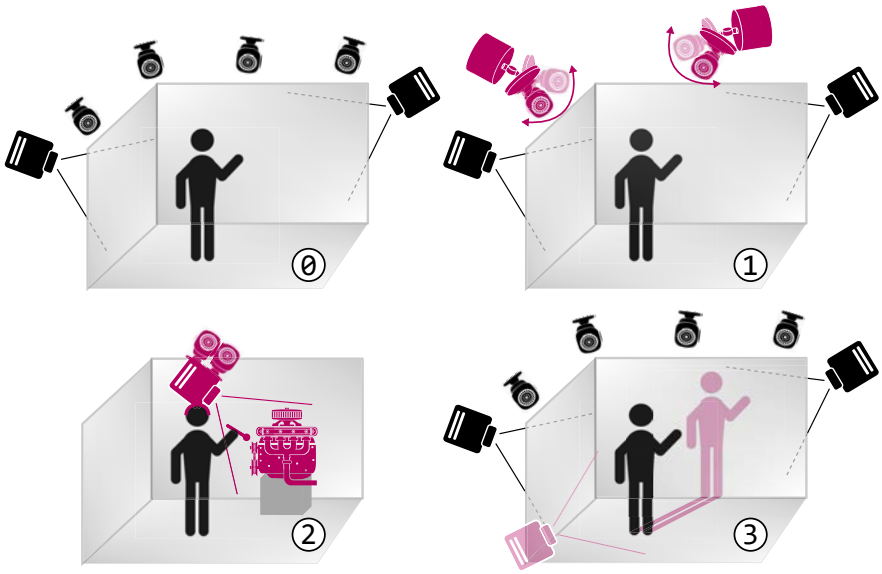


Figure A.2 – Les trois contribution de cette thèse. Les systèmes de projection standard sont illustrés en (0). La contribution (1) vise à augmenter le volume de travail des systèmes de localisation en utilisant, notamment des caméras contrôlées. La contribution (2) propose un nouveau paradigme pour de la réalité augmentée spatiale mobile. La contribution (3) vise à améliorer la perception et l'expérience utilisateur en introduisant leur ombre virtuelle dans les systèmes de projection.

(1) Dans cette axe de recherche (illustré en Figure A.2-haut-droite) nous répondons à certains besoins et caractéristiques des systèmes de localisation. Pour des applications de réalité mixte projective, ces systèmes ont notamment besoin d'être rapide, précis, complets et robustes. De plus ils devraient, autant que possible, s'adapter aux différentes interaction et localiser les objets aussi loin que possible. Finalement, d'un point de vue industriel ces systèmes doivent être au meilleur rapport qualité-prix possible. Ainsi nous proposons une approche permettant d'**augmenter le volume de travail et de réduire les problèmes d'occlusion des systèmes de localisation optique**.

(2) Les acteurs industriels souhaitent généralement collaborer en immersion dans des systèmes de réalité mixte et les systèmes de projection sont de bons candidats pour de telles fonctionnalités. Cependant leur installation n'est pas toujours évidente et requiert beaucoup de place. De plus la taille des systèmes et leur poids les rendent généralement inamovibles. Afin de combler à cette limitation en terme de mobilité, nous proposons de **concevoir une nouvelle approche pour de la réalité augmentée spatiale mobile sur des objets 3D tangibles** (Figure A.2-bas-gauche).

(3) Le présence et la conscience du monde réel peut, parfois, réduire l'immersion de l'utilisateur dans les systèmes de projection. De même il peut être plus difficile pour l'utilisateur d'incarner un personnage virtuel ou de s'approprier un corps virtuel. Dans cette contribution (illustrée en Figure A.2-bas-droite) nous proposons d'**introduire l'ombre virtuelle des utilisateur comme la projection de ces derniers dans l'environnement virtuel**. Cette approche devraient permettre d'augmenter l'expérience utilisateur dans les systèmes de projection immersifs.

A.1 Élargissement du volume de travail des systèmes de localisation optique

Les systèmes de localisation optique sont probablement les plus communément utilisés pour des applications de réalité mixte. Ces systèmes requièrent généralement une configuration stéréo (plusieurs caméras) pour fournir des données de localisation et présentent donc des limites en termes de volume de travail couvert. Nous proposons donc une approche permettant d'augmenter ce volume de travail sans avoir à rajouter de nombreux capteurs. L'approche est basée sur deux méthodes complémentaires (Figure A.3-gauche): *MonSterTrack* (Monocular and Stereo Tracking), permettant de basculer en localisation monoculaire lorsque la stéréo n'est plus disponible et *CoCaTrack* (Controlled Camera Tracking), permettant de contrôler les caméras pour suivre les marqueurs dans le volume de travail.

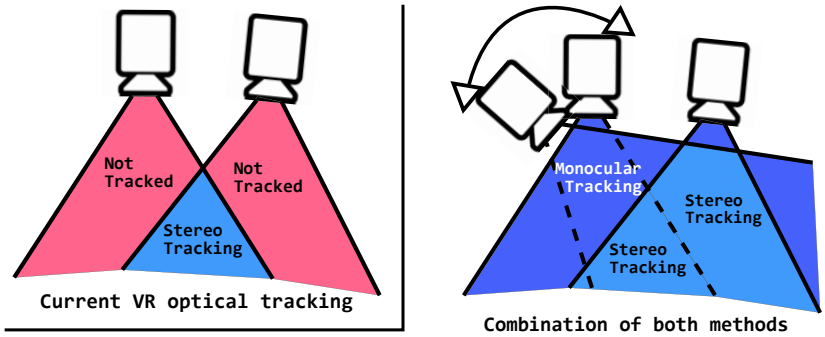


Figure A.3 – Par rapport aux systèmes de localisation optique actuels (gauche), notre approche augmente le volume de travail grâce à deux méthodes: basculer en localisation monoculaire lorsque la stéréo n'est plus disponible (MonSterTrack), ou utiliser des caméras contrôlées pour suivre les marqueurs dans le volume de travail (CoCaTrack). Les deux méthodes peuvent être combinées (droite) pour fournir un plus large volume de travail.

A.1.1 MonSterTrack: localisation hybride stéréo/monoculaire

MonSterTrack ("MONocular and STEReo TRACKing") permet de basculer, occasionnellement, vers la localisation monoculaire lorsque la stéréo n'est pas disponible. Ainsi des algorithmes de localisation 2D-3D sont utilisés si la localisation 3D-3D est impossible.

L'objectif de la localisation est de calculer la pose ${}^{o}\mathbf{M}_{w}$ d'un objet $(\mathcal{F}_o)$ dans le repère de référence $\mathcal{F}_w$. Le calcul de ${}^{o}\mathbf{M}_{w}$ se fait donc par ${}^{o}\mathbf{M}_{w} = {}^{o}\mathbf{M}_{c_1}\,{}^{c_1}\mathbf{M}_{w}$ où ${}^{c_1}\mathbf{M}_{w}$ est connue grâce à la calibration extrinsèque du système [Hartley and Zisserman, 2003]. Deux cas se présentent alors pour calculer ${}^{o}\mathbf{M}_{c_1}$:

- La stéréo est disponible. Dans ce cas la transformation ${}^{o}\mathbf{M}_{c_1}$ peut être calculée grâce à la localisation 3D-3D [Besl and McKay, 1992; Fitzgibbon, 2003; Marchand et al., 2016]. L'objectif est alors de minimiser l'erreur entre les points 3D

reconstruits de l'objet ${}^{c_1}\mathbf{X}_i$ (exprimés dans le repère de la caméra) et leur point 3D correspondants ${}^{o}\mathbf{X}_i$ (exprimés dans le repère de l'objet) transformés dans le repère de la camera grâce à ${}^{c_1}\mathbf{M}_o$:

$$ {}^{c_1}\widehat{\mathbf{M}}_o = \arg\min_{{}^{c_1}\mathbf{M}_o} \sum_{i=1}^{N} ({}^{c_1}\mathbf{X}_i - {}^{c_1}\mathbf{M}_o {}^{o}\mathbf{X}_i)^2. \tag{A.1} $$

- La stéréo n'est pas disponible. Dans ce cas la transformation ${}^{o}\mathbf{M}_{c_1}$ peut être calculée grâce à la localisation 2D-3D [Dementhon and Davis, 1995; Kneip et al., 2011]. L'objectif est alors de minimiser l'erreur entre les points 2D de l'image, $\mathbf{x}_i$, représentant la projection des points 3D de l'objet dans la caméra, et leur point 3D correspondants ${}^{o}\mathbf{X}_i$ (exprimés dans le repère de l'objet) transformés dans le repère de la camera grâce à ${}^{c_1}\mathbf{M}_o$ et projetés dans la caméra grâce à $\mathbf{\Pi}$:

$$ {}^{c}\widehat{\mathbf{M}}_o = \arg\min_{{}^{c}\mathbf{M}_o} \sum_{i=1}^{N} d(\mathbf{x}_i, \mathbf{\Pi}\, {}^{c}\mathbf{M}_o {}^{o}\mathbf{X}_i)^2 \tag{A.2} $$

A.1.2 CoCaTrack: utilisation de caméras contrôlées

CoCaTrack ("COntrolled CAmera Tracking") permet de contrôler les caméras pour qu'elles suivent les objets dans l'espace. Même si une seule des caméras est contrôlées, le volume de travail augmente et, en combinant avec la méthode MonSterTrack, le volume de travail monoculaire est encore plus large.

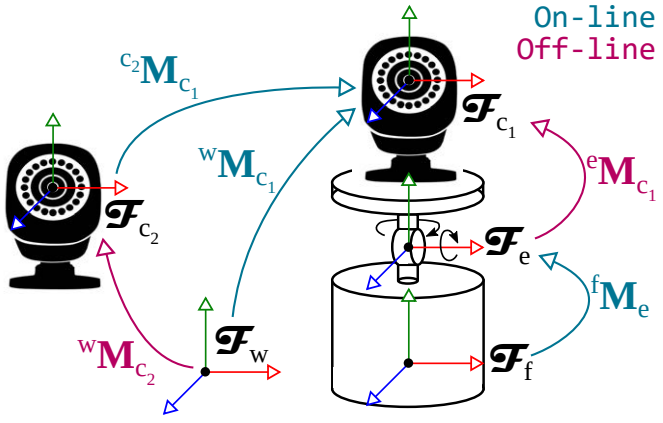


Figure A.4 – Configuration de deux caméras avec un moteur pan-tilt. Les transformations $\mathbf{M}$ définissent l'intégralité du système.

L'ensemble des transformations intervenant dans une installation avec des caméras contrôlées, est illustré en Figure A.4. Nous obtenons alors ${}^{w}\mathbf{M}_{c_1}$ à l'instant t grâce à l'équation suivante:

$$ {}^{c_1(t)}\mathbf{M}_w = {}^{w}\mathbf{M}_{c_1(0)} {}^{c_1(0)}\mathbf{M}_{e(0)} {}^{e(0)}\mathbf{M}_{e(t)} {}^{e(t)}\mathbf{M}_{c_1(t)}. \tag{A.3} $$

${}^{w}\mathbf{M}_{c_1(0)}$ est connue par la calibration extrinsèque à l'instant 0. La transformation ${}^{e(0)}\mathbf{M}_{e(t)}$ est fournie par les mesure d'odométrie. Ainsi la seule transformation inconnues

est ${}^{e}\mathbf{M}_{c_1} = {}^{c_1(0)}\mathbf{M}_{e(O)} = {}^{e(t)}\mathbf{M}_{c_1(t)}$. Cette transformation peut être calculée grâce à une étape de calibration semblable à celle proposée par Tsai and Lenz [1989].

Les caméras sont contrôles grâce à des algorithmes d'asservissement visuel [Chaumette and Hutchinson, 2006]. L'asservissement visuel permet d'estimer la vitesse $\mathbf{v}$ qui doit être appliquée au robot pour que la projection des objets dans la caméra se trouvent au plus proche du centre de l'image. La loi de contrôle est classiquement donnée par:

$$\mathbf{v} = -\lambda \mathbf{L}_{\mathbf{x}}^{+}(\mathbf{x}(\mathbf{r}) - \mathbf{x}^{*}) \tag{A.4}$$

où λ est un scalaire positif et $\mathbf{L}_{\mathbf{x}}^{+}$ est la pseudo inverse de la matrice d'interaction [Chaumette and Hutchinson, 2006].

A.1.3 Preuve de concept

Deux prototypes ont été réalisés afin de valider les deux méthodes, MonSterTrack et CoCaTrack. Les deux prototypes utilisent des caméras infrarouges et des marqueurs qui sont attachés à un objet et qui émettent de la lumière dans l'infrarouge.

Un premier prototype a été réalisé sur un système de réalité virtuelle de type "holobench" grâce à deux caméras Sony PSEye (320x240, 150Hz). Les caméras disposent de lentilles courte focale et fournissent un champ de vision de 87° par 70° chacune.

Le deuxième prototype se base sur les mêmes caméras dont une est embarquée sur un robot pan-tilt. Ce prototype a été réalisé sur un système réalité virtuelle projectif où le contenu virtuel est projeté sur un mur (Figure A.5). Le robot pan-tilt permet des mouvements de $-170°$ à $+170°$ pour l'axe vertical (pan) et de $-60°$ à $+60°$ pour l'axe horizontal (tilt), le tout avec une résolution de 0.03°.

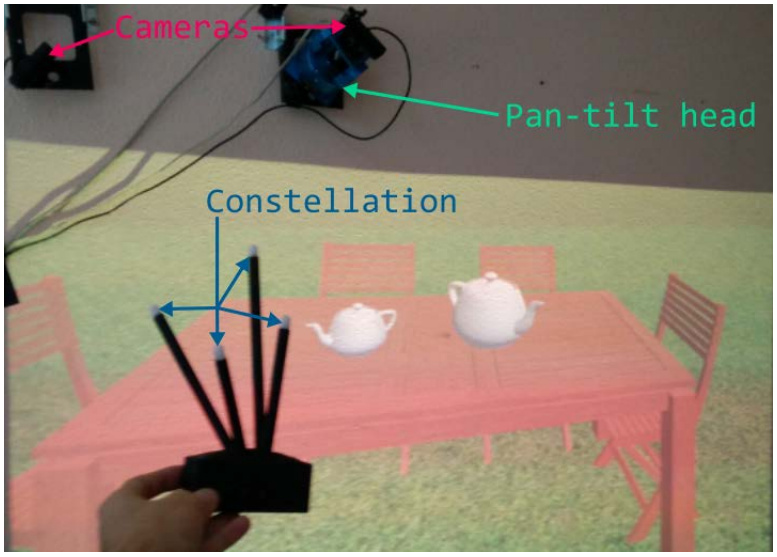


Figure A.5 – Second prototype, combinant les deux méthodes (MonSterTrack et CoCaTrack), réalisé sur un système de projection pour la réalité virtuelle.

A.1.4 Résultats

MonSterTrack et CoCaTrack ont été évaluées sur le deuxième prototype en termes de precision, bruit, latence et gain en volume de travail. Le gain en volume de travail a été estimé à approximativement 100% mais ce gain varie en fonction du positionnement initial des caméras. La combinaison des deux approches permet d'élargir encore plus le volume de travail (Figure A.6).

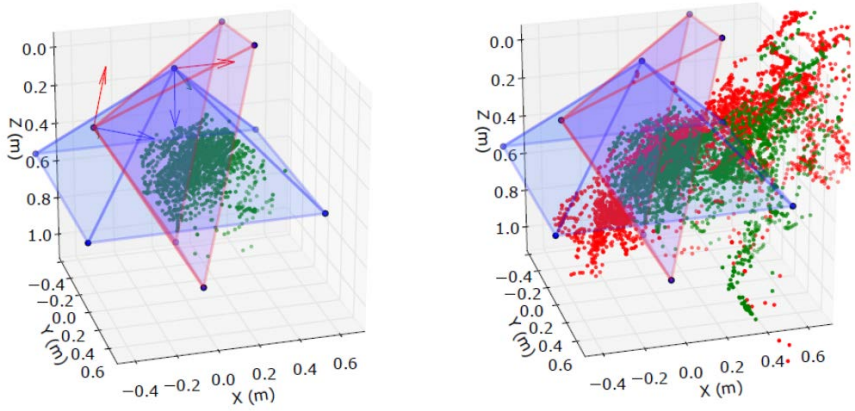


Figure A.6 – Gain en volume de travail de MonSterTrack et CocCaTrack (droite) par rapport à un système stéréo standard (gauche): Les points rouges représentent les objets localisés en monoculaire et les points vert en stéréo. La caméra bleue est contrôlée grâce à un moteur pan-tilt. La caméra rouge est fixe.

En ce qui concerne la précision (1.0mm), le bruit (±0.08mm) et la latence (60ms de bout en bout), les performances sont acceptables pour une utilisation en réalité mixte. Nous constatons, comme attendu, une légère baisse des performances avec le localisation monoculaire par rapport à la localisation stéréo. Cependant cette légère différence ne devrait pas impacter l'expérience utilisateur et une étude plus approfondie de cet impact pourrait être envisagée.

A.2 Réalité augmentée spatiale mobile pour l'interaction 3D avec des objects tangibles

Permettre aux utilisateur de se déplacer tout en projetant du contenu virtuel sur des objets tangibles peut être un atout pour de nombreuses applications industrielles qui requièrent de la mobilité. En effet la plus part des systèmes de projection utilisés en réalité mixte sont statiques et difficilement déplaçables. Nous proposons donc MoSART (Mobile Spatial Augmented Reality on Tangible objects) un concept de réalité augmenté spatiale mobile permettant de projeter du contenu virtuel sur des objets 3D tangibles.

A.2.1 MoSART

L'approche MoSART fournit des interaction mobiles avec des objets tangibles grâce à des systèmes de projection et localisation embarqués sur un casque. MoSART permet

également de facilement et directement manipuler des objects 3D tangibles tout en interagissant avec eux grâce à des outils d'interaction. Grâce au système de projection, MoSART permet de partager l'expérience virtuelle avec des personnes extérieures au systèmes comme par exemple dans le cas de scénarios collaboratifs (Figure A.7) .

Figure A.7 – L'approche MoSART conçue grâce à un casque "tout-en-un" pour la réalité augmentée spatiale mobile (gauche). L'utilisateur peut manipuler directement les objets tangible (milieu) et des outils tangibles peuvent être utilisés pour modifier les objets (droite).

Les principaux composants de MoSART sont: une projection embarquée, une localisation embarquée, des objets tangibles et des outils tangibles. Un processus de "projection mapping" est utilisé afin de projeter le contenu virtuel au bon endroit sur l'objet tangible. Le "projection mapping" consiste donc à faire correspondre l'image virtuelle 2D à un object 3D tangible. Afin d'atteindre cette objectif l'application a besoin de connaître le forme de l'objet et le modèle de projection. En ce qui concerne la forme de l'objet, un scan 3D est effectué sur l'objet réel afin d'en obtenir un objet virtuel ([Newcombe et al., 2011]). Le modèle de projection du projecteur est nécessaire pour connaître la projection d'un point 3D dans le repère image du projecteur. Le modèle du projecteur est calculé grâce une étape de calibration effectuée au préalable ([Yang et al., 2016]) et permet de réaliser une projection inverse (3D vers 2D). Une fois que la forme de l'objet et le modèle de projection sont connus, la scène virtuelle 3D correspond à la scène réelle. Le système de localisation fourni ensuite la position de l'objet par rapport au projecteur et l'image projetée est modifiée afin d'être affichée correctement sur l'objet réel.

Un prototype de MoSART a été conçu et comprend, un système de localisation basé sur le système présenté en partie A.1 et un pico-projecteur, le tout embarqué sur un casque. Des techniques d'interaction ont été proposées afin d'interagir avec des objet tangibles grâce à des outils tangibles dédiés (Figure A.8).

A.2.2 Outils d'interaction

Deux outils d'interaction ont été conçus pour MoSART basé sur les travaux de Marner et al. [2009] et Marner and Thomas [2010]: un panneau interactif et un stylet interactif illustrés en Figure A.8. Le panneau interactif est majoritairement utilisé comme un écran et permet d'afficher directement une interface visuelle 2D. Son rôle peut varier en fonction des applications et il peut par exemple jouer le rôle d'une loupe. En ce qui concerne le stylet interactif il est principalement utilisé comme outil de sélection afin de

Figure A.8 – Outils d'interaction tangibles de MoSART: le panneau interactif (gauche) et le stylet interactif (droite).

sélectionner les objets affichés sur le panneau interactif en les touchant. D'autres comportement peuvent être envisagés pour le stylet interactif tel qu'un feutre ou pinceau.

A.2.3 Cas d'usage illustratifs

Le concept MoSART a été illustré sur deux cas d'usage: le prototypage virtuel (Figure A.9) et la visualisation médicale (Figure A.10). Pour du prototypage virtuel l'utilisateur va d'abord pouvoir sélectionner la couleur de son choix sur le panneau interactif grâce au stylet. Ensuite le stylet est utilisé comme un pinceau et l'utilisateur a la possibilité de peindre directement sur l'objet 3D tangible. En ce qui concerne la visualisation médicale, l'application a été conçue à des fins de formations. Un buste de mannequin est utilisé comme objet tangible et MoSART affiche alors le contenu virtuel directement sur le buste. L'utilisateur peut ensuite manipuler le buste afin de mieux appréhender sa composition. Avec ce genre d'approches, MoSART pourrait être utilisé afin de projeter du contenu directement sur un patient.

Figure A.9 – Peinture virtuelle grâce aux outils tangibles de MoSART.

Figure A.10 – Visualisation médicale sur un buste tangible grâce à MoSART. L'utilisateur peut visualiser les os (gauche) avec ou sans les organes (droite).

A.3 L'ombre virtuelle des utilisateurs dans les systèmes de projection

Visualiser son corps ou incarner un corps virtuel en réalité mixte a été un des principaux défis. Dans les systèmes de type casque, l'utilisation d'avatars 3D représentant les utilisateurs leur permet de s'approprier et contrôler ce corps. Cependant, dans les systèmes de projection, les utilisateurs ont conscience et voient leur propre corps, rendant l'utilisation d'avatars plus délicate. Nous proposons donc d'introduire l'ombre virtuelle des utilisateurs afin de leur fournir une représentation virtuelle dans les systèmes de projection.

A.3.1 Création des ombres virtuelles

L'ombre virtuelle et dynamique des utilisateurs est introduite dans les systèmes de projection grâce à l'utilisation de modèles 3D humanoïdes. Pour l'expérience nous avons choisi une modèle 3D masculin et féminin. Les modèles ont été choisi afin de correspondre aux proportions du corps humain. Ainsi la correspondance entre le corps physique et l'ombre est simplement réalisée en changeant l'échelle du modèle virtuel. Afin de rendre les ombres parfaitement dynamiques et de faire correspondre la position des ombres à la position réelle de l'utilisateur (Figure A.11), des algorithmes de cinématique inverse sont utilisés. La cinématique inverse permet d'estimer la position du corps virtuel par rapport à la position réelle de l'utilisateur. Pour ce faire les pieds, mains, tête et ceinture de l'utilisateur sont localisées grâce à un système de localisation optique infrarouge. Enfin, la mise en place d'un éclairage virtuelle dans la scène permet de crée les ombres virtuelles qui sont enfaite les ombres des différents modèles 3D.

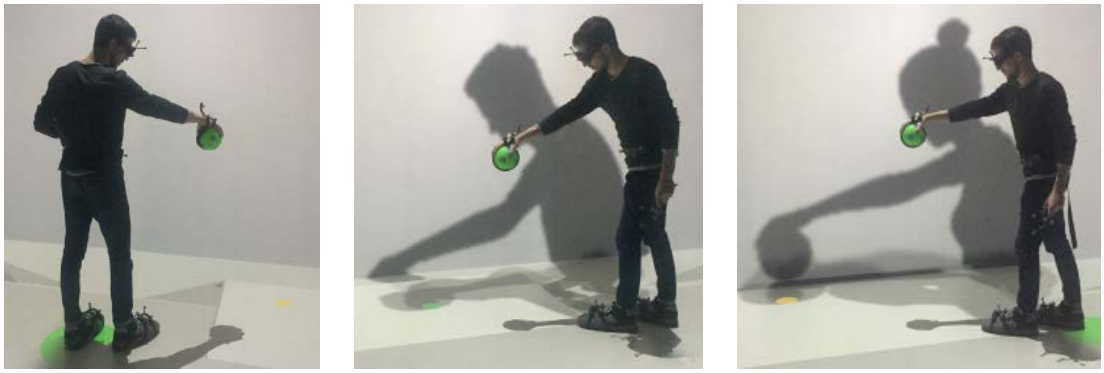


Figure A.11 – Les ombres virtuelles sont introduites dans les systèmes de projection grâce à des modèles 3D et permettent de projeter les utilisateurs dans le monde virtuel et de leur fournir une représentation virtuelle de leur corps.

A.3.2 Étude utilisateur

Un étude utilisateur a été conduite afin d'évaluer l'influence des ombres virtuelles sur leur perception spatiale et sur leur sentiment d'incarnation virtuelle. Pour cela, les utilisateurs étaient immergés dans un système de type CAVE et avaient pour tâche le positionnement d'une balle réelle sur des surfaces planes virtuelles. Différentes conditions d'ombres virtuelles leur étaient proposées: une ombre virtuelle masculine, féminine et pas d'ombre virtuelle.

Les résultats ont montré que les utilisateurs avaient une meilleur notion de l'espace et respectaient plus les limites physiques de l'environnement virtuel lorsqu'un ombre virtuelle était présente. De plus l'environnent virtuel leur semblait plus réaliste en présence de leur ombre virtuelle. En terme d'appréciation de la tâche, les utilisateurs ont eu l'impression que l'ombre les aidait à positionner la balle, ils se sont sentis plus à l'aise et trouvaient la tâche plus simple en présence de l'ombre (Figure A.12). Finalement les utilisateurs avaient globalement un bon sens d'incarnation, d'appropriation et de contrôle du corps virtuel (Figure A.12). Nous avons néanmoins remarqué que l'appropriation du corps virtuel était plus poussée lorsque la morphologie de l'ombre était proche de celle de l'utilisateur.

Conclusion

Dans ce manuscrit de thèse, nous avons visé à améliorer les systèmes de projection et leur utilisation pour des applications industrielles en réalité mixte. La réalité mixte permet de visualiser, modifier et valider la conception de nombreux projets industriels. Les systèmes de projection permettent la collaboration directe avec des utilisateurs externes grâce à de la projection in-situ. Cependant ils nécessitent beaucoup de place pour être installés, sont cher et généralement inamovibles. Dans ce cadre nos objectifs étaient: (1) **améliorer les systèmes de localisation en augmentent leur volume de travail**, (2) **proposer un nouveau paradigme pour de la réalité augmentée spatiale mobile** et (3) **augmenter la perception et l'expérience utilisateur**

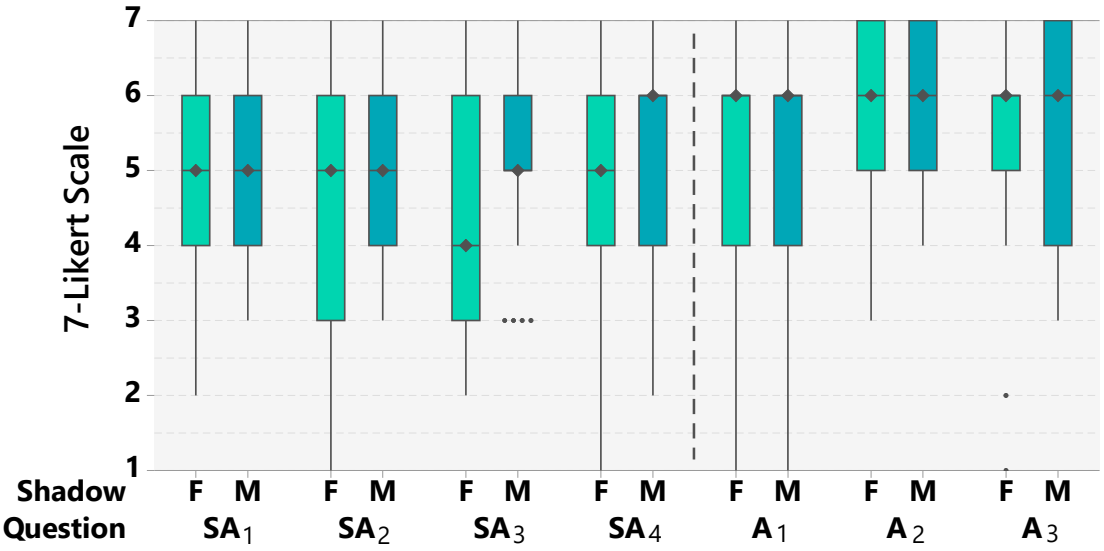


Figure A.12 – Résumé des notes d'appréciation (SA) et contrôle de l'ombre (A). En général les participants ont apprécié la présence de l'ombre et on remarqué avoir un bon contrôle du corps virtuel représenté par cette dernière.

dans les systèmes de projection.

Afin d'améliorer les systèmes de localisation nous avons proposé une approche basé sur deux méthodes complémentaires. La première méthode permet de **basculer sur la localisation monoculaire lorsque la localisation stéréo n'est pas disponible** et que l'objet n'est visible que par une caméra. La deuxième méthode fait usage de **caméras contrôlées qui sont capables de suivre les objets dans le volume de travail** et donc de les localiser plus longtemps. Ces deux méthodes peuvent être combinées pour fournir un volume de travail encore plus grand. Notre approche a été testée sur différents systèmes de réalité virtuelle. Le système final dispose de performance acceptables pour de nombreuses applications de réalité mixte tout en augmentant le volume de travail et réduisant certains problèmes d'occlusion.

Concernant la mobilité des systèmes de projection, nous avons introduit *MoSART*, un **casque "tout-en-un" pour de la réalité augmenté spatiale mobile sur des objets tangibles**. La preuve de concept de MoSART a été conçue avec un système de localisation et un système de projection embarqués sur un casque. Des outils d'interaction ont été proposés tels qu'un stylet et un panneau interactifs. Le projecteur affiche du contenu virtuel sur les outils d'interactions et les objets 3D tangibles qui sont localisés. L'utilisateur peut alors modifier le comportement des objets tangibles grâce aux outils d'interaction. Deux cas d'utilisation de MoSART on été proposé afin de valider le concept: le prototypage virtuel et la visualisation médicale. MoSART permet une interaction direct, mobile et simple avec des objets 3D tangibles. De nombreuses application de réalité augmentée mobile pourraient tirer profit de MoSART pour des scénarios collaboratifs ou avec un seul utilisateur.

Finalement, nous avons proposé d'**introduire l'ombre virtuelle des utilisateurs dans les systèmes de projection**. Notre approche vise à améliorer l'incarnation virtuelle et la perception des utilisateurs dans de tels systèmes. Nous avons conduit

une étude utilisateur afin d'évaluer l'influence de l'ombre virtuelle et dynamique sur leur comportement. L'ombre virtuelle est directement corrélée aux mouvements des utilisateurs afin de leur faire ressentir que c'était la leur. Trois conditions d'ombres leur étaient proposées: une ombre masculine, une ombre féminine et pas d'ombre. Les résultats ont prouvé que les utilisateurs avaient une meilleure compréhension de l'environnement virtuel et de ses limites physiques lorsque l'ombre était présente. De plus il s'appropriaient d'avantage l'ombre lorsque cette dernière avait une morphologie proche de la leur. Ajouter les ombres virtuelles a également permis d'augmenter le confort et l'immersion des utilisateurs.

Les travaux présentés dans ce manuscrit de thèse pourraient faire l'objet de recherches additionnelles. Premièrement, dans l'optique de couvrir un plus grand espace de travail nous pourrions adapter les caméras contrôlées afin de fixer différentes caméras au même moteur. De même les composant matériels pourraient faire l'objet d'études afin de fournir, potentiellement, de meilleurs résultats. Ensuite, concernant la réalité augmentée mobile, une étude du système pourrait être envisagée afin d'augmenter sa portabilité. Par ailleurs les occlusions générées sur la projection pourraient être détectées et la projection pourrait être adaptée. Finalement, des ombres virtuelles alternatives pourraient être explorées afin d'étudier le comportement de l'utilisateur face à des morphologies d'ombres moins réalistes. Aussi, les ombres étant générées par les conditions d'éclairage, ces dernières pourraient être explorées afin de fournir des configurations d'ombres adaptées à une meilleur immersion, incarnation et appropriation du corps virtuel.

List of Figures

1.1 Virtuality continuum . . . 1
1.2 Simplified classification of mixed reality systems . . . 2
1.3 Examples of mixed reality display . . . 3
1.4 Mixed reality industrial use cases . . . 3
1.5 Global framework of mixed reality projection-based systems and research axes of the PhD thesis . . . 4
1.6 Illustration of the PhD thesis contributions . . . 6

2.1 Classification of MR visual displays . . . 12
2.2 Examples of virtual reality near-eye displays . . . 13
2.3 Video see-though and Optical see-through technologies . . . 14
2.4 Examples of optical see-though head-mounted displays . . . 15
2.5 Examples of handheld displays . . . 16
2.6 Examples of stereoscopic glasses . . . 17
2.7 Examples of workbench displays . . . 18
2.8 Projection-based display technology . . . 19
2.9 Examples of surround-screen displays . . . 20
2.10 Examples of spatial augmented reality displays . . . 21
2.11 Examples of handheld projector displays . . . 22
2.12 Examples of head-mounted projector displays . . . 23
2.13 Examples of mechanical tracking systems . . . 25
2.14 Examples of magnetic tracking systems . . . 27
2.15 Position and orientation integration process in Inertial Measurement Units . . . 28
2.16 Examples of hybrid tracking devices . . . 29
2.17 Inside-Out and Outside-In tracking spatial arrangements . . . 30
2.18 Example of natural features used in optical tracking devices . . . 31
2.19 Examples of markers used in optical tracking devices . . . 32
2.20 Overall rigid body tracking workflow . . . 32
2.21 Constellation structure and Constellation 3D cloud . . . 33
2.22 Patterns used for internal camera calibration . . . 33
2.23 Examples of tracking techniques used in surround screen displays environments . . . 35
2.24 Industrial fields using Mixed Reality applications . . . 38
2.25 Mixed Reality training applications for the industry . . . 40
2.26 Example of AR industrial applications for maintenance and task assistance . . . 41
2.27 Design applications of Mixed Reality in the industry . . . 42
2.28 Planning applications of Mixed Reality in the industry . . . 44

3.1 3D scene of our VR application dedicated to the construction field 49
3.2 Our pilot study VR application in use 50
3.3 Workspace covered by the tracking system used in our pilot study 50
3.4 3D position of the head and hand of the users in the CAVE 53
3.5 Influence of the interaction state and the user's past experience with VR on head and hand 3D positions 53
3.6 Influence of the interaction state and the user's past experience with VR on 3D head and hand orientations. 54
3.7 Influence of interaction state and user's past experience with VR on the speed of the head and the hand 55

4.1 Illustration of MonSterTrack and CoCaTrack compared to current optical tracking systems 61
4.2 Global architecture of our approach for maximizing the tracking workspace 62
4.3 Central projection on a focal plane 63
4.4 Impact of different radial distortions on an image. 64
4.5 Standard two-view situation 65
4.6 Standard two-view optical tracking configuration 68
4.7 Different chessboard views used for internal camera calibration 69
4.8 Calibration of the relative transformation between two cameras 70
4.9 Infrared view of the camera before and after the threshold 72
4.10 Example of epipolar configuration 72
4.11 Cameras configuration with two cameras and one pan-tilt head 79
4.12 Frame configuration for controlled camera calibration with 2 camera positions 79
4.13 Prototype 1: Holobench with MonSterTrack 82
4.14 Prototype 2: Wall-sized display with MonSterTrack and CoCaTrack 83
4.15 Workspace gain of MonSterTrack and/or CoCaTrack compared to state-of-the-art stereo optical tracking 84
4.16 Calibration bias 85
4.17 Positional jitter 86
4.18 Rotational jitter 86
4.19 Transition between stereo and monocular tracking modes 87
4.20 Proof-of-concept for UAV localization. 88
4.21 Pose comparison with Vicon's optical tracking 89
4.22 Positional jitter comparison with Vicon's optical tracking 89

5.1 Our MoSART approach enables Mobile Spatial Augmented Reality on Tangible objects 92
5.2 Main components of the MoSART approach 93
5.3 Prototype of MoSART headset 95
5.4 Examples of tangible objects augmented with MoSART 95
5.5 Projection mapping pipeline 97
5.6 Frame configuration for the MoSART calibration process 98
5.7 Tangible interaction tools of MoSART 100
5.8 Tangible interaction tools in use 100
5.9 Texturing objects with MoSART 102

5.10 Virtual painting using MoSART . . . 103
5.11 Medical visualization on a tangible chest with MoSART . . . 104
5.12 Exploration of 3D medical models . . . 104
5.13 Concept of a collaborative setup with two MoSART systems (photomontage) 105

6.1 3D models used to cast the virtual shadow . . . 111
6.2 Tracking devices used during the experiment . . . 112
6.3 Ball offset calibration step . . . 112
6.4 Virtual environment of the experiment . . . 113
6.5 Virtual shadow conditions of the experiment . . . 114
6.6 Depth positioning error . . . 116
6.7 Task completion time . . . 117
6.8 Shadow appreciation and agency questionnaire ratings . . . 118
6.9 Presence and task appreciation questionnaire ratings . . . 118
6.10 Ownership questionnaire ratings . . . 119

7.1 Virtual embodiment in VR projection-based systems through user's virtual shadow . . . 126

A.1 Fonctionnement général d'un système de projection pour la réalité mixte et les axes de recherche de la thèse. . . . 132
A.2 Illustration des contributions de la thèse . . . 133
A.3 Illustration de MonSterTrack et CoCaTrack . . . 134
A.4 Configuration de deux caméras avec un moteur pan-tilt . . . 135
A.5 Prototype 2: système de projection en réalité virtuelle avec MonSterTrack et CoCaTrack . . . 136
A.6 Gain en volume de travail de MonSterTrack et CocCaTrack par rapport à un système stéréo standard . . . 137
A.7 Notre approche: MoSART . . . 138
A.8 Outils d'interaction tangibles de MoSART . . . 139
A.9 Peinture virtuelle avec MoSART . . . 139
A.10 Visualisation médicale sur un buste tangible grâce à MoSART . . . 140
A.11 Les ombres virtuelle des utilisateurs dans les systèmes de projection . . . 141
A.12 Notes di questionnaire d'appréciaiton et contrôle de l'ombre virtuelle . . . 142

Bibliography

Agrawala, M., Beers, A. C., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. (1997). The two-user responsive workbench: support for collaboration through individual views of a shared space. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 327–332. ACM Press/Addison-Wesley Publishing Co. 18

Akşit, K., Kade, D., Özcan, O., and Ürey, H. (2014). Head-worn mixed reality projection display application. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, page 11. ACM. 23

Albani, J. M. and Lee, D. I. (2007). Virtual reality-assisted robotic surgery simulation. *Journal of Endourology*, 21(3):285–287. 40

Aliaga, D. G., Yeung, Y. H., Law, A., Sajadi, B., and Majumder, A. (2012). Fast high-resolution appearance editing using superimposed projections. *ACM Transactions on Graphics (TOG)*, 31(2):13. 20

Amatriain, X., Kuchera-Morin, J., Hollerer, T., and Pope, S. T. (2009). The allosphere: Immersive multimedia for scientific discovery and artistic exploration. *IEEE MultiMedia*, (2):64–75. 20

Ameller, M.-A., Triggs, B., and Quan, L. (2000). Camera pose revisited–new linear algorithms. 76

Arnaldi, B., Fuchs, P., and Tisseau, J. (2003). Chapitre 1 du volume 1 du traité de la réalité virtuelle. *Les Presses de l'Ecole des Mines de Paris*. 1, 131

Arnaldi, B., Guitton, P., Fuchs, P., and Moreau, G. (2006). Le traité de la réalité virtuelle volume 4 – les applications de la réalité virtuelle. *Les Presses de l'Ecole des Mines de Paris*. 43

Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700. 74

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications*, 21(6):34–47. 1, 131

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and virtual environments*, 6(4):355–385. 1

Ban, Y., Narumi, T., Tanikawa, T., and Hirose, M. (2015). Modifying a body perception and an actual motion by visual stimuli of distorted physical exercise using shadow metaphor. In *In proceedings of ASIAGRAPH Conference*, pages 53–58. 109

Bar-Shalom, Y. and Li, X.-R. (1993). *Estimation and Tracking, Principles, Techniques, and Software.* Artech House, Boston. 75

Baratoff, G. and Regenbrecht, H. (2004). Developing and applying ar technology in design, production, service and training. In *Virtual and augmented reality applications in manufacturing*, pages 207–236. Springer. 38

Baricevic, D., Lee, C., Turk, M., Hollerer, T., and Bowman, D. A. (2012). A hand-held ar magic lens with user-perspective rendering. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 197–206. IEEE. 15, 16

Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H., and Srinivasan, M. A. (2004). Haptics in minimally invasive surgical simulation and training. *IEEE computer graphics and applications*, 24(2):56–64. 40

Benko, H., Jota, R., and Wilson, A. (2012). Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 199–208. ACM. 20, 21

Benko, H., Ofek, E., Zheng, F., and Wilson, A. D. (2015). Fovear: Combining an optically see-through near-eye display with projector-based spatial augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 129–135. ACM. 21

Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics. 73, 134

Bimber, O. and Raskar, R. (2005). *Spatial augmented reality: merging real and virtual worlds.* CRC press. 12, 92

Bishop, G., Welch, G., and Allen, B. D. (2001). Tracking: Beyond 15 minutes of thought. *SIGGRAPH Course Pack*, pages 6–11. 26, 27, 28

Bolas, M. and Krum, D. (2010). Augmented reality applications and user interfaces using head-coupled near-axis personal projectors with novel retroreflective props and surfaces. In *Pervasive 2010 Ubiprojection Workshop.* 22, 23

Boud, A. C., Haniff, D. J., Baber, C., and Steiner, S. (1999). Virtual reality and augmented reality as a training tool for assembly tasks. In *Information Visualization, 1999. Proceedings. 1999 IEEE International Conference on*, pages 32–36. IEEE. 39

Bowman, D. A., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. (2004). *3D user interfaces: theory and practice.* Addison-Wesley. 2, 13, 17, 19

Brough, J. E., Schwartz, M., Gupta, S. K., Anand, D. K., Kavetsky, R., and Pettersen, R. (2007). Towards the development of a virtual environment-based training system for mechanical assembly operations. *Virtual reality*, 11(4):189–206. 40

Brown, D. C. (1971). Close-range camera calibration. *Photogram. Eng. Remote Sens*, 37:855–866. 63, 70

Brown, D. G., Coyne, J. T., and Stripling, R. (2006). Augmented reality for urban skills training. In *Virtual Reality Conference, 2006*, pages 249–252. IEEE. 40

Brown, L. D., Hua, H., and Gao, C. (2003). A widget framework for augmented interaction in scape. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 1–10. ACM. 100

Bruder, G., Steinicke, F., Rothaus, K., and Hinrichs, K. (2009). Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *International Conference on CyberWorlds (CW)*, pages 43–50. IEEE. 1

Buriol, T. M., Rozendo, M., de Geus, K., Scheer, S., and Felsky, C. (2009). A virtual reality training platform for live line maintenance of power distribution networks. In *ICBL2009-International Conference on Interactive Computer Aided Blended Learning*, pages 1–13. 40

Butterworth, J., Davidson, A., Hench, S., and Olano, M. T. (1992). 3dm: A three dimensional modeler using a head-mounted display. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 135–138. ACM. 42

Büttner, S., Funk, M., Sand, O., and Röcker, C. (2016). Using head-mounted displays and in-situ projection for assistive systems: A comparison. In *Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments*, page 44. ACM. 41

Calderon, C., Cavazza, M., and Diaz, D. (2003). A new approach to virtual design for spatial configuration problems. In *Proceedings of the 7th International Conference on Information Visualization*, pages 518–523. IEEE. 43

Cao, X., Forlines, C., and Balakrishnan, R. (2007). Multi-user interaction using handheld projectors. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 43–52. ACM. 21, 22

Caruso, G. and Re, G. M. (2010). Interactive augmented reality system for product design review. In *The Engineering Reality of Virtual Reality 2010*, volume 7525, page 75250H. International Society for Optics and Photonics. 43

Chang, Z., Fang, Y., Zhang, Y., and Hu, C. (2010). A training simulation system for substation equipments maintenance. In *Machine Vision and Human-Machine Interface (MVHI), 2010 International Conference on*, pages 572–575. IEEE. 40

Chapin, W. L., Lacey, T. A., and Leifer, L. (1994). Designspace: a manual interaction environment for computer-aided design. In *Conference Companion on Human Factors in Computing Systems*, pages 47–48. ACM. 42

Chaumette, F. and Hutchinson, S. (2006). Visual servo control, part i: Basic approaches. *IEEE Robot. Autom. Mag.*, 13(4):82–90. 74, 80, 81, 136

Cheng, D., Wang, Y., Hua, H., and Talha, M. (2009). Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism. *Applied optics*, 48(14):2655–2668. 14

Chryssolouris, G. (2013). *Manufacturing systems: theory and practice*. Springer Science & Business Media. 38

Chryssolouris, G., Mavrikios, D., Papakostas, N., Mourtzis, D., Michalos, G., and Georgoulias, K. (2009). Digital manufacturing: history, perspectives, and outlook. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 223(5):451–462. 38

Chung, C. A. (2003). *Simulation modeling handbook: a practical approach*. CRC press. 38

Cohen, I. and Medioni, G. (1999). Detecting and tracking moving objects for video surveillance. *IEEE CVPR*. 29

Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM. 2, 3, 19

Czernuszenko, M., Pape, D., Sandin, D., DeFanti, T., Dawe, G. L., and Brown, M. D. (1997). The immersadesk and infinity wall projection-based virtual reality displays. *ACM SIGGRAPH Computer Graphics*, 31(2):46–49. 17, 18

Dangelmaier, W., Fischer, M., Gausemeier, J., Grafe, M., Matysczok, C., and Mueck, B. (2005). Virtual and augmented reality support for discrete manufacturing system simulation. *Computers in Industry*, 56(4):371–383. 43, 44

Dani, T. H. and Gadh, R. (1997). Creation of concept shape designs via a virtual reality interface. *Computer-Aided Design*, 29(8):555–563. 42

De Araùjo, B. R., Casiez, G., and Jorge, J. A. (2012). Mockup builder: direct 3d modeling on and above the surface in a continuous interaction space. In *Proceedings of Graphics Interface 2012*, pages 173–180. Canadian Information Processing Society. 42

De Sa, A. G. and Zachmann, G. (1999). Virtual reality as a tool for verification of assembly and maintenance processes. *Computers & Graphics*, 23(3):389–403. 43

Dementhon, D. F. and Davis, L. S. (1995). Model-based object pose in 25 lines of code. *International journal of computer vision*, 15(1-2):123–141. 77, 135

Diaz, C., Walker, M., Szafir, D. A., and Szafir, D. (2017). Designing for depth perceptions in augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 111–122. 109

Doil, F., Schreiber, W., Alt, T., and Patron, C. (2003). Augmented reality for manufacturing planning. In *Proceedings of the workshop on Virtual environments 2003*, pages 71–76. ACM. 43, 44

Dorfmüller, K. (1999). Robust tracking for augmented reality using retroreflective markers. *Computers & Graphics*, 23(6):795–800. 32

Echtler, F., Sturm, F., Kindermann, K., Klinker, G., Stilla, J., Trilk, J., and Najafi, H. (2004). The intelligent welding gun: Augmented reality for experimental vehicle construction. In *Virtual and augmented reality applications in manufacturing*, pages 333–360. Springer. 41

Faugeras, O. D. and Toscani, G. (1986). The calibration problem for stereo. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 86:15–20. 33, 69

Febretti, A., Nishimoto, A., Thigpen, T., Talandis, J., Long, L., Pirtle, J., Peterka, T., Verlo, A., Brown, M., Plepys, D., et al. (2013). Cave2: a hybrid reality environment for immersive simulation and information analysis. In *IS&T/SPIE Electronic Imaging*, pages 864903–864903. International Society for Optics and Photonics. 19

Fernandes, K. J., Raja, V., and Eyre, J. (2003). Cybersphere: the fully immersive spherical projection system. *Communications of the ACM*, 46(9):141–146. 20

Fiorentino, M., de Amicis, R., Monno, G., and Stork, A. (2002). Spacedesign: A mixed reality workspace for aesthetic industrial design. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 86. IEEE Computer Society. 42

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395. 76

Fitzgibbon, A. W. (2003). Robust registration of 2d and 3d point sets. *Image and Vision Computing*, 21(13):1145–1153. 73, 134

Foxlin, E., Harrington, M., and Pfeifer, G. (1998). Constellation: a wide-range wireless motion-tracking system for augmented reality and virtual set applications. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 371–378. ACM. 26, 28

Foxlin, E., Naimark, L., et al. (2003). Vis-tracker: A wearable vision-inertial self-tracker. *VR*, 3:199. 28, 31

Funk, M., Kosch, T., and Schmidt, A. (2016). Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 934–939. ACM. 41

Gavaghan, K. A., Anderegg, S., Peterhans, M., Oliveira-Santos, T., and Weber, S. (2011). Augmented reality image overlay projection for image guided open liver ablation of metastatic liver cancer. In *Workshop on Augmented Environments for Computer-Assisted Interventions*, pages 36–46. Springer. 41

Gosselin, F., Ferlay, F., Bouchigny, S., Mégard, C., and Taha, F. (2010). Design of a multimodal vr platform for the training of surgery skills. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pages 109–116. Springer. 40

Haritos, T. and Macchiarella, N. D. (2005). A mobile application of augmented reality for aerospace maintenance training. In *Digital Avionics Systems Conference, 2005. DASC 2005. The 24th*, volume 1, pages 5–B. IEEE. 40

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer. 31

Harrison, C., Benko, H., and Wilson, A. D. (2011). Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450. ACM. 23

Hartley, R. et al. (1997). In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):580–593. 66, 67

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press. 33, 63, 65, 66, 68, 70, 71, 73, 76, 77, 134

Hartley, R. I. and Sturm, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2):146–157. 73

Hill, A., Schiefer, J., Wilson, J., Davidson, B., Gandy, M., and MacIntyre, B. (2011). Virtual transparency: Introducing parallax view into video see-through ar. In *10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 239–240. IEEE. 15, 16

Hochreiter, J., Daher, S., Nagendran, A., Gonzalez, L., and Welch, G. (2016). Optical touch sensing on nonparametric rear-projection surfaces for interactive physical-virtual experiences. *PRESENCE: Teleoperators and Virtual Environments*, 25(1):33–46. 20

Holliman, N. S., Dodgson, N. A., Favalora, G. E., and Pockett, L. (2011). Three-dimensional displays: a review and applications analysis. *IEEE transactions on Broadcasting*, 57(2):362–371. 17

Horn, B. K. (1987). Closed-form solution of absolute orientation using unit quaternions. *JOSA A*, 4(4):629–642. 76

Hu, H. H., Gooch, A. A., Thompson, W. B., Smits, B. E., Rieser, J. J., and Shirley, P. (2000). Visual cues for imminent object contact in realistic virtual environment. In *Proceedings of the conference on Visualization'00*, pages 179–185. IEEE Computer Society Press. 109

Hubona, G. S., Wheeler, P. N., Shirah, G. W., and Brandt, M. (1999). The relative contributions of stereo, lighting, and background scenes in promoting 3d depth visualization. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(3):214–242. 109

Hughes, C. E., Zhang, L., Schulze, J. P., Edelstein, E., and Macagno, E. (2013). Cavecad: Architectural design in the cave. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 193–194. IEEE. 42

Hutson, M. and Reiners, D. (2011). Janusvf: Accurate navigation using scaat and virtual fiducials. *IEEE Transactions on Visualization and Computer Graphics,*, 17(1):3–13. 29, 31, 32, 35

Iqbal, M. and Hashmi, M. S. J. (2001). Design and analysis of a virtual factory layout. *Journal of Materials Processing Technology*, 118(1-3):403–410. 43

Israel, J. H., Wiese, E., Mateescu, M., Zöllner, C., and Stark, R. (2009). Investigating three-dimensional sketching for early conceptual design—results from expert discussions and user studies. *Computers & Graphics*, 33(4):462–473. 42

Jiang, B., Neumann, U., and You, S. (2004). A robust hybrid tracking system for outdoor augmented reality. In *Proceedings of IEEE Virtual Reality*, pages 3–275. 28, 29

Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. volume 92, pages 401–422. IEEE. 28

Karitsuka, T. and Sato, K. (2003). A wearable mixed reality with an on-board projector. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 321. IEEE Computer Society. 22

Kilteni, K., Groten, R., and Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21:373–387. 7, 108

Kiyokawa, K., Billinghurst, M., Campbell, B., and Woods, E. (2003). 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. 14, 15

Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 225–234. IEEE. 31

Knapp, J. M. and Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(5):572–577. 15

Kneip, L., Scaramuzza, D., and Siegwart, R. (2011). A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2969–2976. IEEE. 76, 135

Kramida, G. (2016). Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization & Computer Graphics*, (1):1–1. 15

Krüger, W., Bohn, C.-A., Fröhlich, B., Schüth, H., Strauss, W., and Wesche, G. (1995). The responsive workbench: A virtual work environment. *Computer*, (7):42–48. 17, 18

Kurihara, K., Hoshino, S., Yamane, K., and Nakamura, Y. (2002). Optical motion capture system with pan-tilt camera tracking and realtime data processing. *IEEE ICRA*, pages 1241–1248. 34

Kwon, J.-H., Nam, S.-H., Yeom, K., and You, B.-J. (2015). The use of shadows on real floor as a depth correction of stereoscopically visualized virtual objects. In *IEEE International Symposium on Mixed and Augmented Reality-Media, Art, Social Science, Humanities and Design (ISMAR-MASH'D)*, pages 53–54. 109

LaViola Jr., J. J., Kruijff, E., Bowman, D., Poupyrev, I. P., and McMahan, R. P. (2017). *3D User Interfaces: Theory and Practice (second edition)*. Addison-Wesley. 11, 13, 16, 17

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2):155. 77, 99

Li, J.-R., Khoo, L. P., and Tor, S. B. (2003). Desktop virtual reality for maintenance training: an object oriented prototype system (v-realism). *Computers in Industry*, 52(2):109–125. 40

Liu, L., van Liere, R., Nieuwenhuizen, C., and Martens, J.-B. (2009). Comparing aimed movements in the real world and in virtual reality. In *IEEE VR*, pages 219–222. 55

Longuet-Higgins, H. C. (1987). A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O. Firschein, eds*, pages 61–62. 66, 73

Lourdeaux, D. (2001). *Réalité virtuelle et formation: conception d'environnements virtuels pédagogiques*. PhD thesis, École Nationale Supérieure des Mines de Paris. 39

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110. 31

Lu, S.-Y., Shpitalni, M., and Gadh, R. (1999). Virtual and augmented reality technologies for product realization. *CIRP Annals*, 48(2):471–495. 38

Luong, Q.-T., Deriche, R., Faugeras, O., and Papadopoulo, T. (1993). On determining the fundamental matrix: Analysis of different methods and experimental results. 66, 67

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. S. (2012). *An invitation to 3-d vision: from images to geometric models*, volume 26. Springer Science & Business Media. 66, 74

Macchiarella, N. D. and Vincenzi, D. A. (2004). Augmented reality in a learning paradigm for flight aerospace maintenance training. In *Digital Avionics Systems Conference, 2004. DASC 04. The 23rd*, volume 1, pages 5–D. IEEE. 40

Maimone, A., Lanman, D., Rathinavel, K., Keller, K., Luebke, D., and Fuchs, H. (2014). Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, page 20. ACM. 14

Marchand, É. and Chaumette, F. (2002). Virtual visual servoing: a framework for real-time augmented reality. In *Computer Graphics Forum*, volume 21, pages 289–297. Wiley Online Library. 70, 77

Marchand, E., Spindler, F., and Chaumette, F. (2005). Visp for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robot. Autom. Mag.*, 12(4):40–52. 81

Marchand, E., Uchiyama, H., and F.Spindler (2016). Pose estimation for augmented reality: a hands-on survey. *IEEE Trans. on Visualization and Computer Graphics*, 22(12):2633–2651. 73, 77, 99, 134

Marner, M. R. and Thomas, B. H. (2010). *Tool virtualization and spatial augmented reality*. PhD thesis, Citeseer. 99, 138

Marner, M. R., Thomas, B. H., and Sandor, C. (2009). Physical-virtual tools for spatial augmented reality user interfaces. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 205–206. IEEE. 99, 138

Martin, P., Marchand, E., Houlier, P., and Marchal, I. (2014). Mapping and re-localization for mobile augmented reality. In *IEEE International Conference on Image Processing (ICIP)*, pages 3352–3356. IEEE. 30

Martin, S. and Hillier, N. (2009). Characterisation of the novint falcon haptic device for application as a robot manipulator. In *Australasian Conference on Robotics and Automation (ACRA)*, pages 291–292. Citeseer. 25

Mathieu, H. (2005). The cyclope: A 6 dof optical tracker based on a single camera. In *2nd INTUITION International Workshop "VR/VE & Industry: Challenges and opportunities"*. 31, 32

Medeiros, D., Teixeira, L., Carvalho, F., Santos, I., and Raposo, A. (2013). A tablet-based 3d interaction tool for virtual engineering environments. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pages 211–218. ACM. 43, 44

Menck, N., Weidig, C., and Aurich, J. C. (2013). Virtual reality as a collaboration tool for factory planning based on scenario technique. *Procedia CIRP*, 7:133–138. 43

Menck, N., Yang, X., Weidig, C., Winkes, P., Lauer, C., Hagen, H., Hamann, B., and Aurich, J. (2012). Collaborative factory planning in virtual reality. *Procedia CIRP*, 3:317–322. 43

Merhav, S. and Velger, M. (1991). Compensating sampling errors in stabilizing helmet-mounted displays using auxiliary acceleration measurements. *Journal of Guidance, Control, and Dynamics*, 14(5):1067–1069. 37

Meyer, K., Applewhite, H. L., and Biocca, F. A. (1992). A survey of position trackers. In *Presence: Teleoperators and Virtual Environments (ISSN 1054-7460), vol. 1, no. 2, p. 173-200.*, volume 1, pages 173–200. 26

Miles, J. and Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Sage. 115

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–293. International Society for Optics and Photonics. 1, 131

Mine, M. et al. (1995). Virtual environment interaction techniques. *UNC Chapel Hill computer science technical report TR95-018*, pages 507248–2. 50, 51

Mizell, D. (2001). Boeing's wire bundle assembly project. *Fundamentals of wearable computers and augmented reality*, 5. 41

Moeslund, T. B. and Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer vision and image understanding*, 81(3):231–268. 29

Mohring, M., Lessig, C., and Bimber, O. (2004). Video see-through ar on consumer cell-phones. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 252–253. IEEE Computer Society. 14

Moreno, D. and Taubin, G. (2012). Simple, accurate, and robust projector-camera calibration. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 464–471. IEEE. 105

Mourtzis, D., Doukas, M., and Bernidaki, D. (2014). Simulation in manufacturing: Review and challenges. *Procedia CIRP*, 25:213–229. 3, 38, 39, 42, 131

Mujber, T. S., Szecsi, T., and Hashmi, M. S. (2004). Virtual reality applications in manufacturing process simulation. *Journal of materials processing technology*, 155:1834–1838. 39

Mustafah, Y. M., Azman, A. W., and Akbar, F. (2012). Indoor uav positioning using stereo vision sensor. pages 575–579. 5, 29

Nee, A. Y., Ong, S., Chryssolouris, G., and Mourtzis, D. (2012). Augmented reality applications in design and manufacturing. *CIRP Annals-manufacturing technology*, 61(2):657–679. 3, 38, 42

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136. IEEE. 98, 138

Ni, T., Karlson, A. K., and Wigdor, D. (2011). Anatonme: facilitating doctor-patient communication using a projection-based handheld device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3333–3342. ACM. 22, 103

Okumura, K., Oku, H., and Ishikawa, M. (2012). Lumipen: Projection-based mixed reality for dynamic objects. In *Proceedings ot the 2012 International Conference on Multimedia and Expo (ICME)*, pages 699–704. IEEE. 21

Olwal, A., Lindfors, C., Gustafsson, J., Kjellberg, T., and Mattsson, L. (2005). Astor: An autostereoscopic optical see-through augmented reality system. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 24–27. IEEE. 14

Ong, S., Yuan, M., and Nee, A. (2008). Augmented reality applications in manufacturing: a survey. *International journal of production research*, 46(10):2707–2742. 38

Osterlund, J. and Lawrence, B. (2012). Virtual reality: Avatars in human spaceflight training. *Acta Astronautica*, 71:139–150. 40

Paperno, E., Sasada, I., and Leonovich, E. (2001). A new method for magnetic position and orientation tracking. *IEEE Transaction on Magnetics*, 37(4):1938–1940. 26

Payton, M. E., Greenstone, M. H., and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *Journal of Insect Science*, 3(1):34. 52

Peck, T. C., Bourne, K. A., and Good, J. J. (2018). The effect of gender body-swap illusions on working memory and stereotype threat. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1604–1612. 109

Pintaric, T. and Kaufmann, H. (2007). Affordable infrared-optical pose-tracking for virtual and augmented reality. In *Proceedings of Trends and Issues in Tracking for Virtual Environments Workshop, IEEE VR*, pages 44–51. 5, 29, 30, 31, 32, 33, 34, 35, 86, 96

Ponto, K., Tredinnick, R., Bartholomew, A., Roy, C., Szafir, D., Greenheck, D., and Kohlmann, J. (2013). Sculptup: A rapid, immersive 3d modeling environment. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 199–200. IEEE. 42

Puerta, A. M. (1989). The power of shadows: shadow stereopsis. *JOSA A*, 6(2):309–311. 108

Punpongsanon, P., Iwai, D., and Sato, K. (2015). Projection-based visualization of tangential deformation of nonrigid surface by deformation estimation using infrared texture. *Virtual Reality*, 19(1):45–56. 20

Quan, L. and Lan, Z. (1999). Linear n-point camera pose determination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8):774–780. 76

Raab, F. H., Blood, E. B., Steiner, T. O., and Jones, H. R. (1979). Magnetic position and orientation tracking system. *Aerospace and Electronic Systems, IEEE Transactions on*, (5):709–718. 27

Rambach, J. R., Tewari, A., Pagani, A., and Stricker, D. (2016). Learning to fuse: A deep learning approach to visual-inertial camera pose estimation. In *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, pages 71–76. IEEE. 75

Raskar, R., Van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C. (2006). ilamps: geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2006 Courses*, page 7. ACM. 21, 22

Raskar, R., Welch, G., and Fuchs, H. (1998). Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR'98)*, pages 11–20. Citeseer. 20

Regenbrecht, H., Baratoff, G., and Wilke, W. (2005). Augmented reality projects in the automotive and aerospace industries. *IEEE Computer Graphics and Applications*, 25(6):48–56. 39

Ribo, M., Pinz, A., and Fuhrmann, A. L. (2001). A new optical tracking system for virtual and augmented reality applications. In *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference (IMTC)*, volume 3, pages 1932–1936. IEEE. 24, 31, 32

Rios, H., Hincapié, M., Caponio, A., Mercado, E., and Mendívil, E. G. (2011). Augmented reality: an advantageous option for complex training and maintenance operations in aeronautic related processes. In *International Conference on Virtual and Mixed Reality*, pages 87–96. Springer. 40

Rolland, J. and Cakmakci, O. (2009). Head-worn displays: the future through new eyes. *Optics and Photonics News*, 20(4):20–27. 13

Rosten, E., Porter, R., and Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence*, 32(1):105–119. 31

Saakes, D. and Stappers, P. J. (2009). A tangible design tool for sketching materials in products. *AI EDAM*, 23(3):275–287. 43

Sarakoglou, I., Brygo, A., Mazzanti, D., Hernandez, N. G., Caldwell, D. G., and Tsagarakis, N. G. (2016). Hexotrac: A highly under-actuated hand exoskeleton for finger tracking and force feedback. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 1033–1040. IEEE. 25

Schmalstieg, D. and Hollerer, T. (2016). *Augmented reality: principles and practice*. Addison-Wesley Professional. 1, 3, 12

Schwald, B. and De Laval, B. (2003). An augmented reality system for training and assistance to maintenance in the industrial context. 40

Seth, A., Vance, J. M., and Oliver, J. H. (2011). Virtual reality for assembly methods prototyping: a review. *Virtual reality*, 15(1):5–20. 39

Shiu, Y. C. and Ahmad, S. (1989). Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb. *IEEE TRA*, 5(1):16–29. 80

Shu, C., Yupeng, N., Jian, Y., and Qijun, S. (2010). Study on maintenance training system for certain missile based on virtual reality technology. In *Computer Design and Applications (ICCDA), 2010 International Conference on*, volume 2, pages V2–78. IEEE. 40

Siegl, C., Colaianni, M., Thies, L., Thies, J., Zollhöfer, M., Izadi, S., Stamminger, M., and Bauer, F. (2015). Real-time pixel luminance optimization for dynamic multi-projection mapping. *ACM Transactions on Graphics (TOG)*, 34(6):237. 20

Slater, M., Khanna, P., Mortensen, J., and Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications*, 29(3). 109

Slater, M., Usoh, M., and Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments. In *Virtual Environments 95*, pages 8–21. Springer. 108

Smith, S. M. and Brady, J. M. (1997). Susan—a new approach to low level image processing. *International journal of computer vision*, 23(1):45–78. 31

Speranza, F., Tam, W. J., Renaud, R., and Hur, N. (2006). Effect of disparity and motion on visual comfort of stereoscopic images. In *Electronic Imaging 2006*, pages 60550B–60550B. International Society for Optics and Photonics. 15

Stark, R., Israel, J. H., and Wöhler, T. (2010). Towards hybrid modelling environments—merging desktop-cad and virtual reality-technologies. *CIRP annals*, 59(1):179–182. 42

Steinicke, F., Hinrichs, K., and Ropinski, T. (2005). Virtual reflections and virtual shadows in mixed reality environments. In *IFIP Conference on Human-Computer Interaction*, pages 1018–1021. Springer. 109

Steinicke, F., Jansen, C. P., Hinrichs, K. H., Vahrenhold, J., and Schwald, B. (2007). Generating optimized marker-based rigid bodies for optical tracking systems. In *VISAPP (2)*, pages 387–395. 32

Sueishi, T., Oku, H., and Ishikawa, M. (2015). Robust high-speed tracking against illumination changes for dynamic projection mapping. In *Proceedings ot the 2015 International Conference on Virtual Reality (VR)*, pages 97–104. IEEE. 21

Sugano, N., Kato, H., and Tachibana, K. (2003). The effects of shadow representation of virtual objects in augmented reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 76–83. 109

Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM. 13, 14, 15, 26

Sutherland, I. E. (1974). Three-dimensional data input by tablet. *Proceedings of the IEEE*, 62(4):453–461. 76

Taylor, R. H., Mittelstadt, B. D., Paul, H. A., Hanson, W., Kazanzides, P., Zuhars, J. F., Williamson, B., Musits, B. L., Glassman, E., and Bargar, W. L. (1994). An image-directed robotic system for precise orthopaedic surgery. *IEEE TRA*, 10(3):261–275. 5, 29

Tjaden, H., Stein, F., Schömer, E., and Schwanecke, U. (2015). High-speed and robust monocular tracking. *VISAPP*. 68

Tsai, R. Y. and Lenz, R. K. (1989). A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE TRA*, 5(3):345–358. 78, 80, 136

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503. 114

Velger, M. (1998). Helmet-mounted displays and sights. *Norwood, MA: Artech House Publishers, 1998*. 37

Vogt, S., Khamene, A., Sauer, F., and Niemann, H. (2002). Single camera tracking of marker clusters: Multiparameter cluster optimization and experimental verification. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, page 127. 68, 125

Wagner, D., Langlotz, T., and Schmalstieg, D. (2008). Robust and unobtrusive marker tracking on mobile phones. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 121–124. IEEE. 30

Wang, T., Liu, Y., and Wang, Y. (2008). Infrared marker based augmented reality system for equipment maintenance. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 5, pages 816–819. IEEE. 32

Wei, G.-Q. and De Ma, S. (1994). Implicit and explicit camera calibration: Theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):469–480. 63

Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K., and Colucci, D. (2001). High-performance wide-area optical tracking: The hiball tracking system. *presence: teleoperators and virtual environments*, 10(1):1–21. 29, 30, 31

Welch, G. and Foxlin, E. (2002). Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer graphics and Applications*, 22(6):24–38. 5, 6, 24, 25, 26, 27, 28, 45, 132

Wells, M. J. and Venturino, M. (1990). Performance and head movements using a helmet-mounted display with different sized fields-of-view. *Optical Engineering*, 29(8):870–878. 37

Weng, J., Cohen, P., and Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (10):965–980. 64

Wetzstein, G., Lanman, D., Hirsch, M., Heidrich, W., and Raskar, R. (2012). Compressive light field displays. *IEEE computer graphics and applications*, 32(5):6–11. 17

Willis, K. D., Poupyrev, I., Hudson, S. E., and Mahler, M. (2011). Sidebyside: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 431–440. ACM. 22

Xin, M., Sharlin, E., and Sousa, M. C. (2008). Napkin sketch: handheld mixed reality 3d sketching. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 223–226. ACM. 42

Yang, L., Normand, J.-M., and Moreau, G. (2016). Practical and precise projector-camera calibration. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 63–70. IEEE. 99, 138

Ye, N., Banerjee, P., Banerjee, A., and Dech, F. (1999). A comparative study of assembly planning in traditional and virtual environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 29(4):546–555. 43

Yeo, C. T., Ungi, T., Paweena, U., Lasso, A., McGraw, R. C., Fichtinger, G., et al. (2011). The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections. *IEEE Transactions on Biomedical Engineering*, 58(7):2031–2037. 40

You, S., Neumann, U., and Azuma, R. (1999). Hybrid inertial and vision tracking for augmented reality registration. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 260–267. IEEE. 28

Yu, I., Mortensen, J., Khanna, P., Spanlang, B., and Slater, M. (2012). Visual realism enhances realistic response in an immersive virtual environment-part 2. *IEEE Computer Graphics and Applications (CGA)*, 32(6):36–45. 109

Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195. 66, 67, 68

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334. 33, 63, 64, 70

Zheng, Z., Liu, X., Li, H., and Xu, L. (2010). Design and fabrication of an off-axis see-through head-mounted display with an x–y polynomial surface. *Applied optics*, 49(19):3661–3668. 14

Zhou, Y., Xiao, S., Tang, N., Wei, Z., and Chen, X. (2016). Pmomo: projection mapping on movable 3d object. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 781–790. ACM. 20, 21, 36, 126

Zhu, R. and Zhou, Z. (2004). A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(2):295–302. 28

Titre : Contribution à l'étude des systèmes de projection pour des applications industrielles en réalité mixte

Mots clés : Systèmes de projection, réalité mixte, applications industrielles, suivi de mouvements

Résumé :
La réalité mixte apporte certains avantages aux applications industrielles. Elle peut, entre autres, faciliter la visualisation et validation de projets ou assister des opérateurs durant des tâches spécifiques. Les systèmes de projection (PBS), tels que les CAVE ou la réalité augmentée spatiale, fournissent un environnement de réalité mixte permettant une collaboration directe avec des utilisateurs externes. Dans cette thèse, nous visons à améliorer l'utilisation des systèmes de projection pour des applications industrielles en abordant deux défis majeurs: (1) améliorer les composantes techniques des PBS et (2) augmenter l'expérience utilisateur dans les PBS.

En tant que premier défi technique, nous visons à améliorer les systèmes de suivi de mouvements optiques. Nous proposons une approche permettant d'élargir l'espace de travail de ces systèmes grâce à deux méthodes. La première permet de suivre les mouvements à partir d'une seule caméra tandis ce que la deuxième permet de contrôler les caméras et suivre les objets dans l'espace de travail. Le système qui en résulte fournit des performances acceptables pour des applications en réalité mixte tout en augmentant considérablement l'espace de travail. Un tel système de suivi de mouvement peut permettre de mieux exploiter le potentiel des systèmes de projection et d'élargir le champ possible des interactions.

En tant que deuxième défi technique, nous concevons un casque « tout-en-un » pour la réalité augmentée spatiale mobile. Le casque rassemble à la fois un système de projection et un système de suivi de mouvements qui sont embarqués sur la tête de l'utilisateur. Avec un tel système, les utilisateurs sont capables de se déplacer autour d'objets tangibles et de les manipuler directement à la main tout en projetant du contenu virtuel par-dessus. Nous illustrons notre système avec deux cas d'utilisation industriels: le prototypage virtuel et la visualisation médicale.

Enfin, nous abordons le défi qui vise à améliorer l'expérience utilisateur dans les PBS. Nous proposons une approche permettant d'incarner un personnage virtuel et d'augmenter la perception spatiale des utilisateurs dans les PBS. Pour ce faire, nous ajoutons l'ombre virtuelle des utilisateurs dans les systèmes de projection immersifs. L'ombre virtuelle est directement corrélée aux mouvements des utilisateurs afin qu'ils la perçoivent comme si c'était la leur. Nous avons effectué une expérience afin d'étudier l'influence de la présence de l'ombre virtuelle sur le comportement des utilisateurs.

Title : Contribution to the study of projection-based systems for industrial applications in mixed reality

Keywords : Projection-based systems, mixed reality, industrial applications, tracking

Abstract:
Mixed Reality brings some advantages to industrial applications. It can, among others, facilitate visualizing and validating projects or assist operators during specific tasks. Projection-based Systems (PBS), such as CAVEs or spatial augmented reality provide a mixed reality environment enabling straightforward collaboration with external users. In this thesis, we aim at improving the usage of PBS for industrial applications by considering two main challenges: (1) improving the technical components of PBS and (2) improving the user experience when using PBS.

As a first technical challenge, we propose to address the improvement of the tracking component. We introduce an approach that enables increasing the workspace of optical tracking systems by using two methods. As a first method, we propose to use monocular tracking. As a second method, we propose to use controlled cameras that follow the targets across the workspace. The resulting system provides acceptable performances for mixed reality applications while considerably increasing the workspace. Such a tracking system can make it easier to use large projection-based displays and can widen the range of available interactions.

As a second technical challenge, we design an "all-in-one" headset for mobile spatial augmented reality on tangible objects. The headset gathers both a projection and a tracking system that are embedded on the user's head. With such a system, the users are able to move around tangible objects and to manipulate them directly by hand while projecting virtual content over them. We illustrate our system with two industrial use cases: virtual prototyping and medical visualization.

Finally, we address the challenge that aims at improving the user experience when using PBS. We introduce a method that provides virtual embodiment and increases the spatial perception of the users when using PBS. To do so we add the user's virtual shadow in immersive projection-based systems. The virtual shadow is dynamically mapped to users' movements in order to make them perceive the shadow as if it was their own. We then carry out an experiment to study the influence of the presence of the virtual shadow on the user experience and behavior.