Université catholique de Louvain
Faculté des Sciences Appliquées
LABORATOIRE DE TÉLÉCOMMUNICATIONS
ET
TÉLÉDÉTECTION

B - 1348 Louvain-la-Neuve
Belgique

Design, Implementation and Evaluation for Continuous Interaction in Image-guided Surgery

Daniela Gorski Trevisan

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences Appliquées

Examination Committee:

Benoit MACQ (UCL/TELE) - *Supervisor*
Jean VANDERDONCKT (UCL/ISYS) - *Co-supervisor*
Christophe DE VLEESCHOUWER (UCL/TELE)
Laurence NIGAY (UJF/Clips-IMAG, France)
Luciana PORCHER NEDEL (UFRGS/II, Brazil)
Jean-Didier LEGAT (UCL/TELE) - *President*

March 2006

Acknowledgment

First of all, I'd like to express my most sincere gratitude to my supervisors Benoit Macq and Jean Vanderdonckt for giving me the opportunity of carrying out this multidisciplinary study. From Prof. Macq I've learned the ability to interconnect technologies, science and people in order to develop innovative applications and research projects. His enthusiasm has given me the will to continue investing in the research field. Prof. Vanderdconkt through his experience in human computer interaction research allowed me to apply theory in practice and develop transparent interfaces. I'm very thankful to him for always encouraging me and believing in my work.

I would also like to thank the examination committee for all the enriching suggestions that resulted in the improvement of this manuscript. I'm especially grateful to Prof. Luciana Porcher Nedel who spent a great deal of time discussing and addressing many of my questions and doubts. I've learned a lot from her professionalism and work and I really feel lucky of having her around during a critical period of this long journey.

Many thanks for the support from the MERCATOR project under contract WALEO 21/5129; *Fond Spécial de Recherche* (FSR); SIMILAR Network of Excellence - FP6-507609; CRIF-WTCM and the *centre de compétences de l'industrie technologique*, Liege, Belgium.

Thanks also to the Herol group for supporting me in the maxillofacial prototype application. Special thanks to Dr. Hervé Reychler (STOM), Dr. Raphael Olszewski (STOM), Khanh TranDuy (MECA) and Sinisa Milos (CRIF) for their availability, competence and interesting feedback.

Another thank you to the *TELE atelier team*, Robert and Christian, for providing me all the material needed to develop the experimental part of my Thesis. Thanks to Patricia Focant, our secretary and therefore my second hand. Thanks also to all testers volunteers. No need to mention, you know who you are.

I'm also thankful to my fellow researchers and colleagues of the TELE lab for providing a pleasant atmosphere during these last four years. Special thanks to Monica Gemo for helping me with the Medical Studio programing; Patrice Rondao for his efficient comments and suggestions, addressing many of my implementational problems; Julien Pironet for spending a lot of his time debugging program codes; Diego Ruiz for his patience in explaining me the 3D mathematics and Quentin Noirhomme for his expertise with the registration methods.

I couldn't forget to specially mention my colleague and friend Vincent Nicolas. Thank you for supporting me during this time in Belgium. You have always helped me with the many computer problems I have encountered. You're the Medical Studio expert and without you there would be no thesis at all!

I'd like to also express my special thanks to all the people I've met here in Belgium. You have brought color into this gray Belgian climate. Thanks to you Gianfranco for letting me become a member of the *LLN mafia italiana*. Thanks to Klavdia and Marc for supporting my December bad moods; Patrick and Xavier for the great vacations in Provence; Eduardo and Marcela, my Colombian friends, for introducing me to the Belgium life when I first arrived here; and my *colocs* Valerie, Jennifer, Diego and Seb for all the good time we spent together and for all those that are yet to come.

Many special thanks to all my LLN Brazilian friends Alessandra, Fabio, Fernanda, Catarina (oops she is Portuguese), Aline, Valeska, Silvia, Adriane and Marko. You've made me feel at home and closer to our culture. Special thanks to my friends Beatriz and Fred (Brazilian-in-law) for always welcoming me in their house and helping with the many practical life stuff like cooking spaghettis and even helping with the final touches of this work. Thank you Luciana, Bruno, Thais e Pedro. You have become my family here in Belgium and therefore without you I'd be lost.

Thanks to my dear friend Marco Antonio Winckler. You have always been beside me, supporting, encouraging, suggesting and specially believing at my work. Marquinho, I'm lucky to be your friend. Thanks a lot!

At last but certainly not least, I'd like to say a very special thanks to my family. My parents Tania and Cezar, my sister Cristina, my brothers Mauricio and Giuliano and my dear Dadinha. Even faraway from another continent I have always felt your love, supporting and caring. You have given the stability needed to go through with this project. À vocês dedico este trabalho!

Contents

List of Figures

Chapter 1

Content of the Thesis

This Thesis has been conducted within the framework of 3 projects:

- The European Network of Excellence SIMILAR which aims at creating a unified research group between European laboratories involved in the study of Human-Computer Interaction (HCI) and laboratories involved in Signal Processing (SP) research.
- The MERCATOR project which aims at developing a platform for 3D Medical Imaging and its use in Image-guided Surgery (IGS) and Radiotherapy.
- The HEROL project devoted to the development of an computer-guided maxillo-facial surgery application.

The content of this Thesis reflects the work performed at the intersection of research in the HCI and SP domains applied to the development of an computer-guided maxillo-facial surgery application.

1.1 Introduction

More robust interaction modalities, such as gesture, speech or other natural human-computer interactions (HCIs) allow the creation of applications with a seamless integration of computer-based information into the real world. In such systems the user interface moves from the classical computer screen to mixed reality interaction spaces in order to create a smooth and continuous transition between the digital and the real worlds.

This fusion involves the design and development of ”mixed reality systems”, including augmented reality, augmented virtuality, augmented

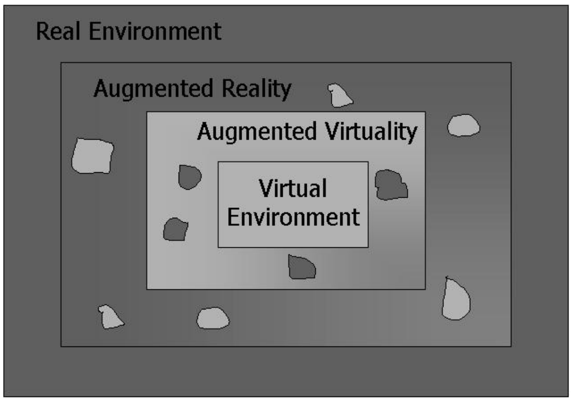


Figure 1.1: *Zones of a reality-virtuality continuum.*

video, and tangible systems. Milgran and Herman [80] argue that real environments (RE) and virtual environments (VE) are, in fact, two poles of a reality-virtuality continuum, defining distinct zones of interactions conforming shown in Figure 1.1. Mixed reality (MR) covers the entire continuum between RE and VE, including augmented reality (AR), and augmented virtuality (AV), but excludes the end-points, perceived as limit conditions. Therefore we can say that AR and AV are components of mixed reality, and that MR systems are any possible combination of real and virtual information.

Currently, one of the most challenging aspects of interactive systems design and development lies in the ability to overlay and to register digital information on the user's workspace in a spatially meaningful way, allowing mixed reality (MR) to be an extremely effective operational medium.

Mixed reality systems can also be classified into the category of attentive user interfaces since the goal of mixed systems research can be to monitor the state of the user and/or the environment, through sensors data acquired in real time, and to adapt or augment the computational interface to significantly improve their performance on the task at hand. For instance, when applied to the medical surgery, it allows users to keep their environmental perception while having intuitively access to more contextual information such as the incisions location, regions to be avoided, diseased tissues and so on.

Consequently, two research areas are closely connected and involved in such kind of applications: human-computer interaction and signal processing.

The Human-computer interaction research area focuses on the design, implementation and evaluation of interactive systems in the context of a users task and work [23]. One of the issues to be addressed in this field is to understand the people behavior while using interfaces. Furthermore, researchers will have a greater understanding of the perceptive, functional and cognitive resources that humans possess and how many of these resources are available during a computationally based task, whether or not their computational systems will include advanced sensors. Once these resources are identified and their activity (or load) measured, designers of computational interfaces can begin to consider these limitations (and perhaps adapt to their status) in the design of new HCI systems.

On the other hand, the Signal processing research area provides strategies to implement such category of system. Implementation of MR systems requires a set of specific services such as video capture, tracking system, methods for registering real and virtual objects and visualization techniques as well. Research in such topics should focus on real time strategies and provide the appropriate accuracy for the application aimed.

1.2 Motivation

Many *one shot* custom-built systems have been developed; indeed, reusability, incremental development and evaluation of Mixed Reality (MR), whether they are Augmented Reality (AR) or Augmented Virtuality (AV) systems, have not reached a reasonable state to build upon. Existing practice does not provide any conceptual guidance for addressing specific challenges of mixed interface design. There is not yet commonly-accepted conceptual structure for discussing and reasoning about this family of systems. Research on software structures, design methods and design support tools for mixed reality systems is still in its infancy. Efforts in this direction can be found, for instance, in [59] and [28] which have mostly focused on AR systems and in [37], [58], [132], [131] and [69] which have mostly focused on tangible user interface systems.

In order to identify reliable foundations for MR systems, there is a need to gather the knowledge acquired so far so as to create a body of reusable knowledge for existing and future systems. This body of knowledge should cover MR design issues, identify key research challenges in terms of design methods and tools and build a corpus of usability knowledge (e.g., usability guidelines, design heuristics, properties).

HCI development relies on the same stages as software engineering development. It organizes design and implementation of an application into six phases: requirements definition, specification, implementation, testing, installation and maintenance [2]. Work in HCI has led to the elaboration of tools or methods to help the designer in each of these steps. Work in Signal Processing and technological developments in devices, communication, etc., has allowed interactive systems to make use of the physical environment that surrounds the user. The user is thus interacting with a combination of the physical and digital worlds.

A direct consequence is that existing tools and methods are no longer sufficient. During the specification phase, task-oriented modeling methods are mainly centered on the temporal organization of the interaction, rather than on the characteristics of the physical/digital information interface. Modeling languages like UML [12] are well-suited to capture application functionality, but, apart from treating users as use case agents, no physical entities are represented or characterized. Scenario and prototype-based design approaches provide more expressive freedom than task and application modeling but do not offer the ability for the designer to systematically compare, explore or generalize design solutions. During the implementation steps, widgets do not yet include physical entities, software architecture models need refinements, connections between real entities and their corresponding digital classes need to be explicit, etc. Finally in terms of usability, relatively little attention has been paid to usability in mixed reality environments [26][86].

Limitations in the current techniques for designing interfaces are the biggest obstacles preventing MR from being realistic in practical uses. Taking as example medical applications, the generic design of interfaces for image-guided surgery (IGS) should deal with pre and intra-operative images displayed in the surgical environment to provide guidance [124]. This is a typical application of augmented reality (AR) systems where the virtual world corresponding to the pre-operative information should be correctly aligned in real time with the real world corresponding to the intra-operative information. For example, in the microscope-assisted guided interventions the surgeons can view virtual features segmented from pre-operative radiology images accurately overlaid in stereo in the optical path of a surgical microscope. Then the description of user interface and interaction for image-guided systems is a non-trivial task. As augmented reality systems they are dynamic environments and due to their continuous and visual nature, defining salient and useful aspects of them is extremely difficult. Normally, the system behavior is able to be broken down into discrete modes of interaction (i.e. mouse clicks, typing, etc.) but the user is engaged with a continuous flow of the environment. This limitation results in two different kind of interaction - one to deal with the physical space and the other for the virtual or digital one. Consequently, interaction discontinues break the natural workflow, forcing the user to switch between digital and real operation modes. Integration and fusion of different sources of information in complex scenarios (such those required in operation rooms) in an effective and meanly way are the focus of this work.

1.3 Objectives and contributions

This Thesis is decomposed into three parts (Design, Implementation and Evaluation) following the cycle of system development [28]. Each part presents a literature review in a chapter called *Related Works* and all other chapters describe the work which has been done by the author. The following subsections describe objectives, contributions and methodology addressed by each part.

1.3.1 Design

Objectives

The main goal to be addressed in this part is to provide useful elements needed to the design of MR systems. It consists in identifying, classifying and proposing a methodology to design MR systems taking into account the usability criteria proposed in Part III.

Summary

We propose a conceptual classification of the design space to support the development of MR systems. The proposed design space is an abstract tool for systematically exploring several design alternatives at an early stage of interaction design, without being biased by a particular modality or technology. Once the abstract design possibilities have been identified and a concrete design decision has been taken (i.e. a specific modality has been selected), a concrete MR application can be considered in order to analyze the interaction techniques in terms of continuous interaction properties. Such methodology was presented in the doctoral consortium session during the Computer Human Interaction Conference (CHI2004) [125].

Contributions

As a general contribution of this part we provide a better understanding of the nature of the phenomena (i.e. mixed interaction spaces) and thinking about choices of design. We propose usability criteria to analyze interaction and extract potential guidelines while developing MR systems [128].

1.3.2 Implementation

Objectives

This part addresses three main goals as follows:

- Provide the elements needed for the software implementation. It consists in studying and adapting existing architecture models to support the MR system specifications;
- Provide adequate services including services for tracking, registration, segmentation and 3D modeling which are required for such category of system;
- Locally validate the complete system. It consists in measuring the accuracy of all implemented methods which are involved in the mixed reality system pipeline.

Summary

In this phase we explore technical aspects related to the components necessary to have a MR system. At most basic level a mixed reality system should contain at least four components: sensors, for determining user, platform or environment state (how to perceive the real world); inference engine or classifier to evaluate incoming sensor information (registration techniques to align virtual information with the real scene); adaptive user interface (visualization of the augmented scene according to the user's view); underlying computational architecture to integrate these components. In reality, a fully functioning system would have many more components, but these are the most critical for inclusion as a MR interaction system. Independently, each of these components is fairly straightforward.

Contributions

As contributions of this part we develop two approaches to provide mixed reality visualizations. The first one is a stereovision-based [126] and the second one is a marker-based methods to track objects in the real world. In both methods we suggest to use a real time rigid registration based on distance map to align real and 3D objects. The marker-based approach has been used to develop a computer-guided application for maxillo-facial surgery [127] [122]. This application was integrated into the Medical Studio[1] open-source platform and tested in several interaction scenarios. It has contributed to demonstrate the flexibility and usability of the platform.

1.3.3 Evaluation

Objectives

This part addresses three main goals, as follows:

- Study, adapt and propose new criteria for the development of usable MR systems.
- Globally validate the interaction techniques. It consists in proposing and applying a usability method to evaluate such criteria.
- Establish connections between the evaluation criteria, implementational strategies and the design principles previously identified in Part I. This allows designers to derive potential interaction problems during an early development cycle of the systems, as well as to have more control about the factors influencing in the continuity of interaction.

Summary

The evaluation step involves the measurement of interaction techniques in terms of continuous interaction. Besides the validation criteria that are applied to a device or a process we should take into account the usability criteria to assess the interaction technique of this new paradigm of interaction. For an augmented system to be successful it must identify at least one

[1] This platform was developed at the Communications and Remote Sensing Laboratory, at Université catholique de Louvain, Belgium. Download available at http://www.medicalstudio.org

of the bottlenecks (i.e. perceptive, cognitive or functional) in real time and alleviate it through a performance enhancing mitigation strategy. These mitigation strategies are conveyed to the user through the adaptive interface and might involve: modality switching (for instance visual, auditory and haptic), intelligent guidance, task negotiation, etc. When the user task state is correctly sensed and an appropriate strategy chosen to alleviate the bottleneck, the interface adapted to carry out the strategy and the resulting sensor information indicates that the aiding has worked - only then has a system *closed the loop* and successfully augmented the user's interaction. In this sense our proposed methodology should be able to evaluate the interaction in terms of: How, where and when will the information be delivered to the user (e.g. perceptual properties); how easy or hard will the information be interpreted by the user (e.g. cognitive properties) and how natural will be the interaction process (e.g. functional properties).

Contributions

As concrete contribution of the experimental phase we provide a set of measured parameters involved in this kind of mixed interaction. We are able to address questions such as: *Does AR improve human performance in computer guided tasks relative to other media? How are the measured variables influencing the final user performance?* We propose a model to describe final user performance and to find the contribution factors of each variable for the continuity properties of the user interaction [122] [123].

1.4 Study cases

One of the most promising application fields of MR is in computer-guided tasks. It is commonly theorized [107], [119] and [138] that AR assistance in complex tasks will increase performance and reduce errors due to the representation of the task properly registered with the workspace. Errors are less likely and workload of translating abstracted instructions onto the reality is reduced. There has been much speculation about what AR can do, but very few empirical research studies exploring the effectiveness of such systems. Even though a number of AR prototypes and test-bed applications have been developed, they are mainly proof-of concept applications or demonstrations.

In this way the study cases have sought to provide key contributions to our understanding of computer-human interaction with MR environments as well as implementational aspects.

Figure 1.2 (left side) shows the general scenario involving the image-guided surgery (IGS). This work is conducted in collaboration with the MRI intra-operative group of the Neurology department at Saint Luc Hospital, in Brussels [124]. In such systems, complex surgical procedures can be navigated visually with high precision by overlaying on an image of the patient, a color coded preoperative (d) plan specifying details such as the locations of incisions, areas to be avoided and the diseased tissues. Therefore Augmented Reality (b) techniques applied to IGS systems allow users to keep their environmental perception while having intuitively access to more contextual information. Another way to provide guidance consists in tracking surgical instruments and in visualizing their positions into pre-operative images (c). This application was not implemented in this work, it uses a proprietary system and it is in current use by the surgeons. We use such application as study case for scenario description and interaction analysis.

The other study case, also shown in Figure 1.2 (right side), is the maxillo-facial surgery conducted in collaboration with the Stomatology and maxillo-facial surgery service at Saint Luc Hospital, in Brussels. In this application we studied the guidance for the osteotomy procedure. The osteotomy procedure consists in cutting the patient's mandible during a maxillofacial surgery. Segmented pre-operative information (h) such as dental nerve and the osteotomy pathline are visualized using two kind of mixed reality systems: augmented reality (f) where information is projected into the video image; and augmented virtuality (g) where tracked tools and objects provide guidance for the surgeon in a 3D scene. This application was completely developed in this work using a marker-based library to track objects [127].

These two applications were used in this Thesis with as goal to provide a practical approach of the proposed concepts.

1.5 Structure of the Thesis

The Thesis is organized following the development cycle of systems comprising design (Part I), implementation (Part II) and evaluation (Part III) phases.

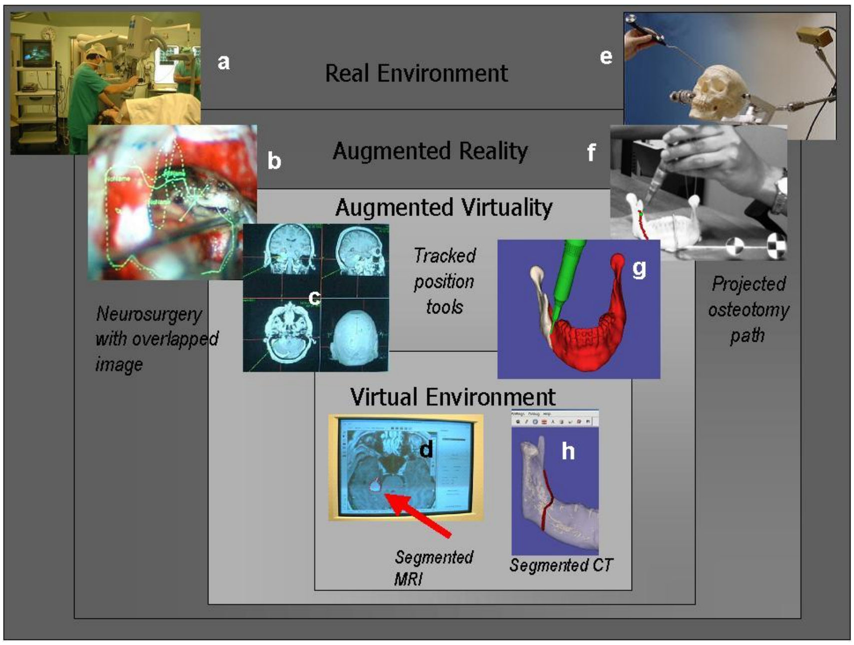


Figure 1.2: *Study cases and related zones in the reality-virtuality continuum.*

Part I introduces the reader to the fundamental basis of interactive mixed reality systems. An overview of the existing frameworks, design spaces and models to support MR design is given in Chapter 2. As a result of this study we defined our own conceptual framework for Mixed Reality Systems presented in Chapters 3, 4 and 5. How to specify a mixed interaction space is discussed in Chapter 3. In Chapter 4 we introduce our definition for the continuity criteria in MR systems and its proprieties: perceptive, cognitive and functional. Furthermore, in Chapter 5, a design space is proposed to explore such criteria. This design space is independently of interaction devices and from an abstract interaction situation, several concrete interaction solutions can be designed. At the end of this part, in Chapter 6, we introduce the two medical study cases and we use them to illustrate the conceptual framework proposed. Many scenarios are extracted from this phase and exploited during the usability tests described in Part III.

Part II is dedicated to the implementation aspects. An overview of the related methods and technologies used to develop mixed reality systems is provided in Chapter 7. We introduce the general components of a MR system and proposed two different methods to implement interactive MR systems: markerless (Chapter 8) and marker-based (Chapter 9). A registration method in real time to align real and virtual objects is also discussed, as well as the system validation.

Part III is purported to assess the interaction technique in terms of continuous interaction. We start in Chapter 10 with a review of existing evaluation criteria and methods for usability evaluation of interactive and image-guided systems. Series of usability experiments with users are described in Chapter 11 to measure our evaluation interaction criteria. Besides, statistical results from these experiments are linked to the subjective tests and used to extract valued parameters to control and assess continuity of interaction in MR systems.

Overall conclusions and perspectives of this Thesis are given in the last chapter.

Part I

Design of Mixed Reality Systems

Introduction

With traditional HCI, designers deal with the concrete design of the user interfaces, i.e. the windows, buttons, modalities, etc. Due to the merging of physical and digital entities, mixed reality system design involves design questions that are not part of this traditional approach. A new aspect of external specification must be taken into account: the abstract design of the interaction, which deals with identification and characterization of the physical and digital entities involved in the system, as well as their relationships. The abstract design should provide a model and/or a tool support for exploring the set of interaction design possibilities, without being biased by a particular modality or technology.

In this Part we address such issues starting with a review of the literature to support the design phase of MR systems. Afterwords we propose a conceptual classification of the design space and then an abstract tool for systematically exploring several design alternatives. Once the abstract design possibilities have been identified and a concrete design decision has been taken (i.e. a specific modality has been selected), a concrete MR application can be considered in order to analyze the interaction techniques in terms of continuous interaction properties which are described in the Chapter 4. Practical use cases of this approach are illustrated with two study cases in the medical field. The medical computer-guided systems shows to be a rich example for studying the conception and design of complex and even critical interaction systems.

Chapter 2

Related Works

Several studies have tried to divide interactive applications into different classes in order to create taxonomy of such systems. The interest of these approaches for the design step of an interactive system is primarily to help the designer to identify which existing applications have a similar behavior and role to the one she/he is intending to develop. The designer may then explore the design solutions used in applications of the same class in the taxonomy, and may even try to reuse parts of the technical solution. As taxonomy for MR systems is still being built, we have looked for many concepts and literature inherited from virtual and tangible systems which are kinds of MR systems. Based on such study we later propose our own conceptual framework for MR system.

2.1 Conceptual frameworks and support tools

Current work has been focused on adding virtual objects to a real environment. However, graphic overlays might also be used to remove or hide parts of the real environment from a user. Another way to augment systems is by including sound in such systems. The user could wear headphones equipped with microphones on the outside. The headphones would add synthetic, directional 3D sound, while the external microphones would detect incoming sounds from the environment. This would give the system a chance to mask or cover up selected real sounds from the environment by generating a masking signal that exactly cancels the incoming real sound [32]. The augmentation type provided by such MR applications is defined in [30] as interaction, user's action and perception. Interaction augmentation is a style of human-computer interface that tries to make computers as transparent as possible. User's actions augmentation increases the number and/or quality of tasks that the user can perform. The perception is augmented by new or more realistic information that is provided to the user.

2.1.1 Transformational framework

According to [102] the conceptual space of MR systems is divided into four transform types, linking different combinations of action and effect, along real and virtual dimensions. For example, if we take the MagicBoard application [9], a RARE (Real Action with Real Effect) transform type might be maked a draw on the board with a pencil. Alternatively, using fingers (real action) to produce a digital copy of a drawing (virtual effect) is a RAVE (Real Action with Virtual Effect) transformation. Another example of RAVE is the *Transfiction* system [76] where video images are analyzed to capture the users' movements and these are integrated into a virtual graphical scene, which reacts in an interactive manner to the behavior of the filmed subject(s).

A VAVE (Virtual Action with Virtual Effect) transformation occurs, for instance, when we select a region of interest in an image (e.g. a tumor) using a mouse, and the selected region changes its color. VARE (Virtual Action with Real Effect) transformations can be found in the experiments explored by [102]. They used the *Mimio* set-up to enable a digitally-potentiated action to trigger a physical effect. A bi-colored (e.g. blue and white) virtual windmill was displayed on the desk surface. Moving the arms of the virtual windmill trigger the spinning of a physical windmill placed nearby, which had corresponding sets of colored arms.

2.1.2 Tangible user interface frameworks

In [65] identifies four kinds of tangible user interface (TUI): spatial, topological, associative and forms applications. A spatial application augments, for example, walls, tables and white boards; a topological application is the one in which acting on physical objects modifies the behavior of the application; in systems of the associative class, physical artifacts serve as handles onto digital entities (URLs, files, etc.); finally forms applications add digital services to natural physical interactions (e.g. scanning a paper).

Drawing from the MVC (i.e. model-view-controller) approach, in [132] they have developed an interaction model for tangible interfaces called

MCRpd, for *model-control-representation (physical and digital)*. This model divides the *view* element into two subcomponents: physical representations (*rep-p*) and digital representations (*rep-d*). Where the MVC model supports the GUI's separation between graphical representation and control, MCRpd highlights the TUI's integration on physical representation and control. This integration is present not only at a conceptual level, but also in physical point - TUI artifacts physically embody both the control pathway, as well as a central representational (information-bearing) aspect of the interface.

The MCRpd interaction model provides a tool for examining several important properties of tangible interfaces and guides the steps toward a conceptual framework for tangible user interfaces. In recent work described in [131] this model has been called *MCRit* an abbreviation for *model-control-representation (intangible and tangible)* and used for hold token-constraint systems.

Token-constraint systems are those based on the TAC (Token and Constraints) paradigm [114] which describe a TUI as a set of relationships between physical objects and digital information. To support the TAC paradigm, TUIMS (Tangible User Interface Management System) has been proposed in [69]. TUIMS allows designers to specify TUIs using a high level description language called TUIML (Tangible User Interface Markup Language). This language predefines five basic models as components [118]: Task, Domain, Representation, TAC and Control. Such combined approach provides a comprehensible design process support, a repository of design data, and a representation of both abstract and concrete aspects of TUIs.

Other advances in this direction can be found in the EQUATOR[1] project. One of the Equator challenges is to develop concepts and methods that help us understand the interweaving of physical and digital interaction, and the context and development of user experiences and systems. They are also building on established work on informal reasoning about space from the AI (Artificial intelligence) and GIS (Geographic information System) communities [46] similar to Allens well known temporal relations [5] which we are using to describe the spatio-temporal relationships in our proposed mixed interaction space [128].

[1]Project Website: http://www.equator.ac.uk/

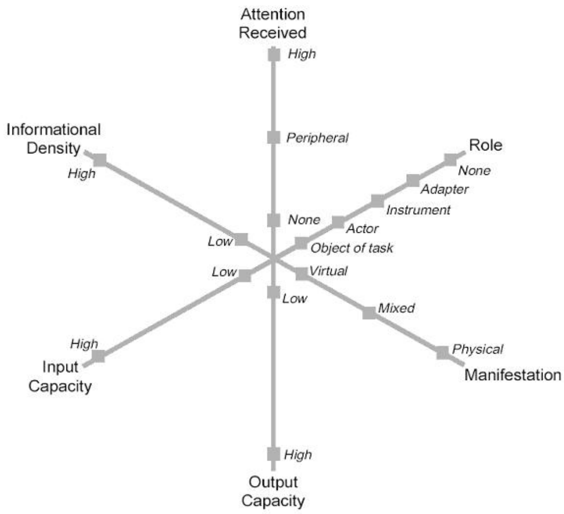


Figure 2.1: *In this space, an entity is described as a plot, from the perspective of a particular actor, carrying out a particular task at some time [44].*

2.2 Design spaces and models

2.2.1 Dimension space

The *dimension space* [44] classifies entities in a mixed reality system according to the attention they receive, and their role, manifestation, I/O capacities and informational density. Using the axes of the dimension space, we can ask questions such as: *what is the purpose of the entity (from the point of view of a user carrying out a task)?*; *how does the entity combine physical and virtual attributes*, and *how does the entity constrain the use of other entities?* Figure 2.1 shows the dimension space in a perspective view. The dimension space is not built to replace methods for describing the mechanisms by which people make sense of and organize their work. However, it includes in the existing methods a capability for generating a design space analysis tailored for hybrid physical-virtual collaborative systems. In terms of the development cycle, it fits into the system design activity between system and software requirements, as part of the process of refining the entities.

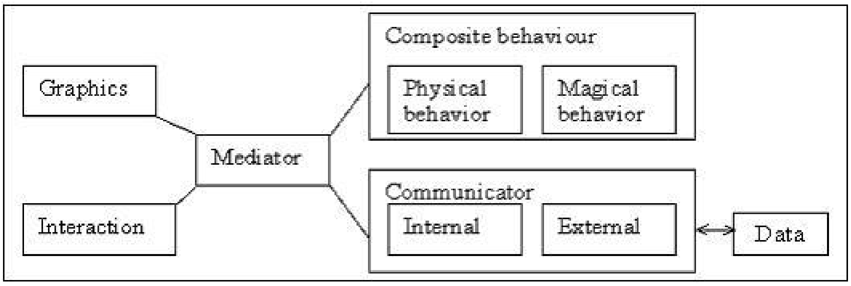


Figure 2.2: *Components of VRID model [121].*

2.2.2 VRID

The Virtual Reality Interface Design (VRID) [121] proposes a model and an associated methodology to design virtual reality interfaces. The model is based on components: graphics, behavior, interaction and mediator. The methodology consists in applying the model at two design phases: the high and low levels. In the high-level design phase, the goal is to specify a design solution, at a high-level of abstraction, using the multi-component object architecture as a conceptual guide. In the low-level design phase, the goal is to provide fine-grained details of the high-level representations, and to provide procedural details as to how they will be formally represented. The outcome of the low-level design is a set of design specifications, which are represented in formal, implementation-oriented terminology, ready to be implemented by software developers.

Figure 2.2 describes the model with the components. The mediator component refers to coordination mechanisms for communication among other components. Physical behavior refers to those changes in an object's state that are observable in the real world. Magical behavior refers to those changes in an object's state which are rarely or not seen at all in the real world.

2.2.3 ASUR

ASUR was initially designed to describe the physical and digital entities that make up a mixed system including user(s), physical and digital artifacts, and physical and informational relationships [27], [26], [30], [29], [85].

ASUR stands for *Adapter*, *System*, *User*, *Real objects*. In user-centered

MR systems are described in terms of entities (A, S, U, R) taking part in the interaction and the relations between those entities. Between the user (U) and the computer system (S), the adapters bridge the gap between the physical world and the digital one. They could be input adapters (A_{in}) (e.g., a mouse, a localization mechanism) or output ones (A_{out}) (e.g., a video projector, audio speakers). Physicality is one key feature of MR systems: real objects are involved in the task. Within the ASUR notation we distinguish physical objects that are tools (R_{tool}) for performing the task, from the ones that are the objects of the task (R_{object}).

Three kinds of relationship between two ASUR entities are identified:

- Exchange of data is represented by an arrowed line between two ASUR entities ($A \rightarrow B$).
- Physical activity triggering an action: a double-line arrow ($A \Rightarrow B$) denotes the fact that when the entity A meets a given spatial constraint with respect to entity B, data will be exchanged along another specified relationship ($C \rightarrow D$).
- Physical collocation is represented by a non-directed double line ($A = B$). This refers to a persistent physical proximity of two entities.

The purpose of ASUR is to help us to understand how to combine the physical and digital worlds by identifying physical and digital objects involved in the system to be designed (the so-called ASUR components), and the boundaries between the two worlds (the ASUR relations). A list of characteristics refines the description of the ASUR components and relations. The use of this notation allows the designer: (1) to identify several design solutions; (2) to describe them in a way that aides the comparison; (3) to study ergonomic properties in a predictable way.

The completeness of such model has been demonstrated in [85] and [75] where for each input as well as output situation all the combination possibilities of ASUR entities were described taking into account two levels of interaction: abstract and concrete. Examples of such representations applied to the CASPER (Computer ASsisted PERicardial puncture, see subsection 4.1.2) system are illustrated in Figures 2.3 and 2.4. For a complete ASUR description, the diagram is completed by the characteristics of each entity and relation. A complete description of the concrete interaction in ASUR can be found in [29].

ASUR model describes physical properties of components and their relationships while UMLi notation describes the structure and behavior

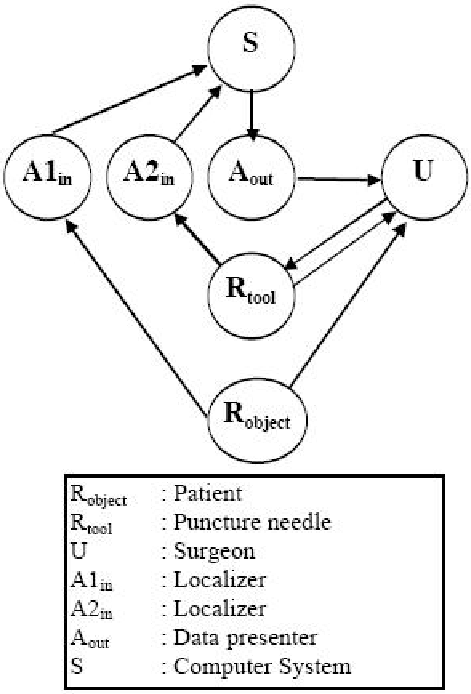


Figure 2.3: *ASUR description of the abstract interaction in CASPER. As opposed to Figure 2.4, the interaction modalities as well as the physical relationships are not yet defined at this stage of the design [85].*

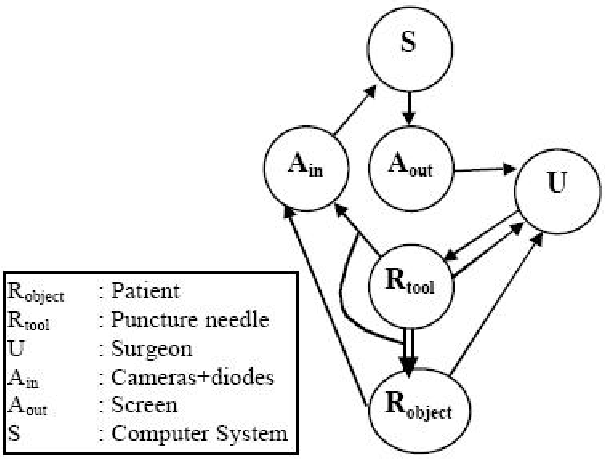


Figure 2.4: *ASUR diagram of the concrete interaction in CASPER [85].*

of user interfaces. An example of how these two notation can be used together for modeling an AR systems is given in [25].

Recently a graphic editor to support the ASUR notation has been developed in [135].

In [18] presents for the first time in the literature a mixed interaction model taking into account concepts from mixed reality and multimodal systems. The mixed interaction model proposes a classification based on the kind of objects and modality of interaction. Objects are defined in terms of their properties which can be physical, digital or represented by the link between them. It can be considered as an evolution of the concepts introduced by ASUR [27] [26] in addition to the support for the multimodal interactions. Open issues in such approach relay on the study of fusion of modalities as well as investigate the utility of the model for evaluation of the interaction usability.

2.2.4 Reference model

Structuring system design in hierarchical layers is a common approach to reduce design complexity in the software engineering practice. Reduction of the complexity can be obtained by defining for each layer what services are provided to the next layer. This way a designer can concentrate on the problems that are relevant to a particular layer while assuming that the lower layers are providing more basic, lower level means of interaction. He also can abstract from the higher level layer problems because he knows what are the means of interaction that the higher level requires. In the reference model presented in [40], the authors tentatively propose five different levels. A physical level where physical interaction takes place, followed by three levels of information processing: the perceptual, propositional and conceptual or task level (Figure 2.5). In the perceptual layer, issues such as the temporal and spatial (in-) coherence of audiovisual information streams and their effects on integration and cross modal interaction are addressed. In general this layer addresses fusion and fission of media. The proposional layer provides information in a form that allows user to transform perceived signals into meaningful signs by recognition of patterns and to associate tasks to situations or states based on a set of memorised rules. The group layer deals with aspects of interaction that are related to the coordination of the activity of more than one user, possibly in real-time. it also deals with more sociological issues of group-work that may impose particular constraints on the application, such as providing the possibility of several degrees of privacy, or allowing non-verbal

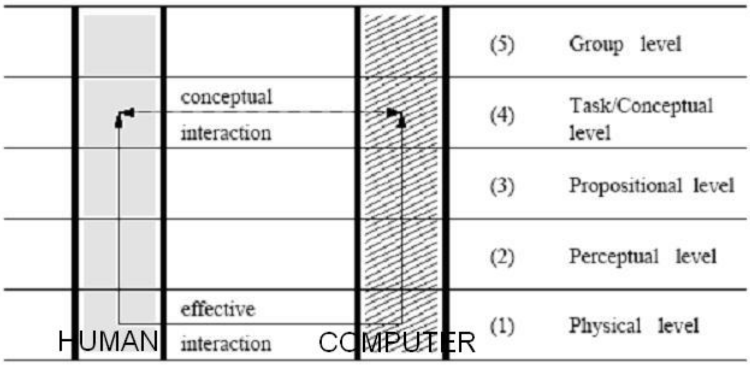


Figure 2.5: *Reference model for continuous human computer interaction [40].*

interaction in group communication.

These layers are closely following the layers in well-known frameworks such as the Skill-Rule-Knowledge-model [97] and are not too far from the levels in Norman's Interaction model [87]. The Reference Model on its own does not provide any specific technique or notation for the description of the behavioral aspects of interaction at the different layers of abstraction. It just provides a framework to guide the way in which a complex problem such as the design of continuous interfaces could be split into sub-problems at different levels of abstraction.

2.3 Comparative analysis

The Table 2.1, shows comparative aspects between the approaches discussed before regarding system type, kind of service provided during design phase, level of interaction specification, user-centered approach and notational support.

From the table we can note that few approaches have considered the user's interaction evaluation during design process. In a general view all methodologies need to provide more explicitly human-computer interaction aspects during development process of these emerging class of systems.

Even if the some of the models proposed are focused on VR systems they have been contribute to obtain a better understanding about interactions in AR systems. In particular, three major features that distinguish VR systems from others, are the (1) real time performance while maintaining minimum realism and presence, (2) objects with two clearly distinct, yet inter-related, aspects, geometry/structure and function/behavior, and (3) the still experimental nature of multi-modal interaction design. These features are also present into mixed reality systems besides those specifics features of AR systems for describing the links between real and virtual world.

As a conclusion we note the literature shows that little attention has been paid to the design of MR systems, i.e. to the phase in which tools, methods and theories are used to support the exploration of the design space, to compare the identified solution, to identify the best suitable solutions with regards to the usability of the final system. This trend is probably linked to the relative youth of the research domain. Up to now, more importance has been given to feasibility studies, exploration of new technologies and, more recently, development of tools to facilitate the engineering of such systems. The next section reports our contribution in conceptualizing and designing mixed spaces that aim at structuring the development or engineering of mixed reality systems.

Approach	Focused on	Service provided	Level of interaction specification	User-centered approach	Notational support
Dimension Space	AR, AV, VR, MR	Evaluation of entity (I/O) purpose	Concrete	no	no
ASUR	AR, AV, VR, MR	Entities (physical and digital) specification Interaction specification and evaluation	Abstract and concrete	yes	yes
VRID	VR	Interface and object's behavior specification	Abstract and concrete	no	yes
Reference model	VR, AR, AV, MR	Interaction specification	Abstract and concrete	yes	Not defined, conceptual framework
Transform	AR, VR, AV, MR	Interaction specification and classification	Concrete	yes	Not defined, conceptual framework
MCRpd	TUI	Physical representation and control	Abstract	no	Not defined, conceptual framework
TUIMS	TUI	Relationships between physical objects and digital information	Abstract and concrete	no	TUIML

Table 2.1: *Comparasion between different approaches supporting the design phase of VR, AR, AV and MR systems.*

Chapter 3

Mixed Interaction Spaces

An interaction space (IS) is assumed to be the complete presentation environment required for carrying out a particular interactive task. The interaction space contains representations of the visual, haptic and auditory elements that a user interface offers to its users, as well as their relationships. It is very often required to deal with questions such as whether particular objects or scenes being displayed are real or virtual, whether images of scanned data should be considered real or virtual, and whether a real object must look realistic whereas a virtual one need not. For example, in some AR systems it is difficult to label the remotely viewed video scene as *real* and the computer generated images as *virtual*. Compare this with an MR system in which the user reaches into a computer generated scene with his or her own hand and grabs an object. There is no doubt, in this case, that the object being grabbed is virtual and the hand is real. Comparing these two examples, it is clear that the reality of the hand and the reality of a video image are quite different, which suggests that a decision must be made about whether it is appropriate to use the term real for both cases.

In this work we adopt the distinction made in [79], that real objects are those that have an actual objective existence and virtual objects are those that exist in essence or effect, but not formally or actually. For a real object be viewed, it can either be observed directly or it can be sampled and then synthesized via some display device. For a virtual object be viewed, it must be simulated, since in essence it does not exist. This entails the use of some sort of description or model of the object.

3.1 Concrete and abstract interaction objects

The presentation of interaction spaces for MR systems is developed from the following concepts:

- Concrete interaction object (CIO). This is an object belonging to the interaction space that any user with the appropriate artifacts (e.g. a see-through head-mounted display) can see. There are two types of CIO, real and virtual. A real concrete interaction object is part of the real interaction space (e.g. live video or a physical object such as a pen or a needle) which can have a representation in the virtual world and so can become a virtual concrete interaction object. The virtual concrete interaction object is part of the virtual interaction space (e.g. text, image, animation, push button, a list box). The virtual CIO can also entail the virtual representation of the real CIO. A CIO is said to be *simple* if it cannot be decomposed into smaller CIOs, or *composite* if it can be further decomposed. Two categories are distinguished: *presentation CIO*, which is any static CIO allowing no user interaction, and *control CIO*, which supports some interaction or user interface control by the user. Both presentation and control CIOs can be part of the RIS and/or the VIS.

- Abstract interaction object (AIO). This is an abstraction of all the CIOs, that is independent of any given computing platform from both presentation and behavioral viewpoints. By definition, an AIO does not have any graphical appearance, but each AIO is connected to one or more CIOs with different names and presentations on various computing platforms.

3.2 Composing mixed reality space

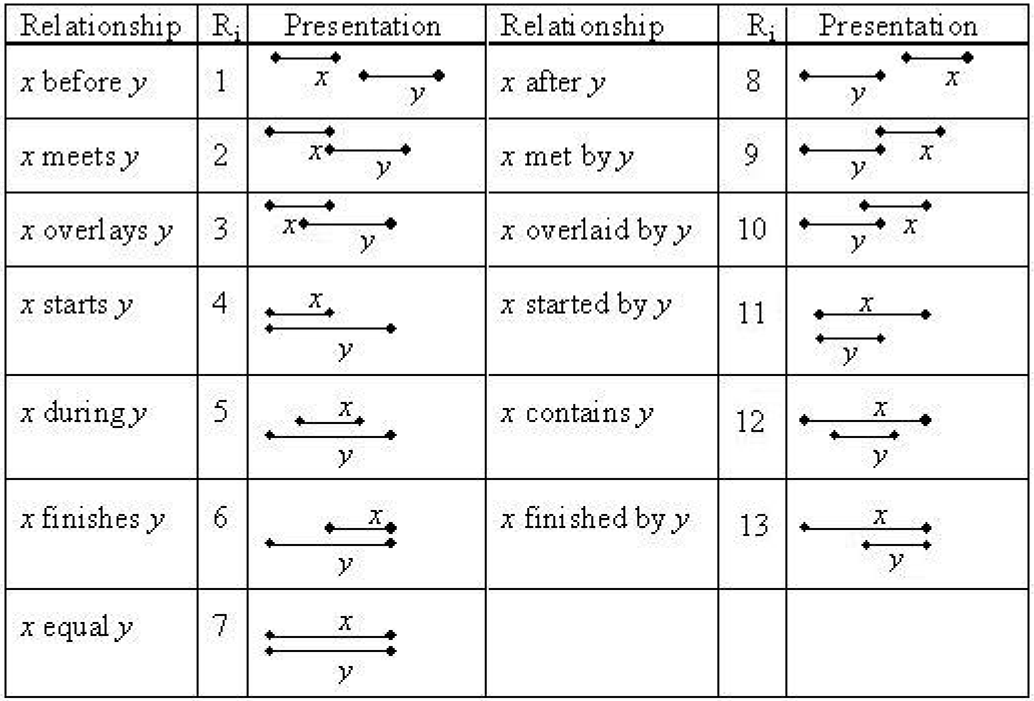
Taking into account reality-virtuality continuum (shown in Figure 3.1) the interaction space is classified as:

- real interaction space (RIS) if it consists only of real components, e.g. real concrete interaction objects such as physical objects;

- virtual interaction space (VIS) if it consists only of virtual concrete interaction objects;

- mixed interaction space (MIS) if it consists of virtual concrete interaction objects added to the real environment, i.e. combined with real concrete interaction objects.

Each MIS is composed of a VIS and a RIS, which are physically constrained by the user's workspace and are displayed on the workspace simultaneously. Each workspace is composed of at least one interaction space (IS), called the basic IS, from which it is possible to derive the other ISs (Figure 3.2). This configuration is necessary for the user to manipulate objects in the virtual world through the VIS or objects in the real world through the RIS. Basically the RIS is composed of information from the real world (e.g. a live video source), and when the user interacts with the real world it has some effect on the virtual world representation (i.e. on the VIS).



Relationship	R_i	Presentation	Relationship	R_i	Presentation
x before y	1	x y	x after y	8	y x
x meets y	2	x y	x met by y	9	y x
x overlays y	3	x y	x overlaid by y	10	y x
x starts y	4	x y	x started by y	11	x y
x during y	5	x y	x contains y	12	x y
x finishes y	6	x y	x finished by y	13	x y
x equal y	7	x y			

Figure 3.1: *Temporal Allen relationships [5].*

3.3 Spatio-temporal relationships

The interaction space may involve a large number of media objects which should be integrated into the MIU (mixed interaction unit). This integration covers the spatial ordering and topological features of the abstract interaction objects. At the abstract level, the designer is interested in expressing only the high-level relationships between AIOs, if any, not the low-level details of the relationships, such as specific distance or time. Spatio-temporal relationships characterize the physical links between AIOs as they are presented in time and space. Since an abstract user interface (AUI) does not preclude the use of any particular modality, we do not know whether a particular AUI will be further reified into a CUI that is graphical, vocal, multimodal or virtual. Therefore, spatio-temporal relationships should be expressed in a way that is independent of any modality.

Allen [5] identified thirteen possible temporal relationships and these are listed in Figure 3.1. Basically, there are two types of temporal relationships: sequential (before the relationship) and simultaneous (which can be equal, meets, overlaps, during, starts, or finishes). Each basic relationship has an inverse, except for the equal relationship which is symmetric. Although Allen relationships have been introduced to characterize temporal intervals, they are suitable for expressing constraints in both space and time, thanks to a space-time value. All simultaneous relationships (such as overlaps, during, starts, and finishes) can be generalized as equal relationship by inserting some delay time when needed. For example, in the x before y relationship, the space-time value between x and y is greater than zero, while in the x meets y relationship the space-time value between x and y is equal to zero. As relationships are abstract at the AUI level, the space-time value is left unspecified until needed at the CUI level. For instance the meets relationship can be used in a vocal user interface to specify that a question is first presented to the user, then immediately after a prompt to recognize the vocal answer of the user. In this case, the relationship represents a time interval.

One-dimensional Allen relationships can be used to specify the temporal composition of n objects. Let A and B be two AIOs. The temporal relationship between A and B is defined as:

$$TemporalComposition(A, B) = R_i,\ where\ i \in \{1, \ldots, 13\}$$

as defined in Figure 3.1.

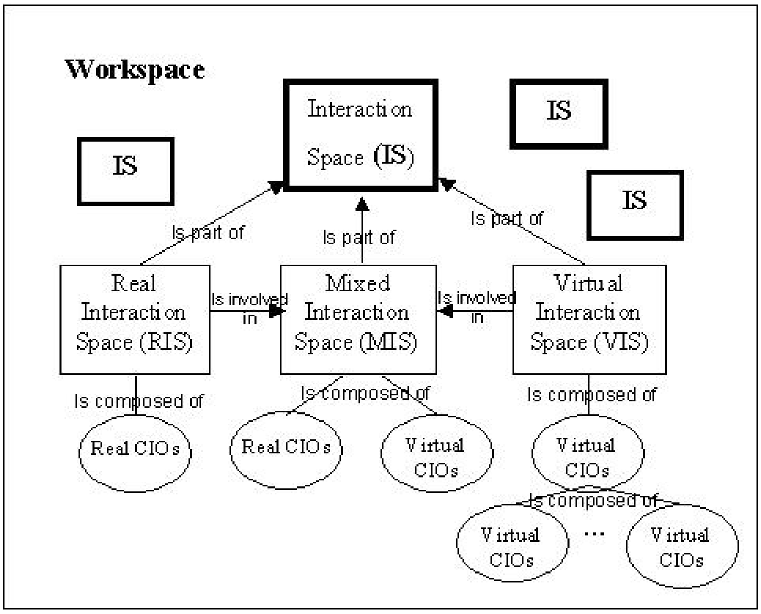


Figure 3.2: *Representation of interaction spaces in a MR workspace*

Allen relationships [5] can be generalized to n dimensions to express similar constraints in an nD space [133]. Here, the 2D generalization is used to express space relationships more precisely in any type of UI involving spatial expressions. To exemplify this, let us assume two AIOs A and B. The spatial relationship between A and B is defined as:

$$SpatialComposition(A, B) = (R_i, R_j),\ where\ i, j\ \in\ \{1, \ldots, 13\}$$

as defined in Figure 3.1.

R_i is the identifier of the spatial relationship between A and B according to the X axis and R_j is the identifier of the spatial relationship between A and B according to the Y axis in the matrix reproduced in Figure 3.3. When a spatial arrangement is expressed only according to one dimension,

$$R_i = \phi\ or\ R_j = \phi$$

.

Figure 3.3 graphically depicts areas of the Allen matrix that are appropriated for representing the spatial composition of certain types of UIs.

The two first lines (i.e. $R_{1,1} - R_{13,1}$ and $R_{1,2} - R_{13,2}$) and columns (i.e. $R_{1,1} - R_{13,1}$ and $R_{1,2} - R_{13,2}$), and the two last lines (i.e. $R_{12,1} - R_{12,13}$ and $R_{13,1} - R_{13,13}$) and columns (i.e. $R_{1,12} - R_{13,12}$ and $R_{1,13} - R_{13,13}$) of the matrix allow the spatial composition of GUIs to be expressed, since widgets in GUIs do not overlap. They can be completely separated from each other (first and last lines and columns, represented in light gray in Figure 3.3). In certain specific cases, they touch each other or share one part of the convex envelope (area represented in dark gray in Figure 3.3). The remaining area, represented in white in Figure 3.3, allows spatial arrangement for semi-transparent and see-through GUIs to be expressed [10] where widgets can overlap thanks to transparency layers [49], for multimedia UIs or for MR UIs [93].

In the next section we discuss how to integrate the mixed interaction space into the specification of user interfaces taking account its static and dynamic aspects.

3.4 Meta model for mixed interaction spaces

An interactive system design mainly relies on the modeling of the system to be developed. A model has four main roles: (1) to facilitate the discussions among the different partners of the design team; (2) to back up the results of the discussions; (3) to systematically explore the set of design possibilities at an early stage; and (4) to keep a trace of the evolutions of the choices and discussions. The model of an interactive system should describe the data manipulated, the services offered and the behavior of the system. External specifications also have to be specified during the design step of an interactive system design.

A User Interface (UI) model is hereby defined as a set of concepts, a representation structure, and a series of primitives and terms that can be used to explicitly capture knowledge about UI and its related interactive application with appropriate abstractions. A model is assumed to raise the level of abstraction with respect to the details or code level by promoting appropriate abstractions of the real world. Any concept of the real world may therefore lead to multiple and ample possibilities of abstractions depending its influence in developing a future UI. Ideally, any model should be declarative (rather than imperative or procedural), editable (preferably through tools) and analyzable (to allow some degree of automation).

How many models do we need? One UI model is probably too complex to handle because it gathers all static and dynamic relations into one

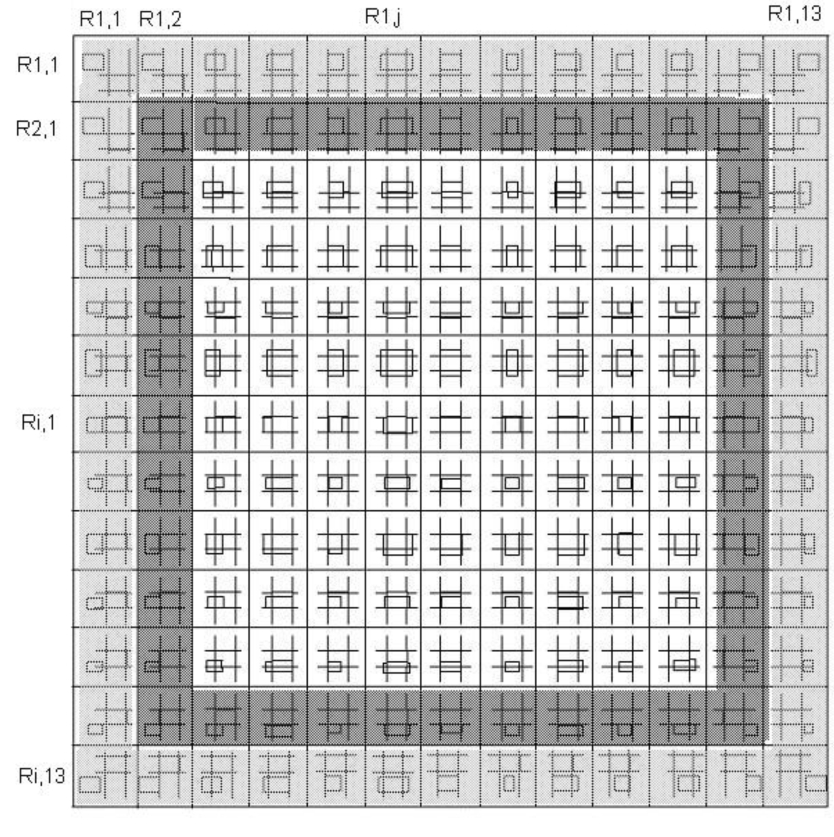


Figure 3.3: *Areas of the Allen 2D Matrix [5] suitable for classic GUIs (grey) and multimodal and mixed UIs (white).*

model, it can pack all abstractions in the same place, thus not reducing complexity to a level that can be mastered. Too many models should be avoided too because the more models we have, the more relations we need to establish and maintain between them.

By reviewing the literature (see Section 2) we see there no many models to support the design of MR systems. We propose to use a model-based approach [89] which aims to identify the specification of the abstract users interface in terms of static structure (Domain of discourse, Task and Presentation model) and dynamic behavior (inter models relations). The domain of discourse provides a global overview of all entities involved in the application and their relationships. The task model describes tasks in terms of goals, action and domain objects. The presentation model is a representation of the elements that should be manipulated by user or by system for accomplishment of the task. Finally the interaction model describes interaction between user and the available objects.

The Domain of discourse includes User, Device, Task, Application and Domain Models. A class diagram can be used to represent all entities of the discourse domain and the static relationships existing among them. Figure 3.4 provides a class diagram for MR systems represented in the UML abstraction [12].

3.4.1 Application model

The application model corresponds to the system class (Fig. 3) and can be composed of one or more computer-based system. Synchronization between different systems can be necessary to exchange information. The system class may synchronize events from devices according the task performed and also integrate different media sources into one or more displays mixing information about both worlds.

3.4.2 Platform model

A device is capable of providing input to the essential computer, receiving output, or both. Then the platform model (represented by device class in Figure 3.4) consists of two categories: input and output devices. The input devices are used to enter commands or information into a computer, such as the keyboard, mouse, touch screen, data glove, etc. For example a sensor is a kind of input device that can track users and objects and it is essential in AR environments. The output devices are any devices for give feedback acoustic, haptic or visual, in a form intelligible to the user, the results of processing carried out by a computer. The device class has information about synchronization type, list of objects that may be synchronized and media type that the device can present of according his physical capabilities (like resolution, size, mobility, etc).

3.4.3 Domain model

The Domain Model defines the objects that a user can view, access and manipulate through a user interface or directly in the physical world (see object class in Figure 3.4). Objects can be real (patients, paper document, a pen, a needle, etc) or digital (images, sounds, etc) which support the task execution. Real or physical object is any object that has an actual objective existence and can be observed directly. Digital objects can be either real or virtual. Digital virtual objects are objects that exist in essence or effect, but not formally or actually. In order for a virtual object to be viewed, it must be simulated, since in essence it does not exist. This entails use of some sort of a description, or modeling of the object like a rendered volume model of brain. However, a live video image of the patient, for example, is a digital real object. A typically screen-based computer system can only render a small amount of the total information on an output device as a head mounted display or a microscope display. For this reason the digital object (such as a path line for guidance) visibility can be observable or browsable. Observability means that the object is always visible and browsability means that the object is visible only on demand. Hence there must be some browsing function (activate by user, for example when the user activates the tracking process or by some agent, for example when the system presents the warning messages automatically during the guidance process) allowing the user to inspect the information in stages. In most situations, like those required by surgical procedures, only the most critical information can be immediately available according to the actual user's task. However the user should be able to access all relevant information eventually. This kind of observability - often called browsability - is based on the general principle of allowing the users to perceive anything they can name, i.e. anything that he or she can provide a description for it [45].

3.4.4 User model

The user model corresponds to the user class (Figure 3.4). This class has an ID, the user function containing information about the user (as experience level, profile, etc) and user position in the SRU - system reference of universe for those applications that require tracking the user. The user interacts with objects in the real world and in the virtual world through some kind of object (object tool) or a device interactive.

3.4.5 Task model

The task model is designed to lead to a goal from the current state of the world and task class in Figure 3.4 represents it. Execution of a task changes the current system state to a new system state and the task focus can be in the real world, in the virtual world or shared between both worlds.

The Task Model describes the tasks an end user performs and dictates what interaction capabilities must be designed. It is typically a hierarchical decomposition of a task into sub-tasks to end up with actions that are no longer decomposed and that can be effectively carried out. The model involves elements such as goals, actions and domain objects. Goals specify when a desired state is met. Sequences of actions define procedures to achieve a goal. Domain objects represent elements that must be available for the user complete each task in the model. By analyzing the temporal relationships of a task model, it is possible to identify the set of tasks that are enabled over the same period of time. Thus, the interaction techniques supporting the tasks belonging to the same enable task set are logically candidate to be part of the same presentation though this criteria should not be interpreted too rigidly in order to avoid too modal users interfaces. Tasks synchronization can be simultaneous or sequential and performed by one user, by various users, by the system, by a third party, or by any combination.

3.4.6 Presentation model

Presentation Model is a representation of the visual, haptic and auditory elements that a user interface offers to its users. It corresponds to the mixed interaction space specification and so far the presentation model should consider physical and spatial integration. The physical integration is controlled by the system and it describes how the user will receive feedback from involved systems and how the media are distributed into outputs devices. It means that each media could be displayed in different displays or integrated within the same display.

3.4.7 Inter-model relations

Following the inter-model relations represented in Figure 3.4 one user may manipulate zero or many devices; s/he may accomplish zero or many tasks that may require zero or many objects and the user can be sensored by zero or many sensors. One task is decomposed into one or many subtask and it may have the focus in the real, virtual or shared world. A task can be accomplished by zero or many user, involved by zero or many system and it can be required by zero or many object. Each device may be manipulated by zero or many users and/or operated by zero or many systems. The device has physical capability to deal with zero or many objects and can present zero or many objects.

A computer system may be a stand-alone system or may consist of several interconnected systems. The system involves one or many tasks and operates zero or many devices. The system is responsible for the synchronization and integration events came from other entities. These relationships will be used later to evaluate interaction in terms of continuity.

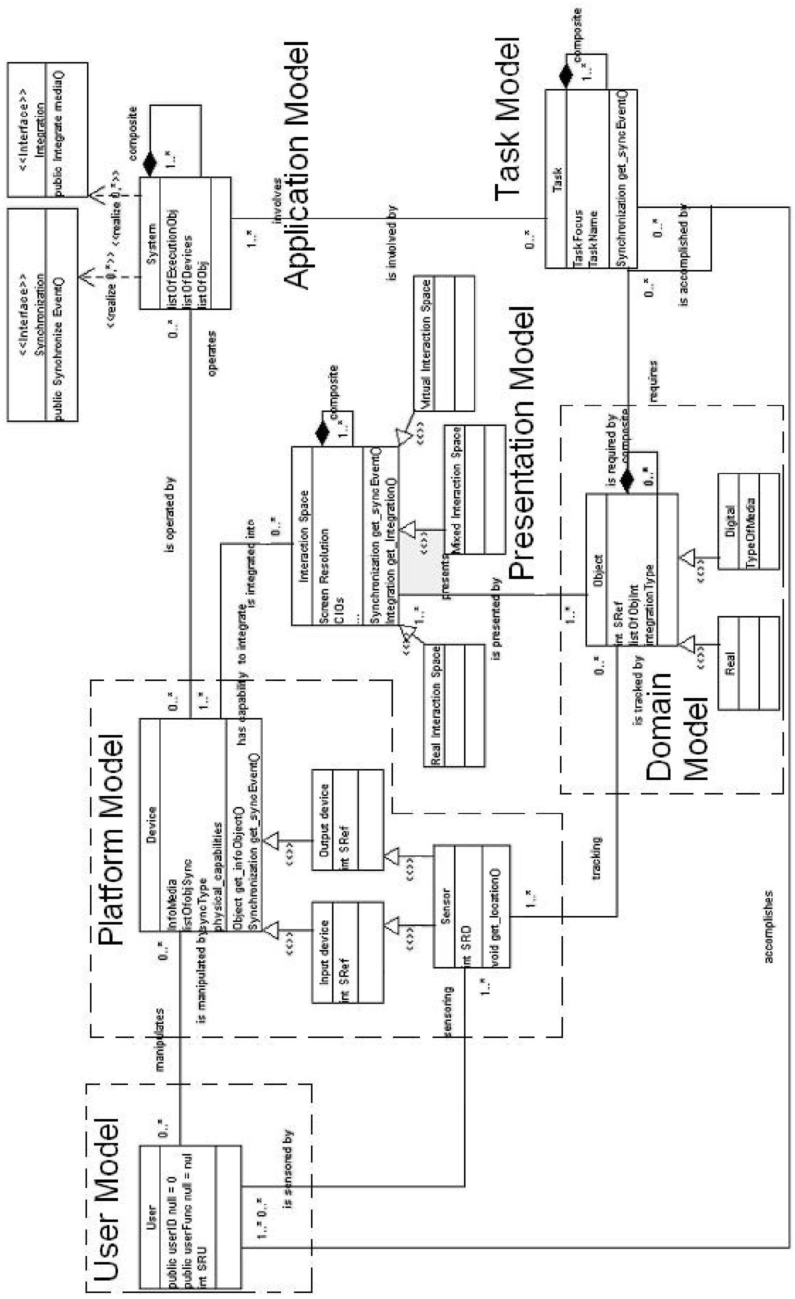


Figure 3.4: *UML diagram class for MR systems.*

3.5 Conclusion

In this chapter we have introduced the definition of mixed interaction spaces as well as its components. Trying to identify relationships between components in the mixed interaction space we propose to use spatio-temporal relations. With that, queries as follows can be extracted:

- pure spatial or temporal query: only a temporal or a spatial relationship is involved in the query. For instance, *which objects always overlap the presentation of live video A?*; *which objects spatially lie above object B in the interaction space?*
- spatio-temporal query: where such a relationship is involved. For instance, *which objects spatially overlap with object A during its presentation?*
- MIS query: spatial or temporal layouts of the application considering interaction focus and insertion context. For instance, *what is the spatial integration (layout of MIS) when the user's interaction focus is shared between A and B objects?*; *which objects are presented when the user's focus interaction is focused on the real world?*; *when the user's focus is on the real world, how is the insertion context of MIS?*; *when the user has the temporal control of presentation where is located the user's interaction focus?*

To help designers identify all components involved in a mixed reality system as well as answer such queries we suggested to use a meta model approach. The models identified were: User, Device, Task, Application and Domain Models. A class diagram in UML can be used to represent all entities of the discourse domain and the static relationships existing among them. Support to dynamic relationships analysis can be obtained from inter models connections as described in Section 6.

Chapter 4

Continuity in MR Systems

Interaction in MR systems is no longer based only on the exchange of discrete messages that could be considered as atomic actions. Instead, the input provided by the user and/or the outputs provided by the computing system are a continuous process of information exchange at a relatively high resolution. As almost, all tools used to interact with the virtual world are separated from those used to interact with the real world it forces the user to switch between operation modes resulting in a discontinuous interaction. Another potential discontinuity can be found for different or not similar representations of the real data in the virtual world. In this chapter we propose to investigate the definition of the continuity properties in the field of mixed reality and interactive systems. We starting by presenting previous work done for defining continuous interaction in interactive and augmented reality systems. Afterwords we introduce our definition for continuity which will be used as usability criteria while evaluating MR systems (see Part III).

4.1 Related works

In the last years continuous interaction has been the interest of works such as those related in [58], [63], [84], [30], [124]. Researches projects are also focused on system design for supporting continuous interaction. For instance the aim of the TACIT project[1] (Theory and Applications of Continuous Interaction Techniques) is to develop theories (including mathematically based models) and methodologies for the design of interfaces in

[1]Website:http://kazan.cnuce.cnr.it/TACIT/TACIThome.html

which continuous interaction techniques are used. They agree *Continuous interaction is a term which has been coined to describe a form of interaction that distinguishes a number of emerging technologies from more traditional interaction techniques. In modern interaction techniques such as gesture recognition, speech recognition, animation and haptic feedback, the user is in constant and tightly coupled interaction with the computing system over a period of time. The interaction is no longer based only on the exchange of discrete messages that could be considered atomic actions. Instead, the input provided by the user and/or the output provided by the computing system are a continuous process of exchange of information at a relatively high resolution. Often the recent history of the interaction plays a role in the behaviour of human and machine. For example, to recognize a gesture the system has to record the trajectory over a certain period of time.*

As results of those studies we can see continuity as being particularly concerned with activity over a period of time. At a low level, this can involve real-time aspects of technologies such as gesture recognition. At a higher level, providing for continuity during a user's interaction with an application can be quite a challenge, as the context of use, environmental conditions and device platform may all change repeatedly.

In the next sections we explore two point of views given by previous work in which we have based our definition for continuous interaction. The first approach is based on spatial, temporal and functional seams emerging mostly from tangibles and AR interactions. The second approach takes account two ergonomic properties of augmented reality systems: continuity and compatibility. Both approaches are described as follows.

4.1.1 Spatial, temporal and functional seams

Much of the ongoing augmented interaction research focuses on integrating the computational components to create systems adapted to their users, tasks or environment. Ishii et al. [58] define a seam as a spatial, temporal or functional constraint that forces the user to shift among a variety of spaces or modes of operation. For example the seam between word processing using a computer and traditional pen and paper makes it difficult to produce digital copies of handwritten documents without a translation step. Seams that force a user to move between interaction spaces are called *functional* seams, while those that force the user to learn new modes of operation are *cognitive* seams.

Many authors [6] have agreed that systems asking users to abandon their acquired skills and to learn a new protocol are likely to encounter strong resistance.

Several research groups have begun to explore Tangible User Interface (TUI) in which real world objects are used as computer input and output devices, or as Hirishi Ishii puts it *by coupling digital information to everyday physical objects and environments* [58]. Tangible interfaces are extremely intuitive to use because physical object manipulations are mapped one-to-one to virtual object operations, and they follow a space-multiplexed input design [37].

So we see that current Tangible interfaces provide very intuitive manipulation of digital data, but limited support for viewing 3D virtual objects. Consequently it may introduce a discontinuity or functional seam between the interaction space and display space. In contrast most AR interfaces overlay graphics on the real world interaction space and so provide a spatially seamless display. However they often force the user to learn different techniques for manipulating virtual content than from normal physical object manipulation or use a different set of tools for interacting with real and virtual objects. So AR interfaces may introduce a cognitive seam but introduce discontinuities in interaction.

In this sense mixing TUI and AR seems to be a good alternative for the chasm between real and digital worlds. A Tangible AR interface provides true spatial registration and presentation of 3D virtual objects anywhere in the physical environment, while at the same time allowing users to interact with this virtual content using the same techniques as they would with a real physical object. So an ideal Tangible AR interface facilitates seamless display and interaction, removing the functional and cognitive seams found in traditional AR and Tangible User Interfaces. This is achieved by using the design principles learned from TUI interfaces, including [63]:

- The use of physical controllers for manipulating virtual content.
- Support for spatial 3D interaction techniques (such as using object proximity).
- Support for both time -multiplexed and spacemultiplexed interaction.
- Support for multi-handed interaction.
- Support for matching the physical constraints of the object to the requirements of the interaction task.

- The ability to support parallel activity where multiple objects are being manipulated.
- Collaboration between multiple participants AR interfaces that follow these design principles will provide completely seamless interaction with virtual content and so will be extremely intuitive to use.

4.1.2 Continuity and compatibility in AR systems

The analysis of continuous interaction provided in [84], [30] is based on two ergonomic properties that characterize the output user interface:

- Observability characterizes *the ability of the system to ensure that the user can perceive the internal state of the system from the perceivable representation of that state*. [23] [45];
- Honesty characterizes *the ability of the system to ensure that the user will correctly interpret perceived information and that the perceived information is correct with regard to the internal state of the system* [45].

They distinguish the observability and honesty of multiple concepts (called compatibility) from the observability and honesty of multiple representations of a single concept. The latter case is called the continuity property.

Additionally Norman's Theory of Action [87] models users' mental activities in terms of 7 steps, which include a perception step and an interpretation step.

Observability and honesty ergonomic properties are thus directly related to these two steps:

- Observability is related to the users' perception (perceptual level) while honesty supports the users' interpretation (cognitive level). Consequently, compatibility and continuity can be applied at both the perceptual and cognitive levels:
- Compatibility at the perceptual level denotes how easy or difficult it is for the user to perceive all the concepts provided at a given time by the system.

Compatibility at the cognitive level is assessed in terms of the cognitive processes involved in the interpretation of all the concepts perceived.

Continuity at the perceptual level is verified if the user directly and smoothly perceives the different representations of a given concept.

Continuity at the cognitive level is verified if the cognitive processes that are involved in the interpretation of the different perceived representations lead to a unique interpretation of the concept resulting from the combination of the interpreted perceived representations.

Table 4.1 summarizes these ergonomic properties when applied for an augmented reality system.

Perceptual Level	Observability	Perceptual Compatibility	Perceptual Continuity
Cognitive Level	Honesty	Cognitive Compatibility	Cognitive Continuity
	1 concept 1 representation	*N* concepts, 1 representation each	1 concept, *N* representations

Table 4.1: *Ergonomic properties of observability, honesty, compatibility and continuity in Augmented Reality systems [30].*

To assess continuity in such terms two examples of AR systems are considered: a computer aided surgery system and a mobile augmented reality system. Both systems are briefly described here below.

CASPER

CASPER [25] [30] [84] is a system for computer assistance during pericardial punctures. The clinical problem is to insert a needle percutaneously to access an effusion in the pericardium, an envelope that surrounds the heart. During this operation, there is the danger of puncturing anatomical structures such as the liver or the heart itself. CASPER provides real-time accurate information about the position and orientation of the needle, superimposed on a pre-planned trajectory. Basically, after having acquired a set of ultrasound images, and based on these medical images, having planned a safe linear trajectory to reach the effusion, guidance is achieved thanks to the use of an optical localizer that tracks the needle position. The interface consists of four areas, hereafter referred to as visor, gauge, ultrasound image and numerical data. The ultrasound image and numerical data are not used by the surgeon during the surgery. The gauge contains the different parts of the preplanned trajectory: the skin and tissues, the effusion and the heart. In addition, a dynamic cursor is displayed on top of this static representation. The cursor denotes the current position of the extremity of the needle along the depth of the pre-planned trajectory. The cursor should thus never enter in the last part of the gauge (heart) and the effusion should be punctured only when the cursor, that represents the extremity of the needle, is in the part of the gauge that represents the effusion. Finally, the visor is composed of three crosses: one stationary representing the pre-planned trajectory to follow; and two dynamics - one encoding the position of a given point of the axis of the needle, according to the pre-planned trajectory to be reproduced; and other corresponding to the position of the extremity of the needle, according to the pre-planned trajectory.

MAGIC

MAGIC - Mobile, augmented reality, Group interaction in Context - is a generic hardware and software mobile platform for collaborative augmented reality. In order to smoothly combine the digital and the real worlds a gateway was created. This gateway plays the role of a door between the physical and digital worlds. As a door belongs to two rooms, the gateway exists in both worlds:

- the gateway is an area of the physical world, delimited by a rectangle displayed in a semi-transparency Head-Mounted Display (HMD),
- the gateway is a rectangular area in the digital world, on the pen computer screen.

Objects in the gateway are visible on the HMD (i.e., in the physical world) as well as on the pen computer screen (i.e., in the digital world). The complete software and its architecture are detailed in [86].

When assessing continuity the perceptual environments involved in the interaction must be identified as well as the different representations of involved objects. While the surgeon is inserting the needle through the body of the patient, the three crosses must be aligned. Indeed, when the three crosses are superimposed the executed trajectory corresponds to the planned one. Moreover, to be sure that the surgical needle has not been distorted during its insertion, the surgeon regularly checks the surgical needle. Consequently, during the introduction of the surgical needle, the surgeon has to look both at the guidance information displayed on the screen, and at the surgical needle in the operating field. The required switch between the screen and the operating field was disturbing to the surgeon. Perceptual continuity is not verified here. Likewise at the cognitive level, two cognitive processes are involved and are very different. Indeed, the representation of the needle on screen (the two mobile crosses) is bi-dimensional, while the position of the real needle is of course three-dimensional. The matching between the real object and its representation is far from direct. Cognitive continuity is transgressed. On the other hand, using MAGIC, the user is looking at the physical environment that is augmented by digital objects. There is only one perceptual environment. Nevertheless the digital objects are picture, bi-dimensional objects, while the physical objects are three-dimensional. As a conclusion, the perceptual continuity is verified but at the cognitive level, the digital objects are not completely integrated in the physical environment.

Recent researches in this field have indicated some potential proprieties of mixed interactions which could be considered during usability evaluation. For instance the fluidity property discussed in [84]. The fluidity between real and digital world can be classified as: no continuity, perceptual continuity, cognitive continuity or perceptual and cognitive continuity.

4.2 Defining continuity

As has been discussed in many works [57] [56] [7] [77] [119] [34], humans have well documented limitations in attention, memory, learning, comprehension, sensory bandwidth, visualization abilities, qualitative judgments, serial processing and decision making. For an augmented system to be successful it must identify at least one of these bottlenecks in real time and alleviate it through a performance enhancing mitigation strategy. These mitigation strategies are conveyed to the user through the adaptive interface and might involve: modality switching (between visual, auditory and haptic), intelligent interruption (for instance in assisted tasks), task negotiation and scheduling. When a user state is correctly sensed, an appropriate strategy chosen to alleviate the bottleneck, the interface adapted to carry out the strategy and the resulting sensor information indicates that the aiding has worked - only then has a system *closed the loop* and successfully augmented the user's interaction.

Taking account such reflection in addition to those presented in the related works section (see Section 4.1) we presented here our definition for continuous interaction in MR systems which may have the goal to be able to evaluate the interaction in terms of:

How, where and when the information will be delivered to the user (e.g. perceptual properties) will reflect how easy or hard the information will be interpreted by the user (e.g cognitive properties) and how natural will be the interaction process (e.g. functional properties).

Then we define continuity as *the capacity of the system to promote a smooth interaction with the user during the task accomplishment considering perceptual, cognitive and functional properties.* Interaction is assumed to be the actions cycle conforming defined by Norman's theory (see Section 4.3).

In the next sections we identify the design factors involved in each property. With that we will be able to estimate the level of continuity provided by one specific mixed interaction scenario suing the design space (see Chapter 5). Each continuity property is detailed bellow.

4.2.1 Perceptual continuity

The perceptual property of continuity is defined as an ability of the system to make all data involved in the user's task available in one perceptual environment in order to avoid changes in the user's focus.

The following design aspects (see Chapter 5) may be identified while evaluating the perceptual property:

- Kind of augmentation: user's perception can be augmented by providing additional information for the user to perform his/her task in an adequate place and device. Perceptual property is addressing the user's perception augmentation (i.e. kind of augmentation axis should be fixed to the *perception* item in the design space).

- Insertion context: it is sorted according to the distance at which each device displaying the interaction space is inserted in the environment relative to the user's position and the user's task focus. It's quite related to the device used to display the interaction spaces. This design aspect is addressing WHERE issue related to device insertion into the environment.

- Spatio-temporal links: fusion mechanisms related to the spatial arrangement and temporal synchronization of objects in the interaction spaces. This design aspect is addressing WHERE-WHEN issues related to insertion of objects in the interaction spaces according to Allen's relationships (see Section 3.3). Examples of usage of this approach can be found in [134] while applied to multimodal systems and in [130] while applied to MR systems.

- Interaction focus: it is quite related to the depth cues which involves perceptive issues such as accommodation, convergence, binocular parallax, motion parallax, occlusion, shades, shadows, perspective, colors and brightness, tactile sense, texture, etc. Such aspects can be used to control the user focus during interaction. This design aspect is addressing HOW issues (i.e. how to display the information) related to the kind of visualization used.

4.2.2 Cognitive continuity

According to the Dubois et al. work [30], cognitive continuity is defined as the ability of the system to ensure that the user will interpret all the perceived information correctly and that the perceived information is correct with regard to the internal state of the system. It can be done for instance by using similar representations of the real and virtual objects and displaying useful information to facilitate the task execution.

A MR scene provides a distortion in space and time (i.e. it depends if the system is on real time or not) to continuously combine the current focus of detail with the overall context. Every viewpoint should include currently relevant information. Providing the continuous context together with the current focus of interest serves as an external memory for decision and reasoning. *Out of sight is often out of mind* [136]. The combination of local experience with the overall structure improves building maps.

The following design principle (see Chapter 5) may be identified while evaluating the cognitive property:

- kind of augmentation: cognitive property is augmenting the user's actions and in this case the kind of augmentation axis should be fixed to the *action* item in the design space. To have augmented actions the number and/or quality of tasks that the user performs should increase. Consequently it involves the final user's performance using a specific technique of interaction.

- Media: It is related to the sensory channels used to interpret the information. Media is sorted by level of complexity and dimensionality, starting with basic modalities such as text (1D) and image (2D) and finishing with more complex and structured modality type such as those found in 3D animation and immersive environments. The level of cognitive continuity that the used modality may provide is quite related to the perceptual cues discussed before and the consistency criteria. If we have found good solutions for the perceptual issues better will be the consistency (i.e. how close to or far from the real concept it is) and then better will be the cognitive continuity of the interaction. It follows the theory of action proposed by Norman (Section 4.3) where the evaluation and interpretation depends on how the user perceives the information.

Other aspect that can have influence in the cognitive continuity but is not directly linked to the design aspects is the knowledge level, for instance. Knowledge level is based on the retrieval and adaptation of knowledge constructed from the use of other devices, systems or even the knowledge of the task (expert and non-expert users).

4.2.3 Functional continuity

Functional continuity is defined as the adaptability of the user to change or learn new modes of interaction. Consequently the functional property is related to the interaction device/tool used. We have identified two different levels of functional properties: interaction and task levels.

Interaction level. . The user's interaction can be augmented by providing similar mode of operation or interaction (e.g. use of tangible interfaces [58]). Augmenting interaction involves the following design principles (see Chapter 5):

- kind of augmentation: it is related to the interaction augmentation and in this case the kind of augmentation axis should be fixed to the *interaction* item in the design space.
- Connection type: it is sorted by level of complexity in registering information. Environments with static links (i.e. where links between the real and virtual world are established during design time) are considerably less complex than environments in which all links are established during the execution time.
- Transform type: is arranged according to the users' level of familiarity with the tuple action/effect. Thus real action with real effect is highly familiar, while virtual action with virtual effect is highly unfamiliar.

Task level. It is related to the continuity of the task through the different contexts (i.e. real, virtual and transition between them) and it involves the following design principles:

The user's interaction focus is sorted by degree of reality-virtuality. The user could be either performing a task in order to manipulate or modify an object in the real world (when the task focus is on the real world), or an object in the virtual world (when the task focus is on the virtual world) [31]. According to where is place the user's interaction focus in a specific time we could identify two kinds of task continuity according to the context (i.e. real and virtual contexts):

- Intra-context task continuity: it is related to the temporal transition between tasks performed in the same context then in the same world or in the same perceptive environment.
- Inter-context task continuity: it is related to the temporal transition in saving and recovering task content from one context to other or from one perceptive environment to another.

Trying to capture the task continuity level during a mixed reality interaction, the usability tests are designed to track the user interaction focus through different scenarios using different kinds of stimulus (e.g. kind of guidance and kind of visualization).

4.3 Norman's model and continuous interaction

One important goal of this work is to measure the continuous interaction during task accomplishment. We have defined interaction according to the theory of action proposed by Norman and Draper [87]. This theory assumes that for people reach the goal of their tasks they have to perform actions. Actions have two distinct aspects: they have to be executed, and their results have to be evaluated. The stages of execution (intention, action specification and execution) are coupled to the stages of evaluation (perception, interpretation and evaluation). With this model Norman shows that problems in the use of objects and interfaces can be explained as the discrepancy between the intention of the user and the actions made available by the system (execution flow), and the discrepancy between the physical representation of the system and the expectation and intentions of the user (evaluation flow).

Conforming presented in Figure 4.1 we have adapted the Norman's model to take account the continuity properties. The user perceives (i.e. perceptual continuity) the information from the mixed interaction space, interprets, evaluates, forms intention and specifies the action sequence (i.e.

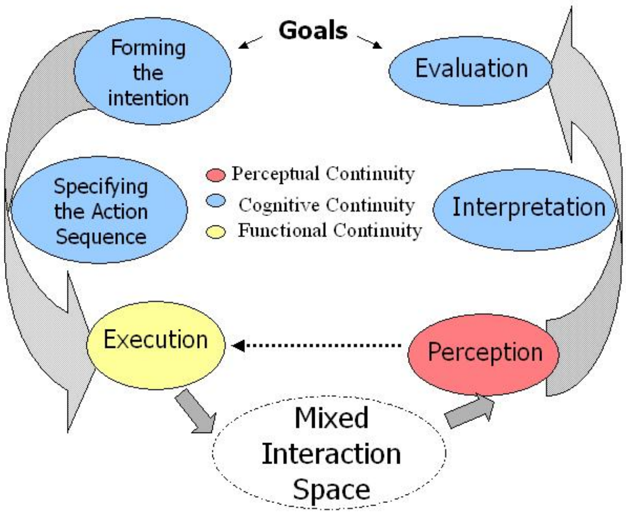


Figure 4.1: *Theory of action adapted from Norman's model [87] and the related continuity properties.*

cognitive continuity) and finally execute the next action in the mixed interaction space (i.e. functional continuity). If the interaction cycle is not involving cognitive continuity, the action can be executed directly after the changes in the mixed interaction space has been perceived.

4.4 Conclusion

In this chapter we have shown why continuous interaction is a keep point while analyzing user interaction in MR systems. Interactions between real and virtual worlds involve manipulations of different kinds of objects and consequently different input/output interaction modes are needed. Such aspects clearly distinguishes MR systems from others such as GUI and VR systems. Based on the literature and previous works we have proposed our definition for continuous interaction in terms of perceptual, cognitive and functional properties. Continuous interaction will be used as our usability criteria for evaluation of MR system as will be described in Part III. In the next chapter such continuity properties will be linked to the mixed realities design principles to help designers identify levels of continuity at early stages of system development.

Chapter 5

Design Space for Mixed Reality Systems (DeSMiR)

To understand how different kinds of mixed realities might produce different user experiences taking account the continuity of interaction that differ from other kinds of computer-mediated interactions requires us to examine the question of how people will deal, not only with the virtual spaces but also with mixed reality environments that combine real, virtual and ubiquitous forms.

Considering this multiplicity of interaction worlds and many kinds of application in which MR is applied we propose a design space to assess continuous interaction at early stages of MR system development. Our design space, called DeSMir, considers six aspects of the conception of MR systems while a specific task is being carried out. These are the *transform type*, *connection type, insertion context*, *media*, *interaction focus* and *kind of augmentation.* We use the Kiviat diagram to display several data variations on several axes for a given item (the task item). Such axis were chosen based on the literature existing to support continuous interaction in VR and MR systems.

DeSMiR is an abstract tool which proposes design principles to identify levels of continuity (conforming defined in Section 4.2 in interaction at early phases of MR interaction design.

5.1 Axes description

The transform type axis is arranged according to users' level of familiarity with the types. Thus real action with real effect is highly familiar, while virtual action with virtual effect is highly unfamiliar. The kind of augmentation axis explores the possible types of augmentation that the system can provide. Augmenting perception (e.g. by adding, removing, etc. virtual objects to/from the real world) is the most common event, while augmenting action is uncommon since it requires more sophisticated equipment.

The connection type axis is sorted by level of complexity in registering information. Environments with static links (i.e. where links between the real and virtual world are established during design time) are considerably less complex than environments in which all links are established during execution time.

The interaction focus axis is sorted by degree of reality-virtuality, while the insertion context axis is sorted according to the distance at which each device displaying the interaction space is inserted in the environment relative to the user's position and the user's task focus.

A device is a piece of hardware used to access and interact with an application. It is capable to provide input to the computer, receive output, or both. The device class consists of two categories: input and output devices. Input devices (such as a keyboard, mouse, touch screen, data glove, etc.) are used to enter commands or information into a computer. For example, a sensor is a kind of input device that can track users and objects and it is essential in AR environments. Output devices are used to give acoustic, haptic or visual feedback, in a form intelligible by the user, on the results of processing carried out by a computer.

For the central zone we have devices inserted from 0 to 45 cm from the user's position and this provides a potential induced continuity. For the public zone we have devices at distances greater than 3.6 m from the user's position and the user's task focus, and this provides a potential induced discontinuity. The media axis is sorted by level of complexity and dimensionality, starting with basic media such as text (1D) and image (2D) and finishing with more complex and structured media type such as those found in 3D animation and immersive environments.

The items situated at the external border of the design space axes (Figure 5.1) have a greater probability of being near the real pole of diagram shown in Figure 1.1. On the other hand, the items situated at the center of the design space are more likely to be near the VE zone of diagram.

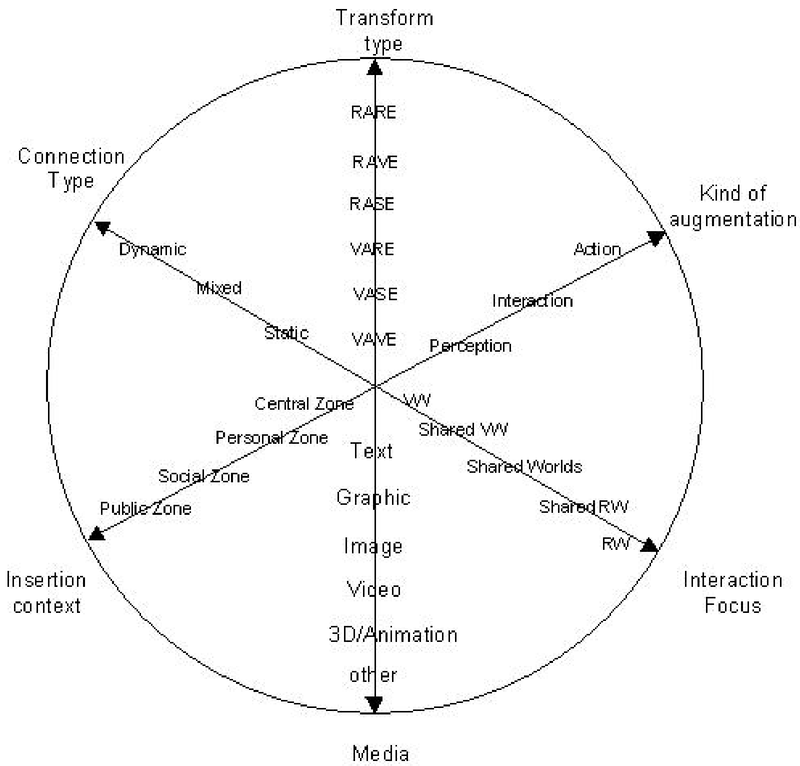


Figure 5.1: *DeSMiR: design space for mixed reality systems*

5.2 Kind of augmentation

To conceptualize the kind of augmentation we are using the definition given in [30].

Interaction augmentation is a style of human-computer interface that tries to make computers as transparent as possible. For instance, by using the concept of tangible interfaces in which the interaction is based on physical objects. For example, [100] shows how users can smoothly exchange digital information between their portable computers and a computerized table and wall, using a technique called hyper-dragging.

User's actions augmentation increases the number and/or quality of tasks that the user can perform. For example, in the MagicBoard application [9], the user can perform cut/paste operations on real drawings, an action which is not possible in the real world.

User's perception is augmented by new or more realistic information that is provided to the user. For example in the Museum project [99] the user can perceive additional information (such as historical data on the paint used for a picture) which is unavailable in the real world.

5.3 Transform type

The term "'transform type"' is used to identify the different types of transformation occurring within the real, virtual and virtuality-enhanced trio identified in the definition of MR environments. We have used the classification given in [102] where the transform type can be:

- real action with a real effect (RARE);
- real action with a virtual effect (RAVE);
- virtual action with a virtual effect (VAVE);
- virtual action with a real effect (VARE).

We have completed this classification by including two other possible kinds of transformations:

- real action with a shared effect (RASE). This kind of transformation can be found in guided surgical systems. By tracking the surgical instruments in the frame of reference of the medical imagery, the surgeon acts in real world but has feedback in both worlds (e.g. changes in the actual patient and in the virtual representation);

- virtual action with a shared effect (VASE). This can be found in systems such as remote environment visualization and manipulation for monitoring and exploration in distant or hazardous locations. For example, in [4] they have developed an augmented virtual world that contains real world images as object textures that are created in an automatic way. They call these reality portals. Using these reality portals with the robotic system, a human supervisor can control a remote robot assistant by issuing commands using the virtual environment as a medium.

5.4 Connection types between worlds

According to the taxonomy proposed by [84], the links between the real and digital worlds can be characterized by two axes: the owner of the link (e.g. the person defining the link) and its nature (e.g. static, dynamic). For example, the designer of the microscope-image-guided surgery system [124] decided to combine digital representation of a tumor with the representation of the current position of the tumor using optical trackers and registration procedures. Such a link is static and it was defined by its designer (during design time). On the other hand, using MAGIC (mobile augmented reality, group interactive, in context), the users dynamically define new digital objects that are combined with physical objects. Examples of applications using this technology can be found in [101].

We propose an extension of this taxonomy to take account of the fact that the link between real and digital worlds can be defined by:

- the designer (e.g. a static link created in the design process);
- the user (e.g. a dynamic link defined during execution time);
- a third part (e.g. an agent system which is capable of making decisions and initiating actions independently, during execution time);
- mixed initiative (e.g. a combination of the above).

5.5 Interaction focus

When there are multiple sources of information and two worlds of interaction (real and virtual) we must make choices about what to attend to and when. Same times, we need to focus our attention exclusively on a single item without interference from other items. Other times, we may need to time-share or divide our attention between two (or more) items of interest, which can be part of the same or different worlds.

For example, in the Museum project [99] the user wears a see-through head-mounted display in which information about an exhibit is displayed. The user is thus able to perceive real objects (the exhibit) and add synthetic information. The object of the task here is the painting in the exhibit. Therefore, the task focus belongs either to the virtual world or to the real world.

The user could be either performing a task in order to manipulate or modify an object in the real world (when the task focus is on the real world), or an object in the virtual world (when the task focus is on the virtual world). By considering all the possibilities of interaction focus while the user is performing a specific task, we have identified five possible combinations (see Figure 5.2):

- Interaction focus in real world without shared attention. In this type of interaction, attention is focused on only one object in the real world. There is no other object competing for the user's attention;

- Interaction focus in virtual world without shared attention. In this type of interaction, attention is focused on only one item in the virtual world. There is no other object competing for the user's attention;

- Interaction focus shared in real world (intra-world interaction focus). In this case, the interaction focus is shared between real objects in the real world;

- Interaction focus shared in virtual world (intra-world interaction focus). In this case the interaction focus is shared between virtual objects in the virtual world;

- Interaction focus shared between worlds (inter-world interaction focus). Here the interaction focus is shared between objects belonging to different worlds (real and virtual).

5.6 Insertion context

Sutcliffe [117] says, *We all like to have personal space surrounding us, generally about 0.5 m, although this is culturally dependent*. Here we want to define

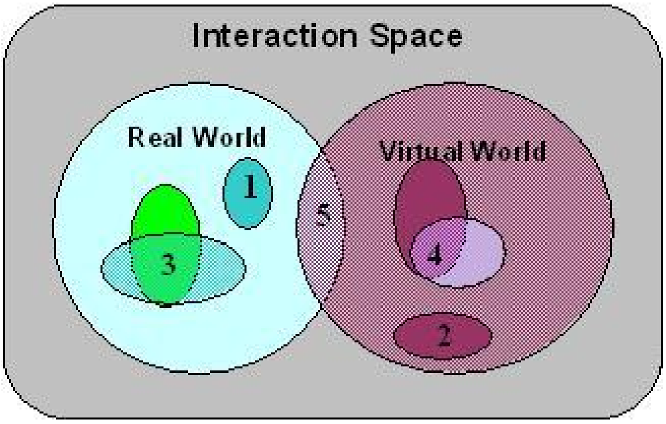


Figure 5.2: *Types of interaction focus in a workspace. The numbers 1 and 2 represent interactions without shared attention; 3 and 4 represent inter-world interaction focus; and 5 represents intra-world interaction focus.*

the space around us according to the user's focus while performing a task. An interaction space can be concretized using any device (screen, HMD, etc.) or any physical object (projected at a table, wall, etc.). We have identified four spatial zones (Figure 5.3) for describing interaction space (IS), depending on the level of periphery:

- Central zone. Corresponds to an insertion distance of 0 to 45 cm from the user;
- Personal zone. Corresponds to an insertion distance of 46 cm to 1.2 m from the user;
- Social zone. Corresponds to an insertion distance of 1.3 to 3.6 m from the user;
- Public zone. Corresponds to an insertion distance greater than 3.6 m from the user.

If the IS is inserted in the central zone of the user's task, she/he does not need to change her/his attention focus to perform the task. If the user's attention focus is changing all the time, then it is probable that the IS has been inserted outside the central zone, in a peripheral context of use. For instance, Figure 5.4 shows a potential source of discontinuity in image-guided surgery systems: while a surgeon is operating on a patient lying on a table (the primary task and, thus, the main focus of attention),

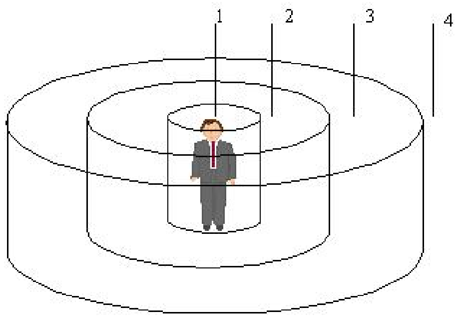


Figure 5.3: *Insertion zones of interaction spaces regarding users task focus.*

additional information is displayed on TV screens and monitors. Those devices are not necessarily located close to the patient's location, thus forcing the surgeon to switch attention from the patient to the various devices and back again. The further the devices are from the main focus of attention, the more discontinuity will be induced. Such kind of discontinuity has been previously studied in [30] in the first version of CASPER system (see Section 4.1.2).

Figure 5.4: *Example of potential discontinuity where information is being displayed in the personal zone [124].*

In the Museum project, which is one application of the NaviCam system [99], the interaction space is inserted in the central context of the user's tasks, he/she does not need to change his/her attention focus to perform the task (Figure 5.5).

A similar approach based on the spatial model has been used by [8] to create highly interactive environments where objects dynamically react to the presence of other objects (e.g. a tool can be activated simply by approaching it).

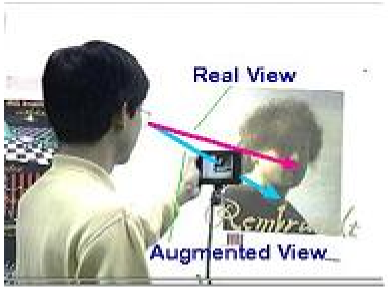


Figure 5.5: *Example of potential continuity during users interaction with NaviCam system [99].*

5.7 Exploring the design space

Once we have identified in what continuity property (see Chapter 4) each axis can be involved in, we can use the design space to map optimal zones for the interaction continuous properties. Figure 5.6 illustrates zones covered by each continuity property.

The optimal cognitive zone is defined by the augmentation and media axes. The *action item* is selected and fixed on the augmentation axis, while the *media item* used in the augmentation is selected on the media axis.

The optimal perceptive zone is defined by the augmentation, interaction context and interaction focus axes. The *perception item* is selected and fixed on the augmentation axis; the *insertion context item* is selected and fixed on the insertion context axis according to the user task focus; while the *interaction focus item* supported by the augmentation is selected on the interaction focus axis.

The optimal functional zone is defined by the augmentation, transform type, connection type and insertion context axes. The *interaction item* is selected and fixed on the augmentation axis; the *insertion context item* is selected and fixed on the insertion context axis according to the user task focus; while the *transform item* supported by the interaction is selected on the transform type axis and the *connection item* supported by the system is selected on the transform type axis.

5.8 Conclusion

A main objective of this theoretical study is to provide a conceptualization of mixed realities that could give us a means to investigate and inform the

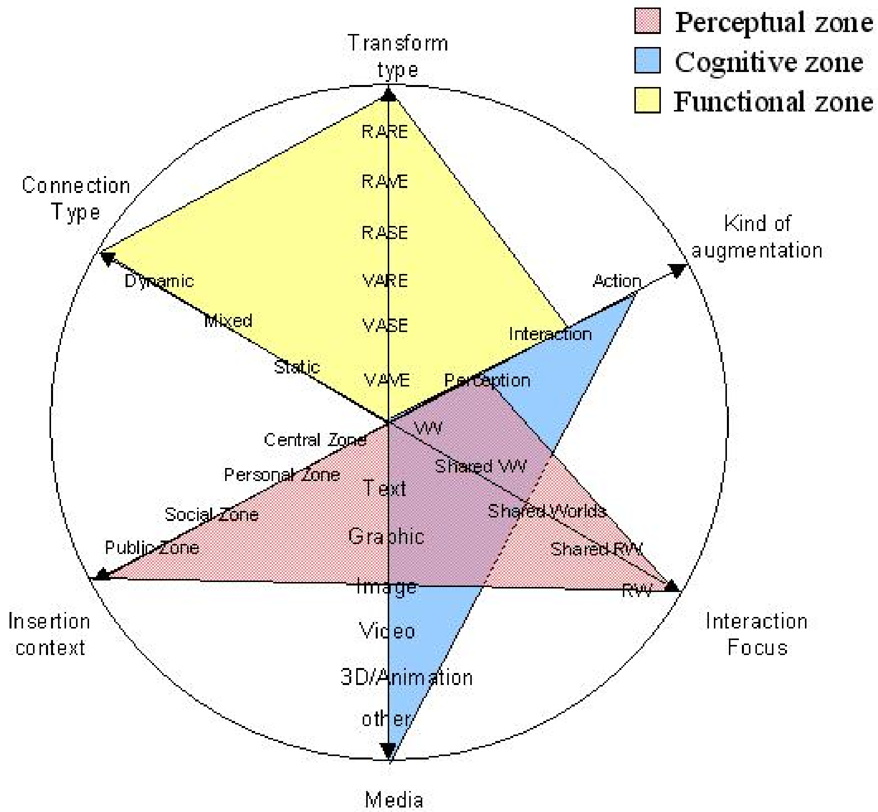


Figure 5.6: *Zones of continuity properties according to the design principles*

design about continuity in mixed reality interactions.

As results of this study we have proposed a set of design principles organized in a design space (DeSMiR) for identify continuous interaction properties at early phases of MR interaction design. The design principles were extracted and adapted from literature and previous works to characterize the different properties of continuous interaction as discussed in Chapter 4. Table 5.1 summarizes the relation between continuity properties and the proposed design principles. By using such relationships, the design space can be used to map optimal zones for the interaction continuous properties.

Continuity Properties	**Related design principles**
Perceptive	Augmentation of user perception Interaction focus Insertion context Spatio-temporal links
Cognitive	Augmentation of user action media
Functional	Augmentation of user interaction Transform type Connection type

Table 5.1: *Proposed design principles and their associated continuous interaction properties.*

We suggest that these design principles can be applied for the design of several kinds of mixed reality applications. However, we mainly address those applications in what smooth connections and interactions with worlds is critical for the system; i.e. image-guide surgery, drivers applications or pilot simulations.

In Chapter 6 we illustrate how to use such design space to identify the level of user interaction continuity (assessed in terms of perceptual, cognitive and functional properties) in an image-guided surgery system.

The usability tests described in Chapter 11 will verify the influence of some of those design principles in the final user interaction.

Chapter 6

Practical Approach

In this chapter we present a practical approach to show how the design space can be used to help designers to think about how to combine real and virtual objects in a mixed interaction space providing efficient assistance during surgical procedure. In the evaluation part (Part III), we will show how decisions and choices in this area are directly linked to the evaluation of interaction continuity.

6.1 Computer-guided surgery

Real world interventions with computer guidance can be implemented with virtual and/or augmented reality. Interventions often need prior planning and there is a lot of information that has been presented to the surgeon. In the conventional way, the information is presented spatially distributed (i.e. on a light box), or temporally distributed (i.e. by looking the images one by one, in sequence). However, what the surgeon needs is to combine the images into an augmented reality scene so that all information needed for the next step in the intervention is combined.

During an intervention, augmented-computer-guided systems have two primary roles: guidance and control. The guidance function employs enhanced reality as an orientation. The control function provides a comparison between the plan and the current patient situation, using intra-operative imaging. The guidance function intends to orient the surgeon towards the established plan: by outlining the path to the target, by highlighting predetermined dangerous areas, and by associating the patient image with structural medical information that allows the surgeon to relate medical knowledge with the structures founded. Examples of such systems can be found in [24], [107], [61],[138]. For guidance, multimodal pre-operative images are registered with any segmented structures (such as tumor target), with any simulations performed (such as the laser treatment), and with anatomical and pathologic atlases (such as functional areas). Guidance functions can be visual, using only a screen to orient the surgeon, or they can employs mechanics or robotics to constrain the surgical action. A guidance based on pre-operative image data is necessary for minimally invasive surgery because the intra-operative views or images are often poor quality. Since target areas are so small, human vision is constrained, and interventional imaging devices, because of their real time nature and spatial constraints, are also less detailed than pre-operative images. A complete taxonomy on augmented surgery is detailed in [24].

On the other hand, a control system is also necessary because numerous deviations from the pre-operative plan can occur, and thus must be recognized by the surgeon. Normally a navigational system presents images with orientation and location defined by the position of a tracked device, which can guide the surgeon to the target with relatively good accuracy unless, the position of the targeted structure and/or the surrounding anatomy has changed significantly during the procedure. In some cases, shifts and deformations of soft tissues occur during surgery because of mechanical factors, physiological motion, swelling, or hemorrhage. These changes may displace organs or their tissue components to such a degree that preoperatively acquired image-based 3D models cannot be registered with the patient's actual anatomy. In this situation either partial correction of the 3D model or full volumetric image update is necessary. Even limited revision of the original images requires some clues about the changes taking place during surgery. While some useful information can be obtained from prior knowledge and experience-based models of deformations, only intra-operative imaging can provide true information. These updated positional data then can be used to modify the original volumetric images, using more elaborate computer-based methods. Thus, intra-operative images are a key component in maintaining anatomic integrity.

Our two study cases focus on the intra-operative guidance function using one modality of pre-operative image (Magnetic Resonance Imaging or Computed Tomography) as described in the next sections. We will no consider intra-operative deviation control.

6.2 Case study 1 - Image-guided neurosurgery

In this study case we described the procedures involved in a neurosurgery using a microscope guidance system. The application describe here was already developed with a proprietary system and it is currently in use at the Saint Luc Hospital in Brussels.

Since the introduction of the operating microscope, neurosurgery has become safer, more precise, and more effective because visualization has improved. The operating microscope has expanded the surgeon's capabilities by extending visualization.

Figure 6.1 shows the flowchart for image-guided surgery from pre-operative MRI scanning (upper right) with transfer through the imaging network to a workstation where surgical planning and 3D modeling is done (bottom right). After the patient is positioned for the surgery in the operating room (OR), calibration and registration are done to place all of the pre-operative images and instruments into a common reference frame (bottom left). As the surgery proceeds, a navigation workstation, a microscope display and a TV monitor are used to guide the procedure based on the original and processed images (upper left). The surgical navigation in the operation room is the most complex interaction design once the surgeon has to deal with surgical procedures and guidance information at the same time.

Basically the scenario evolving a surgery tumour resection of brain can be dividing in 3 main phases: pre-operative, registration and intra-operative procedures. Each task has a set of sub-tasks conforming listed in Table 6.1.

In the pre-operative phase the doctor navigates in a workstation display through by many pre-operative images of patient brain. He inputs shape of tumor to all pictures necessary, locates the fiducials on the images, the entry point and the target. After that, the doctor verifies the 3D reconstructed tumour model in another window with 3 different views (i.e. axial, coronal and sagital).

These information will be used to planning the surgery during the guidance process following.

For that a registration step is needed. Registration is the means by which the virtual is related to the real and this procedure is divided in two steps: microscope registration and pre-operative images registration. Hence registration is done to place all of the pre-op images and instruments into a common reference frame.

Prior to registration, the fiducial head markers as acquired with the pre-operative data, have to be identified in the images.

Patient-to-image registration is conducted to match spatial coordinates. This is performed after the patient is anesthetized and the localizer system has been mounted to the table of the operating room.

The surgeon locates the fiducial on the images. Then, with the microscope, he points at the three ends of a star-shaped arm that is rigidly fixed to the metallic arms that hold the patient's head. The movement of star-shaped arm is monitored by an infrared camera, so that when the surgeon needs to move the patient's head, the registration patient-images remains correct. Then he points the microscope at the fiducial on the patient's head. Since it knows the position of the fiducial on the images and on the patient, the computer is now able to perform the registration. From then on, all microscope positions and orientations can be translated into positions and orientations in the pre-operative image data. The doctor verifies position and direction of microscope together with pre-operative images in the display.

After the registration the doctor verifies position and direction of microscope in the image and then drafts the final craniotomy plan on the patient's head.

In the intra-operative phase the pre-operative tumour image is overlapped with the microscope image and the surgeon see the path line planning on TV monitor. Then, during the surgery the surgeon interacts with the system through clicking buttons and needs to stop his actual task. Some of options available in the system are: to control contrast and activate the tracking system. For example, when a warning message is showed in the display, if the doctor wants to proceed he should press the button.

When the tracking process is activated the surgeon can see the actual microscope position and direction in the pre-operative image.

When the tumour is reached and the tracking process is activated, the path lines planning on microscope system disappear.

Figure 6.2 illustrates such domain of discourse according to the general diagram class for MR systems conforming described in Figure 3.4.

Table 6.1 is describing design principles such as interaction focus and transform type while the user(s) is performing pre-operative, registration and intra-operative tasks.

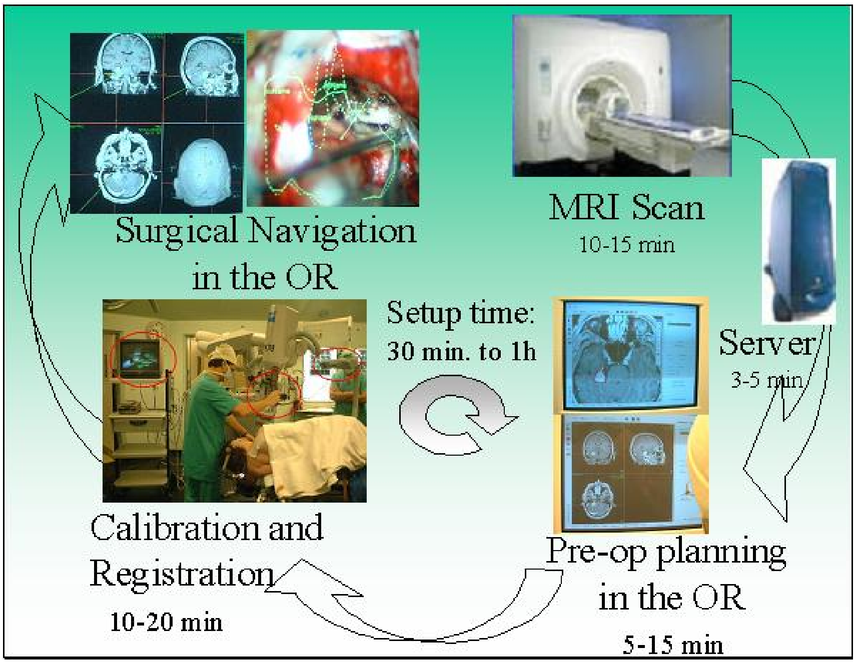


Figure 6.1: *Image-guided surgery scenario where images acquired pre-operatively are registered to the patient images intra-operatively.*

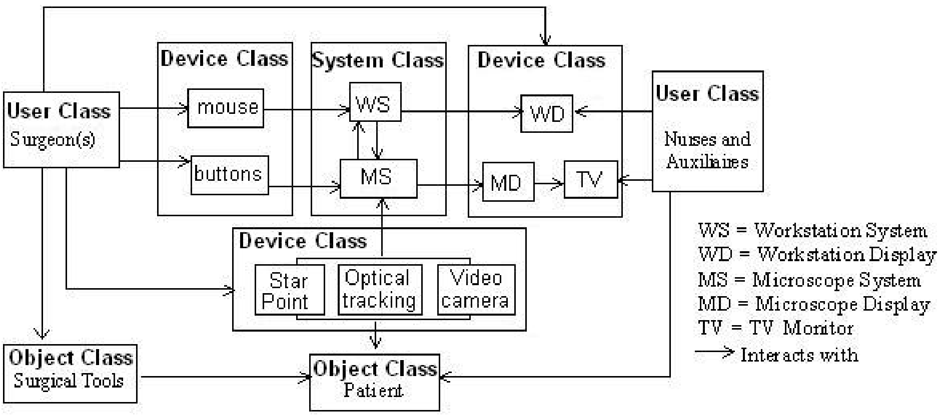


Figure 6.2: *Domain of discourse for Image-guided surgery according to the UML class diagram described in Figure 3.4.*

Sub-task	User	System	Input Device	Output device	Interaction Focus	Transform type
Load Dataset	Surgeon	WS	Mouse		Digital world	VAVE
Visualize Dataset	Surgeon	WS	--	WD	Digital world	
Select ROI	Surgeon	WS	Mouse		Digital world	
Define Entry Point	Surgeon	WS	Mouse		Digital world	
Activate 3D reconstruction	Surgeon	WS	Mouse		Digital world	
Show 3D reconstruction		WS	--	WD	Digital world	
Positioning patient	Surgeon Auxiliares	--	--		Real world	RARE
Positioning devices	Surgeon Auxiliaries	--	Star pointer infrared camera		Real world	
Activate registration patient-to-image	Surgeon	MS	Microscope buttons		Real world	RAVE
Register patient-to-image		MS	Star pointer Infrared camera		Digital world	
Positioning microscope	Surgeon	--	Microscope		Real world	RARE
Activate registration microscope's positions	Surgeon	MS	Microscope buttons		Real and digital world	RAVE
Register microscope's positions		MS	Star pointer Infrared camera		Real and digital world	
Activate updating images	Surgeon	MS and WS	Microscope buttons		Digital world	VAVE
Show microscope's positions and directions		WS		WD	Digital world	
Activate move microscope	Surgeon	MS	Microscope buttons		Real world	RARE
Moving microscope		MS	Microscope		Real world	
Activate IGS	Surgeon	MS	Microscope buttons		Real and digital world	RAVE
Shows images overlapped	--	MS and WS		MD and Monitor TV	Real and digital world	
Activate Tracking System	Surgeon	MS and WS	Microscope buttons		Real and digital world	RAVE
Shows images updated	--	MS and WS	--	WD	Real and digital world	

□ Pre-operative tasks   □ Registration tasks   □ Intra-operative tasks

Table 6.1: *Tasks, entities and design principles applied to the image-guided surgery scenario.*

6.2.1 Analyzing inter-model relations

Taking account the surgical navigation task which take place during the intra-operative phase (see Table 6.1) simultaneously with the surgical procedures, we have the following inter-models relations (see Figure 3.4):

Task and User model: this task is accomplished by one surgeon;

Task and Domain model: this task requires real (live video of the patient) and digital (pre-operative images, microscope position and warning messages) objects (see Figure 6.3 (a));

Domain and Presentation Model: the Presentation unit (i.e., design of interaction space) presents the domain objects (medium) according to the device (modality) that will be used to delivery them. In this case (see Figure 6.3 (a)) we we have overlapping real image from patient and digital path line in a microscope display. The warning message is displayed in context of information, that's mean not overlapping the region of interest (view of patient).

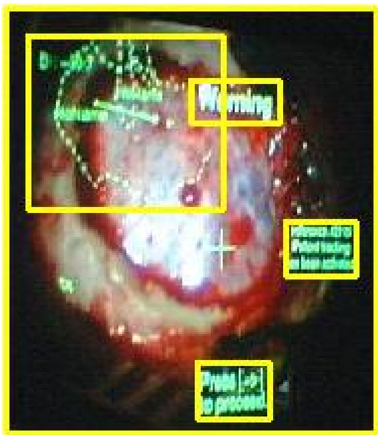
Presentation and Platform Model: the interaction spaces units are integrated into different devices according to their capabilities of render these presentations. For our example the microscope display (which is inserted in central context of user's task) shows one interaction space and the workstation display (which is inserted in peripheral context of user's task) shows another one with complementary information to guide the surgeon.

User and Platform Model: the surgeon manipulates the microscope system through a manual remote control, the workstation system through a mouse and the patient through surgical tools. With this kind of interaction the user needs to stop one interaction to start another one causing a breaking down in the interaction flow and a functional discontinuity. As both displays, microscope and workstation, are required to guide the surgeon during the surgery there is also a perceptive and functional discontinuity during the visualization once that the surgeon needs to change her/his attention focus to access all available information to perform the task.

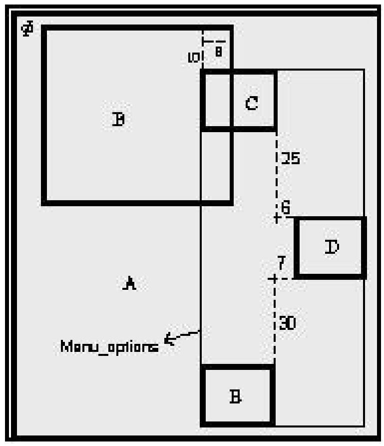
6.2.2 Scenario description

Let us assume the following intra-operative task *show images overlapped* conforming tasks listed in Table 6.1 for the image-guided surgery (IGS) scenario [124]. The display starts with the background presentation of a live video image A (located at point 0 relative to the application origin

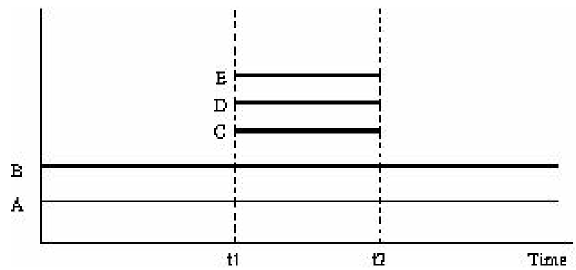
ϕ as represented in Figure 6.3). At the same time, a graphic path line B overlaps image A according to the registration procedures. At time t1, determined by tracking system procedures, the OptionsMenu containing the texts C, D and E are displayed. Object C appears partially overlapping the right side of Object B, 10 cm lower than the top and less than 8 cm from the right of B. Object D appears 25 cm below and 6 cm to the right of Object C. Object E appears 30 cm below the bottom and less than 7 cm to the left of D. Figure 6.3 graphically depicts the initial mixed user interface (a), plus its corresponding spatial (b), and temporal (c) composition.



(a)



(b)



(c)

Figure 6.3: *(a) interface design proposal (b) spatial interface design (c) temporal interface design*

6.2.3 Spatio-temporal composition

The spatial composition of this example, following the guidelines given in Figure 6.3, are described as follows:

$SpatialComposition(\phi, A) = (R_6, R_6)$
$SpatialComposition(A, B) = (R_3, R_3)$
$SpatialComposition(B, OptionsMenu) = (R_{11}, R_{11})$

These spatial compositions are extracted from the inter-relationships existing between the CIOs (i.e. A, B and $OptionsMenu$). Note that these relations pertain to the multimodal (white) area shown in Figure 3.3.

$OptionsMenu:$
$SpatialComposition(C, D) = (R_{13}, R_{13})$
$SpatialComposition(D, E) = (R_{13}, R_1)$

These spatial compositions are extracted from the intra-relationships existing between the CIOs in the $OptionsMenu$ (i.e. C, D and E). Note that these relations pertain to the GUI (gray light and dark) area shown in Figure 3.3.

The temporal compositions for the example, taking into account the relationships outlined in Figure 3.1, are described as follow:

$TemporalComposition(\phi, A) = (R_7)$
$TemporalComposition(A, B) = (R_{12})$
$TemporalComposition(B, OptionsMenu) = (R_{12})$

These temporal compositions are extracted from the inter-relationships existing between the CIOs (i.e. A, B and OptionsMenu).

$OptionsMenu:$
$TemporalComposition(C, D) = (R_7)$
$TemporalComposition(D, E) = (R_7)$

These temporal compositions are extracted from the intra-relationships existing between the CIOs in the OptionsMenu (i.e. C, D and E).

It is important to stress that when each host composition (for instance ϕ,A) ends, all spatio-temporal relationships related to or started by it are also terminated (i.e., B and OptionsMenu).

6.2.4 Design space

In this section we illustrate how the design space can be used to associate design principles to continuous interaction evaluation.

By considering the spatio-temporal composition presented before (Figure 6.3) we can use the design space to identify levels of continuous interaction by associating the appropriate design principles for each continuity property i.e. perceptual, cognitive and functional.

To identify the perceptual continuity level we should fix the kind of augmentation axis to the perception item. The insertion context of the mixed interaction space should be according to the user's task focus. As the surgical task is focused on the patient, in this case the mixed interaction space may be inserted in the central zone of user task to augment perceptive continuity. Besides, the guidance provided by the system is forcing the user shares his interaction focus between worlds. Such design principles configuration give us the perceptual mapped zone conforming illustrated in Figure 6.4. If the concrete interaction device chosen to support the mixed interaction space is able to assure these two design principles, then the perceptual continuity is guaranteed too.

To identify the cognitive continuity level we should fix the kind of augmentation axis to the action item and then selected the media type in the media axis. In this case the cognitive mapped zone (Figure 6.4) is just showing a tendency that more natural media are more easy to interpret.

To identify the functional continuity level we should fix the kind of augmentation axis to the interaction item and then select the kind of transform and the kind of connection type the system is providing. In this case the tracking system should augment the user's interaction and the static connection type between the worlds (i.e. the registration procedure and tracking system were previously established) is allowing the user performs real actions with shared effects (Figure 6.4). Such functional design principles configuration is providing a certain level of functional continuity to the interaction. However other choices could augment the this continuity level such as supporting automatic connection type.

6.3 Case study 2 - Computer-guided maxillofacial surgery

This study case relates maxillofacial surgery. Patients presenting huge deformations in her/his mandible anatomy frequently should suffer an os-

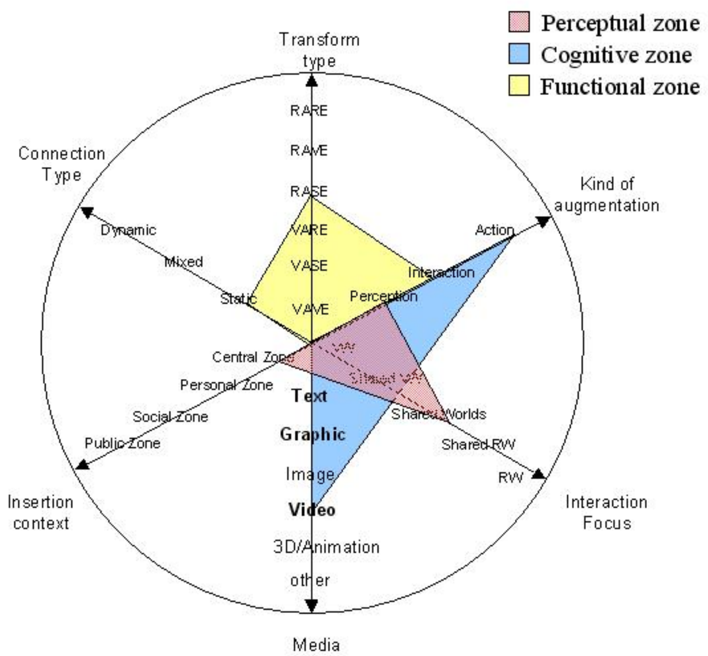


Figure 6.4: *Continuity zones for the Image-guided surgery case study.*

teotomy (operation to cut off part of the bones or to append some kind of prosthesis, according to her/his injury).

In this kind of intervention, the surgeon cuts the patient mandible following an imaginary line based on the pre-operative analysis. Nowadays, this class of surgery procedure takes about 3 hours and the patient recuperation about 2 months. Despite the fact it is a very traumatic procedure, post operatory problems can expose the patient to a new surgery and another 2 months recuperation period.

In collaboration with the Service de Stomatologie et Chirurgie Maxillo-faciale at Saint Luc Hospital, in Brussels, we proposed the development of an application using a mixed reality interface to guide surgeons in this type of surgeries. The goal of this application is to increase the first surgery success, avoiding a second intervention. To achieve this objective, we are mixing real and virtual images by projecting a virtual guidance path-line on the patient mandible live video images. Then, the surgeon should cut the patient mandible paying attention to follow this path-line and avoiding touching the dental nerve. The time involved in this procedure is not important, but the accuracy is mandatory.

The flowchart that represents the system designed by us to support maxillofacial surgery including pre-operative CT scanning and planning to surgical guidance is shown in Figure 6.5. The process begins with the *images acquisition* using a CT scanner. Then, a threshold is used to filter the images, separating the bones and the other soft tissues. After this processing, the images are *segmented* and a 3D model of the skull is *reconstructed* from the segmented images. Using the reconstructed 3D model and the medical doctors' experience, a path-line representing the osteotomy path is then *designed*. The next step involves the hardware *calibration* and the *registration* of real and virtual worlds.

The proposed application is in development and, at the present, we are performing usability tests with volunteers. For this reason, it is not yet based on a real mandible. To simulate the same situation of an operation room, we generated mock-ups of the mandible in gypsum (as shown in Figure 6.8), based on the 3D model reconstructed in the last step. Our real environment (the surgery scenario) involves the mock-up of the mandible, the surgeon hand and the tool used to cut the bones (as shown in Figure 6.5). The virtual environment is composed by the 3D representation of the mandible, the dental nerve, the tool marks and the path-line. When using augmented visualization, after the *registration*, the final images are mixed into a common reference frame and the *visualization* and interaction with the application becomes possible.

As mentioned before, our focus is on guidance for osteotomy procedures. Two main scenarios are explored in this study, considering virtual and augmented guidance.

The virtual guidance provides the 3D of the tracked and/or segmented objects visualization in the screen. The task consists in cutting the object reproducing as the best as possible the path showed in the screen. When the surgeon touches the good position (in the real object), the virtual representation of the surgical tracked tool becomes green. Distance indicator from the pointed position to the internal structure (i.e. in this case the internal structure is related to the segmented dental nerve) is also displayed in a progress bar in the screen (element D in Figure 6.6). It corresponds to one of the scenarios experienced during the usability tests described in the evaluation part (see Part III).

The augmented guidance scenario provides the visualization (in the screen) of the 3D elements superposed to live video images. In this scenario the path-line is projected over the video image. When the user touches the correct position (over the real object), a sphere representing the tracked tool becomes green (Figure 6.8). It corresponds to one of the scenarios experienced during the usability tests described in the evaluation part (see Part III).

6.3.1 Virtual scenario description

Let us assume the following virtual scenario. The display starts with the background presentation of a 3D space A located at point 0 relative to the application origin ϕ as represented in Figure 6.6. At the same time, a 3D object B (i.e. the virtual representation of mandible) overlaps space A according to the registration procedures. At the same time a graphic object D (i.e. the progress bar) overlaps the space A at the bottom of it. At time t1, a 3D object C (i.e. virtual represnetation of the tool) overlaps the space A according to the tracking procedure. At time t2, determined by tracking system procedures, the object C and D have their states changed to C' and D'. Figure 6.6 (a) graphically depicts the initial mixed interaction space plus its corresponding temporal composition (b).

The spatial composition of this example, following the guidelines given in Figure 3.3, are described as follows:

$$SpatialComposition(\phi, A) = (R_6, R_6)$$
$$SpatialComposition(A, B) = (R_8, R_8)$$

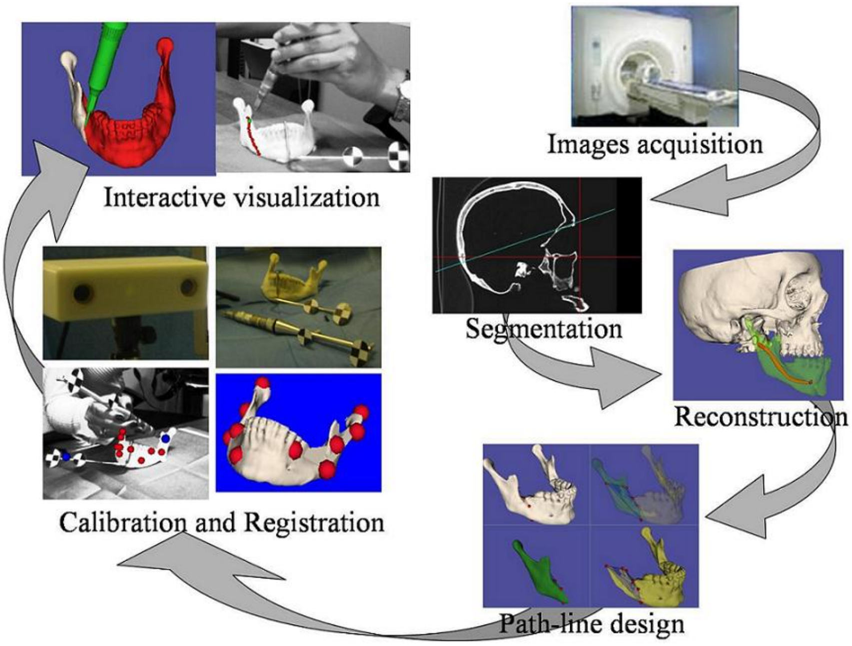
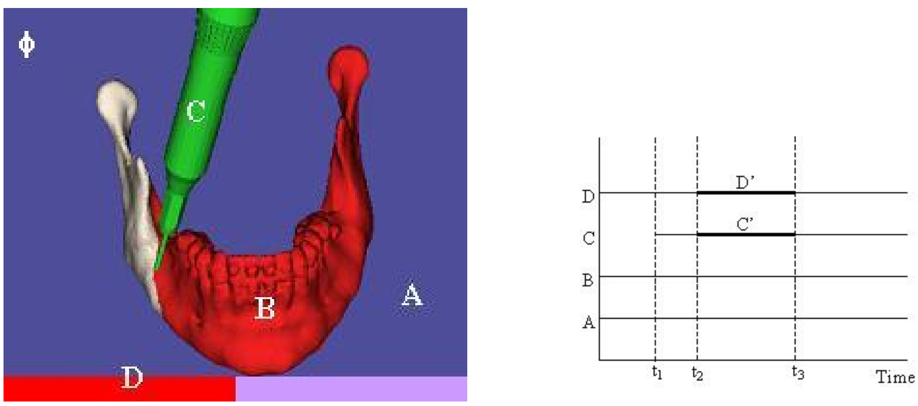


Figure 6.5: *System flowchart to provide surgical guidance during maxillofacial surgery.*



(a) (b)

Figure 6.6: *Virtual guidance scenario for the maxillofacial surgery. (a) Mixed Interaction Space proposal (b) temporal interface design*

$SpatialComposition(A, D) = (R_{11}, R_{11})$

These spatial compositions are extracted from the inter-relationships existing between the CIOs (i.e. A, B and D). The object C has no predetermined spatial composition once its position is calculated dynamically by the tracking system. Note that these relations pertain to the multimodal (white) area shown in Figure 3.3.

The temporal compositions for the example, taking into account the relationships outlined in Figure 3.1, are described as follow:

$TemporalComposition(\phi, A) = (R_7)$
$TemporalComposition(A, B) = (R_7)$
$TemporalComposition(A, C) = (R_{12})$
$TemporalComposition(A, D) = (R_7)$
$TemporalComposition(C, C') = (R_{12})$
$TemporalComposition(D, D') = (R_{12})$

Figure 6.7 shows a potential design space where we have graphic and 3D objects as media type. Graphic media in this case represents the progress bar. The 3D objects represent the mandible and the tracked tool. Such media elements are inserted in the personal zone of the user's task while the main user task focus is placed on the central zone. It is forcing the user to share his interaction focus in the virtual world (i.e. identifying a perceptive discontinuity). The media types should contribute to augment the action of the user (i.e. cognitive continuity) while the tracking tool system should augment the user's interaction (i.e. functional continuity). In this case the static connection type between the worlds (i.e. the registration procedure and tracking system were previously established) is allowing the user performs real actions with shared effects.

6.3.2 Augmented scenario description

The augmented visualization guidance provides the projected position of a 3D location on the left or right video image. This can be used, for example, to provide visual coordinate registration accuracy feedback, or to overlay useful spatial information, such as predicted or hidden locations of interest. In our case we have projected the pathline to perform the osteotomy. Problems related to visualization such as occlusions and depth perceptions were minimized by alerting the user with a visual feedback when the user is touching the correct location. A small sphere indicating

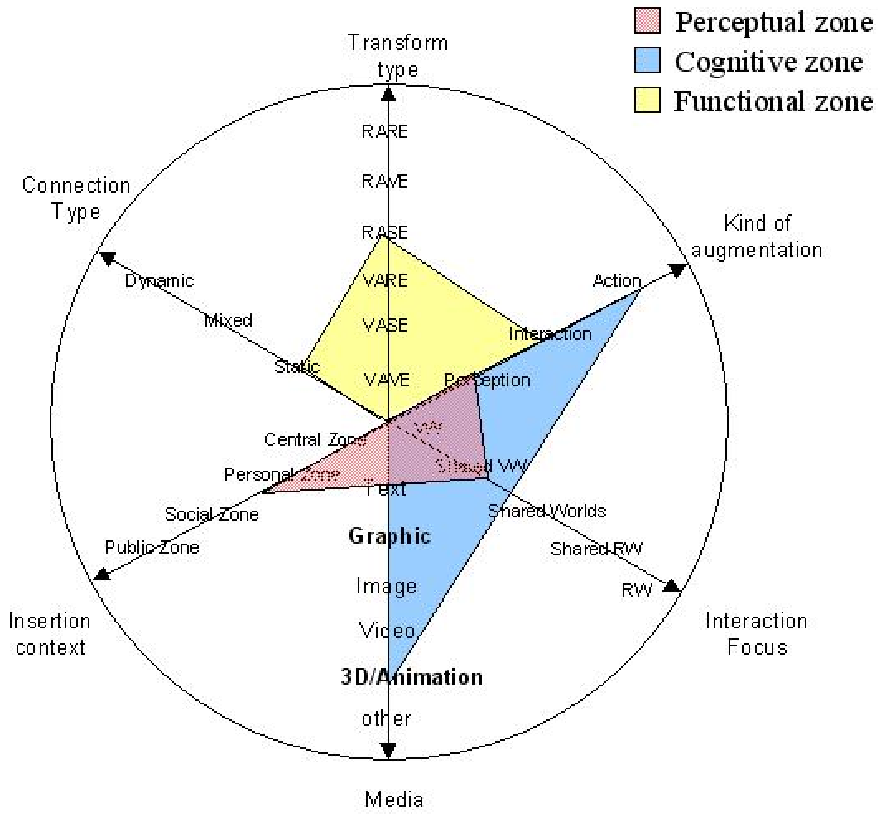
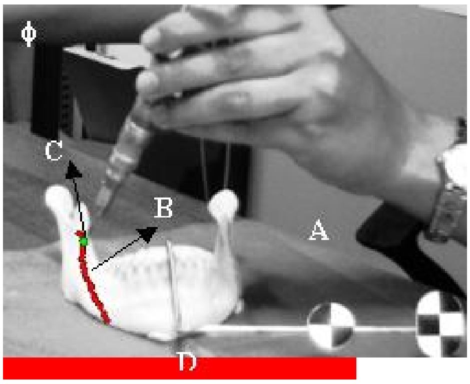


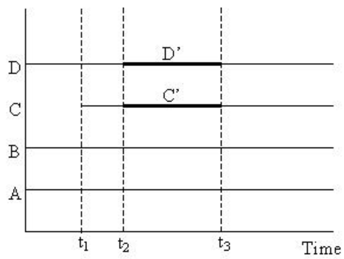
Figure 6.7: *Continuity zones for the virtual guidance scenario.*

the tool tip becomes green. Distance indicator for the pointed position to the internal structure is also displayed in the augmented view displayed in the screen. It corresponds to one of the scenarios experienced during the usability tests describe in the evaluation part (see Part III).

Let us assume the following augmented scenario. The display starts with the background presentation of a video A (located at point 0 relative to the application origin ϕ as represented in Figure 6.8). At the same time, the graphic object B overlaps the space A according to the registration procedures. At the same time the graphic object D overlaps the space A at the bottom of it. At time t1, the graphic object C overlaps the space A according to the tracking procedure. At time t2, determined by tracking system procedures, the object C and D have their states changed to C' and D'. Figure 6.8 graphically depicts the initial mixed interaction space plus its corresponding temporal composition.



(a)



(b)

Figure 6.8: *Augmented guidance scenario for the maxillofacial surgery. (a) Mixed Interaction Space proposal (b) temporal interface design*

The spatial composition of this example, following the guidelines given in Figure 3.3, are described as follows:

$$SpatialComposition(\phi, A) = (R_6, R_6)$$
$$SpatialComposition(A, B) = (R_8, R_8)$$
$$SpatialComposition(A, D) = (R_{11}, R_{11})$$

These spatial compositions are extracted from the inter-relationships existing between the CIOs (i.e. A, B and D). The object C has no predetermined spatial composition once its position is calculated dynamically by the tracking system. Note that these relations pertain to the multimodal (white) area shown in Figure 3.3.

The temporal compositions for the example, taking into account the relationships outlined in Figure 3.1, are described as follow:

$$TemporalComposition(\phi, A) = (R_7)$$
$$TemporalComposition(A, B) = (R_7)$$
$$TemporalComposition(A, C) = (R_{12})$$
$$TemporalComposition(A, D) = (R_7)$$
$$TemporalComposition(C, C') = (R_{12})$$
$$TemporalComposition(D, D') = (R_{12})$$

As mentioned before when each host composition (for instance ϕ,A) ends, all spatio-temporal relationships related to or started by it are also concluded (i.e. B, C and D).

Figure 6.9 shows a potential design space where we have graphic and video as media type. Graphic media in this case represents the progress bar, the video images, the projected path line and the projected sphere representing the tool tip position. Such elements are forcing the user to share his interaction focus between both worlds in the same time that all media elements are inserted into the central zone of user's task. It is contributing to augment the perception of the user once the main user task is also placed on the central zone. The media types should contribute to augment the action of the user (i.e. cognitive continuity) while the tracking tool system should augment the user's interaction (i.e. functional continuity). In this case the static connection type between the worlds (i.e. the registration procedure and tracking system were previously established) is allowing the user performs real actions with shared effects.

Comparing the continuity zones given by each scenario (i.e. virtual and augmented) we note that the cognitive mapped zone given by the virtual scenario (Figure 6.7) is bigger than that given by the augmented scenario (Figure 6.9). It is showing just a tendency that is more easy to interpret 3D objects than projected objects into a video image. The perceptive mapped zone for the virtual scenario (Figure 6.7) is showing a tendency to be discontinuous once the main user task focus is located on the real world and in the central zone, and all perceptive design principles (i.e. insertion context and interaction focus) are not following this direction in the design space. On the other hand the perceptive mapped zone for the

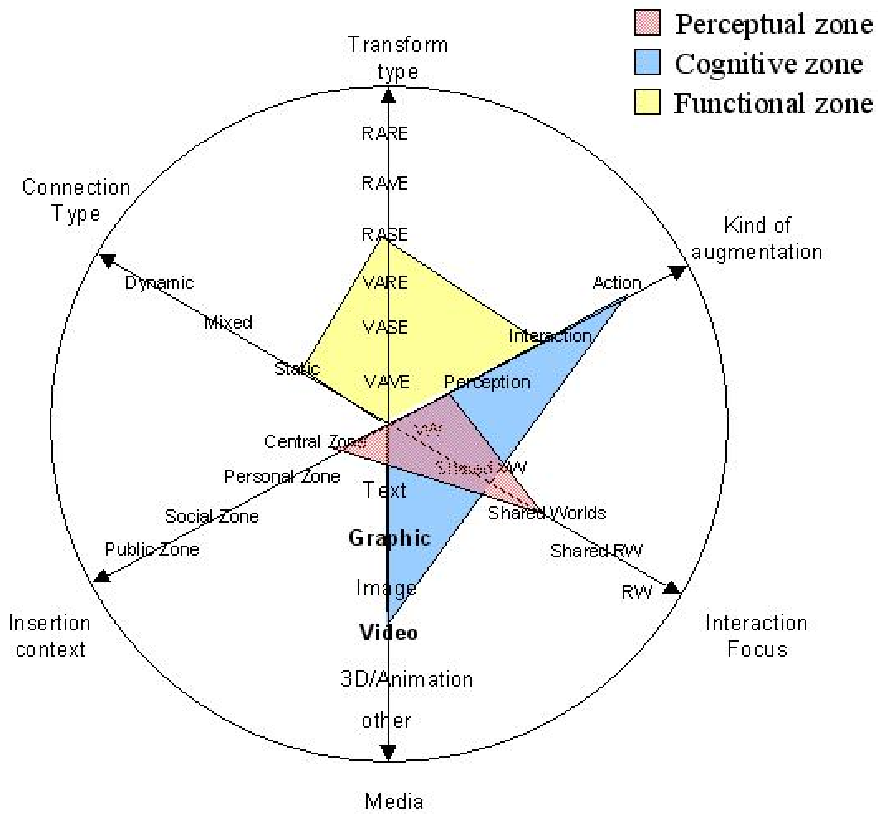


Figure 6.9: *Continuity zones for the augmented guidance scenario.*

augmented scenario (Figure 6.9) is showing a tendency to be continuous once the perceptive design principles are allowing insertion context in the central zone of user task as well as interaction focus on the real world or shared between worlds. The functional mapped zone is the same for both cases since the interaction technique has not changed.

6.4 Discussion

In this chapter we have described two study cases in image-guided surgery to illustrate how the design principles can be used to map zones for continuous interaction in the proposed design space (Chapter 5).

As we have seen with the practical examples, it is important to report we can have the same spatial and temporal configuration for different design spaces configurations and vice-versa. It strengthen the idea we need to take into account the set of design properties to evaluate the mixed interaction and not just isolated aspects.

Finally we should be aware that specific design aspects such as spatial and temporal integration of different media objects have implications for the human perception. However, the information that people assimilate from a modality of interaction (e.g., visual modality) also depends on their internal motivation, what they want to find and how well they know the domain. Such issue is verified using the variable *kind of mock-up* in the usability tests described in Chapter 11.

As contributions of this work we have highlighted:

- Management of a large number of design principles related to the continuous interaction for the mixed interaction design;
- Providing a better understanding of how to design MR systems in an intuitive and effective manner;
- Identifying spatial, temporal and focused layouts of the mixed interaction space under development for verification purposes (see Chapter 11);
- To help designers to envision future interactive mixed systems in terms of continuous interaction.

Part II

Implementation

Chapter 7

Related Works

This chapter describes a number of different contributions that have been made in the development of software toolkits and libraries to support MR applications.

7.1 Existenting software toolkits and systems

The first type of software toolkits for VR applications were hardware abstraction layers, providing generic interfaces to the non-standardised hardware available. Some recently developed software systems are DIVERSE[1], OpenTracker[2], and VrJuggler[3]. Many of these systems are freely available and provide similar types of interfaces to trackers and renderers, indicating that these low level toolkits have converged toward some form of best practice.

To support the definition of more complex software, there is a need to break down an application into objects or modules that communicate through some abstraction mechanism. In this direction one of the most integrated approaches to virtual environments is COTERIE [72], which was developed to help implement applications for distributed virtual environments. It contains language level support for distributing objects over a network, and integrates this with packages that support an interpreted language, threaded processing, tracker abstractions, and 3D animated rendering. The main focus of COTERIE was implementing distributed systems, and so the shared memory aspect is the main core of the system,

[1] http://diverse.sourceforge.net/
[2] http://www.studierstube.org/opentracker/
[3] http://www.vrjuggler.org/

with other features built around this. Applications are built up using multiple threads that communicate via replicated shared objects, with each update blocking the thread and passing through a sequencer process that handles synchronisation. This sequencer inhibits the system from scaling up as well as a nonsynchronised system. An integrated 3D library and tracker abstraction layer allow the user to implement compiled or interpreted programs to work in AR or VR environments.

The Studierstube system [111] is also used to develop distributed applications, but in contrast to Coterie it embeds the application into a distributed scene graph. Studierstube is based on a distributed version of Open Inventor, and to be distributed all components must be expressed using such scene graph interface.

TINMITH [91] is based on processing objects that are connected together to flow data through the system, processing inputs to produce outputs. The sensor model in use is similar to the traditional VR model, where sensors are processed using application code and run time configurations to render data to a head mounted display. Internally, the sensors are abstracted as objects and when new data arrives, these objects notify other listeners that updates are available. These objects are connected together into a directed graph and the final calculated values are used to render to the HMD. The objects are stored in an object repository based on a file system model, and a consistent methodology is used for the design of each set of objects.

There are a number of other systems for developing 3D applications. Although there has been a wide range of research in software engineering for AR and VR applications, the state of the art is still quite primitive compared to the maturity we have in 2D desktop applications.

7.2 Registration approaches for mixed reality systems

The basic structure to support augmented reality applications is based on the perception of the real world, the representation of the virtual world and the registration procedure. The registration procedure enables to know where the virtual object should be positioned in the real world. Basically, the main challenges for developing robust AR systems focus on methods allowing to perceive the real world and register it according to the virtual world. The real world can be acquired by using different methods such as laser scanning, infrared scanning, computer vision with and without fiducial markers, GPS methods for mobile systems, etc. The appropriate method for a specific application will depend on the task, desired resolution, performance and financial resources. We can say that both options, computer vision with and without fiducial markers can be a good compromise between resolution, performance and costs.

The computer vision methods allow acquiring a view of the real world from one or two cameras worn by the user or fixed in the environment. The first class of methods relies on markers previously placed in the environment and the registration is based on them. There is a very well known library called ARToolKit to perform computer marked vision [51]. This software library can be used to calculate camera position and orientation relative to physical markers in real time with a precision of $2mm$.

More sophisticated technology using optical markers have been proposed for medical applications. For instance, Sauer [107] describes a method using a stereo pair of color cameras to capture the live video and a third monochrome camera to capture a ring of infrared LED markers. It is used for head tracking and works in conjunction with retroreflective optical markers which are placed around the surgical workspace. A similar approach is proposed in [115] where the markers used for tracking the patient will be mounted on one of the self-retaining retractors, because the patients position relative to this surgical tool stays fixed during the whole intervention. Because markers are situated in discrete places and solid deformations cannot be perceived correctly, this approach is suitable if restrictions due to marker (such as full time visibility, precision of placement, etc.) are not an issue.

Another class of methods performs the registration right on the virtual world and is sensibly more complex than the first one. This method captures a 2D image of the real environment and then, by detecting features, deduces the 3D position of 3D models. However, its precision is limited by the fact it is only based on a 2D image and false features can be detected and may disturb the registration process. Recent works in this direction [70] have proposed improvements to eliminate drift and jitter. It is done by merging the information from preceding frames in a traditional recursive tracking fashion with that of a very limited number of key-frames created off-line. This solution works in real time, is robust and manages partial occlusions.

The second approach takes into account 3D perceptions of the environment. The accuracy of those methods depends on the technology used to acquire the 3D scene. For instance, with the laser scanning method [47] and [78], the precision is so high that the discrepancy between real model and points obtained from the real scene is less than $2mm$. On the other hand this method has two main drawbacks. The high price of a laser scan camera and the time needed between each image acquisition, which is actually about 1 second per image.

Recent works, using stereovision algorithms, have explored the use of features to track objects in the world. For instance, in [81], the combined use of a 3D model with stereoscopic analysis allows accurate pose estimation in the presence of partial occlusions by non rigid objects like the hands of the user.

7.3 Tools for prototyping AR systems

Rapid prototype versus content expert support in MR system development is another point to be taken into account during the design phase. The rapid prototype aims applications without detailed knowledge about the underlying base technologies. This requires a user-centered approach for authoring in contrast to existing approaches that are primarily driven by technological features and problems of MR systems. In order to support content experts in MR design, the framework may use a component oriented base technology at different levels of granularity for building mixed realities. Researches in this direction can be found in projects such as DART [73] and AMIRE [88].

The emphasis of Designer's Augmented Reality Toolkit or DART has been to support the existing work practices of conceptual thinkers (media designers, HCI practitioners, experience planners, etc.) in exploring the novel domain of augmented reality. The DART provides AR services such as video capture, tracking (via VRPN[4]), and marker based registration (currently, via the ARToolkit [51]) to Director[5] applications. The user can place behaviors onto the Director score using extensions to the Director environment written in LINGO and Xtras (plug-ins for Director written in C++).

AMIRE is a project about the efficient creation and modification of augmented reality (AR) and mixed reality (MR) applications. The AMIRE system allows the Authoring of Mixed Reality (MR) applications without programming using a visual creation tool similar to other tools used for the creation of graphical contents. It is also done by using the marker based registration provided by ARToolKit [51] to augment the real scene.

[4]Virtual Reality Peripheral Network (http://www.cs.unc.edu/Research/vrpn/)

[5]Macromedia Director(http://www.macromedia.com/software/director/)

7.4 Medical Studio

Medial Studio[6] is a software platform which provides a rich set of tools for visualization and manipulation of 2D and 3D medical data sets by using VTK and ITK libraries. It is a cross and component-based platform. More details are given as follows.

7.4.1 Framework implementation

In order to assure cross-compatibility in terms of execution and development, Medical Studio is written in C++. The language is commonly used by the signal and image processing community. Publicly available libraries are used to provide well known functionalities and avoid re-implementation of already validated methods. These libraries have been chosen in function of their specifications, their development language, their cross-compilation possibilities and the community working with them.

VTK [113] is used for all visualization tasks. Gtk [11] is used for all graphical user interfaces. ITK [112] is used for signal and image processing. Dcmtk [1] is used for Dicom 3.0 standard compliance. This list of libraries is not fixed. Thanks to the modular architecture any other library can be linked into a new component providing high extensibility to Medical Studio.

We define a plug-in as a group of components compiled together into one shared library. These plug-ins can group components by any criteria but by convention it is preferable to group them either by type or procedure context. An xml file is created for each plug-in describing its content. With this, the kernel is able to create a repository of all available components without loading them into memory, plug-ins will be loaded lately, only when required letting resources available for data or processing. This facilitates the distribution process as only two files will be needed to add this procedure to the basic Medical Studio platform.

This allows easy customization of final applications. For example all the components for a specific neurosurgery procedure will be grouped into one plug-in. For applications that do not need the neurosurgery plug-in (e.g. maxillo-facial surgery), removing the two files is enough to eliminate the unwanted functionalities.

[6]Medical platform being developed in the Communications and Remote Sensing Laboratory at Universite catholique de Louvain. Website: http://www.medicalstudio.org

The two studies cases discussed in Chapters 8 and 9 were completely integrated with Medical Studio.

7.4.2 Building the application content

The 3D reconstruction, as well as the image processing and registration procedure, are performed using Medical Studio Tools. Medical Studio offers a friendly user interface for 3D reconstruction from segmented volume data sets and easy visualization and manipulation of 2D and 3D data sets like zoom, translations, rotations, placing points, plans. Figure 7.1 is showing how it is possible to change parameters (for instance colors and transparency level) of the 3D volume rendered in the augmented scene. Also, the choice of what objects to visualize or not can be done in run time.

The great advantage to perform this integration is that all these functionalities can be also applied to the augmented scene providing to the surgeon a better understanding of complex and critical surgical procedures. For instance, for the maxillo-facial surgery the bones must be segmented from the CT dataset. To do this, Medical Studio uses the Marching Cubes algorithm [71], which will construct a 3D surface from the CT images given the intensity of the bones. It will act in the same way as thresholding segmentation. The same process can be done with MRI images. In this case we provide the intensity of the skin and the intensity of all brain structures that we want to segment and reconstruct.

Figure 7.2 illustrates the reconstructed 3D model from MRI images highlighting some internal structures of the brain (i.e. the segmented brain structures).

7.5 Conclusions

MR system are based on tracking devices, rendering engine and interaction metaphors/devices. So far some technical issues relate to these technologies should be considered, such as:

- Establishing a perceptible link between physical objects and their representation in the digital world and vice versa. This issue is specific to MR systems.
- Depth perception in the human visual system: stereopsis (disparity of the eyes), accommodation (focusing), pictorial cues (we know

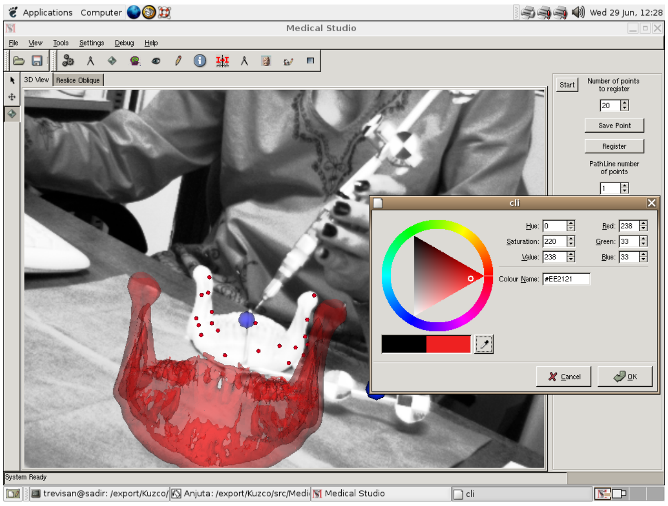


Figure 7.1: *Editing content and visualization parameters in runtime with Medical Studio.*

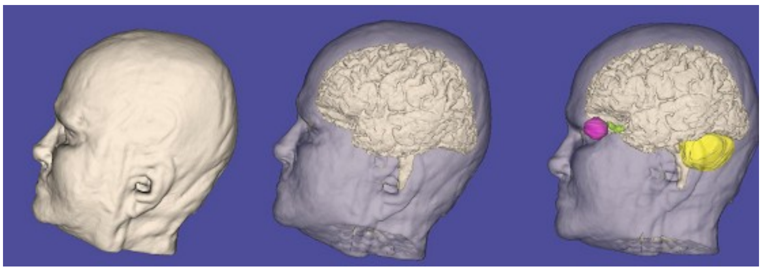


Figure 7.2: *Reconstructed 3D model from MRI data sets with segmented structures.*

how big the things are), interposition (closer occludes farther). Such issues are common to 3D, VR and MR systems.

- Technical barriers: calibration, registration errors, stereoscopic spacing, fixed focus and field of view. This issue is more critical in MR systems once the calibration and registration procedures (specific to MR systems) can have great influence in the final performance of the system.

Based on the features available in many of the frameworks to develop MR systems, discussed in this chapter, we suggest the following layered Integration Space (Table 7.1) to characterize the MR development phase.

4. Content
3. API's components
2. Services to support MR
1. Base Platform

Table 7.1: *Layers of the integration space for MR systems*

1. Base Platform
   (a) Hardware aspects
      i. Computer
      ii. Graphic processor and video hardware
      iii. Other related devices (HMD, motion captors, cameras, etc.)
   (b) Software aspects
      i. Operational system
      ii. System level libraries
2. Services to support MR
   (a) Tracking and Registration
      i. Passive (Based on markers, optical tracking, stereovision, markerless)
      ii. Active (radar, laser, magnetic, acoustic, inertial, eye tracker, structured light system, other)
   (b) Calibration tasks
      i. Video calibration

ii. Sensors calibration
iii. Workspace calibration
iv. Registration calibration (static and dynamic errors)

(c) Video capture

(d) Motion capture

3. API's components

(a) 3D modelling

(b) Simulation level: virtual-human animation, skin deformation, physical clothes, face animation, speech, behaviours, 3D surround sound, 3D stereoscopic, etc.

(c) Graphics render engines

(d) Object tracking

(e) Selection and navigation

(f) GUI (Graphical User Inteface)

(g) TUI (Tangible User Interface)

(h) MUI (Multimodal User Interface)

The content layer is data used in the application program. The other three layers (base platform, services support, calibration tasks, APIs components) are independent of content.

It is not the goal of our research to produce tools for rapid prototyping of MR systems. We are mostly focused on the content expert support for designing continuous interaction in MR systems. For this reason, in this Thesis, we have explored two methods to provide augmented reality scene. The first one is a markerless low-cost approach based on stereovision analysis in conjunction with spatial and color filters to track the object of interest in the real world. The second one is a marker-based approach using the MicronTracker library[7]. Both methods use rigid registration to align virtual and real objects and were integrated into the Medical Studio.

Medical Studio is our platform of integration which supports the layers 2, 3 and 4 of the proposed MR integration space (Table 7.1).

[7]MicronTracker Website http://www.clarontech.com/

Chapter 8

Markerless Approach

In this chapter we describe a markerless approach developed in collaboration with Pierre Del Ville and Michel Cornet d'Elzius in the frame of their undergraduate work. A joint publication can be found here [126]. It is based on stereo vision analysis and spatial and color filtering to select the object of interest on the real scene. The innovation of this method relies on the fact that we are not using markers [51] neither features tracking [70] [81] as has been proposed in previous works. In this sense, such approach can be exploited for a range of mixed reality applications (for instance craniofacial surgery and technical maintenance fields such as oily and greasy engine repair) were the placement of markers or previous detection of features are not possible.

The following section gives an overview of the developed system concerning both the hardware setup used and the software information flowchart.

8.1 Hardware setup

The proposed hardware setup consists of the following components:

- stereoscopic goggles;
- stereo video camera (Bumblebee camera from Point Grey Research Inc.).

The augmented scene can be displayed in an AR video-based goggles (see Figure 8.1. Those goggles are opaque, but as we put the images captured by the stereo cameras right in front of them, they become virtually

Figure 8.1: *Stereo cameras coupled to a 3D stereo glasses and a gypsum phantom representing the patient.*

transparent. In the optimal case the distance between the camera lenses must correspond to the distance between the screens of the AR goggles and the distance between the operator's eyes.

8.2 Software information flowchart

The process of image acquisition and stereovision computation are performed with the PGR libraries (Digiclops and Triclops) from Point Grey Research Inc. As stereovision has already been extensively covered in many papers [109][110] and our technique is centered on Augmented Reality and not on 3D recovering we will not discuss stereovision further. The 3D reconstruction, as well as the image processing and registration procedure, are performed with Visualization ToolKit (VTK) algorithms [113].

The algorithm flowchart shown in Figure 8.2 describes the supported pipeline we designed to perform our AR system. The pipeline described in Figure 8.3 shows the results of each related step described in Figure 8.2.

The stereo pictures (jpg file shown in Figure 8.2) provided by the stereo camera are processed through stereovision algorithms in order to obtain a cloud of points of the real scene (pts file in Figure 8.2). The initial images are kept, to be put back in the final augmented scene. As a result of this

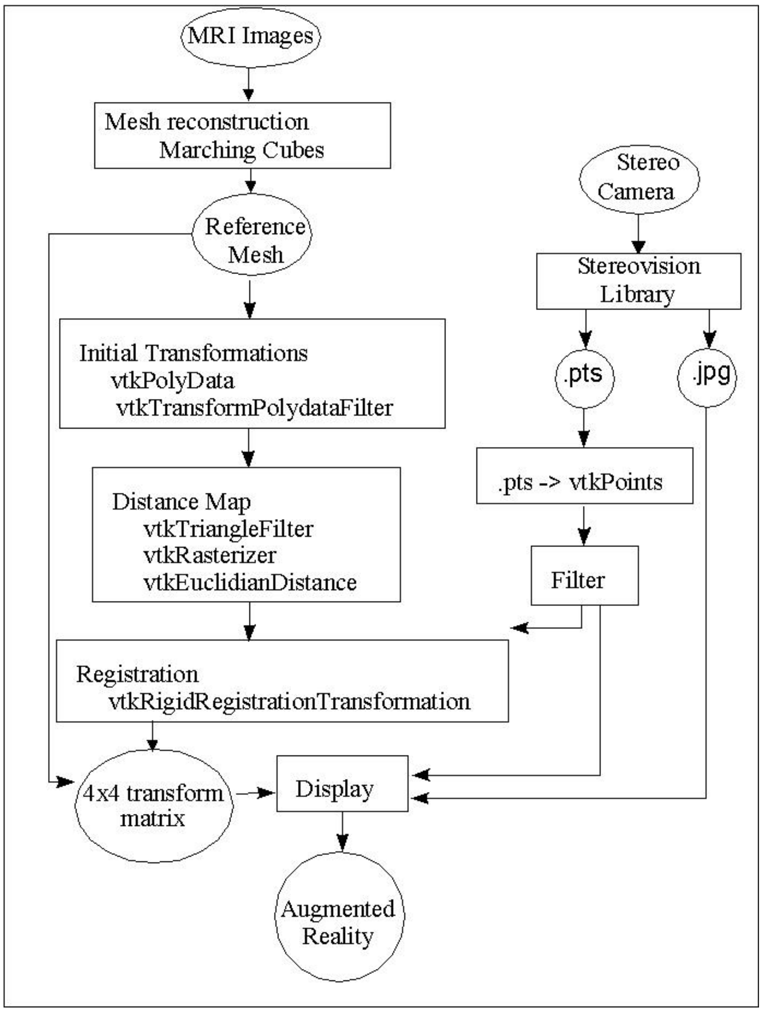


Figure 8.2: *Flowchart algorithm of the proposed method*

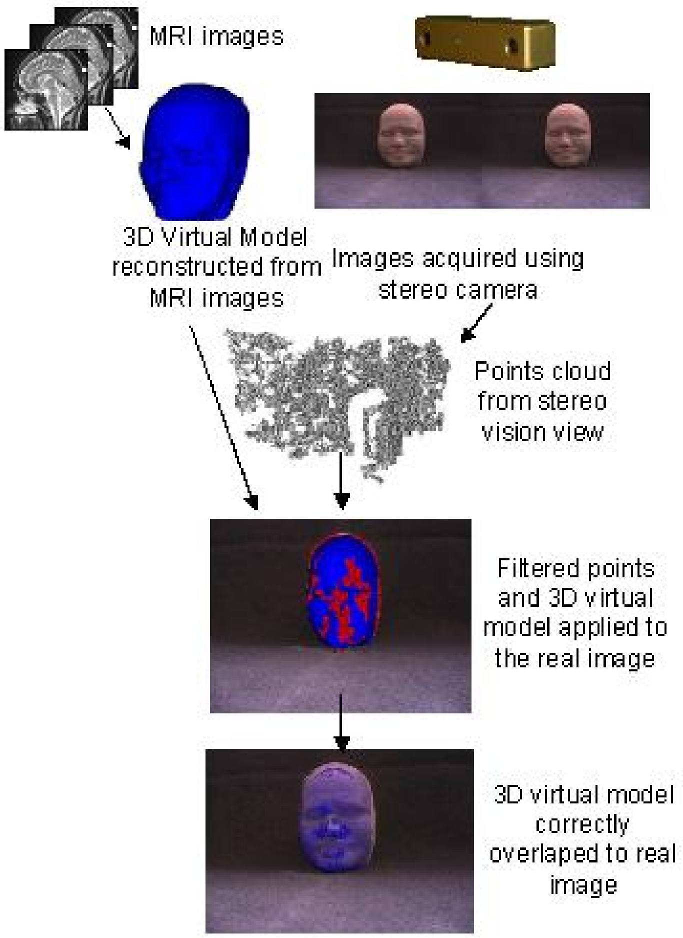


Figure 8.3: *Results from steps shown in Figure 8.2.*

first step, we have obtained a cloud of points that may be transformed into a VTK data format. Once the transformation is done, these points are filtered using color and spatial filters to only select the region of interest that will be used as one of the inputs to the registration algorithm.

The other input data for the registration algorithm is the reference 3D mesh (see Figure 8.2). Basic transformations (rotation and translation) are applied to this mesh to put it in the initial real model orientation specified in the initialization procedure. The orientation difference between real and virtual model cannot exceed a 30° difference. It guarantees the registration algorithm will not fall in a local optimum [20]. After that initialization, the algorithm computes the matching from the last match and no more initializations are required.

The next operation is the conversion of the mesh into a distance map. This is done by rasterizing the model surface and by applying the Euclidean distance transform on it.

Starting from a distance map and points, the registration algorithm minimizes the mean square distance between the points and the virtual model surface. As a result we obtained a 4x4 matrix representing the rigid transformation matching the cloud of points on the distance map.

As we want the virtual model be applied on the reality and not the opposite, we apply the inverse matrix to the reference mesh. At this point we have a matching between the real object and its virtual model. To perform augmented reality we just have to put the virtual model on the initial images coming out from the stereo camera. This is done using camera parameters and stereo processing information such as focal length and distances of the lenses, and distance of the real object from the center of the cameras.

8.3 Standard scene choice

For the set up of our technique, as well as for the determination of results, we chose a standard scene to perform our tests. As the basic idea beneath our project is to design a craniofacial surgery tool, we chose a situation close to a surgical room. For so doing, we built a "phantom" on gypsum of the patient's head. It was made possible by printing a 3D model reconstructed from MRI data sets (Figure 8.2). The phantom head was put on a dark gray-blue sheet.

8.4 Processing imaging filters

A major characteristic (and drawback) of stereovision is its intrinsic lack of accuracy. Because of occlusions and imperfections of stereovision algorithms, the outputs (3D image, depth map, cloud of points) are usually ″noisy″.

Moreover, the cloud of points we obtain after stereovision processing describes the whole scene captured and is consequently not usable for registration with a virtual model corresponding to a small part of it.

Selecting the points belonging to the object of interest is a tricky work. Two main assumptions regarding the real scene are considered:

1. Large parts of the background have a uniform color.
2. The place of the object of interest does not change abruptly from one frame to the other.

To suppress points not belonging to the object (referred as ′noise′) we developed a filter based on color and space assumptions. Those assumptions are common ones, easily deduced from a preliminary observation. For instance, during a surgery the patient′s body is ″wrapped″ in a green (or blue) sheet, we can thus delete points having high green color content.

The criteria (color and distance) have to be hand-determined in each of the scenes where we want to perform AR. Once we have assumed we are in an indoor space where light conditions are constant in time and after an initial initialization, the rest of the method is fully automatic.

8.5 Color filtering

Each point of the cloud computed by stereovision algorithms is composed of 3 coordinates and 3 RGB color intensities. The 3 coordinates correspond to the point position (in meters) in space regarding to the center of the stereo camera and along its three axis.

A ″′spectral″′ analysis highlights the differences in color distribution. Figure 8.4 shows three graphics, each of them corresponds to, from left to right, the Red, Green and Blue color components. The horizontal axis represents the intensity (from 0 to 255) and the vertical axis the number of points with that intensity.

Peaks corresponding to uniform intensities of light can be detected. Those peaks correspond to our first assumption, that large parts of the

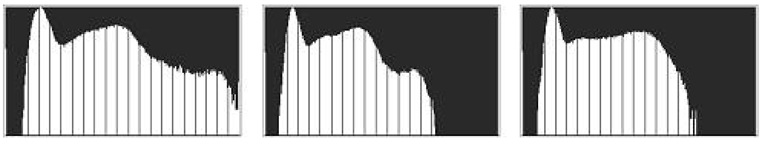


Figure 8.4: *RGB Spectral Analysis of StereoView. Left to right: Red, Green and Blue components.*

background have a uniform color. It is thus possible to isolate pixels with regard to their intensity as well as their dominant color. For instance, the dark blue sheet of the input Stereo images (see Figure 8.2) is responsible for the first peak of all the three drawings of Figure 8.4.

The "'spectral"' analysis permits to determine the criteria of a band pass color filter. Band pass filters are composed of high pass and low pass filters combined.

The high pass filter criteria are determined in order to filter dark backgrounds of blue dominance (the sheet in our example is gray-blue). As every color has RGB elements, we set a threshold in each RGB component to augment filter reliability.

Figure 8.5 illustrates the high pass filter red component. Comparison with the two other colors (not shown here) shows the head has a dominant red color.

The low pass filter criteria are set to filter light blue sheet standing in front of the cranium. Figure 8.6 illustrates the low pass filter blue component. A comparison with the two other elements (not shown here) shows clearly the sheet blue dominance.

Figure 8.5 and Figure 8.6 illustrate, on one of the two stereo pictures, the concept of color filtering.

8.6 Spatial filtering

Spatial filtering is built on observations in a surgical room. The Communications and Remote Sensing Laboratory works in close relations with the Oral and Maxillofacial Surgery Department at Saint-Luc Academic Hospital in Brussels, Belgium. Two authors observed craniofacial surgeries in the aim of determining the surgeon space of interest when performing his surgical act. In all the cases the surgeon stands between 10 and 80 cm of the patient's head which is at the center of his sight during surgical acts. For this reason, the spatial filtering suppresses points with a z coordinate

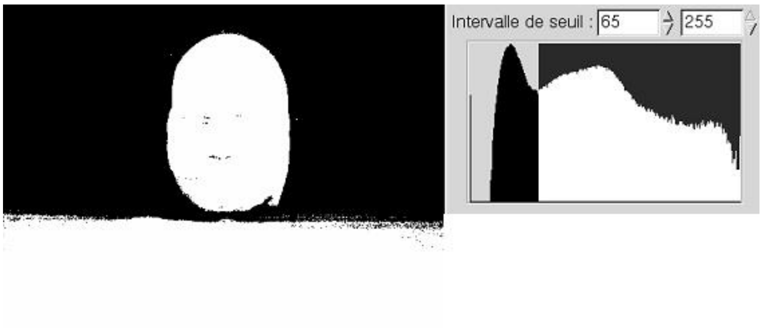


Figure 8.5: *High Pass Filter where a white point stands for the accepted points and a black one for the rejected points.*

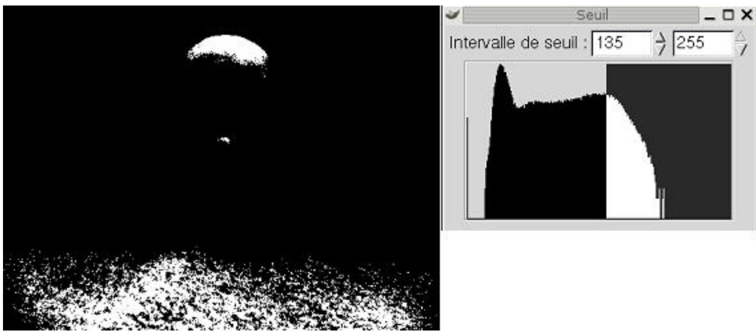


Figure 8.6: *Low Pass Filter where a white point stands for the rejected points and a black one for the accepted points.*

outside the domain $[0.1, 0.8]$.

8.7 Registration procedure

The registration algorithm minimizes the mean square distance between the points from stereovision and the surface from MRI segmentation.

For a given set of digitized points M and an object surface S, the best transformation T is

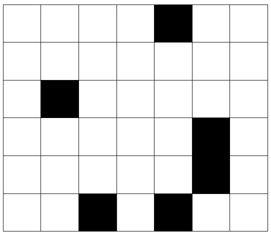
$$T = argmin \left(\sum_{p \in M} d_S \left(T(p) \right) \right) \tag{8.1}$$

$$d_S(q) = min \left(d^2(q, s) \mid s \in S \right) \tag{8.2}$$

The surface is rasterized into a 3D image. Equation 8.2 can be efficiently pre-computed for all points q using an Euclidean distance transform (EDT). For this purpose, the segmented scalp surface is rasterized into a 3D image. Then, the implementation of Saito's EDT [106] found in [19] is applied. It has a $O(N^{4/3})$ complexity, N being the number of points. When q does not fall exactly on the image grid, the value of $d_S(q)$ is interpolated.

Equation 8.1 is then easily solved, as the sum over M only involves a few hundred terms and $d_S(q)$ is available in the form of a lookup table (cfr Figure 8.7 and Figure 8.8). Given the low computational cost of the matching criterion, any minimization technique would be appropriate. We use an iterative scheme similar to Powell's [95] to compute the minimum. Convergence properties are particularly good thanks to the use of an exact Euclidean distance map, as opposed to the often used chamfer metric [103].

Of course, this only converges to a local optimum, so one needs an appropriate initial transform. This can be obtained either by full search at a coarse resolution of translations and rotations, or by using a standard setup during data acquisition. In our case, the first pictures of the real scene are known in advance.



5	4	4	1	0	1	4
2	1	2	2	1	2	5
1	0	1	4	2	1	2
2	1	2	4	1	0	1
5	2	1	2	1	0	1
4	1	0	1	0	1	2

Figure 8.7: *Left figure is a 2D model (elements of the surface are in black); right figure illustrates the distance map of this model.*

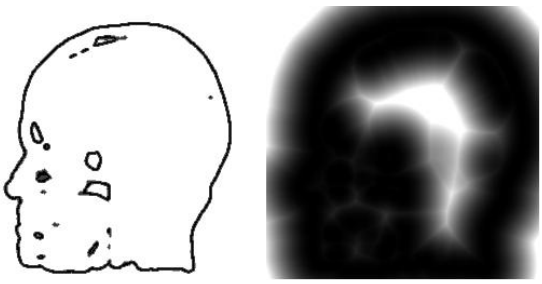


Figure 8.8: *Left figure shows the cranium slice; right figure shows the corresponding distance map where intensity represents distance. Short distances are darker.*

8.8 Augmented view

T is the transformation symbolized by the $4x4$ transformation matrix

$$T() = \begin{pmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where the $3x3$ sub matrix R symbolizes rotations and the vector T symbolizes the translation.

This transformation is applied to the cloud of points to put it on the model surface. As we want to place virtuality on reality, we invert $T()$ and apply it to the model surface. The background picture is one of the 2D initial images from the stereo camera. To correctly visualize the 2D image and the model surface (Figure 8.9), in the 3D scene, we used information such as:

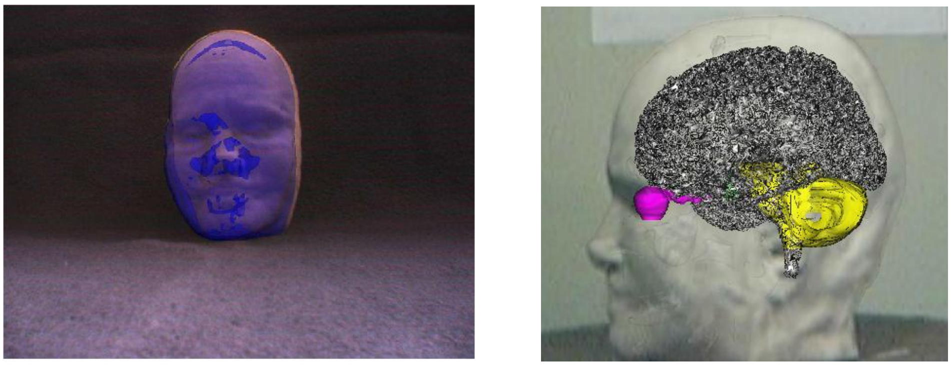
1. Distance of the object from the cameras: this information is given by the coordinates of the filtered cloud of points. Each point coordinate corresponds to the position in meters from the center of the cameras
2. Focal length of the camera lenses

The augmented view should be improved by eliminating some noise (which come from the segmentation process) in the reference 3D model.

8.9 System validation

Series of tests have been performed to determine the spatial precision in translation and rotation as well as the computing time of the method. We do not assess the precision of our whole method. The precision of display is hard to determine, unless we have several operators with AR goggles and real-time temporal resolution.

The precision of the registration algorithm has already been determined in [96] and [20] where the points were precisely digitized from the real model with two receivers. One receiver was fixed on the patient's forehead, the other was a hand-held stylus providing a direct method for points acquisition. In our case, the points are digitized with stereovision algorithms lacking the precision of a direct acquisition. Consequently, the points are noisy and for that we cannot compute the algorithm precision



(a) (b)

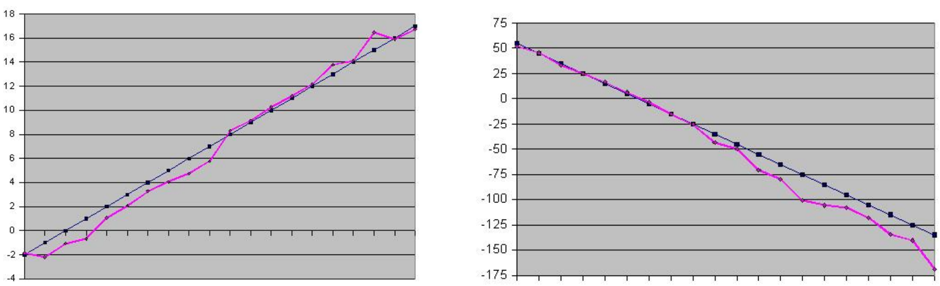
Figure 8.9: *Final augmented real scene: (a)Skin segmented visualization; (b) Brain segmented structures visualization.*

from the mean square distance of the points and the model surface. Then, we have measured spatial precision of the technique as follows.

8.9.1 Spatial precision

We determine the mean square error between the real transformations applied to the real object (i.e., rotations and translations) and the transformations applied to the model surface to match them.

The camera was statically fixed on a tripod at 70 cm of the object and under a $45\,^\circ$ horizontal angle. A ruler and a protractor were disposed under the object to measure the real movements. Nevertheless, we assumed a human hand error of 1 mm in translation and $1\,^\circ$ in rotation.



(a) (b)

Figure 8.10: *Cube rotations (a) and translations (b).*

For the spatial precision tests we considered:

1. A cube in rotation (20 rotations of 1 degree from $-2\,°$ to $+17\,°$). It is depicted in Figure 8.10(a).

2. A cube in translation (20 translations of 1 cm from -5 cm to +15 cm). It is depicted in Figure 8.10(b).

The mean square error found for rotations is $0.7\,°$, and the mean square error found for translations is 1 cm.

Two remarks are important here:

1. Concerning the translations, if the object of attention is at the centre of the sight, the precision is 1 mm, at the edge of the sight the registration is less accurate (clearly visible on Figure 8.10). As an operator using AR technology does not need accuracy unless the object of his attention is at the center of his sight, we can assume our precision in translation is 1 mm.

2. The precision determined corresponds to the precision of the human hand displacing the object. Therefore, more precise method to measure the objects moving in the real world should be taken into account for the next validation tests.

8.9.2 Computing time

Different parameters of the registration algorithm were tested in order to determine which is the most promising to improve the registration algorithm computing time. We found the most promising one to be the number of points to register.

The following Table 8.1 shows influence of subsampling a cloud of 8000 points on accuracy. As computing time linearly depends on the number of points, a subsampling of 400 corresponding to 20 points and a RMS error of 1.99 seems to be the best relation between time and accuracy. Such configuration takes approximately $10s$.

8.10 Conclusions and future works

In this chapter we have described an alternative method to perform registration in markerless Augmented Reality systems without features tracking. The method uses off-the-shelf stereo cameras, virtual reality goggles and MRI data sets to build up the 3D model considering the medical case.

Subsampling	Mean square error	Number of points
1	0.7	8000
2	0.67	4000
5	0.7	1600
10	0.69	800
20	0.84	400
50	0.79	160
100	0.69	80
400	1.99	20
1000	8.23	8

Table 8.1: *Number of points and accuracy.*

The main contribution of this work is the merging of the following aspects:

1. Real scene acquisition is performed using the stereovision technique. It leads to a low-cost solution, in comparison with other solutions such as laser scanning.

2. The selection of the region of interest in the real scene is performed applying color and spatial filters. It contributes to eliminate the noise in the acquired image and speeds up the computation time of the registration algorithm.

3. The registration algorithm performs by minimizing the mean square distance between the points from stereovision and the surface from the virtual model.

The proposed method presents good results for simple objects such as the cube object. The computed precision found (1 mm and $0.7\,°$) corresponds to the human hand precision. More complex objects should be considered for future validations.

Preliminary validation tests with a complex object (cranium) shows the registration does not perform so well as for a simple object. Difficulties encountered include the following:

- Registration algorithm falling into local minima. A potential reason for that is the noise encountered in the virtual model. A better 3D reconstruction is required to eliminate undesired meshes inside the object.

- Cloud of points for more complex objects are noisier than those of simple objects. Some potential reasons for that are the texture and the brightness of the object surface which is too shiny. It could be improved applying some texture simulating the human skin on the phantom (gypsum cranium).

We have envisaged some directions to be followed regarding the development of real time augmented reality applications:

1. Finding the better texture to be applied on the objects we want to register.

2. Developing a dynamic registration for deformable objects using finite elements. For instance in a maxillo-facial surgery there are some structures that can be moved [35].

3. Developing the bounding box, automatically shaping itself around the object of attention in order to reduce noise influence.

4. Improving registration results for more complex objects. This will deal with enhancing cloud of points reliability as well as protecting the algorithm from falling in local minima.

5. Eventually we planned to provide a proper user interface to interactively manipulate the registered virtual objects. It has already been done by integrating this technique into Medical Studio platform as described in Section 7.4. Figure 8.11 shows how the interface looks like.

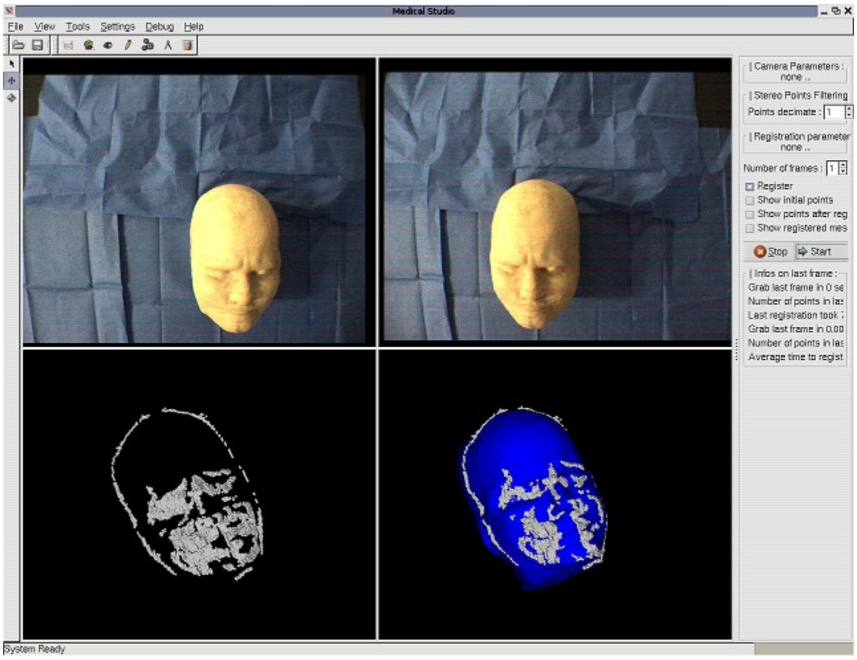


Figure 8.11: *The two top windows are displaying the left and right views acquired with the stereo camera. The left bottom window shows the cloud of points acquired and filtered in real time. The right bottom window shows the cloud of points filtered and correctly registered to the model 3D.*

Chapter 9

Marker-Based Approach

In this chapter we describe a marker-based approach to develop AR system. Such approach is based on marker detection to track objects in the real world and provide visual and intuitive guidance in the intra-operative scenario. Two kinds of guidance are developed: virtual and augmented using the MicronTracker system integrated into Medical Studio. To reach such objectives three main steps should be covered: acquisition of the real scene; tracking of real objects; and visualization of guidances. The following sections describe all these steps.

9.1 System setup

The system setup (Figure 9.1) consists of a fixed stereo camera, printed markers, a phantom of mandible and a LCD screen that could be eventually replaced by stereo glasses.

MicronTracker cameras are customized versions of PointGrey's class-leading BumbleBee stereo cameras. The camera is interfaced to the computer using a standard digital link (e.g., Firewire IEEE 1394 or USB-2). The model used is S60 with 640 x 480 resolution and lenses of 4mm with horizontal angle of 50° and vertical angle of 38°.

MicronTracker System[1] uses available visible light illumination to observe targets stereoscopically much like humans do. Tracked objects are marked with a visible painted or printed target pattern. Software running on the PC processes the camera images to detect and pinpoint visible target patterns in the images. The MicronTracker's target pattern is a set of

[1]WebSite: http://www.clarontech.com/

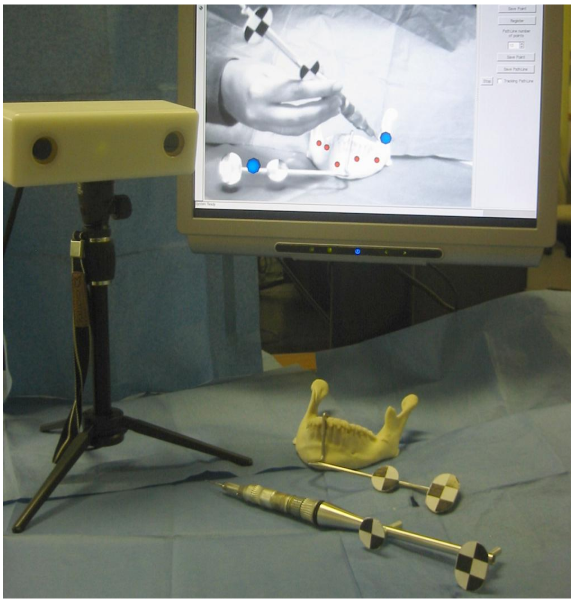


Figure 9.1: *System setup - stereo camera and markers attached to the tracked objects.*

high-contrast regions called a Marker. The regions forming the marker are geometrically defined and arranged according to specific rules dictated by the tracker's designers.

MTC (MicronTracker's C language interface) implements all the identification and pinpointing algorithms and provides a complete object-oriented application programming interface (API) for:

- Detecting and starting up the cameras connected to the computer, then loading their calibration files;
- Loading Marker templates from files;
- Capturing and processing camera frames on demand to report on detected markers and their poses;
- Creating or modifying marker templates;
- Customizing many aspects of the measurement operations to fit specific application needs.

Applications written in any language may call MTC directly to take full advantage of the MicronTracker's capabilities. MTC completely encapsulates and abstracts the lower-level camera control functions, making it unnecessary for the application to make calls to any of the lower-level modules.

MicronTracker system uses Digiclops and Triclops Libraries which are higher level interfaces to the camera control functions, provided by Point Grey Research. Triclops provides support for unstructured stereo vision capabilities.

9.2 Tracking procedure

The two key functions of optical pose trackers are:

- Detection: detect the presence of, i.e., identify, specially marked objects in the sensor's field of measurement (FOM).
- Pose measurement: report on the location and orientation of each detected object.

Both functions need to be performed repeatedly at real-time rates (10Hz or faster).

Optical pose trackers perform these functions by observing marked locations, or targets, on each tracked object from multiple angles of view, then triangulating the lines of sight to the targets to calculate each target's location. The 3-D locations of at least 3 targets are needed to calculate the object's pose, i.e., its position and orientation in space relative to the camera. Pose sensing is sometimes referred to as 6-D measurement, since at least 6 parameters are required to fully describe a pose: 3 space coordinates (X, Y, Z) and 3 orientation angles (e.g., azimuth, elevation, roll).

9.2.1 Tracking tool

The tracking of the tool and the tool tip is possible by fixing markers in the tool. A marker is made of one or more patterned surfaces rigidly attached to, or printed directly on an object, usually either a reference body or a tool of some sort. A marker may contain an arbitrary number of facets. Unlike vectors within a single facet, different facets cannot share Xpoints (and, therefore, vectors). The complete hierarchy is shown in Figure 9.2.

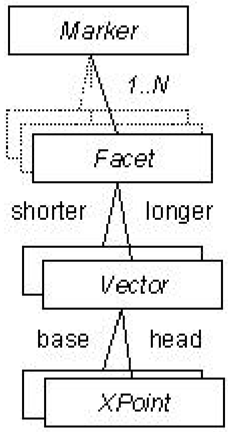


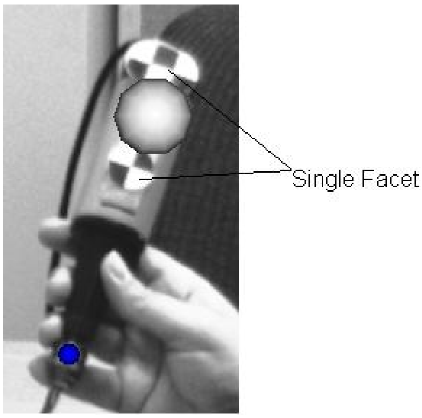
Figure 9.2: *Structure of MicronTracker markers.*

For our case, we have developed single and multi-facet markers (Figure 9.3). There may be two good reasons for designing a marker with more than one facet:

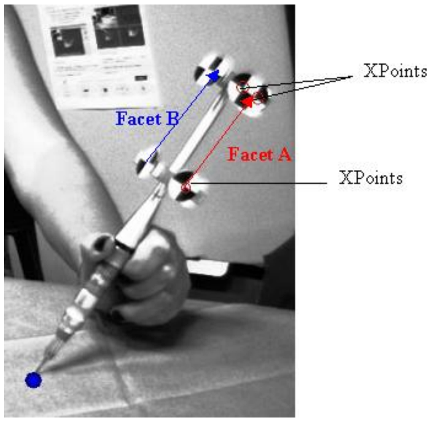
- Increasing the detection angle range by placing facets in different orientations.
- Increasing pose measurement accuracy by increasing the number of Xpoints contributing to the pose calculation.

When the Xnormal is tilted away from aiming at the lens, the Xpoint region's projection on the corresponding sensor shrinks by the cosine of the tilt angle. As the Xpoint region's projected image size falls bellow the minimum needed for detection, MicronTracker cannot continue tracking that Xpoint. Once this happens to one of the facet's Xpoints in one of the two sensors, tracking of that facet is lost. This limits the angular detection range of that facet.

Adding more facets to the marker, each facing in a different direction, increases the angular detection range to the combined angular ranges of all its facets. For example, placing mirror images of a facet on the right and left sides of a tool allows it to be used with the camera either on its left or on its right. Adding another facet on top allows detection from above as well, providing the full angular detection range needed in practice. Figure 9.3(b) shows an example of a marker with three facets (one facet is not visible in the picture) containing 3 Xpoints each.



(a)



(b)

Figure 9.3: *Markers attached to the tracked tool. (a) One facet marker (b) Multi-facet marker.*

The precision provided by such markers design that we have used in our system is discussed in Section 9.5.

9.2.2 Tracking mandible

For the set up of our technique as well as for the determination of results, we chose a standard scene to perform our tests. As our study case is a maxillo-facial surgery tool, we chose a situation close to a surgical room. For so doing, we built a *phantom* on gypsum of the patient's head (Figure 9.4). It was made possible by printing a 3D model reconstructed from CT data sets. Once the calibration procedure was performed the patient's mandible tracking is possible thanks to the markers attached to the mandible. The printed mandible (i.e.phantom) with the respective attached marker is shown in Figure 9.4.

9.3 Calibration procedure

The calibration procedure is performed just once and it comprises the following steps:

1. Select points in the 3D physical space. It is done picking points using the tracked tool tip.

(a) (b)

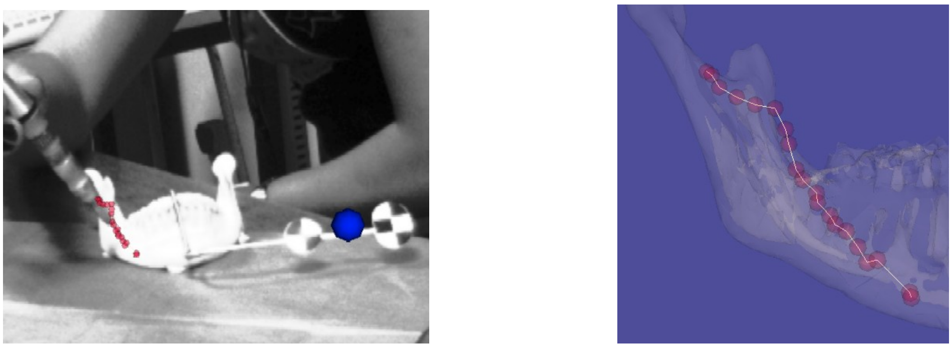
Figure 9.4: *Printed 3D objects based on the rapid prototyping technique. (a) Complete reconstructed phantom; (b) Markers attached to the tracked mandible.*

2. Apply initial transformations (i.e. translations and rotations) to the virtual model. It is a initial condition of the registration algorithm that the difference between real and virtual models can not exceed $30\,^{\circ}$.

3. Registration procedure: find the correct transformations to align the virtual model and the selected points. We use the method described in Section 8.7 for the registration of CT reconstructed mandible model with the physical space (i.e mandible phantom). As a result, we have virtual and real objects aligned and the system is able to provide two kinds of guidance visualization, virtual and augmented, described as follows.

9.4 Kinds of guidance

The main goal of the system is to provide guidance to the surgeon during the osteotomy procedure. The osteotomy pathline is supposed to be provided by the pre-operative module (see Section 6.3). Since this module is not operational yet we simulate the reference pathline by picking points in the real mandible using the tracked tool tip (Figure 9.5 (a)). The picked points are interpolated to define the pathline to be followed by the user. Figure 9.5 (b) shows the 3D representation of the inserted pathline. When the user has finished the task, again we pick the points representing the final pathline. We interpolate such points and we can then calculate the difference between the two 3D lines. How to compute the performance of the interaction is presented in the Section 11.5.1.



(a) (b)

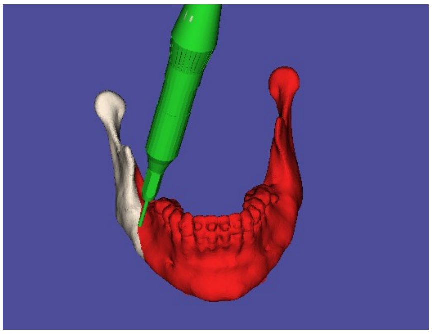
Figure 9.5: *Defining pathline. (a) Augmented reality view. (b) 3D reconstructed view.*

Once the objects were registered and the pathline defined, two distances information can be provided: distance from the tootip position to the pathline (i.e. Euclidean distance from one point to the line segment) and distance from the tooltip position to the segmented mesh (i.e. distance from one point to the mesh). We use such information in both kinds of scenario, virtual 3D and augmented.

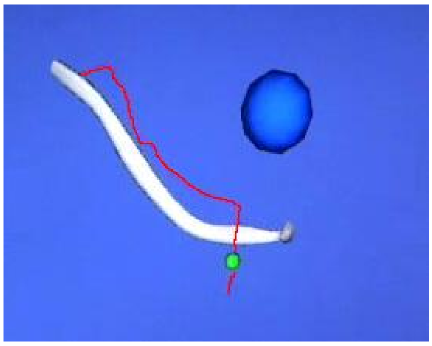
The virtual guidance provides the 3D visualization of the tracked objects in the screen. Changes in the color of the 3D tool representation indicates the guidance. For instance green color means the surgeon is touching the good position (Figure 9.6(a)). Other information such as distance from the tool tip to internal structure as well as where is placed the internal structure in the scene are also supported by the system and can be visualized as shown in Figure 9.6(b).

The augmented scene is provided using MTC library functions to project the position of a 3D location on the left or right video image. This can be used, for example, to provide visual coordinate registration accuracy feedback, or to overlay useful spatial information, such as predicted or hidden locations of interest. In our case, we have projected the selected points to perform the registration procedure, as well as the pathline to perform the osteotomy and the segmented internal structure.

We can also have an augmented 3D scene where the virtual 3D model is correctly visualized in front of the video plane. To implement that



(a)



(b)

Figure 9.6: *Virtual guidance: (a) Visual feedback of the tracked pathline: tool representation becomes green; (b) Visualization of pathline and segmented internal structure. Visual indicator for the tracked pathline (small sphere) and for the tracked distance from tooltip to the internal structure (bigger sphere).*

we have setup the virtual camera VTK with the same parameters of the real camera (i.e.position, orientation, angle, focal distance) and placed the video plane at the focal plane in the 3D space. We have defined our frustum (Figure 9.7) according to the camera parameters (Bumblebee Model S60) where: near clipping plane = $30cm$; far clipping plane = $115cm$; aspect ratio = $70cmm/55cm$ and FOVy = $38\,^{\circ}$

Problems related to the perceptual cues such as occlusions and depth perceptions during navigation were minimized by alerting the user with a color feedback to indicate position status and by setting the opacity of the

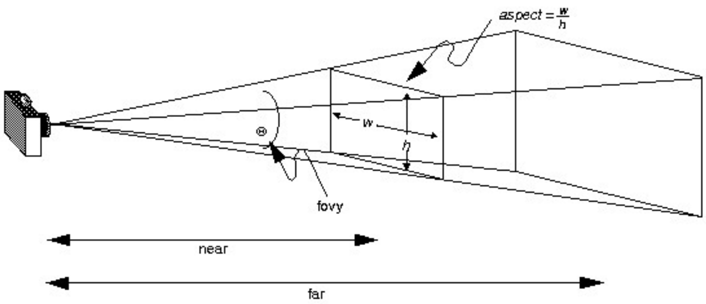
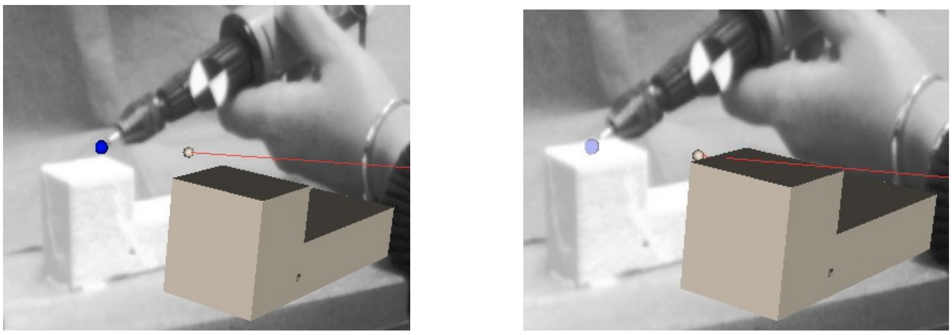


Figure 9.7: *Visualization frustum for the 3D augmented scene.*

tooltip representation to indicate the occlusion of the real object. In this case, when the user is touching the correct position, a small virtual sphere attached to the tooltip becomes green (Figure 9.9) and when the tooltip is not visible by the camera it becomes transparent. After processed the registration procedure we are able to use the 3D virtual representation of the tooltip position and real object to detect surface collision. We calculate the intersection of a line with the mesh. The line segment is defined from the tooltip position to the camera position. If there is intersection of the line with the mesh then the tracked object is not visible (i.e. it becomes transparent), otherwise it is visible (ie. it becomes solid). Figure 9.8 illustrates this visualization feedback of surface detection.



(a) (b)

Figure 9.8: *Surface intersection detection: (a) No intersection detected; (b) Intersection detected.*

Example of such kind of scene is shown in Figure 9.9 where all previous segmented structures can be projected in the video together with the guidance information.

Different scenarios combining kinds of guidance and kinds of visualization are discussed in the usability evaluation chapter (see Chapter 11).

9.5 System precision measurement

A huge problem involved in the evaluation of virtual and mixed reality system is the number of constraints imposed by its components and the dependencies identified among these components. The use of nonconventional devices which the technical performance is not yet as good as desired forbids the correct evaluation of its use. For example, during

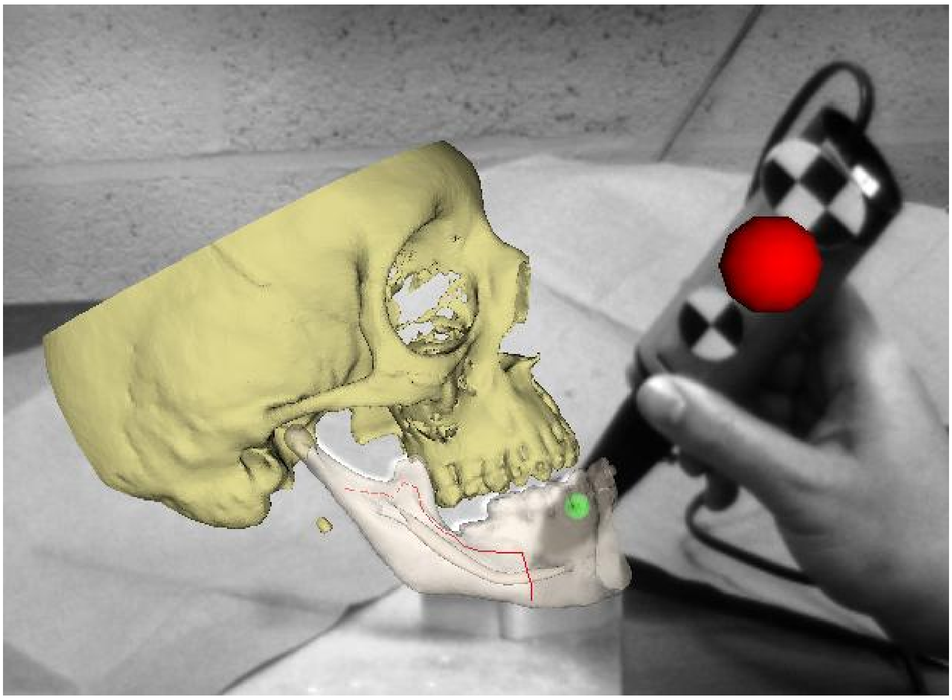


Figure 9.9: *Augmented reality guidance with visual feedback of pathline and internal structure placement and distance.*

the evaluation of a new interactive technique adapted for head-mounted displays, how to identify the origin of the good (or bad) user performance? Is the new technique really good (or bad)? Is the device sufficiently precise for the task proposed? Is the user familiar with this kind of device, if comparing with the conventional ones? As it is yet difficult to individualize the several components of a MR system, it is also hard to answer the questions above.

Trying to minimize the intervention of each component of the system in the final evaluation, we calculated our system performance by estimating the global and local errors.

The local error (e_l) is calculated by computing the sum of the error estimation for all factors that can insert noise in the system: the tracker (e_{tr}), the user interaction (e_{ui}), the registration (e_{reg}) and the printed model (e_{pm}).

$$e_l = e_{tr} + e_{ui} + e_{reg} + e_{pm} \tag{9.1}$$

The global error (e_g) is provided by calculating the distance from the points picked by the user to the registered object. We can say that

$$e_g \approx e_l \tag{9.2}$$

Accuracy, or trueness, by definition, is the closeness of measured values to their expected quantity. In experimentations, the difference between the true value and the value reported by the instrument is called the measurement error.

The most frequently used measure of spatial measurement accuracy is the Root Mean Squared Error (RMSE), calculated by averaging the squares of the individual error values, then taking the square root of the average. If all the errors have the same magnitude, RSME generates the same value as the average error. When some of the errors are larger than others, however, they are given more weight, resulting in the RMSE being larger than the average.

Error estimation for our system is calculated as follows.

9.5.1 Tracking error

There are several reasons for measurement errors in optical pose sensors and their effects are often complex and interdependent. A useful way to discuss them is by dividing into 3 categories: (1) calibration; (2) drift; and (3) jitter.

Calibration errors are introduced in the camera calibration process at the factory, and remain constant for years unless the physical structure of the camera changes, e.g., due to a physical shock. Drift represents the effects of variations in the operating environment, e.g., temperature, lighting, or marker position/orientation. Jitter is a momentary deviation caused by optical or electrical noise in the image capture and analog-to-digital conversion circuitry.

Claron Technology guarantees that the calibration error of each camera shipped falls bellow *0.25mm* RMSE and the jitter's standard deviation under a selection of typical measurement conditions (i.e., taking into account the field of measurement approximately *75 cm*, *150 Lux*, *20 ms shutter*) is *0.007mm* RMSE for static target and *0.07mm* RMSE for moving target.

To control drift problems, we access the hazard messages provided by the MTC library. For instance, drift can be caused by a rapid internal temperature change in camera. In this case a thermal instability hazard is added to measurements until camera becomes thermally stable ($\approx$ 15 minutes following power-up or 2 minutes following activation, whichever is later). Besides, small drift may result when part of an Xpoint (point tracked in a marker, Figure 9.11(b)) is shadowed at certain orientations. This condition is detected and a *Shadow over Xpoint* hazard is displayed on the screen.

A relevant factor for the tracked object accuracy relates to the way we dispose the markers. Due to the lever effect, the farther apart the Xpoints at the ends of a vector, the lower the error in measuring the vector's orientation in space, and, therefore, the lower the error in extrapolating object positions of interest, such as the tool used by us to cut the jaw. The drawing in Figure 9.10 demonstrates this effect by showing a tool position error range for two different placements of the Xpoints. To minimize extrapolation error, one end of the longer vector should be placed as close to the tool tip as possible, and the other as further away from it as possible.

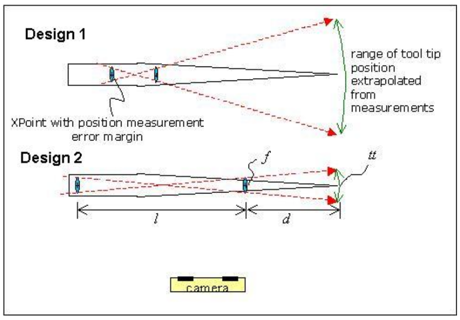


Figure 9.10: *Relation between the tool tip definition accuracy and the markers placement*

Assuming the tool tip is approximately aligned with the long vector of the marker, the vector's length is l (see Figure 9.10), and given a position RMSE at each marker Xpoint e_f and a distance d between f and the tool tip position tt, the tool tip RMSE e_{tt} is approximately:

$$e_{tt} \approx e_f + 1.5 * e_f * d/l \tag{9.3}$$

Such linear dependency has been verified and provided by Claron Technology.

We have tested two designs of markers for our tracked tool, as shown in Figure 9.11. We have computed 1000 poses for each design and the RMS errors for the positions shown in Figure 9.10 are depicted in Table 9.1.

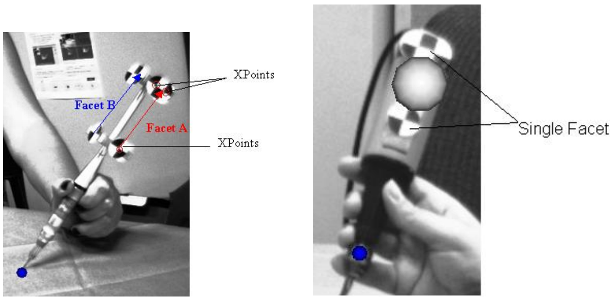


Figure 9.11: *Markers attached to the tracked tool: design 1 – a multi-facet marker (left); and design 2 – one facet marker (right).*

Design	$l(mm)$	$d(mm)$	e_f	e_{tt}	RMSE for the tool tip tracked poses
1	60.55	118.91	0.17	0.5564	0.5680
2	117.1933	58.5695	0.20	0.3470	0.3620

Table 9.1: *Tool tip precision error for 1000 poses*

If d is small relative to l, the tip error is dominated by the Xpoint error (roughly, calibration error + jitter, typically in the range 0.2mm-0.4mm RMS). If d is large relative to l, the error is dominated by the ratio d/l. As a rule of thumb, to ensure that RMS error at the tip is sub-millimeter, keep $l \geq d$.

Analyzing the Table 9.1, we verify that the marker configuration for the Design 2 is most accurate then the one used for Design 1 fitting the expected range (0.2mm-0.4mm). Then, we decided to use such design for our tests.

9.5.2 User interaction error

We call user interaction error, the imprecision of the user while interacting with the system and is directly related to the dexterous ability of each user. To evaluate the precision of such interaction we have performed a sequence of 15 tests with 15 users that provide us the RMS and the standard deviation for the user interaction error.

The test consists in picking the 4 corners of one of the faces of the L mock-up. As the real distance and angles between these points are known (Figure 9.12(left)), we can compare it with the distances and angles given by the acquired 3D points (Figure 9.12(right)).

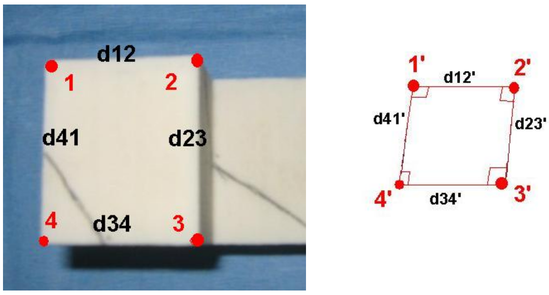


Figure 9.12: *Distances and angles between real and virtual measures are compared: points selected by the users on the real object surface (left); and virtual representation of the selected points (right).*

Table 9.2 shows the RMS error and standard deviation computed from 15 users essays. As these points were acquired by the users while using the tracking system and the tracking error e_{tr} is known, the user interaction error (e_{ui}) should be the values collected during the experiments, minus the tracking error, or yet $0.048mm$.

	RMSE	STD
Distances (mm)	0.41	0.66
Angles (radians)	0.06	0.025

Table 9.2: *RMS error and standard deviation for distances and angles calculated from the points picked by the users to the real corners positions.*

If the user that will operate the system performs such calibration procedure before the registration procedure, the system will be able to compute a personalized performance for this user.

9.5.3 Printed model error

The mock-ups (i.e. the phantoms) used in this work were produced by rapid prototyping based on a principle of manufacture per addition of matter using the Z-corp printer[2].

The final printed object were scanned (with a laser scanner) and compared with the initial 3D model using the Inspector Software to align and

[2]This work is done in collaboration with CRIF-WTCM. Website: http://www.wtcm.be/

register the models. For the L shape, virtual and real mock-ups are overlapped with a precision of *0.20mm*.

The final printed object were scanned (using laser scanner) and compared with the initial 3D model using the Inspector Software[3] to align and registered the models. Figure 9.13 is showing the objects overlapped with a precision of $0.20mm$.

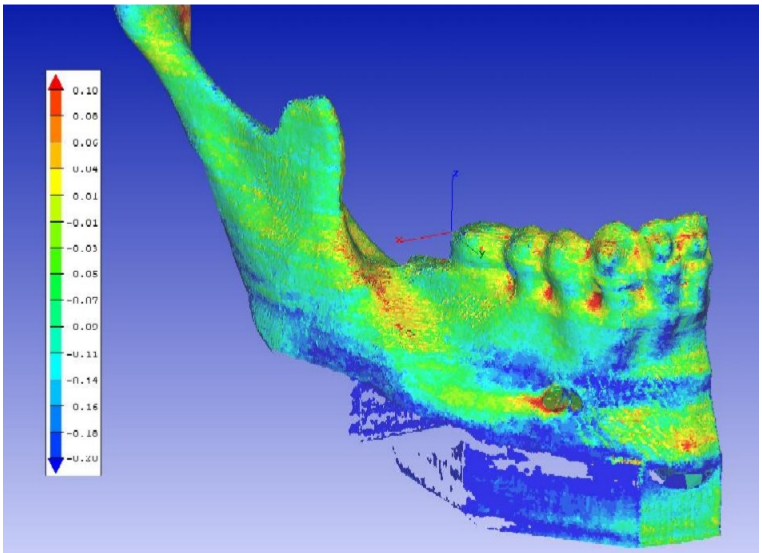


Figure 9.13: *Registration between printed and virtual objects indicates printed model error of* $0.20mm$.

9.5.4 Registration error

The efficiency and accuracy of the registration method used is highly dependent on the choice of a number of parameters both at the acquisition stage and during data processing. Such method was previously validated by Noirhomme *et al.* [96], where registration has sub-pixel accuracy in a study with simulated data, while they obtained a point to surface RMS error of *0.17mm* to *0.38mm* in simulation tests.

While using such algorithm we should pay attention to three parameters that influence the accuracy of the method: the number of points one should acquire to characterize the object surface, the pattern along which

[3] Website: http:// www.metrics.be

those points should be acquired, and the resolution needed for the Euclidean distance map.

For the resolution, we assume the values found by Quentin *et al.* where a resolution similar to the internal slice resolution of the image yields good results. Improving the distance transformation resolution further does not benefit to the accuracy.

For the number of points and which pattern to consider we have performed a sequence of tests and found 20 points along the contour of the mandible and 12 points on the corners of the L has produced acceptable registration matching.

We can verify it comparing the global error estimation with the sum of the local errors. As a result of the registration step we are able to compute the global error by calculating the distance from the points picked by the user to the registered object. Then using equation (1) we can find the registration error

$$e_{reg} \approx e_g - e_{tr} - e_{ui} - e_{pm} \tag{9.4}$$

We have found global registration error from $0.8mm$ to $1.5mm$ which give us a filtered registration error from $0.24mm$ to $0.94mm$. Such results seem to fit the range error of *0.17mm* to *0.38mm* previously computed in simulations [96].

9.6 Conclusions

In this chapter we have presented a marker-based method to provide virtual and augmented guidance for the maxillo-facial surgery. We have developed a methodology for the system validation by estimating local errors from user interaction calibration, calibration system, tracking system, registration algorithm and accuracy of printed model. Using such approach for each user the system will be able to assess a final personalized user performance while executing a specific task. Results of such validation seems acceptable when compared with other systems. Table 9.3 presents our results compared to the MAGI (Microscope-Assisted Guided Interventions) system results [64]. Our system and the MAGI ones are not using the same technology but both involve the same source of errors. In the MAGI system image-to-physical registration is calculated using a set of bone-implanted fiducials which are localized in the image, and in the operative scene using a tracked pointer.

Error	Our system	MAGI system [64]
Calibration	$0.25mm$	$0.26mm$
Tracking	$\approx 0.20mm$ (individuals tracked markers) $\approx 0.36mm$ (tracked tooltip)	$\approx 0.20mm$ (individuals tracked IREDs) —
Registration (3D alignment)	$0.8mm$ to $1.5mm$	$1.61mm$
2D overlay	—	$0.63mm$ to $1.48mm$ (phantom) $0.91mm$ to $1.88mm$ (clinical trials)
Printed Phantom	$0.20mm$	—

Table 9.3: *Comparison of error estimations*

The registration transformation is calculated using the point correspondences. To compensate for intra-operative movement the patient is fitted with a Lockable Acrylic Dental Stent (LADS), which affixes rigidly to the upper teeth. A set of IREDs (infra-red emitting diodes) is attached to the LADS and tracked with the Optotrak 3020 (Northen Digital Inc.). Microscope movement is also tracked. Finally the projection matrix produced by the calibration is used to project the virtual structures into overlay coordinates. A pair of image injection units overlay the images into the eye-pieces of the stereo microscope.

Although a highly accurate tracking system is used (the MicronTracker tracks tooltip to an accuracy of $\approx 0.20mm$), tracking (i.e. taking account the user interaction while picking points) has been identified as the most significant error source at present. Based on the error analysis a number of areas for potential improvement in system performance have been identified. One of these is alternative patterns configurations: in point-based registration errors can be reduced by spreading out the points in all object surface. In this sense an automatic initialization process can be also envisaged to assure that all performed registrations has the same initialization transforms applied.

The current accuracy of our system is thought to be sufficient for a wide range of medical procedures such as planning craniotomy sites (i.e. for the neurosurgery) and osteotomy procedure (i.e. for maxillo-facial surgery). If a less accurate tracking system were employed, the range of applicable procedures would be reduced.

As we are considering that our tracked object and the camera are fixed in the environment we did not account for the system latency. In [54] demonstrated that the hardest source of error to eliminate in a carefully engineered AR system was system latency, where a delay of a single millisecond could cause transient registration errors of the order of 1mm when the user or tracked objects are in motion. Such validation aspect should be considered in future versions of our system for instance for tracking the mandible movement.

Although this analysis is specific for our system, the same concepts can be applied to other augmented reality systems, enabling real/virtual alignment errors to be controlled and reduced significantly.

Part III

Evaluation of Mixed Reality Systems

Chapter 10

Related Works

There are a number of techniques and devices available for the development of applications involving interaction in 3D desktop, 3D immersive, tangible and mixed worlds. To choose the most suitable one for a given situation is a difficult task and evaluation techniques should be developed and applied.

This chapter is related to the design of novel methodologies for experimental evaluation of MR user interfaces; usability studies, evaluation and categorization of MR interfaces and their components. There is a consensus about the lack of explicit theories and few detailed guidelines in computer-human interaction to support the design for usability of this emergent technology and its varied applications. As in the design phase, many usability criteria are inspired and derived from previous usability experience with Virtual Reality and 3D systems. In the following sections we will give an overview from previous to recent results for usability evaluation in such kinds of systems.

10.1 Usability in VR, Tangible and MR systems

Some experiments regarding evaluation of 3D interactions in a VR worlds can be found in previous works, for instance in [92] [3] [141] and more recently in [82] [14] and [120]. We could resume the different results found in this domain by saying that the 3D visualization could help the user interaction if the dominants perceptual cues (i.e. depth cues involving motion parallax, shadows, transparency, etc.) are identified and solved for each application.

Many kinds of interactions are emerging from 3D and virtual technologies taking the direction of more natural interactions. However, the same problems regarding the 3D visualization (i.e. perceptual cues) are still there in addition to other factors specific to this new kind of interaction. So far, evaluation of usability as well as design for usability is needed. Kato et al. in an usability study [63] present five main objectives for tangible interfaces. Object affordances should match physical constraints to task requirements. Moreover, interfaces should support parallel activities when reasonable, as well as providing aids for physically-based interaction techniques based on proximity and spatial relations. Furthermore, objects support spatial and two-handed manipulation.

In [22] they present the usability evaluation results of MixIt, a testbed for authoring in Mixed Reality Environments. The experiment has conducted usability evaluations for simple manipulations using simple geometric transformations (translation, rotation, scaling). As a result, they have noticed that the rotations using the magic ring were especially problematic. Several subjects mentioned the recognition problems resulting from one hand occluding the marker on the other hand, while trying to cross hands. Marker tracking problems related to the lighting also plagued the experiment. When asked if the users found the system to be helping them making mistakes or at least not helping them avoid them, subjects gave balanced answers: 6 said 'no' while 5 said 'yes'. It shows that more efforts in developing natural and intuitive interfaces to interact with Mixed Reality systems need to be investigated yet.

Recent evaluations of tangibles interfaces have showed directions for using more affective input as user's interaction options. For instance in SenToy [55] is a tangible interface device used to influence emotional behavior in the logic game FantasyA. Another interaction type is the Full body interaction. Navigation in virtual and mixed environments usually requires a wired interface, some kind of console, or keyboard. The advent of perceptual interface techniques allows a new option. In the work presented by [67] several different interaction styles are compared, based on an analysis of the space of possible perceptual interface abstractions for full-body navigation and the results of a wizard-of-oz study of user preferences.

10.2 Usability in AR applications

Unlike Virtual Reality, AR enhances the real environment rather than replacing it. Graphics are superimposed on the user's view over the real environment. In a typical AR system, a see-through or video-through head-mounted display (HMD) is used to composite computer generated graphics with the real environment. AR technology has many potential applications, including computer assisted instruction (CAI) [34], industrial training [17], computer-aided surgery [5] computer visualization, engineering design, interior design and modeling [66] [139], and entertainment as the recent Augmented Coliseum game [1].

Some studies have demonstrated the effectiveness of AR applications when compared with traditional ones. For instance, [119] provides evidence to support the proposition that AR systems improve task performance and can relieve mental workload on assembly tasks. Basically the effectiveness of AR applications is obtained by reducing head and eye movement, by reducing cost of attention switching and by supporting spatial cognition and mental transformation. However, the limitations in the current calibration techniques and display and tracking technologies are the biggest obstacles preventing AR from being realistic in practical uses. Designers seeking to make use of the performance gains of AR systems also need to consider how the user manages their attention in such systems and avoid over-reliance on cues from the AR system. The phenomenon of attention tunneling could possibly reduce performance in cases where AR cueing overwhelms the user's attention causing distraction from important relevant cues of the physical environment.

There is extensive research in Head-up display ergonomics for aircraft pilots with particular examination of the issue of attention switching among information sources and the environment [77] [7]. Head-Up Display or HUD is now a standard feature on many aircraft and the idea is even being used in some automobiles. In both cases, information (such as speed, altitude, etc.) is projected in a semi-transparent fashion over the outside view. All of the most important information is available in one place.

Researches in this field suggest that visual attention can be focused on either Head-Up Displays (HUD) or on the world beyond them, but not on both simultaneously [77]. The results suggest that attentional constraints place severe limits on the ability of pilots to process HUD-referenced information and world-referenced information simultaneously. In addition, they provide direct evidence that transitioning from processing HUD information to processing world information requires an attention shift.

[1] Augmented Coliseum Website: http://www.hi.mce.uec.ac.jp/inami-lab/en/projects/AugmentedColiseum/index.html.

Such issues will be taken into account while designing our usability tests for detection of cognitive continuity property.

10.3 Usability for AR in medical applications

See-through AR in surgery is still in the experimental phase of research and its benefits are speculative. In the small number of systems developed worldwide so far, it is assumed, but not always tested that stereo overlays provide accurate and unambiguous depth information to the surgeon (assuming a good calibration). In [61] they have empirically tested this assumption and found that errors in perception still arise despite accurate stereo displays. The most significant findings of the work provided evidence that the presence of a physical surface, placed between the observer and the virtual object, caused a misperception of the virtual object's depth.

This is consistent with the work of Ellis and Menges [33] who also showed an underestimation of judged depth using a virtual object presented behind a physical surface. Any interaction, in the form of overlap between the virtual cone and physical surface will draw the observer's attention to both surfaces. The observer either has to simultaneously fuse the cone and physical surface, or divide visual attention between them. The observer then has to align the physical pointer with the cone (the most distal of the two surfaces), which results in an overestimation of its depth. With the cone tip coincident with the surface ($0mm$) exactly half the cone protrudes in front of the surface, and the eyes need only focus at one point. As the cone moves below the surface, the percentage of cone surface protruding through the physical surface decreases but the separation between the points of focus increases. The size of the error in locating the cone relative to the surface increases until the cone no longer protrudes or contacts the physical surface (the tip is presented more than 15mm from the surface). It is possible that at this point either the observer's attention is no longer drawn to both surfaces simultaneously as they no longer overlap, or that the separation between the two objects is greater than the fusible range and they choose to focus their attention on the cone. With the cone at distances greater than $40mm$ below the surface, outside the predicted fusional area, observers underestimate the depth in accordance with the theory of occlusion and proximal convergence effects.

In [138] the augmented reality navigation system prototype was adapted to Magnetic Ressonance-guided needle placement procedures. The table of the MR scanner and the needle are equipped with retroreflective markers, which are tracked by a head-mounted infrared camera. The video-see-through AR system overlays medical images, virtual guidance information, and the needle graphics onto the live stereoscopic video view.

During insertion, the needle can be observed virtually at its actual location in real-time, supporting the interventional procedure in a very efficient and intuitive way. The usability criteria investigated in this study was the precision and robustness of the AR needle guidance on Natrosol gel phantoms with embedded target structures of 12mm and 6mm diameter. As results the system can significantly reduce procedure time for AR guided interventions, while at the same time increasing the precision. As conclusion these first experimental results suggest that stereo AR should have a role in the future of surgical navigation.

10.4 Effective knowledge

10.4.1 Usability criteria in VR applications

In [42] they have proposed a taxonomy on usability characteristics in virtual environments framework to be particularly well-suited for addressing how individual usability issues fit into a more abstract, larger scale understanding of interaction between users and virtual environments.

In [62] presents usability guidelines for the design of content and interaction in virtual environments. The guidelines aim to highlight important usability considerations and are informed by theoretical research into and empirical studies of user interaction behavior. The guidelines are applicable to virtual environments that: are spatially expansive and populated with a number of objects; have different points of observation that can be accessed through navigation; are generally modeled on real world phenomena, and are single-user systems.

An assignment experiment (Ergonomic Criteria X Usability Problems) has been conducted in [108] to ensure the appropriate understanding of the proposed criteria. The results of this study identify the criteria that are well understood and the ones that need further improvement. It is only a first step with the goal of providing usability inspection methods for Human Virtual Environments Interactions.

10.4.2 Validation criteria in image-guided surgery

Validation requires the application of defined criteria to a device or process. Common examples of validation criteria which may be applicable to image-guided surgery include [60]:

- Accuracy: Goodman [105] defines accuracy as the degree to which a measurement is true or correct. For each sample of experimental data local accuracy is defined as the difference between computed values and theoretical values, i.e., known from a ground truth. This difference is generally referred to as local error. Under specific assumptions, a global accuracy value can be computed for the entire data set from a combination of local accuracy values.

- Precision and Reproducibility or Reliability: Precision of a process is the resolution at which its results are repeatable, i.e., the value of the random fluctuation in the measurement made by the process. Precision is intrinsic to this process. This value is generally expressed in the parameter space. Goodman [105] defines reliability *as the extent to which an observation that is repeated in the same, stable population yields the same result*.

- Robustness: The robustness of a method refers to its performance in the presence of disruptive factors such as intrinsic data variability, pathology, or inter-individual anatomic or physiologic variability.

- Consistency or Closed Loops: This criterion is mainly studied in image registration validation [38], by studying the effects of the composition of *n* transformations that forms a circuit: $Tn1 \circ \ldots \circ T23 \circ T12$. The consistency is a measure of the difference of the composition from the identity. This criterion does not require any ground truth, but the onus is on the user to convince that there is no bias in the error estimation obtained.

- Fault Detection: This is the ability of a method to detect by itself when it succeeds (e.g. result is within a given accuracy) or fails.

- Functional complexity and computation time: These are characteristics of method implementation. Functional complexity concerns the steps that are time-consuming or cumbersome for the operator. It deals both with man-computer interaction and integration in the clinical context and has a relationship with physician acceptance of the system or method. The degree of automation of a method is an important aspect of functional complexity (manual, semi automatic or automatic).

The requirement of an overall validation of image-guided surgery systems (i.e. including all its components) should also be taken into account by estimating errors at each stage of the IGS process and by modeling how errors propagate through the entire IGS process. For this reason some of these criteria such as precision, accuracy, consistency were measured as local criteria resulting from the co-registration and optical tracking implementation (see Chapters 8 and 9). On the other hand the functional complexity and computational time will be considered as a global criteria to evaluate the whole interactional impact of the technique. It lies on our definition of functional property (see Section 4.2.3) to evaluate the continuous interaction.

As conclusion besides the validation criteria that are applied to a device or a process we should take account the usability criteria to assess the interaction technique of this new paradigm of interaction. As any mixed reality application, IGS systems are characterized by a continuous process of exchange of information at a high resolution and multiple worlds of interaction. When there are multiple sources of information and two worlds of interaction (real and virtual) designers must make choices about what to attend to and when. Users need to focus exclusively on a single item at a time without interference from other items. They occasionally need to time-share or divide their attention between two (or more) items of interest that can be part of the same or different worlds.

10.5 General guidelines for AR systems

In [39] they have generated some a priori guidelines based on the major perceptual issues involved with Obscured Information Visualization (OIV) systems but may also generalize to other types of AR and VR displays. Important design guidelines include:

- Distance conveyance - In OIV environments, distance and absolute location can be confusing, so AR renderings should disambiguate information about distance or position.

- Proper motion physics - For dynamic displays, motion parallax is an important cue to human observers, so it is important that the information in depth move in such a way as to convey its proper position. This can be achieved with properly defined geometries and metrically-accurate models of the environment that rotate and move in realistic ways.

- Eliminate unneeded AR motion - Because the human visual system is so sensitive to motion, unneeded motion of rendered material (e.g., the slowly moving self-organization of rendered AR tags) as well as mis-registration should be eliminated or minimized as much as possible.

- Selective or multiple cues - Because the accuracy of different perceptual cues vary with distance, specific perceptual cues (e.g., motion parallax or size-constancy/scaling) should be selected if displays are operating in an environment of a limited range of depths/distances. Multiple perceptual cues should be integrated if the ranges of display depth/distances are variable.

There are many other factors that do not deal directly with perceptual influence on OIV, but are important features to cover, including:

- Define rule space - Another important factor for OIV system development is to define the conditions under which augmented information should be displayed. For example, if the viewer knows that augmented information will only be presented on hallway surfaces or only within the confines of the object, then these rules can help disambiguate otherwise vague location information. Carefully defining a series of rules, such as having different color or symbols to designate particular locations, will go a long way toward building an unambiguous OIV system. Even implicit rules or context clues, such as augmented messages only appear in one's office, also reduce positional uncertainty. So, while more detailed rules increase the cognitive complexity for the user, more sophisticated rule systems can drastically reduce the ambiguity in OIV displays.

- Effectiveness testing - While these general guidelines can be used as a starting point, some form of experimental testing (whether pilot testing or more involved experimental procedures) should be incorporated through the design development. Such an empirical approach can improve overall system design as certain specific perceptual cues may be better suited for certain applications. Multiple cues may interact in a way that could be quantified and subsequently compared to other designs. The instructions that participants get in experiments should be created very carefully so as not to deliberately bias subjects' responses.

Regarding the perceptual cues some guidelines have been proposed. In light of the results from the studies [57] [56], which indicate that fusion breakdown can be used as an extra cue, over and above binocular disparity matching, two perceptual features can be manipulated to facilitate the alignment performance using the model of real-virtual object interaction in stereoscopic AR environments: geometric form of the virtual point and texture density of the surface. Based on the real-virtual object interaction model [56], the more features on either the virtual pointer or the surface, the easier to perceive the fusion breakdown, and thus facilitate stereo matching and decision making process, and eventually make more accurate measurements. However, it should be also noted that there appears to be a limit for the higher texture density advantage, somewhere around 30%. In other words, when the texture density approached this optimal value, it facilitated alignment, but when the texture density exceeded this optimal value, it started to generate more alignment errors, possibly because there were too many surface features to be fused (just like a field of grass with too many small element to identify), or perhaps due to too much glare caused by the increasing proportion of white versus black texture.

The more appropriate perceptual cues should be identified and adapted for each application. Some perceptual cues for Medical applications were discussed in section 10.3. The design development of similar systems could be improved using other formalized usability-engineering processes that provide a structured evaluation technique [41] and [52] conforming discussed in the next section.

10.6 Evaluation methods

Usability engineering is a cost-effective, user-centered process that ensures a high level of effectiveness, efficiency, and safety in complex interactive systems [53]. Usability engineering includes both design and evaluations with users; it is not typically extensive hypothesis-testing-based experimentation, but instead is structured, iterative user-centered design and evaluation applied during all phases of the interactive system development life cycle. Most existing usability engineering methods widely in use were spawned by the development of traditional desktop graphical user interfaces (GUIs).

Since the focus of this part is the usability engineering activity of usability evaluation, in the following sections we briefly explain several types of usability evaluation. These include expert evaluation (also sometimes called heuristic evaluation or usability inspection), user-based statistical evaluation, formative evaluation, and summative evaluation. These introductory and brief explanations are, of necessity, rather abstract, to introduce each type of evaluation and then justify our method choice.

10.6.1 Expert usability evaluation

The process of identifying potential usability problems by comparing a user interface design to established usability design guidelines is called expert usability evaluation (or heuristic evaluation or usability inspection). Those identified problems are then used to derive recommendations for improving that design. This method is used by usability experts to identify critical usability issues early in the development cycle, so that these design issues can be addressed as part of the iterative design process [83]. Often the usability experts rely explicitly and solely on established usability design guidelines to determine whether a user interface design effectively and efficiently supports user task performance (i.e., has high usability). Usability experts may also rely more implicitly on design guidelines while they work through user task scenarios (typically created during domain analysis, another usability engineering activity) during their evaluation. Each evaluator first inspects the design alone, independently of other evaluators' findings. All evaluators then combine their data to analyze both common and conflicting usability findings. Nielsen [83] recommends that three to five evaluators performing an expert evaluation will find a majority of the most severe usability problems. He has also shown empirically that fewer evaluators generally identify only a small subset of problems and that more evaluators produce diminishing results at higher costs. Results from an expert evaluation should not only identify problematic user interface components and interaction techniques, but should also indicate why a particular component or technique is problematic. Results of this type of evaluation typically are not applicable across a variety of different application, since the purpose of the evaluation is to assess specific components or techniques for a specific application. This is arguably the most cost-effective type of usability evaluation, because it does not involve users.

10.6.2 User-based statistical evaluation

The process of performing relatively small and quick empirical studies to determine what specific design factors are most likely to affect user task performance we call user-based statistical evaluation. This can be especially effective when designing a user interface to support new and novel hardware, domains, and user tasks. Such evaluations typically focus on lower-level cognitive or perceptual tasks, where the importance of these tasks would be suggested by earlier activities in the usability engineering process. These studies are usually targeted at a specific part (e.g., a component or feature) of a user interface design, as opposed to the user interface as a whole. They may involve tasks that are atomic components of higher-level representative user tasks, and the tasks are often generic rather than application-specific. These evaluations are very similar to traditional human factors experiments and are guided by a well-crafted experimental design to assess user performance by varying design factors. Users perform tasks that are narrowly focused and carefully designed to study a specific user interface component or feature. Such evaluations help refine various user interface components or features, in preparation for more comprehensive and application-specific formative evaluations. Studies [53] indicate that the components designed and refined through quick and iterative user-based statistical evaluations produce mature user interface components and features that are well-suited to support overall application tasks and user task flow. Results of this type of evaluation typically are not applicable across a variety of different applications, since the purpose of the evaluation is to refine components or features for a specific application.

10.6.3 Formative usability evaluation

The process of assessing, refining, and improving a user interface design by having representative users perform task-based scenarios, observing their performance, and collecting data to empirically identify usability problems [53] is called formative usability evaluation. This observational evaluation method can ensure usability of interactive systems by including users early and continually throughout user interface development. The method relies heavily on usage context (e.g., user tasks, user environment, user profiles), as well as a solid understanding of human-computer interaction. A typical cycle of formative evaluation begins with creation of user scenarios based on domain analysis activities. These scenarios are specifically designed to explore and evaluate user tasks, information, and work flows. Representative users perform these tasks as evaluators collect both qualitative and quantitative data. Qualitative data include critical incidents [104], a user event that has a significant impact, either positive or negative, on users' task performance and/or satisfaction. Quantitative data include metrics such as how long it takes a user to perform a specific task, the number of errors a user makes during task performance, measures of user satisfaction, and so on. Collected quantitative data are then compared to appropriate baseline metrics, sometimes redefining or altering evaluators' perceptions of what should be considered baseline. Both qualitative and quantitative data are equally important since they each provide unique insight into a user interface designs strengths and weaknesses. Finally, evaluators analyze these data to identify user interface components or features that both support and detract from user task performance, and to suggest and prioritize user interface design changes. As with the first two types of evaluations, results of this type of evaluation typically are not applicable across a variety of different applications, since formative evaluation is designed to assess a specific application.

10.6.4 Summative usability evaluation

The process of statistically comparing several different systems or candidate designs, for example, to determine which one is *better*, where better is defined in advance, is called summative evaluation (also called outcome or impact evaluations). In contrast to formative evaluation, it is typically performed after a product or some part of its design is more or less complete. In practice, summative evaluation can take many forms. The most common are the comparative field trial, and more recently, the expert review [116]. While both the field trial and expert review methods are well-suited for design assessment, they typically involve assessment of single prototypes or field-delivered designs.

10.7 Mental workload assessment

Several theories and models have been developed to explain information processing mechanisms required during multiple task performance. The tasks being performed determine the sensory inputs. The level of interference created between concurrent tasks depends on many factors and affects the workload experienced for task completion. Although similarities do exist among current information processing theories, conclusive evidence has not been produced to determine which theory most appropriately describes the cognitive processes involved in completing concurrent tasks requiring different resources. Multiple-task interference occurs due to resource competition. Identifying the resources required to complete a task is of fundamental importance to the measurement of workload and performance. The approach used by an individual to process multiple inputs, however, influences subjective assessment levels.

10.7.1 Kind of demands

A satisfactory consensus of the factors that determine mental workload levels, have yet to be clearly presented in the literature. *Mental workload is intrinsically complex and multifaceted* [48]. Evidence exists that there is a relationship between mental workload and stress, but the two terms are not synonymous. Mental workload refers to the processing that is required to perform a task given the resources available to the individual. Although stress also refers to the relationship between environmental demands and availability of resources, demands originate from the total environment. Stress is any circumstance that threatens or is perceived to threaten adaptation and ability to cope with demands [68]. Demands can be physical, psychosocial or organizational. Mental workload research attempts to specify the limits of the processing capacity of the individual, whereas stress research examines the factors in the work environment that produce negative effects [43].

10.7.2 Methods and tools

There is no direct method for measuring mental workload; instead, several indirect measures have been developed (objective and subjective). Common categories of indirect mental workload measurement techniques include: (1) physiological measures (e.g. heart rate, heart rate variability, and brain activity), (2) performance measures (e.g. reaction time, number correct, and number detected), and (3) subjective procedures that include subjects estimations of workload [129]. None of these techniques are a pure measure of mental workload, with each contaminated by other factors. It is usually recommended that the most effective approach is to combine two or more of the above techniques. In the following subsections we give a brief description of the two most known and used methods to assess subjective workload during task execution.

NASA-TLX

The NASA Task Load Index (NASA-TLX) provides an overall subjective workload score based on a weighted average of ratings on six subscales or dimensions: mental demands, physical demands, temporal (time) demands, own performance, effort, and frustration, as described by [50]. Each of the six scales is used to represent the underlying characteristics of subjective workload. At the conclusion of a task, subjects provide ratings on each of the six dimensions. Ratings from the scale are then weighted on the basis of data generated by the subject concerning the contributions of each dimension to the total workload associated with performance of a task. Rating scale definitions are provided in Figure 12.2. The weighted ratings are then combined to create an overall index of subjective workload. Although the physical demand of the task is included in the overall rating provided by the NASA-TLX, the index is not used as a measure of physical workload but instead recognizes the potential influence of physical activity on the perception of mental workload.

Experimental results indicate that the NASA-TLX is a sensitive instrument for assessing subjective workload in a number of tasks [21] and [50]. It has also been shown to be applicable in a number of different task environments, such as flight simulation, actual flight tasks, air defense, remotely controlled vehicles and signal detection [21]. Consistent scores are produced among people who are performing the same task, reducing between-rater variability [137].

SWAT

SWAT provides an easily administered subjective scaling method to be used in the cockpit or other crew stations to quantify the workload associated with various activities [98]. SWAT postulates a multidimensional model of workload comprising three-point dimensions or factors: 1) time load, 2) mental effort load and 3) psychological stress load. The method involves a two-step procedure. In the first step, the scale development phase, hypothetical activities are rank-ordered according to perceived workload. Each activity is specified in terms of a particular distribution of load across the three dimensions. These data are transformed, by means of conjoint measurement, into an interval scale of workload ranging from 0 to 100. In the second step, the event scoring phase, an activity or event is rated by assigning a value of 1 to 3 on each of the three dimensions. The scale value associated with this combination (obtained from the scale development phase) is then assigned as the workload value for that activity.

The three dimensions in this scale were developed somewhat intuitively. Boyd [16] determined that the dimensions are not independent and that ratings tend to increase for all three dimensions even if only one dimension is altered. The dimensions are weighted based on the definition of workload being employed by the individual. The scaling algorithm produces interval-scaled estimates of the workload dimensions as well as estimates of their combined effects [98]. SWAT is unique among other workload rating scales because it demonstrates interval-scale properties [98]. Although results are similar to those obtained using the NASA-TLX, they are slightly less consistent across individuals [137].

10.8 Conclusion

To assesses the continuity property we will explore performance measures using objectives tests linked with subjective tests. Because it is difficult to simulate all conditions existenting in the real situation (for instance the stress during a surgery) we did not consider physiological measures and then user's states such as stress were not taken account.

As reported in [52] empirical comparative approach employing representative users, instantiated in the summative evaluation process, is very effective for analyzing strengths and weaknesses of various well-formed, candidate design set within appropriate user scenarios. However, it is the most costly type of evaluation because it needs a large number of users to achieve statistical validity and reliability, and because data analysis can be complex and challenging. Unlike the other three types of evaluation we presented, results of this type of evaluation typically are applicable across a variety of different applications, since they give comparative outcomes for different kinds of user interface components, features, and/or interaction techniques spanning a number of diverse user interfaces. Then we have chosen the summative evaluation process as will be discussed in Chapter 11.

For the workload measurement we have chosen the NASA-TLX method to compare with the performance measurements provided by the summative evaluation approach. The choice of NASA-TLX was based on the facility of applying this test as well as of interpreting the results.

Chapter 11

Usability Tests

11.1 Introduction

In this chapter, we specifically report the method used by us to evaluate a mixed reality application in the medical domain. This method was adapted from the testbed application method [15], currently used to evaluate virtual reality applications, to be applied to mixed reality system, in general. The method presented below was tested in the evaluation of a mixed reality maxillofacial surgery system, also developed by us.

Our main objective with this usability test is to identify theoretical and practical basis that explains how mixed reality interfaces might provide support and augmentation for interactive applications. In other words, we wish to establish a relation between the design space principles ([128] and Chapter 5) and the variables tested during the experiments. To accomplish that, we intend to use the results provided by the tests to find a model that allows us to identify the contribution factor of each of the analyzed variables in the final user interaction.

11.2 Testbed application

11.2.1 Study case presentation

The study case we are using in this application is related to the maxillofacial surgery. Patients presenting huge deformations in her/his mandible anatomy frequently should be operated to cut off part of the bones or to append some kind of prosthesis, according to her/his injury.

In this kind of intervention, the surgeon cuts the patient mandible following an imaginary line, as the ones shown in Figure 11.1 Nowadays, this class of surgery procedure takes about 3 hours and the recuperation about 2 months. Despite the fact it is a very traumatic procedure, post operatory problems can expose the patient to a new surgery and another three months recuperation period.

In collaboration with the Stomatology and maxillo-facial surgery service at Saint Luc Hospital, in Brussels, we proposed the development of an application using a mixed reality interface to guide surgeons in this type of interventions, as presented in Chapter 8. The goal of this application is to increase the first surgery success, avoiding a second intervention. To achieve this objective, we are mixing real and virtual images by projecting a virtual guidance path-line on the patient mandible captured video. Then, the surgeon should cut the patient mandible paying attention to follow this path-line and avoiding touching the dental nerves. The time involved in this procedure is not important, but the accuracy is mandatory.

(a) (b) (c)

Figure 11.1: *Three views of the cutting line for mandible surgery: left (a), upper (b), and right (c) views, respectively.*

11.2.2 Apparatus

In mixed reality applications, as well as in virtual reality ones, the features of the input and output devices used affect the user performance during the execution of a task [74], [94]. Attributes like resolution, field of vision

Figure 11.2: *Experimental scenario tools.*

and degrees of freedom should be considered when comparing users' performance.

We have performed our tests in a PC computer with an AMD Athlon 64 3.0Ghz processor, 1GB DDR-800 of RAM memory, a nVidia GeForce FX 5700 graphics card, a 120Gb Western Digital 7200RPM SATA hard drive and with the Ubuntu 5.04 Linux operating system distribution.

To capture the interaction of the user with the mock-up we have used a live video capture system based on a stereo camera from Claron Technology model S60 (Figure 11.3(a)).

The tool used to cut the mock-ups is a mini-driller with controlled speed (Figure 11.3(d)).

The mock-ups used in the experiments are two: a simple one, representing a 3D L geometry (Figure 11.2(b)) and a more complex, the printed 3D mandible (Figure 11.2(c)). Both present a wire passing inside, simulating the dental nerve, were built with the same material (gypsum) and printed using the Z-corp[1] machine.

To detect if the user touched the dental nerve or not during the experiment, we have developed an electronic circuit (Figure 11.2(a)) connecting the wire and the tool. It emits a sound alarm and a visual feedback (using a LED) when the tool touches the wire.

[1]Z-corp Website: hhttp://www.zcorp.com

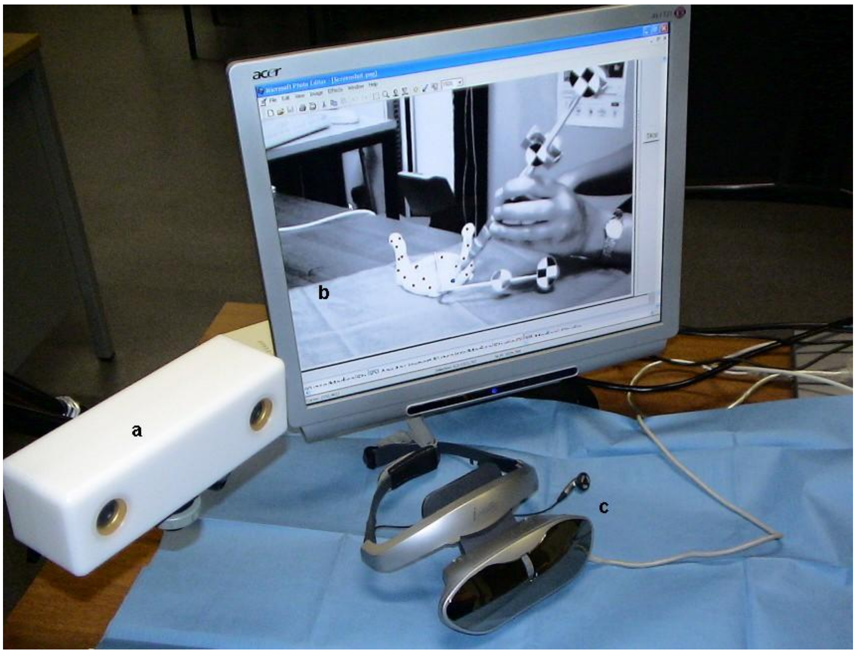


Figure 11.3: *Experimental setup: displays and stereo camera*

11.2.3 Scenarios considered

For the evaluation tests, we considered three different scenarios:

- No guidance scenario
- Virtual guidance scenario
- Augmented guidance scenario

Table 11.1 summarizes the 4 proposed scenarios. For technical limitations, at this stage of the work we are just considering the *L* mock-ups. Despite we also made some tests using mandible mock-ups, these are not yet sufficient to correctly evaluate interaction in our system.

No guidance scenario

In this scenario none computer guidance is provided to the user. Three different views of the path-line (Figure reffig:problem) are shown to the user in a paper printed version and the user should cut a real mandible

Scenario	Kind of visualization	Kind of guidance	Kind of display
1	none	(1)	—-
2	virtual	(3)	screen
3	augmented	(2)	screen
4	augmented	(3)	screen

Table 11.1: *List of the scenarios used during the tests sessions. The numbers in the kind of guidance column correspond to: (1) no guidance; (2) the visualization of the path-line with distance indicator; (3) the visualization of the path-line with its distance indicator plus the internal structure (dental nerve) visualization with its distance indicator.*

just checking this piece of paper. This scenario corresponds to the real currently situation and will serve as reference scenario to measure the value added by the guidance provided by the other scenarios.

Virtual guidance scenario

In this scenario the 3D virtual guidance is provided in the screen, as illustrated in Figure 11.4. When the user touches the correct position (on the real object) the virtual sphere that represents the extremity of the tracker tool-tip becomes green. Another dynamic sphere attached to the tracked tool indicates the distance from the tool-tip to the internal structure. Additional information such as where is placed the internal structure (dental nerve) is provided by using transparency visualization effect. Manipulations with the virtual object such as rotation, scale and zooming are allowed at any time.

Augmented guidance scenario

The augmented scenario presents two alternatives according to which information are displayed in the scene. In the first variant the path-line is projected over the video image. When the user touches the correct position (on the real object), the projected sphere representing the tracked tool-tip becomes green (Figure 11.5). To indicate if the tool is in front of or behind the mandible mock-up we change the intensity of the green sphere. If the tracking tool is behind the mock-up and consequently occluded in the viewed image, the visualization of a semi-transparent sphere guarantees a kind of guidance for the user. The same principle is used in rela-

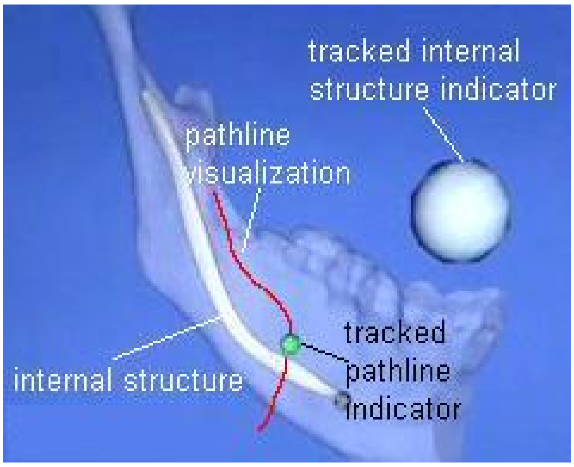


Figure 11.4: *Example of virtual guidance scenario: path-line to be tracked, tracked distance from the tool-tip to the internal structure, visualization of the path-line and internal structure (dental nerve) as well.*

tion to the path-line. As shown in Figure 11.5, the draw representing the path-line is dotted when it indicates an occluded area (behind the visible real surface). This kind of representation was suggested by the surgeon once dotted lines have been currently used in other procedures such as pre-operative planning.

The second variant of the augmented guidance scenario counts with two more additional information: the distance indicator from the tool-tip to the internal structure (dental nerve) and the projection of this internal structure on the video image. In a previous design scene proposed in Part I, Figure 6.8, the distance indicator was displayed as a progressive bar located at the bottom of the image. Trying to drive all the user attention to the task to be accomplished, we decided to change our distance indicator strategy. Now, when the marker is recognized by the system, two spheres are displayed in the screen: one attached to the tool-tip; and other, bigger, displayed in the middle of the tool (see Figure 11.6). The bigger one indicates the distance from the tool-tip to the internal structure using three colours:

- Gray means go ahead;
- Blue means pay attention (i.e. you are less than 3mm from the nerve);

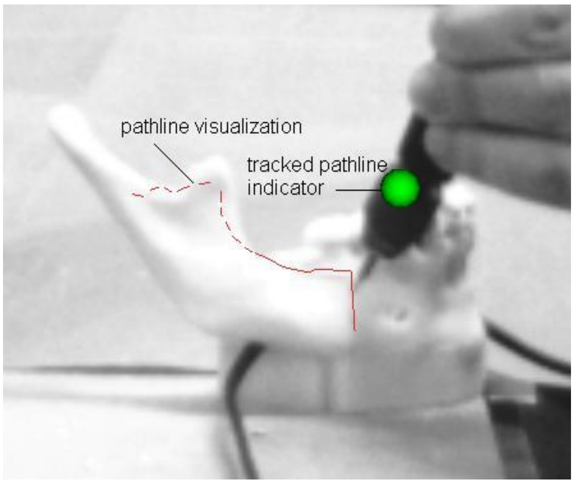


Figure 11.5: *Augmented guidance scenario: path-line visualization and tracked path-line indicator*

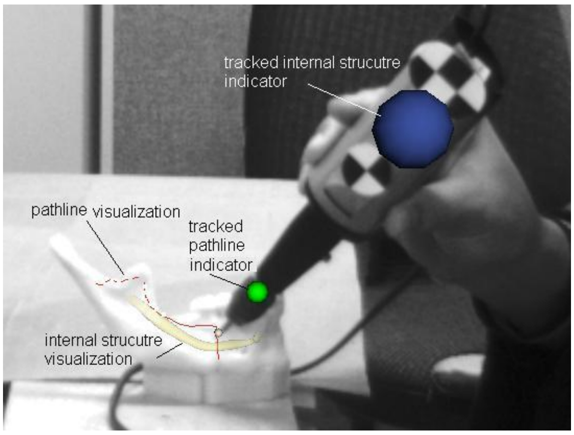


Figure 11.6: *Augmented guidance scenario. Path-line and internal structure (dental nerve) projection; tracked path-line position; and distance indicator from the tool-tip to the internal structure.*

- Red means stop (i.e. you will probably touch the nerve)

During the pre-tests (see Section 11.4.2) such values are re-calibrated.

11.3 Evaluating the system usability

11.3.1 Hypotheses

In this section we describe the hypotheses we intend to proof with the tests. We considered only single variable hypotheses, all of them relevant in the evaluation of the continuity criteria. Each hypothesis is based on the user performance which, in our case, is defined as the *precision for the task execution.*

H1 - The kind of guidance will influence in the task performance.

- H1.1 - precision of task execution will increase according to the volume of guidance (information) given
  - H1.1.1 - none guidance will provide the worst task precision.
  - H1.1.2 - pathline guidance provides lower task precision than pathline + internal structure + distance indicator.
  - H1.1.3 - pathline + internal structure + distance indicator provides the best task precision.
- H1.2 - time of task execution will increase according to the volume of guidance (information) given
  - H1.2.1 - none guidance has the shortest time of task execution.
  - H1.2.2 - pathline guidance has shorter time of task execution than pathline + internal structure + distance indicator guidance.
  - H1.2.3 - pathline + internal structure + distance indicator guidance has the biggest time of task execution.
- H1.3 - task performance (global) will increase according to the volume of guidance (information) given
  - H1.3.1 - none guidance provides the worst task performance.
  - H1.3.2 - pathline guidance provides lower task performance than pathline + internal structure + distance indicator.
  - H1.3.3 - pathline + internal structure + distance indicator provides the best task performance.

H2 - The kind of visualization used will influence in the task performance.

- H2.1 - precision of task execution will increase according to the kind of visualization
  - H2.1.1 - 3 views provides the worst task precision
  - H2.1.2 - virtual view provides lower task precision than augmented view
  - H2.1.3 - augmented view provides the best task precision
- H2.2 - time of task execution will increase according to the kind of visualization
  - H2.2.1 - 3 views has the shortest time of task execution
  - H2.2.2 - virtual view has shorter time of task execution than the augmented view
  - H2.2.3 - augmented view has the biggest time of task execution.
- H2.3 - task performance (global) will increase according to the kind of visualization
  - H2.3.1 - 3 views provide the worst task performance
  - H2.3.2 - virtual visualization provides lower task performance than augmented visualization
  - H2.3.3 - augmented visualization provides the best task performance.

11.3.2 Independent variables

Independent variables are the experiment variables that are manipulated to generate different conditions to compare. In general, good examples of independent variables in interfaces are interface style and help level [23]. In virtual reality environments, examples are the size of the selectable objects and the visualization strategy used. In this mixed reality experience, we established the following variables to test the hypotheses:

Kind of guidance. Used to prove the first hypothesis, concerns the strategy used to help the user to better accomplish the task (to cut a mock-up following a precise path-line).The available guidance possibilities are: (1) no guidance at all; (2) the visualization of the path-line with a visual feedback of the tracked line; (3) the visualization of the path-line with a visual feedback of the tracked line and the internal structure (dental nerve) visualization + the visual feedback of the tracked internal structure;

Kind of visualization. This variable allows the direct test of the second hypothesis. The three possible visualization strategies considered are: (1) three paper printed views of the path-line over the mandible; (2) the virtual visualization; (3) the augmented visualization;

11.3.3 Dependent variables

Dependent variables are the measures taken as the performance indicative or level of acceptance of the technique by the users. These measures can be objective, as the time taken to accomplish a task, or subjective, collected from post-test questionnaires answered by the subjects.

The dependent variables used in our experiment were:

Precision on task execution. We measure the user precision by calculating the distance between the path-line and the line drawn by the user;

Time to perform the task;

Task performance: Precision/Time;

Workload. Based on the result given by the subjective pos-test NasaTLX (see Appendix A).

11.4 The experiment

11.4.1 Task description

The task is always the same for the three scenarios described before in Section 11.2.3 and consists in cut the mock-up reproducing as best as possible the indicated trajectory without touching the internal hidden structure (dental nerve). There is no limit time to accomplish the task.

Before to start the execution of the main task, the user has free time for training, consisting in performing sub-tasks that allow the user to understand how the system will provide guidance in each scenario. In the same time, we can decompose the task complexity and isolate variables involved in the final interaction. Then, the main task was decomposed into two sub-tasks: user perception calibration and motor calibration.

User perception calibration

In the perception calibration task the user should select over the mock-up surface arbitrary points indicated in the virtual or augmented scene. This task is done once without guidance (i.e. the user perceives the point in the virtual or augmented scene and performs the task) and then with guidance (i.e. when the user is pointing the good position, a small sphere representing the tracked tool-tip becomes green). This task is performed in both guided scenarios: virtual and augmented. In the augmented scenario, the tracked tool-tip becomes transparent and the path-line visualization is dotted to indicate the surface occlusion. Time and precision (i.e. performance) for the user task accomplishment are computed.

User motor calibration

In the motor calibration procedure, the user should cut a trajectory drawn over the mock-up surface. In this case, the mock-up resistance is the same used to execute the main task. Time and precision (i.e. performance) for the user task accomplishment are computed.

11.4.2 Pre-Tests

Based on the three scenarios stated in Section 11.2.3 we have designed 12 different scenarios (as illustrated in Table 11.2) which were tested during the pre-tests with 5 volunteers. As a result of these preliminary tests, some visualization parameters were adjusted, such as, level of object transparency, position of the distance indicators in the scene and the values range used to track distances (i.e. path-line and internal structure distance). For the tracked path-line distance we have found the range of $2mm$ ($1mm$ to the left and $1mm$ to the right) a good one for setting colours guidance. For the tracked internal structure distance we have delimited the range from $2mm$ to $5mm$ as the *pay attention zone* and less than $2mm$ as the *stop zone*. In this case, the surgeon has suggested giving the *red* visual feedback before touching the internal structure once it would be too late in the real scenario.

Scenario	kind of visualization	Kind of guidance	Kind of mock-up	kind of display
1	none	(1)	L	—
2	none	(1)	Mandible	—
3	virtual	(3)	L	screen
4	virtual	(3)	Mandible	screen
3	virtual	(3)	L	goggles
4	virtual	(3)	Mandible	goggles
5	augmented	(3)	L	screen
6	augmented	(3)	Mandible	screen
7	augmented	(3)	L	goggles
8	augmented	(3)	Mandible	goggles
9	augmented	(2)	L	screen
10	augmented	(2)	Mandible	screen
11	augmented	(2)	L	goggles
12	augmented	(2)	Mandible	goggles

Table 11.2: *List of the scenarios used during the pre-tests session. The numbers in the "kind of guidance" column correspond to: (1) no guidance; (2) the visualization of the path-line with a visual feedback of the tracked line; (3) the visualization of the path-line with a visual feedback of the tracked line and the internal structure (dental nerve) visualization + the visual feedback of the tracked internal structure.*

11.4.3 Subjects and procedure

Sixteen users volunteered for experience the tasks: 14 males and 2 females, with a mean age of 24 years. Each participant tested only one of the designed scenarios, while each scenario was tested 4 times.

Before each session, the mock-up should be calibrated and registered. Session length varied according to the index of difficulty given by the kind of guidance, kind of visualization, and the level of the individual motor coordination of each user.

The subjects were tested individually. They began the session with a familiarization period and performing some simple tasks (such as perception calibration and motor calibration) taking into account the kind of scenario in use. Considering the virtual 3D scenario, users were carefully informed how to use the manipulation commands (i.e. translation, rotation and zoom), but they were left free to use it or not during experimental trial.

At the end of the test, each user was asked to answer the NasaTLX questionnaire as illustrated in Appendix A [50].

At the end of the test, each user was asked to answer the NasaTLX questionnaire, .

11.4.4 Logging

For each user and each task, we are saving the time spent to accomplish the task, the original path-line and the path-line drawn by the user.

11.4.5 Evaluation strategy

We are focusing on a strategy to evaluate the implementational approach taking into account the design principles addressing the continuous interaction in MR systems conforming summarized in Table 11.3.

The continuity properties were detailed in Chapter 4.2 as our evaluation criteria for mixed interaction systems and the design principles (2nd column in Table 11.3) were described in Part I.

During the application development process many strategies of implementation are possible (Table 7.1). It is probable (especially in the case of complex and large applications) that authors would need information related to the design principles (described in Part I) to guide the implementation process. Some strategies are the same those applied to classic GUI and VR systems (for instance perceptual cues, insertion context) and others are specific to MR systems (for instance registration, calibration and tracking process, transform type, kind of augmentation).

Once concrete implementational decisions has been done (as describe in the 3rd column Table 11.3), the related design principles can be identified and evaluated.

To validate and identify the influence of the design principles for the final interaction we propose an evaluation strategy based on simulation and on user's tests, as summarized by us in the 4th column of Table 11.3. Basically the simulation technique can be used to estimate system reliability which is related to technological and technical factors. For instance the technology used to track objects and the registration method used to align real and virtual objects. However, the simulation technique is not able to capture the user focus while performing a task. User tests may evaluate other aspects related to final interaction performance and help to identify the impact of independent variables (such as kind of display, kind of visualization, kind of guidance, kind of mock-up) while assessing the continuity criteria. Users tests can indicate subjective preferences and workload indexes as well. Next section describes how continuity properties have been assessed through user usability tests.

Continuity Properties	Related design principles	Implementation strategies	Evaluation strategies
Perceptive	Augmentation of user perception		
	Interaction focus	Choice of visualization types (perceptual cues)	Kind of visualization
	Insertion context	Choice of output device/tools	Zone of insertion corresponds to the user's task focus?
	Spatio-temporal links	Services to support MR (Layer 2 in Table 7.1)	Reliability of the system
Cognitive	Augmentation of user action		
	media	kind of guidance: text, graphic, 2D, 3D, video, etc. (Layer 3 and 4 in Table 7.1)	workload (guidance + subjective test)
Functional	Augmentation of user interaction		
	Transform type	Choice of input interaction modality	Level of interaction familiarity
	Connection type	Registration and tracking techniques	Level of connectivity (manual, automatic, semi-automatic)

Table 11.3: *Evaluation strategies and their related design principles and implementaional strategies.*

11.5 Results

Some much known metrics has been used to compute performance and movement time in many experiments. For instance, the Fitts' law [36] is an effective method of modeling rapid, aimed movements, where one appendage (like a hand) starts at a specific start position, and moves to rest within a target area. The Fitts' law can be used to assist in the design of user interfaces, to predict the performance of operators using a complex system, to assist in allocating tasks to operators, as well as to predict movement times for assembly line work.

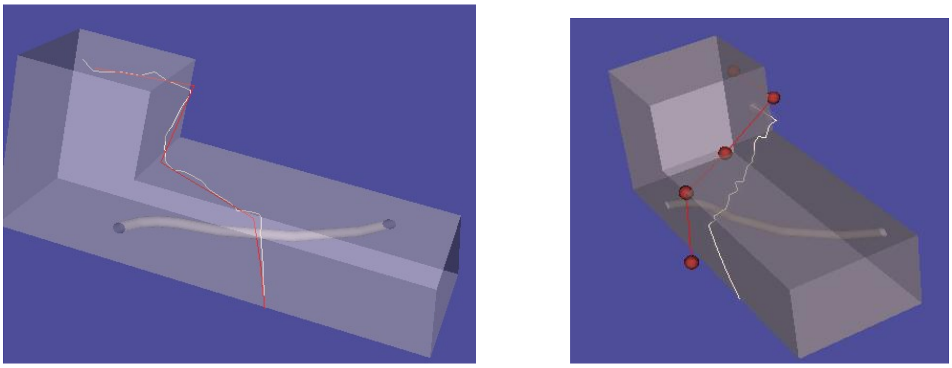
However, Fitts' law predicts movement in just one dimension. Fitts' original experiments tested human performance in making horizontal moves toward a target. Both the amplitude of the move and the width of the terminating region were measured along the same axis. It follows that the method is inherently one-dimensional. So, when dealing with two or three dimensional target acquisition tasks, new interpretations of target width must be considered. Another major deficiency is the absence of a consistent technique for dealing with errors.

Researchers have developed a method to handle errors, but it has been largely ignored because of its complexity. In summary, using such approach for any particular design problem we are working on the Math won't tell us which part of the parameter space we are operating in. We are not interested just in find a number representing the final user performance; we want to be able to decompose such result by identifying all parameters involved in the final performance.

11.5.1 Computing precision

For each tested scenario, the Euclidean distance between the reference and the final path-line performed by the user were calculated, providing the final user precision. Both path-lines (i.e. reference and final, Figure 11.7) were acquired in the same way, i.e., digitalizing points over the surface trajectory with the tracked tool. The task accomplishment involves the cutting of the real mock-up. However, to calculate the Euclidean distance between the reference and the user performed path-lines, we are first projecting the line designed by the user on the mock-up surface to fit the requirements of the reference line, also designed on the mock-up surface.

Our results are expressed in terms of the mean distance and standard deviation calculated for each scenario. We are not considering the time variation because the precision of the task is more important than time



(a) (b)

Figure 11.7: *Result of user interaction (white line) for variables configuration given by scenario 2 (left) and by scenario 1 (right).*

for our tests. Smaller is the distance, better will be the interaction technique. Analyzing the chart of Figure 11.8, we conclude that scenario 2 has provided the best result.

As we are not using all functionalities that could be provided by AR systems, such as the use of a head mounted display, we are not interested in a comparison between 3D visualization and AR. Instead, we are more interested in identifying how the independent variables involved in these two types of techniques (3D and AR) are influencing the interaction. Despite the 3D visualization offers better spatial perception, AR visualization presents more realistic representation of the real scene by projecting virtual objects over the video live images.

11.5.2 Detecting independent variables influence

Once the final user performance was calculated for each scenario, we can compare the scenarios 2, 3 and 4 with scenario 1 (neither guidance nor visualization) and extract the performance gain of each one. Such gain is attributed to the influence of the guidance and visualization on the task accomplishment.

The performance gain given by the visualization can be deduced from the assessed perception calibration task (see Section 11.4.1). We assume that the remaining performance gain is divided between user motor performance and guidance. User motor performance was assessed during the motor calibration tests (see Section 11.4.1) and we assume it as a constant

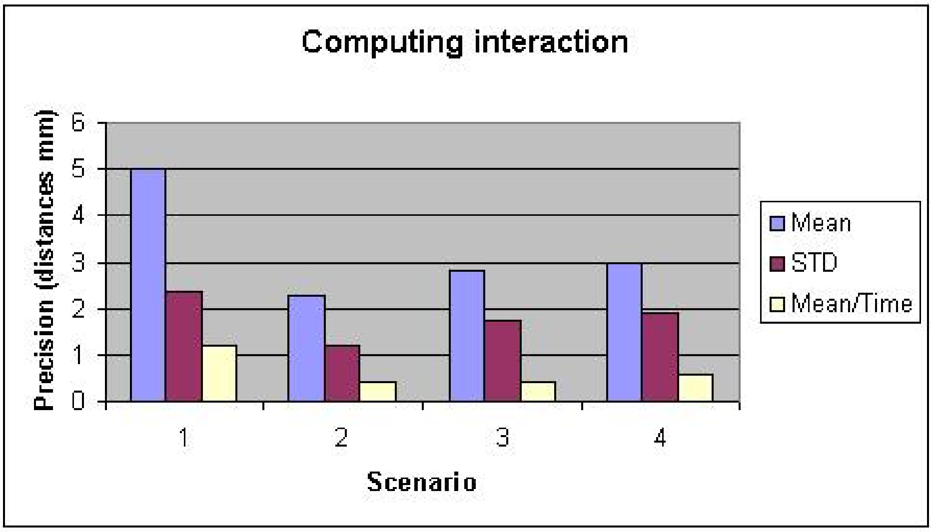


Figure 11.8: *Precision of the user interaction in the 4 proposed scenarios.*

for all scenarios. We found a mean of $1.31mm$ with a deviation standard of $0.47mm$.

Concerning the workload, we tried to verify its relation with the visualization and guidance variations. From Figure 11.9, we can conclude that the higher workload rating coincide with the kind of guidance 3, where the user has more guidance information. The workload rating was normalized to be plotted together with the other two assessed variables.

As the users were not experts in the application domain and very poor information was provided in scenario 1, users were not very aware about the path to follow and less mental effort was spent for the task execution in this scenario. It reflects the low workload level found in scenario 1 (Figure 11.9).

11.5.3 Defining the model

Once we have created our data set we want to identify how the captured variables are related. To do that we have used the Spearman's test where significant correlation was found between precision and visualization ($p = 0.0241$), and between precision and guidance ($p = 0.0306$). No significant correlation was found between visualization and guidance.

Based on such analysis the user interaction can be expressed by the

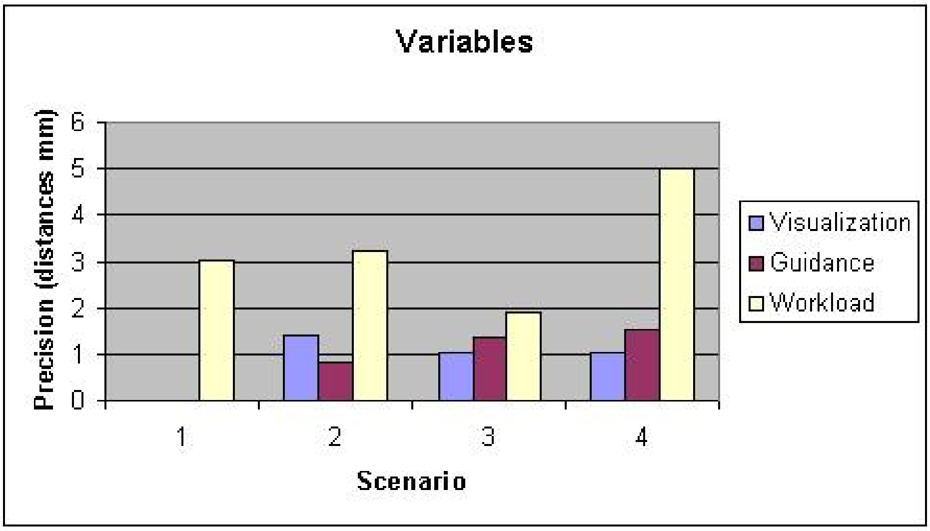


Figure 11.9: *Independent variables influence in each scenario.*

following expression:

$$I \approx \alpha V_i + \beta G_j + \gamma + \epsilon \tag{11.1}$$

where the indexes i, j represent the values that the variables can assume:
V_i = kind of visualization (i= 1, 2, 3)
G_j = kind of guidance (j= 1, 2, 3)
γ = constant for the user motor calibration
α, β = contribution factors for the final interaction I
ϵ= error factor

To solve such linear system we applied a SVD (Singular Value Decomposition) method which uses the least squares and pseudo-inverse computation [90].

Least square is a mathematical optimization technique that attempts to find a *best fit* to a set of data by attempting to minimize the sum of the squares of the ordinate differences (called residuals) between the fitted function and the data.

In mathematical terms, we want to find a solution for the equation

$$[\underline{A}^T \ \underline{A}]\, C = \underline{A}^T I \tag{11.2}$$

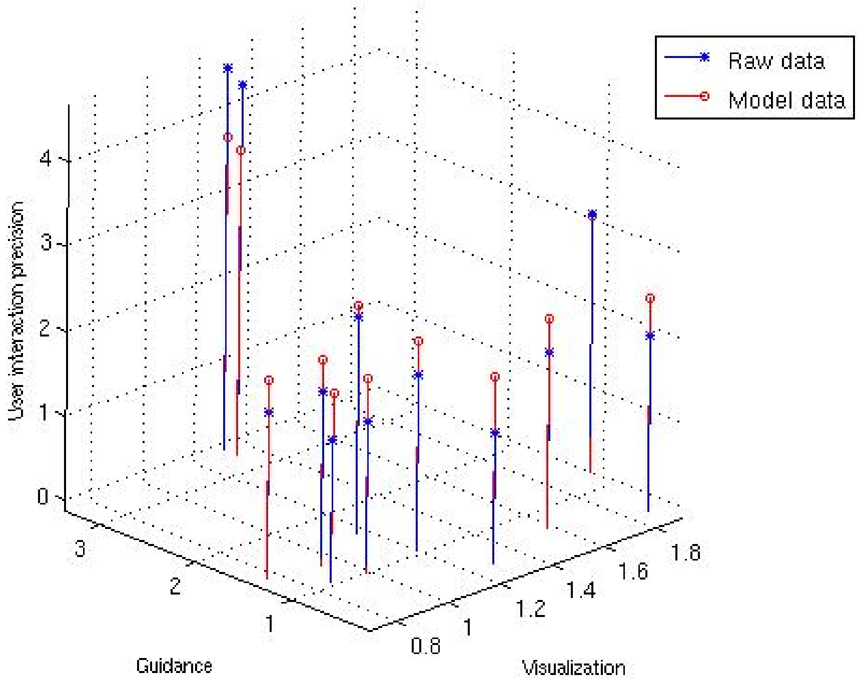


Figure 11.10: *Approximation given by the proposed model.*

where $\underline{A}$ is an m-by-n matrix of variables (with $m > n$) and C and I are n- respectively (vector of constants α, β,γ) m-dimensional (vector of final user interactions) column vectors.

We have found α = 0.6, β = 0.5, γ = 1.31 (value previously measured in Section 11.4.1)
Figure 11.10 plots the real measured data versus those given by the proposed model. We can observe a good approximation between the data set.

Consequently, we can conclude that visualization has more influence than guidance for our experiment. It reinforces the results shown in Figure 11.8, where the virtual visualization with the same amount of guidance has provided better precision.

11.5.4 Discussion

By analyzing the interaction between our assessed variables, it is apparent that the workload augments according to the amount of guidance provided. As the workload index was measured with very subjective tests, more concrete and objective evaluation should be envisaged to assess such influence. Previous work [140] have demonstrated physiological responses to audio and video degradations can be detected and that they do not always correlate with subjective results. In this way, analysis of biological signals to capture user' state (such as stress) could be combined with the NasaTLX results. On the other hand, the precision of the interaction technique varies according to the guidance and visualization choices. In this way a model to describe the final user interaction based on such variables influence was proposed. As a result, future interactive mixed systems can use such model to estimate interaction at early stages of system development or adapt the model to better fit your own system variables.

As a result we have found that visualization has more influence than guidance in the final user interaction. Considering the final evaluation strategy proposed in Table 11.3 we could say that perceptual continuity (i.e. kind of visualization provided) has been more important than cognitive continuity(i.e. kind and amount of guidance provided). However the most significant modeled factor has been the motor calibration which represents the functional continuity property.

At the moment we have just tested our model taking into account the error factor found (≈ 0.4) to best fit our first data set. Performing more user tests will augment our data set allowing us to verify the reliability of the proposed model.

11.6 Conclusion

In this chapter we studied the user reaction in relation to IGS applications using different kind of augmentation. Our main objective with this usability test were to identify theoretical and practical basis that explains how mixed reality interfaces might provide support and augmentation for interactive applications.

To accomplish that, we presented the strategy we developed to evaluate the usability of mixed reality systems, precisely, the use of mixed reality in a maxillo-facial surgery. First of all, we briefly presented the problem we are trying to solve and the application developed to be used as a testbed application. Basically, in this chapter we focused on the final user interaction measurement and the errors inherent from the computer system and devices used were previously discussed in Section 9.

We used the results provided by user tests to find a model that allows the identification of the contribution factor of each independent variable in the final user interaction. With that future interactive mixed systems can use such model to estimate interaction at early stages of system development or adapt the model to best fit your own system variables. It will contribute to reduce the time for user interaction estimation.

Chapter 12

Conclusions and Perspectives

This Thesis has addressed the design, implementation and evaluation of MR systems in computer-guided surgery applications. The key point driving the links between the cycles of system development is the analysis of interaction in terms of continuous properties.

Our definition for continuous interaction and its properties (see Chapter 4) are based on the perceptual and cognitive aspects introduced by [45] and [30]. However we go forward extending the definition and identifying what design properties (see Chapter 5) are involved as well as what are their impact when assessing the continuous interaction (see Chapter 11).

The main objective of the Design part was to provide a conceptualization of mixed realities that provides a means to identify the key design parameters involved in the continuous interaction. As results of this study we have proposed a design space (DeSMiR) for continuous interaction in mixed reality applications.

The goals of this design space is to help the designers to manage a large number of options for the mixed interaction design under development of MR systems; to have a better understanding of how to design MR systems in an intuitive and effective manner; to identify spatial, temporal and focused layouts of the mixed interaction space under development for verification purposes; to envision future interactive mixed systems as well. We go further developing series of usability scenarios for a maxillofacial surgery application to demonstrated how the design principles can be used to characterize the different properties of continuous interaction.

As contributions of the Implementation part we developed two approaches to support mixed reality applications: markerless and marker-based. In both approaches we suggest to use a real time rigid registration method based on the distance map computation to align real and 3D objects. Validation of both approaches were also performed.

The achieved spatial accuracy (rotation and translation) given by the markerless approach are encouraging for real time applications. Future work in this direction includes the reduction of noise influence in the cloud of points by using bounding boxes to automatically shaping themselves around the object of interest, limiting the number of points. Another work consists in improving registration results for more complex objects. This will deal with enhancing cloud of points reliability as well as protecting the algorithm from falling in local minima.

The marker-based approach was used to develop a computer-guided application for maxillo-facial surgery and its accuracy was compared with only one another system, taking account the few availability of systems using the same or comparable techniques.

Although a highly accurate tracking system is used (the MicronTracker tracks tooltip to an accuracy of $\approx 0.20mm$), tracking (i.e. taking account the user interaction while picking points) has been identified as the most significant source of errors at present. Based on the error analysis a number of areas for potential improvements in system performance have been identified. One of these is to define patterns configurations: in point-based registration errors can be reduced by spreading out the points in all object surfaces. In this sense an automatic initialization process can be also envisaged to assure that all performed registrations have the same initialization transforms applied.

The current accuracy of our system is thought to be sufficient for a wide range of medical procedures such as planning craniotomy sites (i.e. for the neurosurgery) and osteotomy procedure (i.e. for maxillo-facial surgery). If a less accurate tracking system were employed, the range of applicable procedures would be reduced.

As we are considering that our tracked object and the camera are fixed in the environment we did not account for the system latency. Such validation aspects should be considered in future versions of our system for tracking the mandible movement.

Although this analysis is specific for our system, the same methodology can be applied to other augmented reality systems, enabling real/virtual alignment errors to be controlled and reduced significantly.

Both approaches were integrated into the Medical Studio open-source platform. It has contributed to demonstrate the flexibility and usability of the platform in supporting the content design for MR applications. Future works in this direction are focusing on the integration of these MR techniques as components in the SIMILAR multimodal platform - OpenInterface[1]. This platform allows the rapid development of multimodal applications by using the ICARE [13] fusion and fission components. With that, we could think to include other kind of guidance, for instance sounds, to indicate when the surgeon is touching the right position. Other interaction modalities such as speech and gestures could be also envisaged. In this sense, the rapid prototyping of new scenarios will facilitate the validation of other design principles proposed in this Thesis as well as the extension of them.

In the evaluation phase we have performed a series of usability tests to assess variables involved in this kind of mixed interaction. We proposed a model to describe final user performance and find the contribution factors of each variable for the continuity properties of the user interaction. With that we are able to identify why and how mixed reality systems can improve human performance in computer guided tasks relative to other media.

On the basis of our work and research we will argue the methodological approach presented in this work could be applied for the design and evaluation of several kinds of mixed reality applications; especially those in which smooth connections and interactions between virtual and real worlds are needed. It can be the case of driver applications, pilot simulations and computer-guided surgery systems.

However such approach is not exhaustive and many open issues still exist. Other implementational and evaluation strategies can be envisaged according to new techniques and technologies.

12.1 Perspectives

Supporting the development cycle of MR systems is still in its infancy. Rapid evolution in technology and implementational strategies to support MR applications are leaving a gap in the field of design and evaluation of human interaction in such class of systems. Several extensions involving design, implementation and evaluation of mixed reality systems are worth investigating.

[1]Website:http://www.openinterface.org

Design

As we discussed supporting the system design phase could anticipate interactions problems at early stage of development. For his reason development of tools and methodologies which can facilitate such task remains a relevant field of research, specially when we are talking about MR systems. Such systems provide the opportunity to design application which supports the continuity of user interaction. In this work we have identified some of the design principles involved in the continuous interaction applied to the computer-guided surgery. However extensions focusing on other kinds of applications such as edutainment could be also envisioned.

Future works could be focused on adaptation of software engineering techniques to decrease the distance existing now between design and implementational phases. Such adaptation should be mainly focused on how to capture and represent virtual and real interactional aspects.

Interaction modeling languages like dialoging nets, attributed grammars, state-transition diagrams, dialog templates, statecharts and petrinets could be investigated to specify the interaction behavior in such class of system.

Implementation

While the implementational strategy presented in this thesis achieve accurate results on controlled marker-based environment, the uncontrolled environment case is still problematic. The uncontrolled case involves the objects tracking in an environment where is not possible to place markers, form of tracked object is changing and the precision is a relevant factor.

In some cases of navigational surgery, shifts and deformations of soft tissues occur during surgery because of mechanical factors, physiological motion, swelling, or hemorrhage. These changes may displace organs or their tissue components to such a degree that pre operatively acquired image-based 3D models cannot be registered with the patients actual anatomy. In this situation either partial correction of the 3D model or full volumetric image update is necessary. Even limited revision of the original images requires some clues about the changes taking place during surgery. While some useful information can be obtained from prior knowledge and experience-based models of deformations, only intra-operative imaging can provide true information. These updated positional data then can be used to modify the original volumetric images, using more elaborate computer-based methods. Thus, intra-operative images are a key component in maintaining anatomic integrity.

Developing a dynamic registration for deformable objects using finite elements could be a direction to be investigated [35] to solve such issue.

Evaluation

Interaction evaluation in MR systems is a complex task. As we have discussed in this Thesis a MR system involves many source of objects in a mixed space of interaction and one of the main goals of mixed and augmented reality systems is to provide additional information to help the user accomplish the task in a continuous and natural way. So far, development of techniques to track the user attention focus in a mixed interaction space remains an open issue. In this Thesis we started this step by providing a strategy based on usability scenarios to measure objectively the user performances (we assume that continuity and focus of attention relate directly to performances). However other approaches could increase the results. For instance a 3D eye tracking able to detect in deeply and overlapped objects could be used to track the user focus with more accuracy. It will allow to have more control about the user interaction focus as well as to compare the results with the usability evaluation provided in this work.

Other research field consists in detecting and measuring the workload. As the workload index was measured with very subjective tests in this Thesis, more concrete and objective evaluation should be envisaged to assess the workload in MR systems. Previous work [140] have demonstrated physiological responses to audio and video degradations can be detected and that they do not always correlate with subjective results. In this way, analysis and fusion of biological signals to capture user state (such as stress) could be combined with the NasaTLX results.

Part IV

Appendices

Subjective Test

The evaluation you are about to perform is a technique developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. A set of six rating scales will be display (Figure 12.1). Please read the descriptions of the scales carefully (Figure 12.2). If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

You will evaluate the task by marking each scale at the point which matches your experience.

After this step you will be presented with a series of pairs of rating scale titles (for example, Physical Demand vs. Frustration conforming shown in Figure 12.3) and asked to choose which of the items was more important to your experience of workload in the task(s) that you just performed. Each pair of scale titles will appear separately on the screen. Select the Scale Title that represents the more important contributor to workload for the specific task(s) you performed in this experiment.

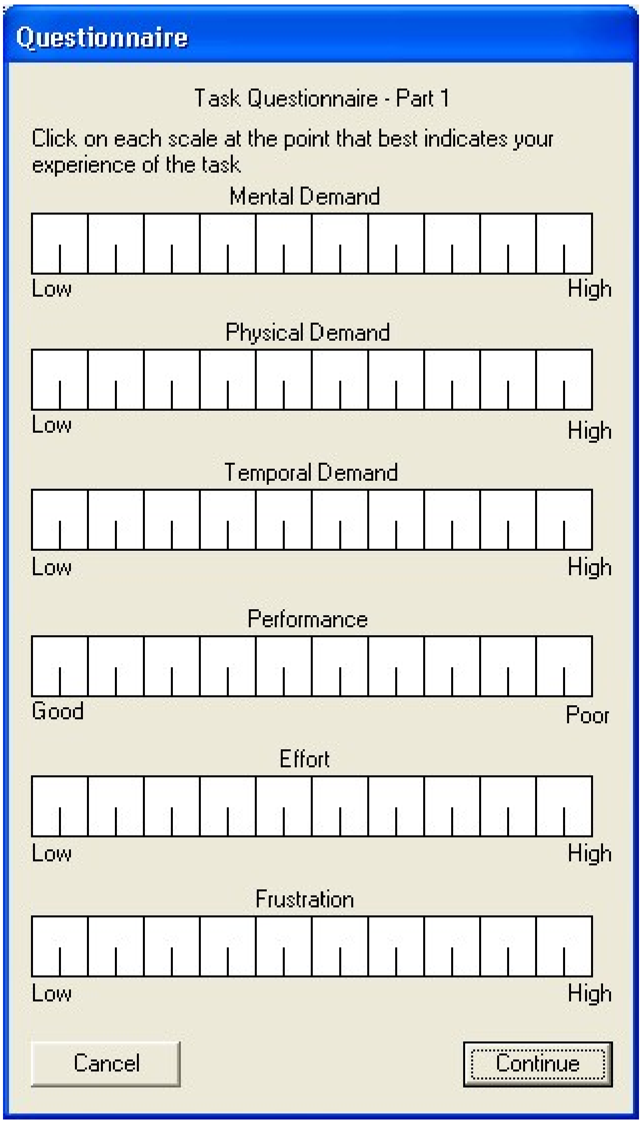


Figure 12.1: *Scaling variables for NasaTLX subjective test.*

Title	Endpoints	Descriptions
Mental Demand	Low/High	How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Figure 12.2: *NASA Task Load Index (NASA-TLX) rating scale definitions [50]*
.

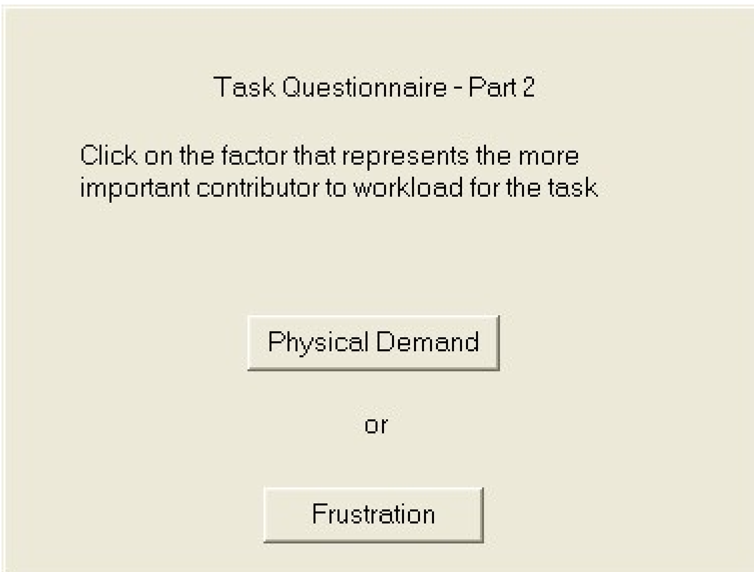


Figure 12.3: *Choosing contributors factors.*

Publications

European Commission Deliverables for the SIMIILAR Network of Excellence - FP6-507609

Daniela Trevisan, Luciana P. Nedel, Jean Vanderdonckt and Benoit Macq. Usability evaluation method for mixed reality systems in surgery. In Deliverable 69: Results of iterative framework testing, Laila Dybkjr (eds) SIMILAR Network of Excellence - FP6-507609. December 2005. To be published at www.similar.cc.

Daniela Trevisan, Jean Vanderdonckt and Benoit Macq, Pablo Lamata de la Orden, Samuel Rodrguez Bescs, Enrique J. Gmez Aguilera, Chapter 5 Usability evaluation issues in mixed reality systems in surgery. In Deliverable 18: Current practice description and evaluation: Towards a framework for usability evaluation, Laila Dybkjr (eds) SIMILAR Network of Excellence - FP6-507609. 2005. Available at www.similar.cc. pp. 21-25.

Daniela Trevisan, Jean Vanderdonckt and Benoit Macq, Pablo Lamata de la Orden, Samuel Rodrguez Bescs, Enrique J. Gmez Aguilera, Chapter 5: State-of-the-art in mixed reality systems in surgery. In Deliverable 16: Usability Evaluation Issues in Natural Interactive and Multi-modal Systems - State of the Art and Current Practice, Laila Dybkjr (eds) SIMILAR Network of Excellence - FP6-507609. December 2004. Available at www.similar.cc. pp. 26-38.

Daniela Trevisan, Monica Gemo, Jean Vanderdonckt and Benoit Macq. Description of an Image Guided Application Using Contextual Focus Driven Interaction. In Deliverable 17: Descriptions of SIMILAR Applications, Laila Dybkjr (eds) SIMILAR Network of Excellence - FP6-507609. October 2004. Available at www.similar.cc. pp.58-67.

Demos Session and Workshops

A. Benoit, L.Bonnaud, A. Caplier, P. Ngo, L. Lawson, D. G. Trevisan, V. Levacic, C. Mancas, G. Chanel. Multimodal Focus Attention Detection in an Augmented Driver Simulator. eNTERFACE'05: The SIMILAR NoE Summer Workshop on Multimodal Interfaces, pp.34-43 July 18 - August 12, 2005, Facult Polytechnique de Mons, Belgium. Available at www.enterface.org

Trevisan, Daniela G., Nicolas, V., Olsweszki, R. and Macq, B. Intra-Operative guidance. Demonstration session during " 100 ans de Stomatologie et 25 ans de Chirurgie Maxillo-Faciale " at Clinique Universitaires Saint Luc, Brussels, 25 Mai 2005.

Trevisan, Daniela G., Gemo, M., Vanderdonckt, J. and Macq B. Focus-based design of mixed reality systems, Third International workshop on TAsk MOdels and DIAgrams, Philippe Palanque, ACM Press, New York, NY, USA, 2004, pp.59-66.

Proceedings of the international ACM - IUI-CADUI'04 Workshop on Exploring the Design and Engineering of Mixed Reality Systems, Dubois, E., Gray, P., Trevisan, D., Vanderdonckt, J. (eds), also available at ceur-ws.org/Vol-91/, ISSN1613-0073, Funchal, Island of Madeira, January 13th, 2004, 79 pages.

Trevisan, D., Vanderdonckt, J., Macq, B. Analyzing Interaction in Augmented Reality Systems, Proceedings of ACM Multimedia'2002 International Workshop on Immersive Telepresence ITP'2002 (Juan Les Pins, 6 December 2002), G. Pingali, R. Jain (eds.), ACM Press, New York, 2002, pp. 56-59.

Trevisan, Daniela G., Nicolas, V., Delville, P., d'Elzius, Michel C. and Macq, B. Towards Markerless Augmented Medical Visualization. In MICCAI International Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery, Rennes, France 30 September 2004.

Vanderdonckt, J., Limbourg, Q., Michotte, B., Bouillon, L., Trevisan, D. G. and Florins, M. USIXML: a User Interface Description Language for Specifying Multimodal User Interfaces. In W3C Multimodal Workshop, Sophia Antipolis, France, 19-20 July 2004.

Limbourg Q., Vanderdonckt J., Michotte B., Bouillon L., Florins M., Trevisan D., USIXML: A User Interface Description Language for Context-Sensitive User Interfaces, in Proceedings of the ACM AVI'2004 Workshop "Developing User Interfaces with XML: Advances on User Interface Description Languages" (Gallipoli, May 25, 2004), K. Luyten, M. Abrams, Q. Limbourg, J. Vanderdonckt (Eds.), Gallipoli, 2004, pp. 55-62.

Trevisan, D., Vanderdonckt, J., Macq, B., Raftopoulos, C., Modeling Augmented Reality System, Image Guided Surgery Case Study, Proceedings of the 1st Glasgow Context Group Colloquium GCG'2002 "Building Bridges: Interdisciplinary Context-Sensitive Computing" (University of Glasgow, 9 September 2002).

Journal and Reviewed Conference Publication 2005

Trevisan, Daniela G., Vanderdonckt, J. and Macq, B. Conceptualizing Mixed Spaces of Interaction for Designing Continuous Interaction. Virtual Reality Journal, Spring-Verlag, London, 2004, Vol.8 N.2, pp.83-95.

Trevisan, Daniela G. and Lawson, L. Chapter 3: OpenInterface. In the SIMILAR Dreams book, UCL Presses Universitaires de Louvain, 2005. ISBN 2-930344-91-1. pp.43-62.

2004

Trevisan, Designing smooth connections between worlds. In Extended abstracts of the 2004 conference on Human factors and computing systems CHI2004, Session Doctoral Consortium, ACM Press, New York, NY, USA, 2004, pp.1043-1044.

Vanderdonckt, J., Bouillon, L., Chow Kwok Chieu, Trevisan, D. ”Model-based Design, Generation, and Evaluation of Virtual User Interfaces”. In proceeding of the 9th ACM International Conference on 3D Web Technology Web3D Symposium 2004, Monterey, California, USA, April 2004,pp.1273-1277.

Florins, M., Trevisan, D., Vanderdonckt, J. The Continuity Property in Mixed Reality Systems and Multiplatform Systems: a comparative study. In proceedings of 5th International Conference on Computer-Aided Design of User Interfaces CADUI’04, Funchal, Portugal, January, 2004, pp.1268-1272.

2003

Trevisan, D., Vanderdonckt, J. Macq, B. (2003). ”Model-Based Approach and Augmented Reality Systems”, Proc. of 10th International Conference on Human-Computer Interaction HCI International’2003 (Heraklion, 22-27 June 2003), C. Stephanidis (ed.), Lawrence Erlbaum Associates, Mahwah, 2003.

Trevisan, D., Vanderdonckt, J. Macq, B., (2003). ”Continuity as Usability Property”, Proc. of 10th International Conference on Human-Computer Interaction HCI International’2003 (Heraklion, 22-27 June 2003), C. Stephanidis (ed.), Lawrence Erlbaum Associates, Mahwah, 2003;

Trevisan, D., Vanderdonckt, J., Macq, B., Raftopoulos, C., Modeling Interaction for Image-Guided Procedures, Proceedings of International Conference on Medical Imaging SPIE2003 (San diego, 15-20 February 2003), K. Hanson, C.-T. Chen, E.L. Siegel (eds.), International Society for Optical Engineering, 2003. vol.5029 pp. 108-118.

Bibliography

[1] DCMT software. DICOM OFFISComputer Science Institute http://www.dicom.offis.de/dcmtk.php.en/.

[2] Ansi/ieee standard 729-1983. In *Software Engineering Standards*, New York, 1989. IEEE Computer Society.

[3] F. A., H. K., and Z. R. Effective techniques for selecting objects in immersive virtual environments. In *Proc. ACM UIST'96 Symposium on User Interface Software and Technology (UIST)*. ACM Press, 1996.

[4] K. P. Akesson and K. Simsarian. Reality portals. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, New York, 1999. ACM.

[5] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1993.

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

[7] R. Becklen and D. Cervone. Selective looking and the notice of unexpected events. In *Memory and Cognition*, pages 601–608, 1983.

[8] S. Benford, C. Greenhalt, G. Reynard, C. Brown, and B. Koleva. Understanding and constructing shared spaces with mixed reality boundaries. *ACM Transaction on Computer-Human Interaction*, 5(3):185–223, 1998.

[9] F. Berard. The magic table: computer-vision based augmentation of a whiteboard for creative meetings. In *Proceedings of the IEEE Workshop on Projector Camera Systems PROCOMS'03*, 2003.

[10] E. A. Bier, M. C. Stone, K. Fishkin, B. W., and T. Baudel. A taxonomy of see-through tools. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 358–364. ACM, 1994.

[11] J. Blandford, M. Clasen, T. Janik, et al. GTK software. http://www.gtk.org/.

[12] G. Booch, J. Rumbaugh, and I. Jacobson. Unified modeling language user guide. Addison-Wesley Longman, First edition 1998.

[13] J. Bouchet and L. Nigay. Icare: A component-based approach for the design and development of multimodal interfaces. In *Proc. of 11th International Conference on Human-Computer Interaction HCI International*, Vienna, 2004. Lawrence Erlbaum Associates.

[14] D. Bowman, J. Gabbard, and D. Hix. A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424, 2002.

[15] D. Bowman, E. Kruijff, J. LaViola, and I. Poupyrev. *3D User Interfaces - Theory and Practice*. Addison-Wesley, 2005.

[16] S. Boyd. Assessing the validity of swat as a workload measurement instrument. In *Proceedings of the Human Factors and Ergonomics Society 27th Annual Meeting*, pages 124–128, Santa Monica, CA, 1983. Human Factors Society.

[17] T. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the International Conference on System Sciences*, 1992.

[18] C. Coutrix, L. Nigay, and P. Renevier. Modele d'interaction mixte : la ralit mixte a la lumiere des modalites d'interaction. In *Deuxiemes journees francophones Mobilite et Ubiquite UBIMOB'05, Grenoble, 31 mai - 3 juin 2005*, 2005.

[19] O. Cuisenaire. Distance transformation, fast algorithms and applications to medical image processing. *Ph.D. dissertation, Universit catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.*, Oct. 1999.

[20] O. Cuisenaire, M. Ferrant, Y. Vandermeeren, E. Olivier, and B. Macq. Registration and visualization of transcranial magnetic stimulation on magnetic resonance images. In *Proc. MICCAI '01*, volume 2208, 2001.

[21] D. L. Damos. *Multiple-task performance*. Taylor and Francis., Washington, D.C., 1988.

[22] M. Dias, J. Jorge, J. Carvalho, P. Santos, and J. Luzio. Usability evaluation of tangible user interfaces for augmented reality. In *The Second IEEE International Augmented Reality Toolkit Workshop*, Tokyo, Japan, 2003. ACM Press.

[23] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human Computer Interaction*. Prentice Hall Europe, 1998.

[24] E. Dubois. Chirurgie augmente : un cas de ralit augmente ; conception et ralisation centre sur lutilisateur. In *PhD Thesis*, page 275. Universite Joseph Fourier Grenoble 1 University, 2001.

[25] E. Dubois, P. da Silva P., and P. Gray. Notational support for the design of augmented reality systems. In *Proceedings of the 9th international workshop conference DSV-IS'02, Forbrig, P., Limbourg, Q., Urban, B., Vanderdonckt, J. (eds.), Rostock - Germany, June 2002*, pages 95–114. ACM, 2002.

[26] E. Dubois, P. D. Gray, and L. Nigay. Asur++: A design notation for mobile mixed systems. In *Mobile HCI '02: Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction*, pages 123–139, London, UK, 2002. Springer-Verlag.

[27] E. Dubois, P. D. Gray, and L. Nigay. Asur++: a design notation for mobile mixed systems. *Interacting With Computers, Special Issue of Mobile HCI*, 15(4):497–520, 2003.

[28] E. Dubois, P. D. Gray, D. Trevisan, and J. Vanderdonckt, editors. *MIXER '04, Exploring the Design and Engineering of Mixed Reality Systems, Proceedings of the IUI-CADUI*04 Workshop on Exploring the Design and Engineering*

of Mixed Reality Systems, Funchal, Island of Madeira, January 13th, 2004, volume 91 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.

[29] E. Dubois, L. Nigay, and J. Troccaz. Consistency in augmented reality systems. In M. R. Little and L. Nigay, editors, *Conference proceedings of EHCI'2001*, pages 111–122, Toronto, Canada, 2001. Springer Verlag.

[30] E. Dubois, L. Nigay, and J. Troccaz. Assessing continuity and compatibility in augmented reality systems. *Journal of Universal Access in the Information Society*, 1(4):263–273, 2002.

[31] E. Dubois, L. Nigay, O. Troccaz, J.and Chavanon, and L. Carrat. Classification space for augmented surgery, an augmented reality case study. In A. Sasse and C. Johnson, editors, *Proceedings of Interact99*, pages 353–359, Edimburgh,UK, 1999.

[32] N. I. Durlach and A. S. Mavor. *Virtual Reality: Scientific and Technological Challenges*. National Academy Press, Washington, D.C., 1995.

[33] S. Ellis and B. Menges. Localization of virtual objects in the near visual field. *Human Factors*, 40:415– 431, 1998.

[34] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, 1993.

[35] M. Ferrant, S. Warfield, C. R. G. Guttmann, R. V. Mulkern, and R. K. F. A. Jolesz. 3d image matching using a finite element based elastic deformation model. In *Proc. MICCAI '99*, pages 202–209, 1999.

[36] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[37] G. W. Fitzmaurice and W. Buxton. An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *CHI*, pages 43–50, 1997.

[38] J. Fitzpatrick. *Detecting failure, assessing success.* CRC Press, 2001.

[39] C. Furmanski, R. Azuma, and D. M. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR'02)*, pages 215–224, Darmstadt, Germany, 2002.

[40] F. G. and M. Massink. A reference framework for continuous interaction. *International Journal Universal Access in the Information Society*, 1(4):237–251, 2002.

[41] J. Gabbard, J. S. II, D. Hix, M. Lanzagorta, M. Livingston, D. Brown, and S. Julier. Usability engineering: Domain analysis activities for augmented reality systems. In A. J. Woods, J. O. Merritt, S. A. Benton, and M. T. Bolas, editors, *Proceedings of the SPIE - Stereoscopic Displays and Virtual Reality Systems IX*, volume 4660, pages 445–457. International Society for Optical Engineering, 2002.

[42] J. L. Gabbard and D. Hix. A taxonomy of usability characteristics in virtual environments. In *Deliverable to Office of Naval Research Grant No. N00014-96-1-0385*, 1997.

[43] A. W. K. Gaillard. Comparing the concepts of mental load and stress. *Ergonomics*, 36(9):991–1005, 1993.

[44] T. C. N. Graham, L. A. Watts, G. Calvary, J. Coutaz, E. Dubois, and L. Nigay. A dimension space for the design of interactive systems within their physical environments. In *DIS '00: Proceedings of the conference on Designing interactive systems*, pages 406–416, New York, NY, USA, 2000. ACM Press.

[45] C. Gram and G. Cockton. *Design Principles for Interactive Software*. Chapman and Hall, London, 1996.

[46] M. Grigni, D. Papadias, and C. Papadimitriou. Topological inference. In *Proceedings of the International Joint Conference of Artificial Intelligence IJCAI*. AAAI Press, 1995.

[47] W. Grimson, T. Lozano-Perez, W. I. Wells, G. Ettinger, S. White, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *Transactions on Medical Imaging*, 1996.

[48] P. Hancock and N. Meshkati. *Measuring mental workload: problems, progress, and promises*. Elsevier Science, Amsterdam, NL, 1988.

[49] B. L. Harrison, H. Ishii, K. J. Vicente, and W. Buxton. Transparent layered user interfaces: an evaluation of a display design to enhance focused and divided attention. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 317–324. ACM, 1995.

[50] S. G. Hart and L. E. Staveland. *Development of NASA-TLX (Task Load Index): results of empirical and theoretical research*. Elsevier Science, North-Holland, 1988.

[51] M. B. Hirokazu Katoa and I. Poupyrev. *ARToolKit*. ATR MIC Labs, MIC Research Labs, ATR International, Hikaridai, Seika, Souraku-gun, Kyoto 619-02, Japan, 2000.

[52] D. Hix, J. Gabbard, J. S. II, M. A. Livingston, T. H. Hllerer, S. J. Julier, Y. Baillot, and D. Brown. A cost-effective usability evaluation progression for novel interactive systems. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences - HICSS'04*, page 90276.3, 2004.

[53] D. Hix and H. R. Hartson. *Developing User Interfaces:Ensuring Usability through Product and Process*. John Wiley and Sons Inc., 1993.

[54] R. L. Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):413432, 1997.

[55] K. Hook, A. Bullock, A. Paiva, M. Vala, R. Chaves, and R. Prada. Fantasya and sentoy. In *Computer Human Interaction (CHI03), extended abstracts on Human factors in computer systems*, pages 804–805, Ft. Lauderdale, Florida, USA, 2003. ACM Press.

[56] M. Hou. A model of real - virtual object interactions in stereoscopic augmented reality environments. In *Seventh International Conference on Information Visualization - IV'03*. IEEE Computer Society, 2003.

[57] M. Hou and P. Milgram. Investigation of surface effects on alignment of real and graphic objects in stereoscopic augmented reality environments. In *Proc. of HFES 45th Annual Meeting*, pages 1877–1881, Minneapolis, MN, 2001.

[58] H. Ishii and B. Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *CHI*, pages 234–241, 1997.

[59] ISMAR. Ieee and acm international symposium on mixed and augmented reality. IEEE Computer Society, 2004.

[60] P. Jannin, J. Fitzpatrick, D. J. Hawkes, X. Pennec, R. Shahidi, and M. Vannier. Validation of medical image processing in image-guided therapy. *IEEE Trans. Med. Imaging*, 21(11):1445–1449, 2002.

[61] L. Johnson, P. Edwards, L. Griffin, and D. Hawkes. Depth perception of stereo overlays in image-guided surgery. In *Proceedings of SPIE Medical Imaging 2004: Visualization, Image-Guided procedures and Display*, volume 5372, pages 263–272. International Society for Optical Engineering, 2004.

[62] D. K. K., S. A., and M. N. A design advice tool presenting usability guidance for virtual environments. In *Workshop on User Centred Design and Implementation of Virtual Environments*, York, UK, 1999.

[63] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and T. K. . Virtual object manipulation on a table-top ar environment. In *IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 111–119, 2000.

[64] A. P. King, P. J. Edwards, M. R. Pike, D. L. G. Hill, and D. J. Hawkes. An analysis of calibration and registration errors in an augmented reality system for microscope-assisted guided interventions. In *Medical Image Understanding and Analysis 1999*, 1999.

[65] S. R. Klemmer. Papier-mch: toolkit support for tangible interaction. In *Proceedings of the User Interface Software and Technology UIST'03*. ACM, 2003.

[66] G. Klinker, D. Stricker, and D. Reiners. *Reality for Exterior Construction Applications*. Lawrence Erlbaum Associates, 2001.

[67] T. Konrad, D. Demirdjian, and T. Darrell. Gesture + play: full-body interaction for virtual environments. In *Computer Human Interaction (CHI03), extended abstracts on Human factors in computer systems*, pages 620–621, Ft. Lauderdale, Florida, USA, 2003. ACM Press.

[68] R. Lazarus. Theory-based stress management. *Psychological Inquiry*, 1(1):3–13, 1990.

[69] N. Leland, O. Shaer, and R. Jacob. Tuims: Laying the foundations for a tangible user interface management system. In *Workshop on Toolkit Support for Interaction in the Physical World, Pervasive 2004.*, 2004.

[70] V. Lepetit, L. Vacchetti, D. Thalman, and P. Fua. Fully automated and stable registration for augmented reality applications. In *the Second IEEE and ACM International Symposium on Mixed and Augmented Reality - ISMAR'03*, page 93102, 2003.

[71] W. Lorensen and H. Cline. Marching cubes, a high resolution 3d surface contruction algorithm. *Computer Graphics*, pages 163–169, 1997.

[72] B. MacIntyre and S. Feiner. Language-level support for exploratory programming of distributed virtual environments. In *ACM Symposium on User Interface Software and Technology*, pages 83–95, Seattle, WA, 1996. ACM Press.

[73] B. MacIntyre, M. Gandy, S. Dow, and J. Bolter. Dart: A toolkit for rapid design exploration of augmented reality experiences. In *Proceedings of the User Interface Software and Technology UIST04*. ACM, 2004.

[74] I. MacKenzie. *Virtual Environments and Advanced Interface Design*. Oxford University Press, NY, 1995.

[75] B. Mansoux, L. Nigay, and J. Troccaz. Interaction between a surgeon and a computer assisted surgery system: an interactive design space. In *SURGETICA 2005 Gestes mdico-chirurgicaux assists par ordinateur : Outils et applications*, 2005.

[76] X. Marichal, B. Macq, D. Douxchamps, T. Umeda, and A. Consortium. Real-time segmentation of video objects for mixed-reality interactive applications. In *Proceedings of SPIE Visual Communication and Image Processing VCIP'2003*, volume 5150, pages 41–50. International Society for Optical Engineering, 2003.

[77] R. McCann, D. Foyle, and J. Johnston. Attentional limitations with head-up displays. In *Seventh International Symposium on Aviation Psychology*, Columbus, OH, 1993.

[78] J. Mellor. Enhanced reality visualization in a surgical environment. *Massachusetts Institute of Technology, Artificial Intelligence Laborarory Technical Report*, (1544), Jan 1995.

[79] P. Milgram and H. J. Colquhoun. A taxonomy of real and virtual world display integration. *Mixed Reality, Merging Real and Virtual Environments, Ohmshda and Springer-Verlag*, pages 5–30, 1999.

[80] P. Milgran and C. J. Herman. A taxonomy of real and virtual world display integration. In *Mixed Reality: Merging Real and Virtual Environments*, pages 1–16, Berlin, 1999. Springer-Verlag.

[81] H. Najafi and G. Klinker. Model-based tracking with stereovision for ar. In *Proc. ISMAR '03*, pages 313–314, 2003.

[82] L. P. Nedel, C. M. D. S. Freitas, L. J. Jacob, and M. Pimenta. Testing the use of egocentric interactive techniques in immersive virtual environments. In *Interact 2003*, pages 471–478. IFIP, 2003.

[83] J. Nielson. *Usability Engineering*. Academic Press, 1993.

[84] L. Nigay, E. Dubois, P. Renevier, L. Pasqualetti, and J. Troccaz. Mixed systems: combining physical and digital worlds. In *Proc. of 10th International Conference on Human-Computer Interaction HCI International'2003 (Heraklion, 22-27 June 2003), C. Stephanidis (ed.)*, pages 1203–1207. Lawrence Erlbaum Associates, 2003.

[85] L. Nigay, B. Mansoux, and J. Troccaz. Abstract and concrete interaction with mixed reality systems: The case of the mini screen, a new interaction device in computer-assisted surgery. In *MIXER*, 2004.

[86] L. Nigay, P. Salembier, T. Marchand, P. Renevier, and L. Pasqualetti. Mobile and collaborative augmented reality: A scenario based design approach. In *Mobile HCI '02: Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction*, pages 241–255, London, UK, 2002. Springer-Verlag.

[87] D. A. Norman and S. W. Draper. *User Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Elbraum Associates, Hillsdale, NJ, 1986.

[88] G. P., H. M., P. V., R. S., R. C., and Z. J. Amire: Authoring mixed reality. In *The First IEEE International Augmented Reality Toolkit Workshop, 29 September, 2002, Darmstadt, GERMANY.*, 2002.

[89] F. Paterno. *Model-based Design and Evaluation of Interactive Applications*. Springer-Verlag, 2000.

[90] G. Peters and J. H. Wilkinson. The least squares problem and pseudo-inverses. *Comput. J.*, 13:309–316, 1970.

[91] W. Piekarski and B. H. Thomas. An object-oriented software architecture for 3d mixed reality applications. In *International Symposium on Mixed and Augmented Reality*, Tokyo, Japan, 2003.

[92] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: Non-linear mapping for direct manipulation in vr. In *ACM Symposium on User Interface Software and Technology*, pages 79–80, 1996.

[93] I. Poupyrev, D. S. Tan, B. M., H. Kato, H. Regembrecht, and N. Tetsutani. Developing a generic augmented reality interface. *IEEE Computer*, 35(3):44–50, 2002.

[94] I. Poupyrev, S. Weghorst, B. M., and T. Ichikawa. A framework and testbed for studying manipulation techniques for immersive virtual reality. In *ACM Symposium on Virtual Reality Software and Technology*, pages 21–28, Lausanne, Switzerland, 1997. ACM.

[95] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives,. *The Computer Journal*, pages 155–162, 1964.

[96] N. Quentin, F. Matthieu, V. Yves, O. Etienne, M. Benoit, and B. Olivier. Registration and real-time visualization of transcranial magnetic stimulation with 3-d mr images. *IEEE Transactions on Biomedical Engineering*, 51:1994–2005, Nov. 2004.

[97] J. Rasmussen. *Information processing and human-machine interaction*. Elsevier Science, Amsterdam:North-Holland, 1986.

[98] G. B. Reid and T. E. Nygren. *The subjective workload assessment technique: a scaling procedure for measuring mental workload*. Elsevier Science, Amsterdam: North-Holland, 1988.

[99] J. Rekimoto and K. Nagao. The world through the computer: computer augmented interaction with real world environments. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 29–36. ACM, 1995.

[100] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of CHI'99*, pages 378–385. ACM, 1999.

[101] P. Renevier and L. Nigay. Mobile collaborative augmented reality: the augmented stroll. In *Proceedings of EHCI'2001*, volume 2254, pages 315–334. Springer-Verlag, 2001.

[102] Y. Rogers, M. Scaife, S. Gabrielli, H. Smith, and E. Harris. A conceptual framework for mixed reality environments: designing novel learning activities for young children. *Presence*, 11(6):677–686, 2002.

[103] A. Rosenfeld and J. Pfaltz. Distance functions on digital pictures. *Pattern Recognition*, 1(1):33–61, 1968.

[104] J. Rubin. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley and Sons Inc., 1994.

[105] G. C. S. Introduction to health care technology assessment. In *National Library of Medicine/NICHSR*. Accessible at http://www.nlm.nih.gov/nichsr/ta101/ta101.pdf, 1998.

[106] T. Saito and J.-I. Toriwaki. New algorithms for euclidian distance transformations of an n-dimensional digitized picture with applications. *Pattern Recognition*, 27:1551–1565, Nov. 1994.

[107] F. Sauer, U. J. Schoepf, A. Khamene, S. Vogt, M. Das, and S. G. Silverman. Augmented reality system for ct-guided interventions: system description and initial phantom trials. In *Medical Imaging 2003: Visualization, Image-Guided Procedures and Display, Proceedings of the SPIE*, volume 5029, page 384394, 2003.

[108] D. L. Scapin and C. Bach. Adaptation of ergonomic criteria to human-virtual environments interactions. In *INTERACT*, 2003.

[109] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV 47(1/2/3)*, pages 7–42, June 2002.

[110] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003),Madison, WI*, 1:195–202, June 2003.

[111] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavari, L. M. Encarnac, O. M. Gervautz, and W. Purgathofer. The studierstube augmented reality project. *Presence*, 11(1), 2002.

[112] W. Schroeder and L. Ibanez. The itk software guide. The Insight Consortium.

[113] W. J. Schroeder, editor. *The visualization Toolkit User's Guide*. Kitware inc., 2001.

[114] O. Shaer, N. Leland, E. Calvillo, and R. Jacob. The tac paradigm: Specifying tangible user interfaces. *Personal and Ubiquitous Computing*, 8(5):359–369, 2004.

[115] R. Splechtna, A. Fuhrmann, and R. Wegenkittl. Aras - augmented reality aided surgery system description. In *VRVis Technical Report*, 2002.

[116] F. Stevens, L. Frances, and L. Sharp. User-friendly handbook for project evaluation: Science, mathematics, engineering, and technology education. In J. Frechtling, editor, *NSF*, pages 93–152, 1997.

[117] A. Sutcliffe. *Multimedia and Virtual Reality: Designing Multisensory User Interfaces*. Lawrence Erlbaum Associates, 2003.

[118] P. Szekely. Retrospective and challenges for model-based interface development. In *Computer Aided Design of User Interfaces CADUI'96, F. Bodart and J. Vanderdonckt (eds)*, pages 1–27. Springer-Verlag, 1996.

[119] A. Tang, C. Owen, F.Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the conference on Human factors in computing systems CHI2003*, pages 73–80, Fort Lauderdale, Florida, USA, 2003. ACM Press.

[120] V. Tanriverdi and R. J. K. Jacob. Interacting with eye movements in virtual environments. In *Proceedings of Computer Human Interaction (CHI03)*, pages 265–272. ACM Press, 2000.

[121] V. Tanriverdi and R. J. K. Jacob. Vrid: a design model and methodology for developing virtual reality interfaces. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST 2001, November 15-17, 2001, Banff, Alberta, Canada. ACM, 2001*, pages 175–182. ACM, 2001.

[122] D. Trevisan, L. P. Nedel, J. Vanderdonckt, and B. Macq. Usability evaluation method for mixed reality systems in surgery. In *Deliverable 69: Results of iterative framework testing*. SIMILAR Network of Excellence, FP6-507609 http://www.similar.cc, 2005.

[123] D. Trevisan, J. Vanderdonckt, B. Macq, P. L. de la Orden, S. R. Bescs, and E. J. G. Aguilera. Usability evaluation issues in mixed reality systems in surgery. In *Deliverable 18: Current practice description and evaluation: Towards a framework for usability evaluation*, pages 21–25. SIMILAR Network of Excellence, FP6-507609 http://www.similar.cc, 2004.

[124] D. Trevisan, J. Vanderdonckt, B. Macq, and C. Raftopoulos. Modeling interaction for image-guided procedures. In *Proceedings of International Conference on Medical Imaging SPIE2003 (San diego, 15-20 February 2003), K. Hanson, C.-T. Chen, E.L. Siegel (eds.)*, volume 5029, pages 108–118. International Society for Optical Engineering, 2003.

[125] D. G. Trevisan. Designing smooth connections between worlds. In *Extended abstracts of the 2004 conference on Human factors and computing systems CHI2004, Session Doctoral Consortium*, pages 1043–1044. ACM Press, 2004.

[126] D. G. Trevisan, V. Nicolas, P. Delville, M. C. d'Elzius, and B. Macq. Towards markerless augmented medical visualization. In *MICCAI International Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery*, 2004.

[127] D. G. Trevisan, V. Nicolas, R. Olsweszki, and B. Macq. Intra-operative guidance. In *Demonstration session during 100 ans de Stomatologie et 25 ans de Chirurgie Maxillo-Faciale*, Clinique Universitaires Saint Luc, Brussels, 2005.

[128] D. G. Trevisan, J. Vanderdonckt, and B. Macq. Conceptualizing mixed spaces of interaction for designing continuous interaction. *Virtual Reality Journal*, 8(2):83–95, 2004.

[129] P. S. Tsang and G. Wilson. *Handbook of human factors and ergonomics*. Wiley, New York, 1997.

[130] P. Turner and E. Davenport, editors. *Managing multiple spaces*. Kluwer, 2005.

[131] B. Ullmer, H. Ishii, and R. Jacob. Token+constraint systems for tangible interaction with digital information. *ACM Transactions on Computer-Human Interaction*, 12(1):81–118, 2005.

[132] B. Ulmer and H. Ishii. Emerging frameworks for tangible user interfaces. In *Human-Computer Interaction in the New Millenium, John M. Carroll, ed*, pages 579–601. Addison-Wesley, 2001.

[133] M. Vazirgiannis, Y. Theodoridis, and T. Sellis. Spatio-temporal composition and indexing for large multimedia applications. *Multimedia Systems*, 6:284–298, 1998.

[134] F. Vernier. La multimodalite en sortie et son application a la visualisation de grandes quantites d'information. In *PhD Thesis*, page 271. Universite Joseph Fourier Grenoble 1 University, 2001.

[135] J. Viala, E. Dubois, and P. Gray. Guide-me : Environnement graphique de manipulation de la notation asur. In *Proceedings of the french conference Mobilit et Ubiquit 2004, Canals, G., Giboin, A., Nigay, L., Pinna, A.-M., Tigli, J.-Y. (eds), Nice, France, June 2004*, pages 74–78. ACM, 2004.

[136] K. J. Vicente and J. Rasmussen. Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4):589–606, 1992.

[137] M. A. Vidulich and P. S. Tsang. Techniques of subjective workload assessment: a comparison of swat and the nasa-bipolar methods. *Ergonomics*, 29(11):1385–1398, 1986.

[138] S. Vogt, F. Wacker, A. Khamene, D. R. Elgort, T. Sielhorst, H. Niemann, J. S. Lewin, and F. Sauer. Augmented reality system for mr-guided interventions: Phantom studies and first animal test. In *Proceedings of SPIE Medical Imaging 2004: Visualization, Image-Guided procedures and Display*, volume 5367, pages 100–109. International Society for Optical Engineering, 2004.

[139] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger. Augmented reality in architectural construction, inspection and renovation. In *ASCE Third Congress on Computing in Civil Engineering*, Anaheim, CA, 1996.

[140] G. M. Wilson. Psychophysiological indicators of the impact of media quality on users. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 95–96, New York, NY, USA, 2001. ACM Press.

[141] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.