Personalized Perception and Interaction with Medical Information in Mixed Reality Environments

Meng Ma

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality

Personalized Perception and Interaction with Medical Information in Mixed Reality Environments

Meng Ma

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Prof. Bernd Brügge, Ph.D.
Prüfer der Dissertation:	1. Prof. Dr. Nassir Navab
	2. Prof. Hongen LIAO, Ph.D. Tsinghua University, VR China
	3. Prof. Pascal Fallavollita, Ph.D. University of Ottawa, Canada

Die Dissertation wurde am 19.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 05.08.2016 angenommen.

Abstract

Medical information is employed in different scenarios, such as education, training, diagnosis, and surgery. Perception and interaction of this information is fundamental, and augmenting the perception through personalized interaction with multi-model data is user specific. The good of this thesis if to investigate personalized perception through development of novel interaction method with the different medical information suitable for students, patients, and doctors.

At first, a Magic Mirror framework is presented to create a mixed reality environment with an *in-situ* augmented reality view. It tracks the user and achieves a registration between the medical information and the user. It inherently allows the user to perform interactive registration to enhance personalized perception. This opens interesting possibilities for personalized augmented reality in medical education and rehabilitation. User studies were designed to respectively assess the precision and acceptability of the Magic Mirror system for anatomy education. We also take advantage of the interactive mixed reality to generate personalized augmented reality applications for the learning of organs and muscle and for rehabilitation exercise. In addition, functions, e.g. "Organ explosion" and "Self-control virtual view", are designed and implemented, and serious games are also developed in this framework.

We then introduce the Pointing At Several Targets (PAST) interaction method that offers an approach to calculate a virtual eye center as the origin for the *eye-rooted* pointing ray without tracking the eye gaze in an egocentric setting. Consequently, the user-specific pointing geometry is recovered for personalized interaction with ambient information without visual feedback. The proposed method is mathematically analyzed and shows that the calibration is feasible with a measured accuracy of the recovered pointing ray below $0.9°$. Our user study shows that the calibration is independent of the environment and the pointing ray is recovered with an accuracy of $0.67° \pm 0.71°$ without any visual feedback. Based on the recovered pointing gesture, a novel user interface is introduced, allowing the surgeon to perform touchless interaction with the various medical systems, and switch effortlessly among them, without accessing or modifying the software and hardware of targeted systems.

Finally, in conjunction with the proposed Magic Mirror framework and recovered pointing gesture, a framework for collaborative Mixed Reality (MR) is introduced. The human-human communication is enhanced via pointing gesture within an MR framework.

Keywords: Medical Information, Personalized Perception, Personalized Interaction, Augmented Reality, Pointing Gesture, Mixed Reality

Zusammenfassung

Medizinische Informationen werden bei Ausbildung, Training, Diagnose und Operationen benutzt. Die Wahrnehmung von unterschiedlichen Medien und Gegenständen und die Interaktion damit sind grundlegende menschliche Aktivitäten und sie sind benutzerspezifisch. Wissenschaftler arbeiten an Systemen, welche die Darstellung und Interaktion mit medizinischen Informationen den Bedürfnissen von Studenten, Patienten, und Doktoren anpassen.

Zuerst wird der magische Spiegel vorgestellt, der reale Bilder mit digitalen Informationen überlagert. Dazu verfolgt das System die Pose des Benutzers in Echtzeit und passt mit Hilfe eines Kalibrierungsalgorithmus die Visualisierung von medizinischen Datensätzen spezifisch auf die Proportionen des jeweiligen Benutzers an. Dies bietet interessante Möglichkeiten für benutzerspezifische erweiterte Realität in der medizinischen Ausbildung und für Rehabilitation. In zwei Benutzerstudien wurde die Genauigkeit der überlagerten Informationen und die Akzeptanz des Systems für Anatomie Ausbildung überprüft. Das System wurde verwendet um Benutzern die menschlichen Organe und Muskeln zu lehren, sowie zur Rehabilitation. Zusätzliche Funktionen wie zum Beispiel "Organ Explosion" und "Selbst-gesteuerte virtuelle Ansicht" wurden entworfen und implementiert. Ernsthafte Spiele wurden auch auf Basis des magischen Spiegels entwickelt.

Eine neuartige Mensch-Maschine Schnittstelle mithilfe einer tragbaren Tiefenkamera wurde ebenfalls entwickelt. Sie ermöglicht es Benutzern durch einfaches zeigen auf Gegenstände mit dem System zu interagieren. Dazu schlagen wir eine neuartige Kalibrierungsmethode vor, welche die Position des Blickpunkts eines Benutzers ermittelt. Ziele werden durch eine gerade Linie anvisiert, welche definiert ist durch den Blickpunkt und den Zeigefinger des Benutzers. Diese intuitive Art der Interaktion mit Gegenständen in der Umgebung erlaubt die Kontrolle von Programmen ohne weitere visuelle Rückmeldung. Bei einer mathematischen Analyse der Kalibrierungsmethode konnten wir zeigen, dass der Fehler der Zielstrahl Berechnung unter $0,9°$ liegt. Mithilfe einer Benutzerstudie haben wir die Genauigkeit der vorgestellten Kalibrierungsmethode gemessen. Benutzer konnten die vorgegebenen Ziele mit einer Genauigkeit von $0,67° \pm 0,71°$ ohne visuelle Rückmeldung erreichen. Die intuitive Interaktionsmethode erlaubt es Ärzten eigenständig mit diversen medizinischen Geräten berührungslos zu interagieren. Die Implementierung des Benutzerinterfaces ermöglicht zudem eine Integration in bestehende Systeme ohne Änderungen an Geräten und den benutzten Programmen.

Zum Schluss wird der magische Spiegel mit der berührungslosen Interaktion kombiniert und ein System vorgestellt, das von mehreren Benutzern gleichzeitig bedient werden kann. Die zwischenmenschliche Kommunikation wird durch die Unterstützung von Zeigegesten verbessert.

Schlagwörter: Medizinische Informationen, personalisierte Wahrnehmung, Personalisierte Interaktion, Augmented Reality, Zeigegeste, Mixed-Reality

Acknowledgments

Contents

Introduction

Medicine is the science and practice of the diagnosis, treatment, and prevention of disease, and most medical information is about our human body. Medical information are employed in different scenarios, such as education, training, diagnosis, operation, and so on. As the computing technologies develop, more and more information is digitized from the physical world, and then researchers work on how to process and show them back to the user, enabling the user to perceive and interact with the information naturally and effectively. Perception and interaction with different media and objects are the fundamental human activities, and they are, meanwhile, user specific [148]. Mixed Reality (MR) tries to blur the boundary between the real world and the virtual world, generated based on the digital information, to improve the perception and interaction experience [85]. Human centered computing aims at adapting computers to human minds and habits, and an important research direction in this area is the design of user interfaces. A perfect user interface should automatically adjust to the engaged user's personality. In this thesis, a MR framework for personalized perception of medical information is proposed for medical education and rehabilitation exercise, and pointing gestures are employed for the user interacting with ambient media and objects in an egocentric setting. Finally, these two techniques are combined to create a framework for collaborative MR.

1.1 Motivation

The foundation for the motivation of this work are the recent developments in Mixed Reality and the excellent projects that have been done in NARVIS (Navigated Augmented Reality Visualization System) laboratory . Different MR applications were developed for medical education, training, and intra-operative navigation, and these techniques enable more intuitive perception of the medical information and instinctive interaction with different medical information [111]. In NARVIS, *in-situ* visualization in medical augmented reality was firstly developed for a video see-through AR system, which enables the stereoscopic view on visualized CT data registered with the real anatomy of a patient, and then Tobias Blum [16] created a portable AR system for medical education with the help of Microsoft Kinect. All of these are the foundation of my research work for the personalized augmented reality. In addition, NARVIS lab provides an interdisciplinary platform for our research activities, and thanks for the computer scientists who worked in NARVIS, a strong collaboration with surgeons has been established here. Hence, I understood the special requirements for the human computer interaction in operating room and started my research on the personalized user interface for the intra-operation medical systems, using the personal pointing gesture. In the following paragraphs a short overview of the current challenges and motivations for each topic are presented.

1.1.1 Personalized perception of medical information

Challenges in anatomy education   Anatomical education is an important content of every curriculum and starts already very early in school, to form a good understanding of the body and improve the general health awareness [38]. In today's medical schools, students are required to understand both function and spatial context of human anatomy. Traditional medical education learning is classified into three categories: cadaver, model, and book-based. The cadaver-based learning has seen decline due to practical and cost issues. As a result, anatomical education has shifted towards diagram, physical, and virtual models. However, some drawbacks exist with these learning paradigms. For example, it is difficult to interpret the spatial and physical characteristics of anatomy by observing two-dimensional images, diagrams, or photographs. Many physical models also lack detail levels to fully understand the specific anatomy. Numerous digital anatomy models exist and these resources are very valuable. The disadvantage is that the user still has to mentally map/link anatomy onto the human bodies. Providing an adequate learning experience to different learners is also a challenging issue as the traditional learning system generally does not adapt content to suit individual learner needs. With the advent of the plethora of exciting technological advancements there should be no reason not to include these for the creation of new education paradigms for medical learning.

Challenge in motor rehabilitation   Motor rehabilitation restores a patient's functional health that is damaged due to musculoskeletal or nervous system impairments. It consists in iterative repeating exercises to strengthen the affected body area. It requires movements to be performed in a very specific way, otherwise the gain may not be adequate and the desired results will be less noticeable [102, 155]. Often, since the rehabilitation must be performed over quite a long period it is common for patients to get disinterested, and consequently, they perform the exercises in a casual and incorrect manner. Usually, motor rehabilitation occurs by regular meetings between patient and physiotherapist with frequency depending on the availability of both. The time that patients spend on therapy in clinics is very short compared to the potential time that they can spend on it outside of the clinical setting (e.g. at home). Thus, we believe that it becomes crucial for the patient to perform exercises at home to speed up the rehabilitation process. However, both therapeutic instructions and corrections by the therapist are not verbally provided to the patient at home and it potentially leads the patient to perform the exercises incorrectly.

Mixed Reality for education and serious gaming   MR systems have the advantage that digital information can be embedded and/or superimposed upon reality. This allows for a more close-to-reality presentation of medical knowledge and offers opportunities for new and interactive learning context. The user can spatially relate corresponding medical information to the reality objects, e.g. human body. The main reason for MR not being used in many systems today is that developing MR systems is more challenging than developing virtual reality systems. The integration of real and virtual objects requires accurate calibration, more advanced visualization and special user interfaces. Personalization for promoting a multi-modal learning environment is also a growing area of interest, such as the development of user modeling and personalized processes which place the user at the center of the learning environment. The "Magic Mirror" is that the user stands in front of a screen and via a camera, the image of the user is shown on the screen such that it acts like a mirror. While such a

system restricts the motion of the user, it allows for an inexpensive and robust solution that is better suited for medical education and serious gaming.

1.1.2 Natural interaction in operating room

Challenges of user interaction in operating rooms   As medical technology is continuously developing, more medical systems are integrated inside the operating room (OR) for diagnosis and intra-operative navigation. Today, the conventional display technology in the operating room is still the 2D display and the information from different medical systems is shown on different displays. Figure 1.1 showcases the photo of a modern day operating room at Klinikum der Universität München - Campus Großhadern that consists of six unique displays for different medical information output. Surgeons rely on these medical systems to capture, browse, and manipulate patient information, physiological data, and medical images, but the current control input of these systems is still mouse, keyboard, touch screen, and joystick. To interact with these medical systems and their control inputs, the user has to press either a physical/virtual button using their own finger or a visual cursor controlled by another device. Yet, one of the most important rules in the OR is to maintain a strict boundary between what is sterile and what is not [119]. A sterile processing of the input devices is impossible; hence the surgeons cannot directly touch these input devices after they are scrubbed and gloved. However, research shows that the surgeon needs to directly control the medical systems to mentally 'get to grips' to what is going on, something which is not achievable by proxy [78]. In addition, the input devices of different systems are totally independent due to security and privacy rules. The management of multiple interfaces in the modern operating room is also a challenge for the nurse and surgeon.
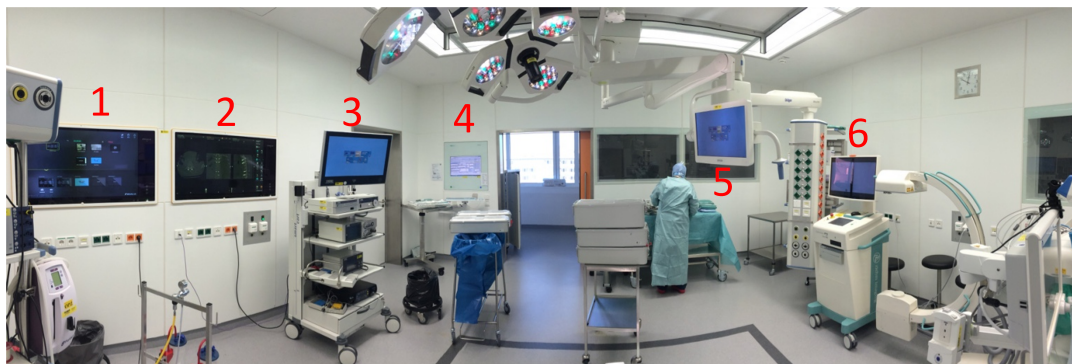
Figure 1.1   There are six displays in a normal operating room in Klinikum der Universität München - Campus Großhadern for different medical systems. The function of these displays are 1&2: DICOM images from the archive or live video streaming from the handle grip camera; 3: Arthroscopy Monitor; 4: SAP Hospital Information System; 5: Ample monitor sharing the same screen as 1 or 2; 6: Fluoroscopy monitor.

Pointing gesture for interaction   Human centered computing aims at adapting computers to human minds and habits, and an important research direction in this area is the design of natural and ergonomic human-computer interaction techniques. Pointing gestures are fundamental to human behavior [97] and are used consistently across cultures [99]. The gestures begin at an early developmental stage [20] and let humans reference proximal objects as well as abstract concepts in the world. Today, pointing gestures are not only part

of our gestural language but are inherently used for interaction [109]. A few authors have tried to estimate the user's eye position and recover the *eye-rooted* pointing geometry to enable interacting with ambient media and objects in the real world. One existing technique consists in positioning a camera next to the monitor or on the ceiling to observe the user and implement the outside-in tracking. In this case, the computing device calculates the eye position by detecting the user's gaze and face or wearable trackers. Unfortunately, the user has to stand within a limited working zone defined by the camera, as shown in Figure 1.2(a). As an alternative to the traditional setting of external sensors observing the user, the egocentric setting mounts a portable camera on the head or clothing of the user and perceives the interaction from an egocentric perspective [44, 90]. Other researchers have taken advantage of an egocentric setting employing inside-out tracking [62, 106]. They directly take the camera center as the pointing ray origin, as shown in Figure 1.2(b). These methods work well when the user performs interaction with objects and media which are reachable. However, what if the user tries to point towards objects which are not reachable (see Figure 1.2(c)). There is a significant angle between the ray starting from the camera and that from the eye. Hence, there is an important offset between O_{eye}, the object which is pointed at in the user's view, and O_{camera}, the object observed in the camera view.

(a) (b) (c)
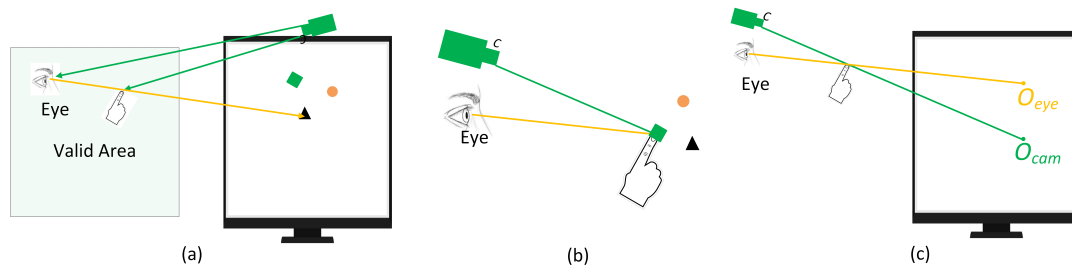
Figure 1.2 The *eye-rooted* pointing geometry in different setups. (a) There is a limited working zone when positioning a camera to observe the user. (b) The camera position can be chosen as the origin point when the desired object is reachable. (c) There is a significant angle between the ray starting from the camera and that from the eye when the target is unreachable.

1.1.3 Collaborative characteristics

Collaboration between people becomes more and more common when using computer. Human-Computer Interface (CHI) is giving away to a Human-Human Interface mediated by computers [14]. From the real world to a MR environment, there are functional seams and cognitive seams for the user during the cooperation. The normal interaction methods should be kept in the MR application, if possible. There are new technical challenges to be addressed when there are more than one user in a common workplace. In normal anatomy teaching or rehabilitation scenario, there is a teacher or doctor to teach and monitor the student/patient for the learning and exercise. There are also several surgeons to control several medical systems during the operation. The challenge is to support the collaborative MR when we try to introduce the new techniques.

1.2 Objectives

The objectives of this thesis are to investigate the personalized Magic Mirror framework, in order to improve the perception and interaction with digital medical information, as well as the development of a pointing gesture recovery in an egocentric setting for natural interaction with ambient media and objects. Finally, these two technologies are integrated to create a framework for collaborative MR. In the following listings the prime objectives are briefly highlighted.

- Develop a personalized MR Magic Mirror for medical education and motor rehabilitation, which improves the perception based on the personal information and natural interaction while motivating the user via gaming elements.

- Propose a method to accurately recover the *eye-rooted* pointing ray without tracking the eye gaze in an egocentric setting, enabling the user directly performing pointing gesture to interact with ambient media and objects without visual feedback.

- Introduce a novel user interface that allows the surgeon to personally perform touchless interaction with the various medical systems in the operating room, and switch effortlessly among them, all of this without modifying the systems' software and hardware.

- Propose a collaborative MR framework for medical teaching and rehabilitation, which combines the personalized MR Magic Mirror and pointing gesture interaction in an egocentric setting.

1.3 Contribution

In the course of this thesis several solutions and methods have been contributed to personalized mixed reality and pointing gesture interaction, both for medical education, rehabilitation exercise and touchless operation with various medical systems.

Personalized perception in Magic Mirror framework   A Magic Mirror framework is presented creating a MR environment with an *in-situ* augmented reality (AR) view. The system can track the current user and finish the registration between the medical datasets and the user based on the personal information (e.g., height, body size, gender, and age). It inherently allows the user to perform interactive registration and directly map the medical information onto his/her body. This opens interesting possibilities for personalized mixed reality for medical education and rehabilitation exercise. Through the participation of 7 clinicians and 72 anatomy students, two user studies were designed to respectively assess the precision and acceptability of the Magic Mirror system for anatomy education in the classrooms of tomorrow. We also take the advantage of the interactive mixed reality to generate a personalized learning procedure. In addition, "Organ explosion" and "Self-control virtual view" systems are designed and implemented for anatomy leaning. Serious games are also developed in the framework for health-care education and rehabilitation exercise.

Pointing gesture recovery in an egocentric setting   A method is proposed to calculate a virtual eye center as the origin for the *eye-rooted* pointing ray technique without tracking the eye gaze in an egocentric setting. The user specific pointing geometry can then be recovered for interaction with ambient media and objects without visual feedback. The pointing ray starts from the virtual eye center and goes through the fingertip to the desired target information. During the initial calibration, the user is asked to perform pointing gestures towards several targets displayed within the field of view. A user specific virtual eye center is then estimated based on the detected fingertip in the coordinate system of the wearable device. Then the *eye-rooted* pointing ray is recovered when a user is performing a pointing gesture. The proposed method is mathematically analyzed and the method is also evaluated in a user study consisting of 10 participants performing pointing interaction with objects shown on the display while wearing a RGB-D sensor.

Natural interaction with multi-system in OR   Contrary to the state-of-the-art, a novel user interface is introduced, allowing the surgeon to personally perform touchless interaction with the various medical systems, and switch effortlessly among them, all of this without modifying the systems' software and hardware. The advantages are:

- needs no modification of the software/hardware of the existing medical systems and does not need the medical software to be re-certified

- combines pointing with personal gesture command

- switches smoothly among different medical systems and operands

- offers multi-user, multi-system framework

Android devices with a special application are connected to the computers on which the medical systems are running, simulating a normal USB mouse and keyboard. Based on the pointing gesture recovery technique, the surgeon's pointing gesture can be calculated. The target position in the real world, the desired cursor position in the display and the surgeon's personal gestures, are detected and analyzed. The result is sent to the Android devices via WiFi. The application running on the Android system generates the corresponding mouse or keyboard events according to the targeted medical system allowing completion of the interaction. In essence, one personal user interface is conceived to control all systems in an operating room.

Collaborative MR framework for teaching and rehabilitation   A multi-user collaborative MR framework is introduced by integrating the MR Magic Mirror and the pointing gesture in egocentric setting. The human-human communication is partly implemented via pointing gesture and MR framework. The framework is examined in two scenarios, anatomy teaching with a student and rehabilitation exercise with a physiotherapist. The teacher directly points to the MR view and controls it to lead the teaching procedure, and the physiotherapist in charge of the exercise in the same way.

1.4 Organization

In this thesis, issues related to improve the personalized perception and interaction are discussed and presented, and it is organized as follows. The state of art of perception and interaction with medical information, including medical education, rehabilitation, and intra-operative interaction, is presented in Chapter 2. Furthermore, the history and the state of mixed reality and natural user interface using gestures is discussed. In Chapter 3, the Magic Mirror framework is presented and several methods are investigated to improve the user's perception in the Magic Mirror setting. Personal information is detected to select the best medical data for registration and personalized registration is performed to improve the AR view. Several functions and applications are developed for medical education and rehabilitation. In Chapter 4, a Point At Several Targets (PAST) method is proposed to calculate an origin for the *eye-rooted* pointing gesture and recover the pointing ray in an egocentric setting. The proposed method is mathematically analyzed and evaluated in a user study. A novel user interface for various medical systems in the operating room is presented based on the recovered pointing gesture. Then, a collaborated MR framework encompassing the Magic Mirror conception and the pointing gesture is proposed in Chapter 5. Finally, conclusions will be drawn and possible future research direction will be discussed in Chapter 6.

History and State of the Art

Computing is at one of its most exciting moments in history, playing an essential role in supporting many important human activities. The explosion in the availability of information in different media forms and through multiple sensors and devices means, on one hand, that the amount of data we can collect will continue to increase dramatically, and, on the other hand, that we need to develop new paradigms to search, organize, and integrate such information to support all human activities. Human-centered computing (HCC) is a set of methodologies that apply to any field that uses computers, in any form, in applications in which human directly interact with systems or information that use computer technologies [72]. HCC is an emerging field that aims at tightly integrating human sciences and computer sciences for timely cognitive and interactive support. It focuses on humans from beginning to end. HCC also aims at radically changing computing with new methodologies to design and build systems that support and enrich people's lives.

Perception and interaction are the communication between user and computer and important parts of HCC. On the user-side, communication is in task-language and on the system side, in core language. In this chapter we want to provide the reader with the necessary and relevant background information about perception and interaction with medical information during procedures, such as medical education, rehabilitation exercise, and intra-oprative interaction. Then we will turn our attention to topics that are related to mixed reality and natural user interface using gestures.

2.1 Perception and interaction with medical information

Different medical knowledge is collected through many years of practice and different imaging methodologies, and the knowledge has to be transferred to fresh medical students through medical education and the general population through public education. Additionally, during surgery and for each patient, quite a lot of information is collected for diagnosis and navigation. Good communication between the patient and the surgeon is also very important for making the patient comply with the treatment. Rehabilitation exercises also have to be performed after surgery, if need be. During the above-mentioned procedure, people act different roles to perceive and interact with medical information. In this section, we discuss the states and challenges for the perception and interaction with medical information in medical education, public and patient education, rehabilitation exercise and in operating rooms.

2.1.1 Medical education

While medical education has already performed for a long time, we start from Abraham Flexner's report in 1910 [49]. Flexner visited all 155 medical schools that existed in North

| (a)Atlas. | (b)Physical model. |
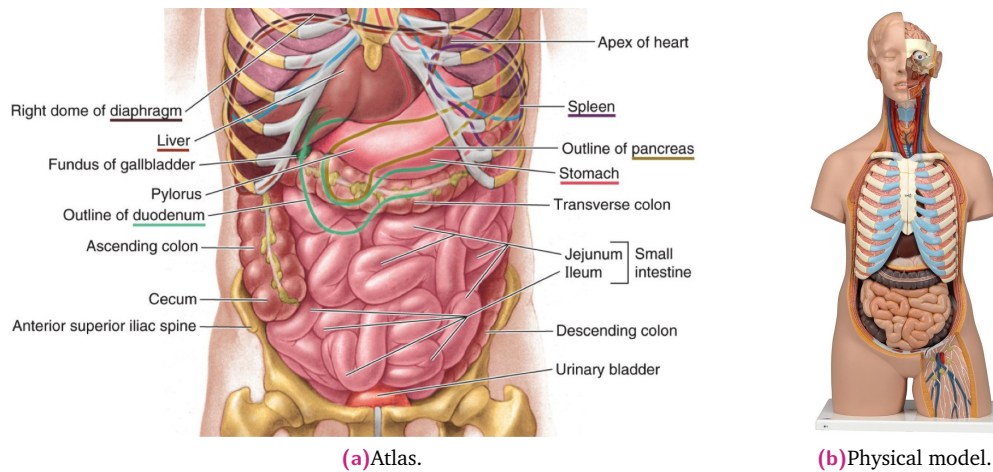
Figure 2.1 Traditional methodologies for Medical learning.

America in 1909 and he identified four major problems of medical education: lack of standardization, lack of integration, lack of inquiry, and identity formation. The report had a huge impact and shaped modern medical education, where patient care, teaching, and research are combined. However, academic hospitals don't spend enough time on teaching today due to enormous pressure to publish and for economic reasons [135]. Another problem is that today the research is focusing on very small subtopics, which do not relate to medical education [93]. On the other hand, only a little research is done for teaching. An example is gross anatomy where most topics are already know and only very few interesting research exists. In today's medical schools students are required to understand both function and spatial context of human anatomy.

Traditional methodologies

Traditional medical education learning is classified into three categories: cadaver, model, and book-based. Although technology has advanced significantly in the last decades, classical school education still mostly uses the same methods to convey anatomical knowledge [38]. Typically, the information is collected in printed books like anatomy atlases, displayed in the form of charts and diagrams. Those diagrams provide a simple and well-known method to illustrate form and appearance of organs, having the advantage that the user is accustomed to such methods of display. However, there exist several downsides to this method. First of all, the view is limited to a selected few different angles the author chose to present. This may not be sufficient in some cases to fully convey how an organ is situated relative to its surroundings since occlusions limit the possibilities to visualize these spatial relations [1]. Another problem is that often the organs are only depicted schematically by leaving out details or distorting tissue colors, thus giving only a coarse impression on how the organ actually looks like in reality. For example, it is difficult to interpret the spatial and physical characteristics of anatomy by observing two-dimensional images, diagrams, or photographs. Many physical models also lack detail levels to fully understand the specific anatomy (see Figure 2.1).

Anatomy education is also performed by the dissection of cadavers. The value of dissection classes as a teaching format lies in the fact that it provides a 3D view on human anatomy including tactile learning experiences. It enables elaboration of knowledge already acquired

Figure 2.2 Anatomy learning with cadaver. It is a very important training but too expensive, and there are always a lot students around one cadaver.
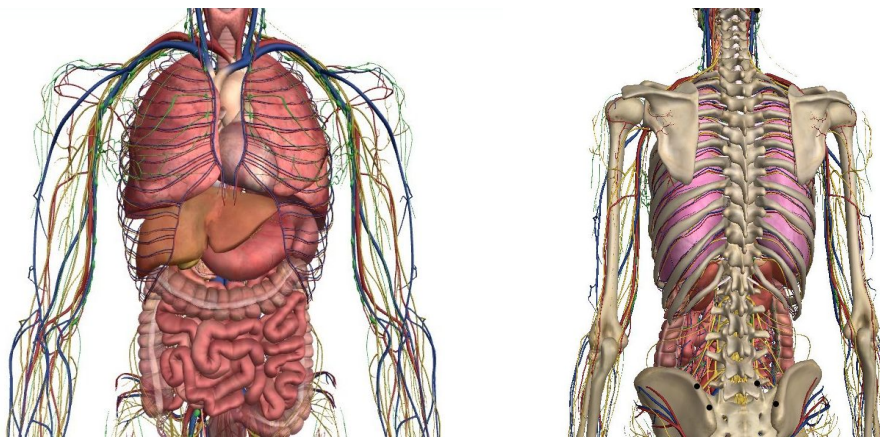
Figure 2.3 Anatomy 3D models from Zygote[1].

in lectures and study books and it provides an overall perspective of anatomical structures and their mutual relations in a whole organism [98]. The cadaver-based learning has seen decline due to practical and cost issues (see Figure 2.2). And so far, no objective empirical evidence exists concerning the effectiveness of dissection classes for learning anatomy [51].

Computer-based learning material

It is developed by experts and students can use these materials if there is no available expert in the hospital. Computer-based education can be very powerful for anatomy teaching, where 3D visualization is of great benefit (see Figure 2.3). Many virtual model databases exist and E-learning is commonly used today. These resources are very valuable and are more interactive and interesting than textbooks. In addition, the personalization in E-learning systems has been the subject of much recent research and allows teachers to select parameters and combine them flexibly to define different personalized strategies according to the specifics of the courses. In recent times, more and more digital methods evolve which are trying to improve on these methods. Many of them offer interactive anatomy models, usable either as an on-line service or as a standalone application. There also exist organizations specifically offering teaching bundles for use in classrooms, making it possible to use alternative teaching

---

[1]www.zygotebody.com
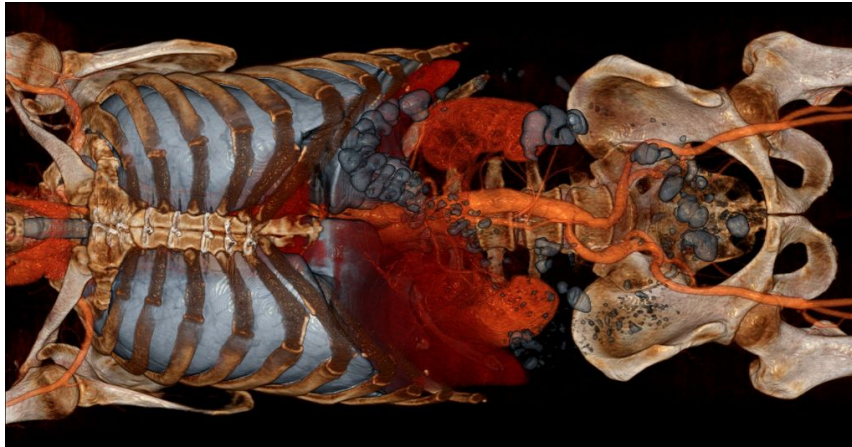
Figure 2.4 Recent CT volume visualization.

methods at different school levels assisted by videos and interactive tools. Those interactive applications offer large improvements over the classical method. It is possible to view organs and structures from any desired angle, control magnification and often even select specific organs and systems to be displayed or hidden. Many commercially available systems use VR for medical education and psychomotor skills training [92, 10]. These systems have indeed proven to be valid and useful. By combining computer models of anatomical structures with custom software we can showcase to students new ways of interacting with anatomy that could not be achieved during cadaveric dissections or in static diagrams and models for increasing their learning satisfaction [7]. Visualization of medical data can also be used for education. Medical datasets are very large and the visualization must correctly represent the reality. Today images with a very high level quality can be generated by high performance graphics for computer gaming (see Figure 2.4).

Through all those materials and interactive elements, the users can control the region of interest and hopefully better understand the information they are looking for. These resources are very valuable and are more interactive and interesting than textbooks, however, with the disadvantage of having users still mentally map/link anatomy onto their bodies. Humans usually have to build the knowledge network in the brain after perception of the knowledge. The information is very general but the learning procedure is quite personal. Personalized information is much easier for the user to accept. Fresh medical students still bear a very long and hard study during the medical knowledge transferring. At the same time, different technology is developed to simplify the learning tasks.

2.1.2 Public and patient education

For obvious reasons, methodologies for medical education are not suitable for public education, as it is not necessary to teach a high-level of anatomy understanding. However, real-life demonstrations are limited to the skin-surface, which is why most people learn about anatomy in the form of charts or plastic models of the inner organs. Still, there is a certain fascination about looking into the inside of one's own body, something that can usually only be achieved to a limited degree by the use of X-Ray imaging or similar imaging modalities. Those methods are usually not applicable for educational methods, mostly because the devices are too expensive

(a)Using Atlas.   (b)Multimedia.

Figure 2.5 Patient education scenarios.

to use without medical indication and for certain methods there are risks like irradiation or injection of tracer materials.

Patient education is the process by which health professionals and others impart information to patients and their caregivers that will alter their health behaviors or improve their health status.It can improve the patient's understanding of medical condition, diagnosis, disease, or disability. Patient education is an important and often underestimated responsibility of a health professional [161]. It is the responsibility of a doctor to inform and motivate patients to ensure that they understand the diseases, the treatment options available and the consequences of no treatment or non-compliance [45]. This makes it possible for the patients to understand it, eventually allowing them to take responsibility for their own health [2]. Good information and communication increases the patient's possibility to contribute in the decision-making process, leading to higher levels of patient satisfaction, loyalty to the physician and favouring treatment outcomes [68, 2, 26]. Traditionally, patient-health professional communication in a clinical setting comprises of a face-to-face narrative interaction often combined with static images or real-time sketching (see Figure 2.5a). It is inevitable that patients' comprehension of explanations is often lacking due to the complexity of the medical information [26]. Evidence suggests that patients often do not understand what is being said to them when information is given during a medical encounter due to cultural and educational gaps between clinicians and patients [11]. This discourages the patient and leaves them overly reliant on their doctor's advice.

Improving and facilitating the information process and understanding by the patient has been a focus of research in the medical field. Ong et al. [120] presented a literature survey of doctor-patient communication, summarizing the key topics of information exchange, interpersonal relationship building, and shared medical decision making. Wilcox et al. [164] proposed a design for patient-centric information displays to deliver useful information to a patient during an emergency room visit, since the patients are frequently underinformed. Previous work has also demonstrated that technology can positively impact communication. Hsu et al. [66] conducted a longitudinal study the impact of in-room access to electronic health records on doctor-patient interactions during outpaient visits. Recent reviews show growing evidence

Figure 2.6 AnatOnMe in use in a simulated physical therapy consultation. The physical therapist explains a knee injury to his patient using on-body projection of medical imagery.

with regards to the benefit of multimedia tools in enhancement of patients' satisfaction and improvement of knowledge retention [11]. A patient education computer program using 3D multimedia videos is shown in Figure 2.5b. In addition, 3D animations have been suggested, for educating audiences that are preliterate or have limited literacy skills, such as children with mental handicaps [147] and to be useful for patients with a lower learning pace as it leads to a better understanding of the disease [76]. Ni et al. [113] presented the design, development, and evaluation of AnatOnMe, a projection-based handheld device designed to facilitate medical information exchange (see Figure 2.6). Adopting a user-centered design approach, they interviewed medical professionals to understand their practices, attitudes, and difficulties in exchanging information with patients, as well as to identify their workflow, tasks, and design requirements for a technology intervention.

Gonzales and Riek [57] presented an application that personalized the content presented on the device to the patient's diagnosis in an easy-to-understand language, rather than hard-to-understand medical terminology, and encourages patient-physician interaction on the main topical areas of a patient's diagnosis. Ihrig et al. [70] evaluated the view of physicians performing multimedia supported preoperative education within a randomized controlled trial, enrolling 8 physicians and 203 patients, and patients and physicians profited from multimedia support for education and counseling without prolonging patient education.

2.1.3 Rehabilitation exercise

Rehabilitation is the process to regain full function following injury[2]. This involves restoring strength, flexibility, endurance and power and is achieved through various exercises and drills (see Figure 2.7). Rehabilitation is as important as treatment following an injury but unfortunately is often overlooked. Usually physiatrists are involved in the exercise procedure and optimal physical activities are carried out in an effort to reach specific health goods. Its purpose is to return to normal musculoskeletal function or to decrease pain caused by injuries or other health problem.

---

[2] http://www.sportsinjuryclinic.net/rehabilitation-exercises

Figure 2.7 Rehabilitation Exercise. Usually physiatrists are involved in the exercise procedure and optimal physical activities are carried out in an effort to reach specific health goods.

The communication between a medical assistant and a patient is very important during the exercise, e.g. make sure that the patient knows the current state of his/her disability and follows guidelines during the rehabilitation exercise. The rehabilitation process is commonly performed in a rehabilitation center, and the physiotherapist identifies and evaluates the movements and motor functions that are being affected. In order to achieve kinetic gains as quickly as possible, patients are encouraged to perform the same exercise movements several times [103]. Patients' learning by movement repetition is crucial for their successful therapy. The time that the patient expends in the rehabilitation center is relatively short compared with the time he/she spends at home with no supervision [30]. However, there are some issues with home exercises:(i) the patient cannot be motivated by the therapist, and (ii) wrong movements might be performed without timely correction.

Today, interactive solutions for rehabilitation are being developed using different sensors and motion tracking systems including gloves, magnetic, fiducial or infrared markers, video and depth cameras, and VR and AR systems are introduced to motivate the patient and facilitate a more accurate exercise [52, 25]. The activity of patients' muscles can be visualized and the movement is checked based on the contracted muscles [165]. Cho et al. [25] developed a virtual reality proprioception rehabilitation system for stroke patients to use proprioception feedback in upper limb rehabilitation by blocking visual feedback. The markerless tracking feature of Kinect enables a natural user interaction for rehabilitation applications, which alleviates existing issues for patients having difficulty to hold any sensor or marker. Using Kinect, various research are being explored for the development of assistive systems that help interact with patients during their rehabilitation exercises [29, 53, 4, 18].

2.1.4 Intra-operative interaction with medical information

As medical technology is continuously developing, more medical systems are used not only in diagnosis but also during surgery. In a modern Operating Room (OR), the surgeon completes the operation with the help of different medical systems and quite a few interactions have to be performed by the surgical team to control the corresponding medical systems and retrieve the desired information (see Figure 1.1). The key to successful minimally-invasive procedures is immediate access to the patient's medical imaging data during the intervention. In contrast to open surgery, physicians are not able to see their target, surrounding risk structures or their instruments inside the patient and therefore rely heavily on intra-operatively acquired medical
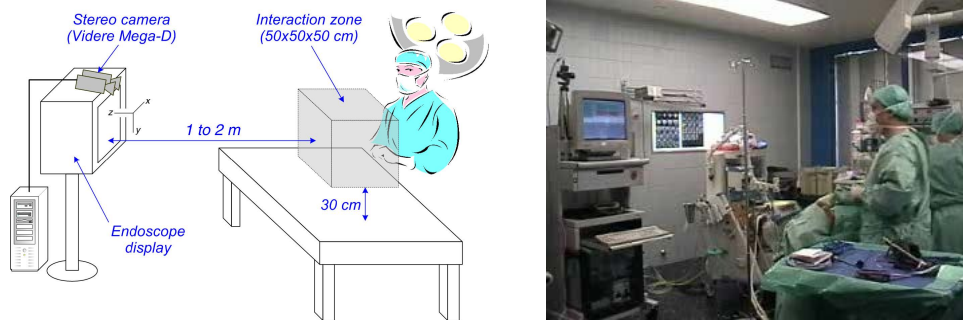
Figure 2.8 Touchless Mouse. A non-contact mouse was developed using a stereo camera, providing the surgeon standard mouse functions to control a single medical system.

images. Because of special working conditions in the OR, i.e. sterility, limited space and time pressure, physicians face challenging human-computer interaction tasks. These tasks include the control of medical image viewers, interactive registration of images, and interaction with medical robots. In clinical routine, sterile covers enable the direct use of interaction devices, e.g. joystick, touchscreen or control panel. In addition, foot pedals are used to control software directly. In contrast, the interaction with software is very often delegated to a non-sterile assistant using speech or gesture commands [69]. It is a big challenge for the surgical team to manage and naturally interact with all the medical information in OR.

Giving the surgeon direct control of the interaction of different medical systems while maintaining sterility within the operating room, has captured the imagination of many research groups, and different solutions have been proposed. One approach is to insert a barrier between the sterile gloves and the non-sterile input device [71]. This solution is very simple but involves certain inherent risks due to the potential damage to the barrier. Computer vision techniques are also being used to introduce interaction methods in the operating room that avoid the need for contact with an input device. A non-contact mouse was developed in [58] using a stereo camera that provided the surgeon standard mouse functions to control a single medical system. Even if they only developed simple gestures to control the cursor movement and clicking, the interactions were not performed smoothly. It also required each medical system to be augmented by a pair of stereo cameras which increases the hardware complexity in the OR (see Figure 2.8). Washs *et al.* [159] developed more sophisticated gestures for image browsing. Quite a few touchless systems [142, 39, 149] for the OR were developed after the Kinect sensor was released, as it has lower barriers and could be easily adapted. Ebert *et al.* [39] presented a system which allowed the control of the open source Picture Archiving and Communication System (PACS) OsiriX by means of gesture and voice commands. The system in [142] helped navigate a predefined stack of MRI/CT images, using a simple constrained gesture set to move forward or backward through the image slices. The work in [149] built compound gestures that combined dominant and non-dominant hand, which were used for selecting, and other particular functions and modes. Once Leap Motion (California, United States) commercially appeared, a custom workstation was also set up in order to manipulate the imaging software by hand gestures [133]. All these systems have a similar setup, since a depth sensor is placed next to the screen to perform outside-in tracking while observing the surgeon to detect the gestures which are subsequently mapped to different functions.
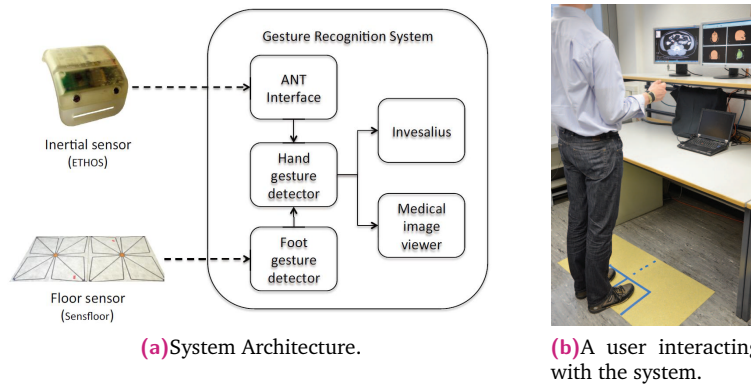
(a)System Architecture.

(b)A user interacting with the system.

Figure 2.9 Touch-less interaction with medical images using hand&foot gestures.
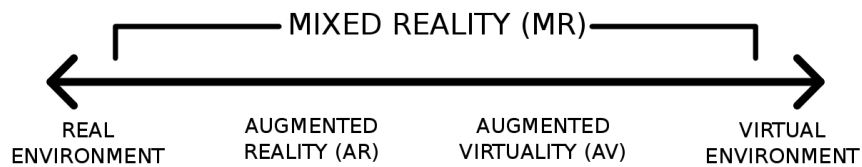
Figure 2.10 Reality-Virtuality (RV) Continuum

Jalaliniya et al. [74] presented user interfaces for gesture-based interaction with medical images via a wristband sensor or capacitive floor sensors (see Figure 2.9). However, interaction with medical images or systems in a surgical setting often extends beyond simple navigation, requiring a much richer set of image manipulation options. Along with the larger gesture sets enabled by this expressiveness comes concern over the learnability of the system - natural user interfaces are not natural any more [116, 139]. Several strategies have been also explored for touchless interaction with medical images, through speech recognition and egocentric interaction. Speech recognition is used in [40], however voice-recognition software involves special challenges in noisy operating rooms and often are exhausting when repeated frequently during surgeries throughout the day. The egocentric interaction paradigm [124] redefines the relationship between the human, computer system, and world, hence a wearable personal prototype for surgeons was developed based on Google Glass [73]. A set of socio-technical issues, or human factors, arose when the above systems were used in the context of an actual surgery, including collaboration, engagement and disengagement.

2.2 Mixed Reality Environment

Mixed reality is the merging of real and virtual worlds to generate new environments and visualizations where physical and digital objects co-exist and interact in real timeand it encompasses both augmented reality and augmented virtuality (AV), as shown in Figure 2.10. MR environment is defined as one in which real world and virtual world objects are presented together, that is, anywhere between the extrema of the RV continuum [104]. Usually the MR environment is created via a MR/AR application, which combines the real world and virtual world through an AR view. MR/AR applications provide a basis for embedding content perceptually within the environment, either by direct augmentation or via a 3D model of it.

Both these application types use the world as the user interface, thereby potentially minimizing the need for spatial inferencing and cue-based triangulation for simple short-distance target finding tasks. The advantages of MR environment are to improve the perception of the ambient information with the corresponding interactive methods [118]. In this section, we first examine the states of augmented reality and Medical AR application, then simply discuss collaborative mixed reality and serious gaming.

2.2.1 Augmented Reality

Augmented Reality supplements the real scene with virtual elements, creating a MR environment. It has been widely used in areas such as training and education, manufacturing and repair, annotation and visualization, robot path planning, entertainment, and military aircraft guidance [87]. In AR, artificial, computer generated content is used to enrich or annotate the real world. AR techniques are inherently egocentric. Augmentations lie either directly in the view of the user (using see-through displays) or on top of a camera view [82]. An AR system has three key characteristics: (a) it mixes real and virtual imagery, (b) registers the digital data to the real world, and (c) provides interactivity in real-time [6]. AR in formal and informal MR environments with an emphasis on affordances, usually utilizes context-aware technologies, which enable participants to interact with digital information embedded within the physical environment [37]. The AR system presents digital media to users after they point the camera at an object (e.g., QR code, 2D target) and creates immersible learning experiences with the physical environment, providing a novel and potentially transformative tool for learning, training, navigation and so on [34, 117].

One of the main challenges is accurate registration, where the virtual content is accurately positioned onto the real world. Two fundamental methods for this are commonly utilized: 1) sensor-based registration and 2) computer vision based registration. Typical computer vision based AR systems use fiducial markers for tracking. Recently, research has focused in markerless environments. These approaches have their pros and cons. Sensor-based registration is globally applicable, but depends on sensor accuracy. For markerless recognition, visual features from areas and targets to be recognized need to be pre-extracted. Another challenge is the correct perception of augmentations, related to environments, capturing, augmentation, display, and individual user differences [85].

Tracking via RGB-D sensors

A RGB-D sensor provides synchronized color and depth images and is frequently employed as an input and tracking device in the AR system. There are RGB-D sensors with different techniques, such as stereo cameras, time-of-flight (TOF) cameras, and structured light. There are several popular sensors available, with different size and power consumption, and the popular devices are Intel RealSense 3D Camera, Asus Xtion LIVE PRO for laptops, and Microsoft Kinect versions 1 and 2 for desktop. Each sensor's specification is shown in Table 2.1.

The Kinect V1 sensor introduced by Microsoft in 2010 is evidently a breakthrough that brings low-cost 3D depth sensing and human body pose estimation technologies to the public [168]. Since Kinect's launch date, it has been exploited in all kinds of computer gaming and mixed reality applications. Hence, we only examine the Kinect V1 for technologies of the depth sensor and human skeletal tracking. The Kinect V1 sensor incorporates several advanced sensing

Table 2.1 Specification of RGB-D Sensors. RealSense is very light, while Kinect v2 is heavier and has much higher power consumption.

|	RealSense	Xtion	Kinect V1	Kinect V2
Weight (pound)	0.077	0.5	4	4.5
Size (inch)	5.2x0.25x0.75	7.1x1.4x2	11x2.3x2.7	9.8x2.7x2.7
power	2.5W USB	2.5W USB	12.96W	115W
technology	TOF	structured light	structured light	TOF
depth resolution	628x468	640x480	640 x480	512z424
color resolution	1920x1080	640x480	640x480	1920x1080
tracked objects	face and hand	human body	human body	human body
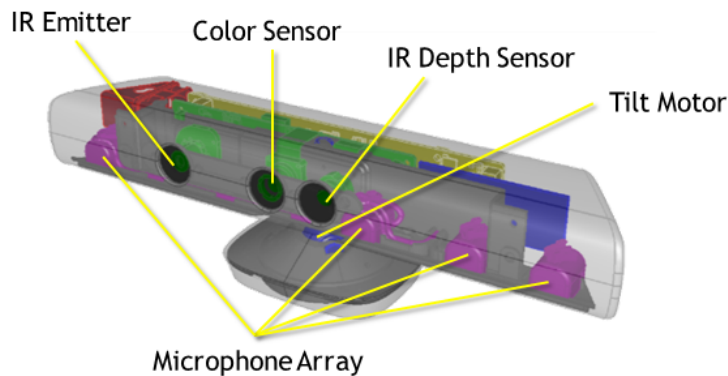
Figure 2.11 Kinect for Windows Sensor Components. There is an RGB camera, an infrared emitter and IR depth camera, a multi-array microphone, and a 3-axis accelerometer.

hardware. Most notably, it contains a depth sensor, a color camera, and a four-microphone array that provide full-body 3D motion capture, facial recognition, and voice recognition capabilities (see Figure 2.11). Here we only focus on the main technologies, depth sensor and human skeletal tracking, and some technologies to collect more personal information.

Depth sensor Figure 2.11 shows the arrangement of the infrared (IR) projector, the color camera, and the IR camera. The depth sensor consists of the IR projector combined with the IR camera, which is a monochrome complementary metal-oxide semiconductor (CMOS) sensor. The depth-sensing technology is licensed from the Israeli company PrimeSense[3]. Although the exact technology is not disclosed, it is based on the structured light principle. The IR projector is an IR laser that passes through a diffraction grating and turns into a set of IR dots. The relative geometry between the IR projector and the IR camera as well as the projected IR dot pattern are known. If we can match a dot observed in an image with a dot in the projector pattern, we can reconstruct it in 3D using triangulation. Because the dot pattern is relatively random, the matching between the IR image and the projector pattern can be done in a straight forward way by comparing small neighborhoods. The depth values from Kinect do not have the same accuracy. Depth is determined through triangulation, similar to stereovision. The depth error increases with the distance squared. A point cloud can be generated from the depth image to implement tracking and 3D reconstruction. The influential KinectFusion system [112] demonstrated real-time depth camera tracking and dense scene reconstruction by registering incoming depth images to a volumetric representation of the scene. Schulman

---

[3]www.primesense.com
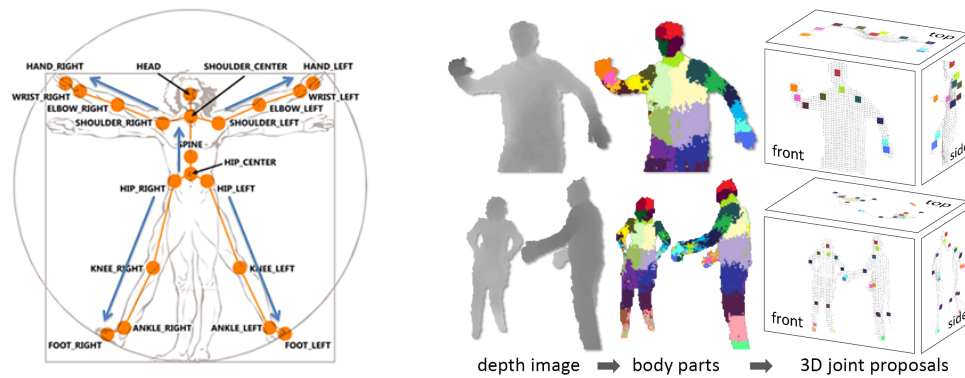
(a)Using a skeletal representation of various body parts.

(b)Kinect uses per-pixel, body-part recognition as an intermediate step to avoid a combinatorial search over the different body joints.

Figure 2.12 Skeletal tracking. From a single input depth image, a per-pixel body part distribution is inferred. (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals). Local modes of this signal are estimated to give high-quality proposals for the 3D locations of body joints, even for multiple users.

et al. [138] introduced an algorithm for tracking deformable objects from a sequence of point clouds, which is based on a probabilistic generative model that incorporates observations of the point cloud and the physical properties of the tracked object and its environment. Zhou and Koltun [169] presented an approach for tracking camera pose in real time given a stream of depth images via useful contour cues.

Human skeletal tracking The innovation behind Kinect hinges on advances in human skeletal tracking. The demands for commercially viable skeletal tracking are enormous. Simply put, skeletal tracking must ideally work for every person on the planet, in every household. A dauntingly high number of dimensions describe this envelope, such as the distance from the Kinect sensor and the sensor tilt angle. Entire sets of dimensions are necessary to describe unique individuals, including size, shape, hair, clothing, motions, and poses. Household environment dimensions are also necessary for lighting, furniture and other household furnishings, and pets. However, in this skeletal tracking technology a human body is represented by a number of joints representing body parts such as head, neck, shoulders, and arms(see Figure 2.12a). Each joint is represented by its 3D coordinates. The goal is to determine all the 3D parameters of these joints in real time to allow fluent interactivity and with limited computational resources. Rather than trying to determine directly the body pose in this high-dimensional space, the Kinect team met the challenge by proposing per-pixel, body-part recognition as an intermediate step (see Figure 2.12b) [140]. With the 3D human skeletal tracking algorithm, it enables quite a few personalized AR applications and interaction between users and visual and physical objects without the need of touching and markers.

Personal information As technologies improve, the perception and interaction with the digital information is more and more natural. These include the user's personal behaviors, thus personal information should be collected to improve the user's experience in the MR environment. The user is the final stage of the perceptual pipeline and the perception of the digital content presented in the MR environments can be highly influenced by individual differences between users. These differences may require noticeable modifications in the way the virtual
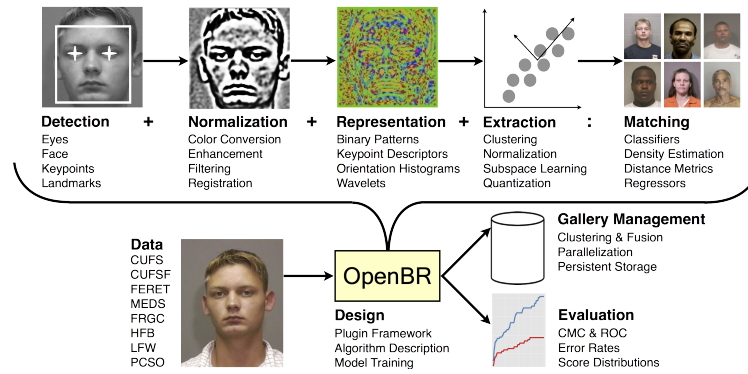
Figure 2.13 Overview of OpenBR Capabilities.

information in presented [85]. Personal information usually includes gender, age, weight, size, and body shape. Based on the skeletal tracking, computing has to be done to detect more information from the image device. The Open Source Biometric Recognition (OpenBR)[4] is a communal biometrics framework supporting the development of open algorithms and reproducible evaluations [84]. The algorithm is based on the statistical learning methods and used in the field of demographics and aging using face recognition concepts. (see Figure 2.13). It is easy to use and fairly precise with 92.8% accuracy of the gender estimation, 9.9 years RMS error of males and 12.8 years RMS error of female. In addition, ReconstructMe[5] is a powerful and affordable real-time 3D reconstruction system. Quite a few personal information can be detected from the 3D reconstruction of the human body.

Magic Mirror concept

The concept of a MR magic mirror has been proposed to address the challenges in AR related to perception [59]. A mirror is the most popular way to learn about ourselves. We all spend a lot of time looking in the mirror. It is first thing we look at in the morning to see what we need to do, such as shaving, putting on make-up, and tooth-brushing. We also use it to check if we should go on a diet and figure out what to wear. The magic mirror is a user interface technique that mimics a normal mirror and presents non-physical visual feedback in addition to the normally optical effect. Here the user stands in front of a screen and via a camera, the image of the user is shown on the screen such that it acts like a mirror.

The concept of a MR magic mirror has been shown previously, mainly for advertisement and marketing. Previous systems augmented onto the user virtual shoes [42, 94], shirts [41] and knight's armors [47]. For tracking the position of the user, previous systems have used tracking markers or a tracking shirt with a rectangular highly textured region [64]. For shoes, a vision-based approach has been presented [42]. Based on the tracking technologies of Kinect, Bodymetrics[6] created a system in which a 3D avatar is generated for each user whose body shape and size are measured with Kinect. The user can control the 3D avatar through Kinect and see how the virtual clothes look on the avatar in different poses. Fitnect[7] also provides a system for users to try on different clothes, but instead of using an avatar, the

---

[4]http://openbiometrics.org
[5]http://reconstructme.net/reconstructme-sdk/
[6]http://www.bodymetrics.com/
[7]http://www.fitnect.com/

virtual clothes are directly displayed on the user's body in augmented videos. Facecake[8] gives a similar augmented reality system called Swivel in which users can virtually try on a variety of fashion products including handbags, clothes, necklace and sunglasses.

All the medical knowledge is about ourselves, but we cannot look through our cloth and skin to directly observe what is inside of the human body. While a magic mirror system restricts the motion of the user, it would allow an inexpensive and robust solution for medical education and training. Blum et al. [15] presented a magic mirror for teaching anatomy. The system uses a depth camera to track the pose of a user standing in front of a large display. A volume visualization of a CT dataset is augmented onto the user, creating the illusion that the user can look into his/her body.

2.2.2 Medical AR Applications

Many commercially available systems use virtual reality (VR) for medical education and psychomotor skills training. These systems have indeed proven to be valid and are of use in clinical training and teaching [89, 151, 157]. In addition to the advantages of VR, AR systems also have the advantage that information can be embedded and/or superimposed upon reality. This allows for a more close-to-reality presentation of medical knowledge and offers opportunities for more interactive learning and training. The user can also spatially relate virtual objects to the reality.

Education

The General Medical Council recently proposed standards for effective teaching and learning of medical students [56]. They stated that: "...medical schools should take advantage of new technologies . . . to deliver teaching." Augmented reality research has matured to a level that its applications can now be found in both mobile and non-mobile devices [7] and research on AR has also demonstrated its extreme usefulness for increasing student motivation in the learning process [21, 35].

In medical, an AR system for patient-doctor communication has recently been presented [113]. This system used a hand-held projector to project anatomy onto the skin of the patient. While this system is relatively inexpensive, it could only display anatomy which is close to the skin in correct perspective. Furthermore, it did not automatically track the position of the projector and the patient. The doctor must then figure out manually how to position the projector and the patient in such a way that the anatomy is augmented correctly.

Understanding human anatomy is essential for practicing medicine since anatomical knowledge supports the formulation of a diagnosis and communication of that diagnosis to patient and colleagues [51]. AR technology could offer an additional teaching method for anatomy education, depending on how it is implemented. Strong points are the visualization capabilities including the 3D rendering of anatomical imagery. By combining computer models of anatomical structures with custom software we can showcase to students new ways of interacting with anatomy that could not be achieved during cadaveric dissections or in static images and diagrams for increasing their learning satisfaction [7, 100, 77]. Augmented reality systems for visualization of anatomy have been shown before. [32] presented a system

---

[8]http://www.facecake.com/swivel/

that augmented a 3D model of anatomical airways onto a patient phantom using a head mounted display (HMD). Another system that used a HMD to visualize human anatomy onto a phantom has been shown by [80]. Their system allows students to open the abdomen of the phantom and visualizes different organs on the phantom. Other sensory experiences could be implemented as well, such as tactile feedback. AR provides real-time manipulation of these visualizations and direct feedback to students. With that, AR technology could comply with some affordances of traditional dissection classes. Several AR systems have already been developed specifically for anatomy education [152, 24, 17]. Blum et al. [17] described 'Mirracle' which was an AR system that can be used for undergraduate anatomy education. The set-up of that system is as follows. The trainee stands in front of a TV screen that has a camera and the Kinect attached to it. The camera image of the trainee is flipped horizontally and is shown on the TV screen, mimicking a mirror function. Part of an anonymous patient CT dataset is augmented on the user's body and shown on the TV screen. This creates the illusion that the trainee can look inside their body.

Training

A gesture-based user interface allows real-time manipulation and visualization of the CT data. The trainee can scroll through the dataset in sagittal, transverse and coronal slice mode, by using different hand gestures [81]. Consistent with modern learning theories advocating for active participation with immediate application of knowledge [128], virtual interactive presence (VIP) technology enables learners to immerse themselves in a real surgical environment. The free-form feedback from the participants in a study organized by Ponce et al. [128] highlights the potential for an improved educational experience through greater involvement of a resident surgeon. Resident surgeons expressed greater comfort with having the attending surgeon outside the operative room because they experienced a feeling of autonomy while still having sufficient oversight. The attending surgeon commented that instruction was improved since virtually "touching" anatomy helped to better communicate degrees of motion. The use of the system to convey hand gestures or point to a particular anatomy was thought to be more effective. Both the resident and attending surgeons thought that the identification of anatomy was better with use of this technology. This pilot study revealed that the VIP technology was efficient, safe, and effective as a teaching tool. The attending and resident surgeons agreed that training was enhanced, and this occurred without increasing operative times. Furthermore, the attending surgeon believed that this technology improved teaching effectiveness [128].

Real-time navigation

Surgeons have welcomed AR to assist them only in specific work-flow tasks and in surgical planning. Consequently, it was discussed that augmented reality is a promising solution to improve the accuracy of surgical procedures, decrease the variability of surgical outcomes, reduce trauma to the critical anatomical structures, increase the reproducibility of a surgeons' performance, and reduce radiation exposure [111]. Medical augmented reality has been successfully applied in various disciplines of surgery, such as neurosurgery, orthopaedic surgery, and maxillofacial surgery [43]. Navab et al. [110] extended a standard mobile C-arm fluoroscope by a video camera and mirror construction, delivering an intuitive real-time intra-operative visualization of X-ray images co-registered with live video. Stetten et al. [141] presented a Real-Time Tomographic Reflection (RTTR) as an image guidance technique for needle biopsy. RTTR is a new method of in-situ visualization, which merges the visual outer

surface of a patient with a simultaneous ultrasound scan of the patient's interior using a half-silvered mirror. The ultrasound image is visually merged with the patient, along with the operator's hands and the invasive tool in the operator's natural field of view. Fichtinger et al. [48] proposed an intraoperative CT-based medical AR system for visualizing one CT slice onto the patient in-situ using a specific arrangement of a half transparent mirror and a monitor rigidly attached to a CT scanner. Feuerstein et al. [46] augmented laparoscopic video images with intraoperative 3D cone-beam CT by tracking both C-arm and laparoscope using the same external optical tracking system. Wendler et al. [163] proposed freehand SPECT to augment the 3D reconstruction of radioactive distributions via live video by using a calibrated optical tracking and video camera system. In general, AR has not been widely accepted in clinical practice due to its cumbersome system setups, the requirement of a line-of-sight for tracking, on-site calibration, and a lack of real-time accurate registration between medical data, patient and surgical instruments. Additionally, real-time non-rigid registration methods, display techniques and user-interfaces must be addressed and delivered for surgeon acceptance inside the operating room.

2.2.3 Collaborative mixed reality

The world is becoming more complex, so problem solving often needs a team of experts working together. To do this effectively there is a need for collaborative tools. MR technology allows users to view and interact in real time with both physical objects and virtual information seamlessly, hence MR systems can also be used to create unique collaborative experience. MR is ideal for collaborative interfaces to develop radically new types of collaborative experiences. MR technologies could be used to merge the shared perceived realities of different users, as well as enrich each user's own individual experience in collaborative work [95]. For example, 3D virtual information is shared and operated by co-located users, or remote assistance is implemented by annotating the live video of the remote worker. The main objective is to augment the face-to-face collaborative experience and enable remote users to have a feel that they are co-located [96].

Several studies have explored the effectiveness of using MR for complex tasks. For example, MR has been shown to improve the effectiveness of individual manual assembly tasks [8] and the performance time and mental effort in collaborative design tasks [162]. There are many examples of collaborative MR applications. The Studierstube system [146] targeted face-to-face presentations and allowed users to walk around virtual 3D scientific data. Höllerer et al. [65] enabled indoor AR users to visualize the locations and paths of outdoor AR users, and all users to create shared annotations. Research on using MR for collaboration among crime scene investigators indicated that using shared visual feedback promoted mutual understanding, led to consensus, and supported hypothesis testing [127]. Datcu et al. [31] presented an AR system and the results showed that MR environment could successfully support information exchange in teams operating in the security domain. Gauglitz et al. [55] introduced a tablet-based system that incorporates a touchscreen interface through which a remote user could navigate a physical environment and create world-aligned annotations for supporting remote maintenance (see Figure 2.14). In summary, MR technology is becoming mature enough to support a variety of collaboration scenarios. However, there are still a number of open issues that need further research. One major issue is to remove functional seams and cognitive seams for the user during the cooperation.

Figure 2.14 In touch with the remote world. A remote user points out an element in the environment (here: a car's engine bay) by drawing an outline around it. The annotation is world-stabilized and displayed in augmented reality to the local user (bottom left), who is holding a tablet.

2.2.4 Serious gaming

The concept of "serious gaming" is often used to describe the use of digital gaming technology to address a specific set of learning objectives, or behavioral goals. As such, they often seek to build upon the increasingly pervasive role games play as an entertainment medium to provide an engaging and entertaining way to communicate educational content [137]. Using games to teach skills used in real life or to convey knowledge has in the last years become generally known under the term serious gaming. The goal is to teach the player something practical in an entertaining and "fun" way.

More and more researchers and educators have made an effort to integrate serious games into classrooms [28]. Serious games create an immersible context wherein students are allowed to experience things and repeat experimentation that are unlikely to be realized in their daily lives. It is believed that their use in educational settings can increase the probability of providing students with authentic learning [22]. Moreover, it is hoped that the use of such games can improve student learning motivation [122], facilitate knowledge acquisition [105], increase task engagement [3], and foster specific abilities such as problem solving and collaboration [136] by appropriately visualizing abstract ideas and key principles of given topics in the game environment [23]. In the context of anatomy education, Arcade[9] offers a large collection of anatomy-related arcade games. The games range from crosswords and jigsaw puzzles to more sophisticated ideas like "Poke-A-Muscle" game where the user has to select the right muscles as quickly as possible. These games are accompanied with educational videos to provide a base knowledge that can be tested in the games. The games certainly are not suitable for educating medical students, but they are a fun way to learn basic human anatomy. A recently successful game is the Surgeon Simulator 2013[10]. Here, players have to perform different surgeries from the viewpoint of the acting surgeon. In a

---

[9]http://anatomyarcade.com/
[10]http://www.surgeonsim.com/

3D environment where organs and tools behave in a physically believable way, players have to remove various organs to eventually perform a heart transplant. Another free on-line service more tailored towards patient education is Surgery Squad[11]. Here, players can perform different types of surgery in the browser. They are guided by a virtual surgeon explaining the medical background to the surgery and every step involved. They interactively learn about the surgical procedure by listening to the instructions the virtual surgeon gives and subsequently performing them, mostly by using mouse clicking or dragging over certain areas. These interactive surgeries have been created to educate patients or other interested people on how the surgeries or other procedures are being performed, with a minimal baseline of anatomical knowledge.

Serious games can also be developed for rehabilitation training in MR environments. The gaming system offers a motivational component to the patients in performing exercises, introducing entertainment or distracting them from the potential pain. In addition, the camera sensor can monitor the patient and provide correction hint/feedback using game elements during rehabilitation exercise. It has the potential to be deployed to the patient household and eliminate the frequent visits of patients to rehabilitation clinics.

2.3 Natural user interface using gestures

Natural User Interface (NUI) is an emerging computer interaction methodology which focuses on human abilities such as touch, vision, voice, motion and higher cognitive functions such as expression, perception and recall. NUI seeks to harness the power of a much wider breadth of communication modalities which leverage skills people gain through traditional physical interaction. NUI is part of the larger field of Human Computer Interaction (HCI) which is the design and implementation of interactive computing systems that users can interact with. The user's formulation of the desired task needs to be articulated in the input-language. The task is phrased in terms of certain psychological attributes that highlight the important features of the domain for the user which, if mapped clearly onto the input language, simplify the articulation of a task. Direct manipulation can also facilitate the articulation [36]. The responses of the input are translated to stimuli for the system. Once a state transition has occurred within the system, the execution phase is completed and the evaluation begins by translating the system's responses into stimuli for the output component. Finally, the response from the output is translated to stimuli for the user.

The traditional user interaction consists of the direct manipulation of graphic objects, such as icons and windows, using some type of pointing device. Mouse and keyboard are widely used and well-known for these applications, and recently, users can interact more intuitively using touch-screens. However, the main limitation for these techniques is that the user has to go and grab the device or touch it. It is widely believed that as the computing, communication, and display technologies progress further, the existing user interaction solutions may become a bottleneck for the effective selection and retrieval of desired data. Recognizing gestures for interaction can help achieve the ease and naturalness desired for all human computer interaction. Users generally use hand gestures for expressing their feelings, communicating and notification of their thoughts. The use of hand gestures provides an attractive and natural

---

[11]http://www.surgerysquad.com/

alternative to some traditional cumbersome devices for human computer interaction. There has been growing interest in the development of new approaches and technologies for bridging the human-computer barrier. Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with our computing devices [130]. Gestural interaction is the new excitement in the halls of industry. Advances in the size, power, and cost of microprocessors, memory, cameras, and other sensing devices now make it possible to control by wipes and flicks, hand gestures, and body movements. A new world of interaction is here: The rulebooks and guidelines are being rewritten, or at least, such is the claim. And the new interactions even have a new marketing name: natural, as in "Natural User Interface". Numerous methods for vision-based hand gesture recognition have been developed and evaluated. There are many applications about information processing [108, 88] and visualization as the premier solutions, followed by robotics and sign language [145, 121]. Since the release of commercial depth sensors, there have been numerous inspiring successes in finger tracking and hand gesture recognition for human computer interaction [170, 86]. Ruppert et al. [134] presented a touchless gesture interface which enables the control of a mouse pointer of a computer using *InVesalius*.

Challenges with normal hand gestures   Swindell[145] indicated an important problem that occurs in many scenarios -" *the user is aware of where the computing device is but could not directly begin interaction with the device since excess time is lost to build-up connection with the device*". It is known that the geometry of the hand and the mapping from the hand movement to the selected object are hard to reconstruct [129]. In addition, most gestures are neither natural nor easy to learn or remember. Few are innate or readily predisposed to rapid and easy learning. Even the simple head shake is puzzling when cultures intermix. Westerners who travel to India experience difficulty in interpreting the Indian head shake, which at first appears to be a diagonal blend of the Western vertical shake for "yes" and the horizontal shake for "no." Similarly, hand-waving gestures of hello, goodbye, and "come here" are performed differently in different cultures. To see a partial list of the range of gestures used across the world, look up "gestures" and "list of gestures" in Wikipedia. More important, gestures lack critical clues deemed essential for successful human-computer interaction. Because gestures are ephemeral, they do not leave behind any record of their path, which means that if one makes a gesture and either gets no response or the wrong response, there is little information available to help understand why. The requisite feedback is lacking. Moreover, a pure gestural system makes it difficult to discover the set of possibilities and the precise dynamics of execution. These problems can be overcome, of course, but only by adding conventional interface elements, such as menus, help systems, traces, tutorials, undo operations, and other forms of feedback and guides.

2.3.1 Pointing gesture

Pointing gestures are fundamental to human behavior [97] and are used consistently across cultures [99]. The gestures begin at an early developmental stage [20] and let humans reference proximal objects as well as abstract concepts in the world. Today, pointing gestures are not only part of our gestural language but are inherently used for interaction [109]. Regarding the ray's origin, pointing techniques can be classified into two groups: *hand-rooted*, where the ray originates at the user's hand, and *eye-rooted*, where the ray starts at the user's eyes. Whenever a *hand-rooted* technique is used, the objects that are along the pointing ray

might differ from those that the user focuses on [5]. The more interesting scenario is the latter. A number of studies have demonstrated that pointing gestures result in quite effective interactions in virtual environments [5] and with very large, high resolution displays [158].

Virtual Pointer The *eye-rooted* pointing technique is developed as a virtual pointer for object selection in virtual environments. The simplest approach, occlusion selection, used a selection ray which was defined as beginning from the user's eye position and pointing to the user's hand position [50, 126]. Forsberg et al. [50] defined the origin to be one of the positions of the user's two eyes or an average of them. Alternatively, Argelaguet et al. [5] proposed a hybrid technique in which the ray started from the user's eye position and its direction was controlled by the hand orientation. All eye positions in these papers was known as it was manually defined during scene rendering.

Distant Pointing The *eye-rooted* pointing technique also allows users to perform interaction with a remote display by directly pointing at the object they want to interact with. The remote pointing position is defined by the ray casting from the user's eye to their fingertip. Liang and Green [91] first presented a highly interactive 3D model by computing a ray going from the user's eye through a hand-held bat while two tracking devices monitored the user's head and the bat. Nickel and Stiefelhagen [114] explored computer vision methods to recognize the body, hand, and finger positions in 2D/3D space to calculate the pointing direction. A vision-based human-computer interaction system was presented by Reale et al. [131], integrating components using multiple gestures, including eye gaze, head pose, hand pointing, and mouth motion. Vogel and Balakrishnan [158] employed the *Vicon* motion tracking system to analyze the hand-based ray casting and relative pointing with clutching. Banerjee et al. [9] developed *MultiPoint*, a set of perspective-based remote pointing techniques that allowed users to perform remote manipulation of graphical objects on large displays while extra cameras and motion trackers were employed to detect the eye and finger. Huang et al. [67] developed *Dart-it*, a lightweight system that enables perspective-based remote direct-pointing on any remote display using only one RGB-D camera. However, since the recovered pointing ray requires reliable eye and fingertip tracking, all the above methods require external camera or motion trackers for their real-world deployment.

2.3.2 Natural interaction in an egocentric setting

A few authors have tried to estimate the user's eye position and recover the *eye-rooted* pointing geometry to enable interacting with ambient media and objects in the real world. One existing technique consists in positioning a camera next to the monitor or on the ceiling to observe the user and implement the outside-in tracking. In this case, the computing device calculates the eye position by detecting the user's gaze and face or wearable trackers. Unfortunately, the user has to stand within a limited working zone defined by the camera, as shown in Figure 1.2(a). As a solution, authors in [106, 62] suggest that the camera be kept to a similar view in an egocentric setting.

The interaction between the user and any device can begin with a signal sent using their hands or their eye-gaze estimated within FOV of the wearable camera.

Gesture interaction   Several articles focused on natural gesture interaction methodologies, in which sensors detect and track the pose of the user's arm, hands or fingers to recognize the gesture interaction. De Campos et al. [33] proposed a framework combining a coarse-to-fine method for shape detection and 3D tracking that could identify pointing gestures and estimated their directions in the wearable domain. A gesture recognition algorithm was developed for an interactive museum using 5 different gestures by García-Rodríguez and García-Chamizo [54]. Colaço et al. [27] presented *Mime*, a compact, low-power 3D sensor for unencumbered free-form, single-handed gestural interaction with HMDs.

Relative pointing   Often a visual feedback is provided to hint where the potential object of interest might be. The user has to learn how to use the natural gesture movement to control the feedback cursor. Mistry et al. [106] proposed to bring information out into the tangible world by using a tiny projector and camera mounted on a hat in order to help the user indirectly point to and manipulate the information in the real word.

Natural pointing   Solutions for natural pointing in an egocentric setting in different scenarios have been previously proposed. (1) In the case where the object is reachable, the sensor can directly detect which object the user is pointing at without the need for defining a pointing ray and/or its origin. *OmniTouch* [62] is a wearable depth-sensing and projection system that enables interactive multi-touch applications on everyday reachable surfaces. SemanticPaint [156] allows users to segment the scene simply by reaching out and touching any desired object or surface. (2) Ha et al. [61] developed a video-see-through augmented reality (AR) system allowing the user to manipulate virtual 3D objects with their bare hand. A bare hand-based interaction with reachable virtual objects in an egocentric viewpoint was also developed by Jang et al. [75]. Here, the user's view was controlled by the computer and the user's eye position was defined during the rendering procedure. Hence, the origin of the pointing ray was manually defined. (3) *Pupile* [83] provides an accessible, affordable, and extensible open source platform for eye tracking and gaze-based interaction. The eye position is taken as the origin point and the pointing direction is also directly estimated from the eye gaze. (4) Meng et al. [101] defined a virtual eye center as the pointing origin after moving the fingertip along the pointing rays. The disadvantage of this method is that the user has to keep his/her head still and the calibration procedure is not user friendly.

Usually the head, eye, hand, and/or finger are detected and tracked to recover the pointing ray. Also, the interaction techniques are implemented using different techniques, for example: (i) with or without visual feedback, (ii) based on direct or relative positioning, (iii) acting on dynamic or static scene, (iv) functioning within reachable or unreachable objects in the work space, and (v) extended in the real world or only within the virtual world. Table 2.2 summarizes the characteristics of the above-mentioned methods based on the objects to detect and track and the properties of the interaction techniques.

Table 2.2 The objects to track and the properties of the interaction in different works

Work	Objects to Track				Properties of the Interaction					
	Head	Eye	Hand	Finger	Visual Feedback	Direct Positioning	Dynamic Scene	Unreachable Objects	Real World	Virtual World
[126]	✗	✗[a]	✓[b]	✓[b]	✓	✓	✓	✓	✗	✓
[5]	✗	✗[a]	✓	✓	✓	✓	✓	✓	✗	✓
[91]	✗	✗[a]	✓[b]	✓[b]	✓	✓	✗	✓	✓	✓[c]
[114]	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗
[9]	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗
[67]	✗	✓	✗	✓	✗	✓	✗	✓	✓	✗
[33]	✗	✗	✓	✓	✓	✗	✓	✓	✓[d]	✓[d]
[27]	✗	✗	✓	✓	✓	✗	✓	✓	✓	✓
[106]	✗	✗	✗	✓	✓	✗	✓	✓	✓	✗
[62]	✗	✗	✓	✓	✗	✓	✓	✗	✓	✗
[61]	✗	✗[a]	✓	✓	✓	✓	✓	✓	✓[e]	✓
[75]	✗	✗[a]	✓	✓	✓	✓	✓	✗	✗	✓
[83]	✗	✓	✗	✗	✓	✓	✓	✓	✓[d]	✓[d]

[a] The eye position is defined during the rendering procedure.
[b] Special tool is tracked instead of hand and finger.
[c] 3D virtual world is shown in 2D display.
[d] The method could be used in this scenario with proper hardware.
[e] The real world is shown via video-see-through HMD.
[f] The eye position is calculated via the tracking of the finger and scenario.

Personalized Augmented Reality

3

Providing adequate learning experience to different learners is a challenging issue as the traditional learning systems generally cannot adapt content to suit the individual learner's needs. Personalization for promoting a multi-modal learning environment is a growing area of interest, such as the development of user modeling and personalized processes which place the student at the center of the learning development. AR systems can present a virtual representation of the subject matter and create a direct connection between the information the user wants to learn and their own body and activities at the same time. Hence, it could help understand and memorize complex information, and either supplement conventional learning or even supersede it altogether. Previous AR/MR systems on visualization of anatomy used expensive systems involving HMDs or markers. Tobias Blum from our chair presented this great original idea to create an excellent Magic Mirror system for anatomy education, by employing a Kinect for tracking and a TV display for visualization to generate an AR view [16]. The system presented is both inexpensive and easy to use. It takes advantage of the Magic Mirror concept to present medical information. It presents medical data augmented onto the user's body and shows additional 2D and 3D information according to the need. The Magic Mirror concept provides the user "the superman ability" to look into his/her own body. It enables the medical information to be perceived naturally linking to a real human body. Natural gesture is chosen as the interaction methodology, and interaction with the AR view of the user's own body provides a personalized perception in the Magic Mirror framework. Blum passed all his works about the Magic Mirror system to me and it was my great honor to continue this excellent project. Based on his works, I redefined the software framework, proposed solutions to improve the accuracy of the registration, and developed different medical applications based on the framework.

The Magic Mirror framework with medical information is firstly presented in Section 3.1, including the hardware setup, the system framework and some important system features. An example application for anatomy education is created using the proposed framework, and a survey with 72 medical students was done to evaluate the conception and find the direction to improve the framework. In Section 3.2, the user specific information is collected to improve the registration of the AR view and enhance the personalized perception. Section 3.3 takes the advantage of the interactive mixed reality to generate a personalized learning procedure, and systems are developed and evaluated for anatomy leaning, especially for muscle learning. At last some serious games are developed in the Magic Mirror framework for public education and rehabilitation scenarios.

3.1 Magic Mirror framework

The Magic Mirror [59] is a user interface technique that mimics a normal mirror and presents non-physical visual feedback in addition to the normally optical effect. Here the user stands in

front of a screen and via a camera, the image of the user is shown on the screen such that it acts like a mirror. While previous systems have augmented objects onto the user, this system extends the Magic Mirror concept for medical education and rehabilitation. It creates the illusion that the user can look inside their own body. A depiction of this can be seen in Figure 3.1. To achieve this visualization, the Magic Mirror augments the volume visualization of a CT dataset onto the user. To allow a correct augmentation of the medical data, the pose of the user is tracked. The Magic Mirror concept came out as a framework and it generates a personalized perception of the medical information for every user. In addition, the framework takes the user's natural gesture as input to create an interactive MR environment.

Knowledge about human anatomy is an important issue for everyone working in the field of medicine. It is also an important part of the general education and relevant for many other professions related e.g. to health-care or sports. Human anatomy is very complex and it does not only involve knowledge about a single organ, but also about issues such as chemical processes, human motion and spatial relations inside the body. Therefore, teaching human anatomy is very difficult and often large effort is spent on teaching it e.g. by letting students perform dissection courses, creating illustrations and plastic models of anatomy or by utilizing 3D computer graphics. Our Magic Mirror framework is firstly employed to display anatomical structures overlaid onto the body of the user to intuitively teach human anatomy.

3.1.1 Magic Mirror prototype

The augmented reality Magic Mirror prototype focuses on a few important organs of the abdomen, namely the liver, lungs, pancreas, stomach, and bones. The system prototype has a mirror-like effect to the user by projecting a "looking glass" on the body and displays the skeleton of the user, rendered from CT data and anatomy 3D model (see Figure 3.1). The framework tracks users' movements using a depth camera and an algorithm to detect the pose of the user from the depth image. This is realized using the Microsoft Kinect which was originally developed to allow controlling computer games by motion. By using the Magic Mirror metaphor, the user is led to believe that he or she is able to look inside their own body. At the same time, medical information (CT, MRI data and a fully segmented dataset of cross-sectional photographs of the human body) are augmented in real-time. The current system also allows visualization of static anatomy on the user and offers a simple user interface to select CT, MRI or photographic slices [16, 111].

Hardware setup

An illustration of the hardware setup can be seen in Figure 3.2a. The first component of the system is a display device. In different setups of the system large TV screens or video projection onto a planar surface has been used. The second component is a color camera, which is mounted next to the display surface and which is looking at the user and perceive the visual information. The third component is a depth sensor which is placed next to the color camera and which has a similar field of view and viewing direction as the color camera to collect the user's skeleton information. The current system uses the Microsoft Kinect V1, which consists of a color and a depth camera that are assembled into one housing (see Figure 2.11 and Figure 3.2b). The depth sensor is an infrared camera that uses structured light, which is emitted by an additional infrared projector to estimate depth values for each pixel. The

Figure 3.1 The system prototype has a mirror-like effect to the user by projecting a "looking glass" on the body. By using the Magic Mirror metaphor, the user is led to believe that he or she is able to look inside their own body.

user's skeleton and personal information can be generated from the Kinect sensor based on the machine learning. The system employs the color camera to create a mirror-like effect to the user, and all the non-physical visual feedback is generated based on the user's skeleton and personal information via rendering the corresponding medical information onto the human body.

Software framework

The system framework of the prototype is illustrated in Figure 3.3. To access the Kinect the system uses Microsoft Kinect SDK or OpenNI[1], which is an open source software framework that allows retrieving color and depth images from the Kinect. The depth sensor is used for two purposes. First, the depth values are projected to the color image providing depth information for each pixel in the color image. Second, a skeleton tracking algorithm uses the depth image to track the pose of multiple joints of a user who is standing in front of the camera. For skeleton tracking the Magic Mirror uses NITE, a software by PrimeSense, that performs gesture recognition and skeleton tracking based on depth images. NITE can be used with the Kinect through the OpenNI framework.

The Magic Mirror AR view is based on the information perceived via Kinect and corresponding medical information. As shown in Figure 3.3, there are four modules: Skeleton & Personal Information Processing is important to achieve personalized perception. Color & Depth Image Processing is the basic module to generate the mirror-like effect and merge the non-physical visual outcome. 3D Medical Information Rendering processes the corresponding 3D medical images or models and generates virtual elements for the Magic Mirror AR view and it can employ OpenGL and OpenGl Shading Language, and any other 3D rendering libraries, e.g.
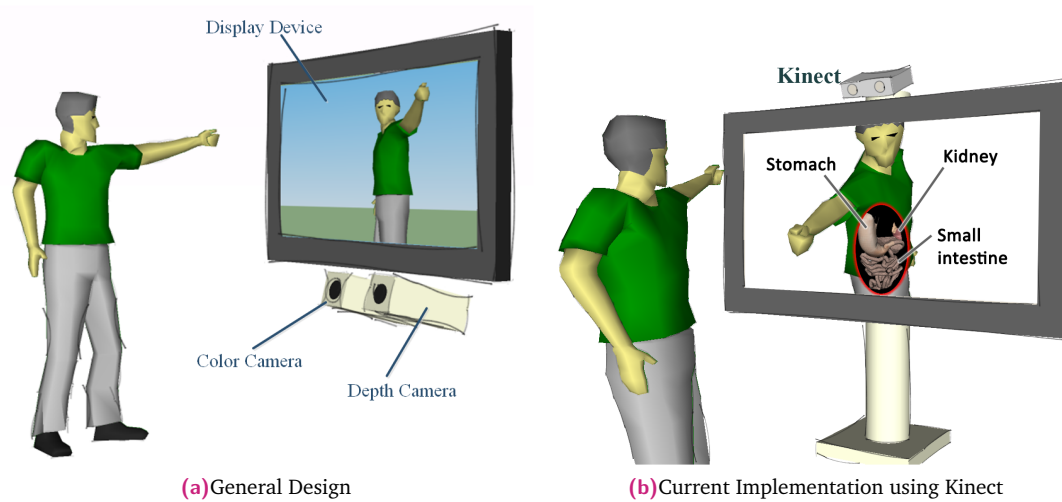
---

[1] https://en.wikipedia.org/wiki/OpenNI

| (a)General Design | (b)Current Implementation using Kinect |

Figure 3.2 Magic Mirror Hardware setup. (a) The system consists of a display, a color camera and a depth sensor. The color camera perceives the visual image, which is shown in the display to generate a mirror effect. The depth sensor observes the user and collects natural input for interaction with the system. (b) The color and depth camera are replaced by one Microsoft Kinect.
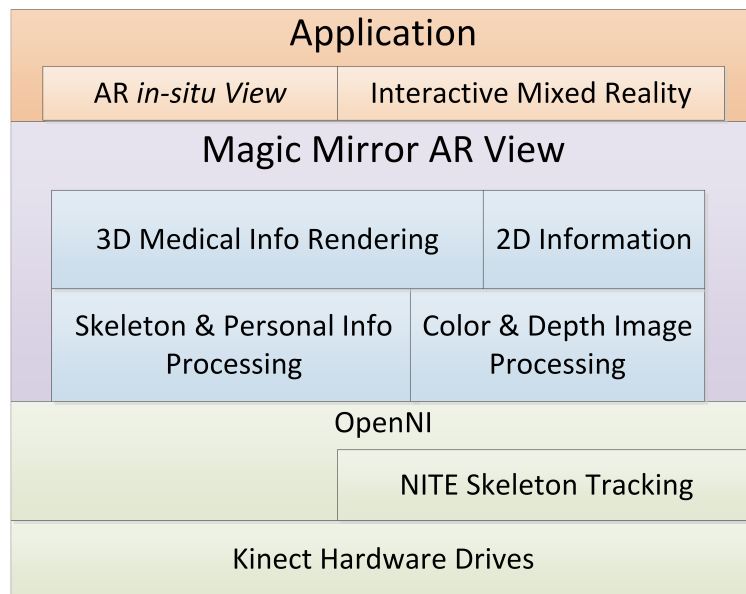

Figure 3.3 Magic Mirror system framework. The lowest two layers represent the open source library to access the Kinect raw image data and skeleton information. The middle layer is the modules used to create Magic Mirror AR view of the user. The top level is the application with basic system features.

Coin3D. 2D Information includes window management and basic user interface elements, such as 2D text and image information, and it can be implemented using Qt.

The applications based on the *Magic Mirror AR View* has two important features, **AR *in-situ* view** and **Interactive Mixed Reality**. The blue rectangles represent the basic visualization functions of the system, which will be discussed in details later. The prototype system discussed in this section is a personalized Magic Mirror system for anatomy learning (see Figure 3.1).

Medical Information

Aside from text and 2D atlases, 3D volumes and models are also important methodologies to represent medical information. The common way in medicine is to use 3D volumes, where we have one pixel or voxel for every point in 3D. This is the kind of data we get from CT or MRI. The other ones are polygonal or surface models, where only the surface (e.g. skin or the surface of the organ) is stored. For models that have been created for education often surface models are used because they look better. While the system can use a CT or MRI scan, it doesn't make sense to acquire a CT or MRI scan of the user if it is not required for medical reasons. Therefore, we augment the medical volume from the Visible Korean Human dataset (VKH) [123] onto the user. This dataset consists of a CT scan, an MRI volume and a photographic volume which has been acquired by stacking up cryosections. Most CT and MRI medical images are saved in the DICOM standard, which is the format that is used in all hospitals. Unfortunately, most software for research does not support DICOM. The Magic Mirror system takes a "∗.*MHD*" file as the medical data. One drawback of the volume data is that the 3D dataset cannot be deformed in real-time. So if the user bends, this is not reflected in the visualization of the medical data and also movements of the limbs are not visualized correctly. While later, possible solutions to address this issue are discussed, for the current system, which focuses on the abdominal area, this is a minor problem.

Visualizing structures other than bones from the CT is more challenging. In a first attempt, the segmentation that is available for the CT volume was used to visualize different organs in the abdominal area. The quality of the visualization was low, as the segmentation does not have sub-pixel accuracy and transfer-functions on CT intensities cannot provide a visualization with realistic colors and textures of organs. Therefore instead of using the volumetric data, additional polygonal models were integrated. The Anatomium[2] dataset provides polygonal models of many organs of the human body. A scene graph including multiple organs was extracted from the dataset. Using Coin3D this scene graph is augmented onto the user. The simultaneous visualization of bones from CT and a polygonal model of the small intestine is shown in Figure 3.5.

3.1.2 Personalized perception and interaction

To create a Magic Mirror AR view for anatomy leaning, special medical data should be selected from the database and transformed based on current user's personal information and interaction (see Figure 3.4). Then the data is processed to generate the non-physical visual effect, merging with the mirror-like color image based on the depth information. The user learns anatomy knowledge via perception of the AR view of their own body. Interaction can
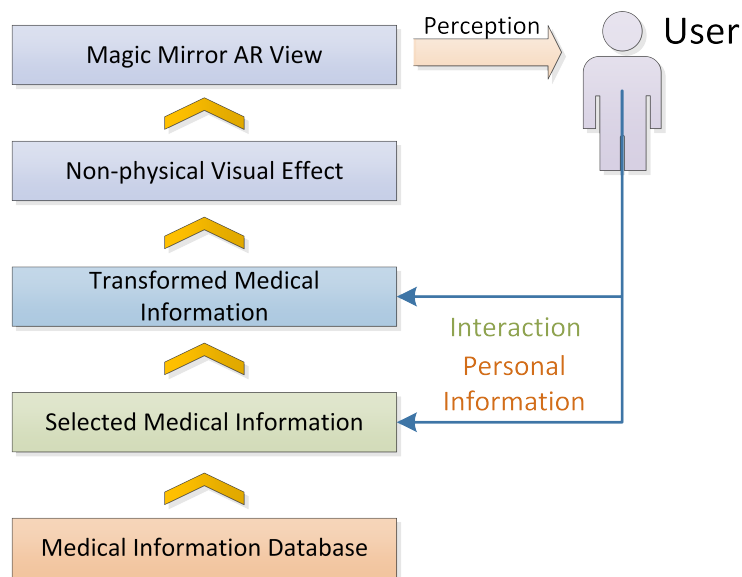
---

[2]www.anatomium.com

Figure 3.4 For each non-physical visual effect, corresponding medical information is selected from the medical information database, transformed, and rendered according to the function setting and the user's interaction and personal information. Finally, the Magic Mirror AR view is generated.

also be performed by the user intentionally or unintentionally to control the selection and transformation of the medical information, enabling an interactive AR view.

Calibration of the user

The skeleton tracking algorithm has to calibrate the user. For this, the user has to take a certain pose and hold it for several seconds. This calibration estimates the individual distances between joints for each user. This allows estimating the size of the user. The 3D volume and models are scaled to the size of the user and augmented onto the user. The NITE skeleton tracking algorithm, which is used in the Magic Mirror system, also requires the user to calibrate before the tracking works. In order to build a system that people can use without further instruction, we need to attract their attention and develop a self-explaining system. The NITE skeleton tracking provides a pixel map, where all pixels that belong to a person are labeled, before the user is calibrated. We use this to attract the attention of people passing by. As soon as someone walks into the field of view of the camera, that person is augmented with green color. The user has to take a calibration pose to become tracked. The system augments a silhouette of a person taking the calibration pose onto the display and asks the user to take the same pose. After calibration ends the system prints a message telling that a user has been recognized and the AR view can then be generated.

AR in-situ visualization of human anatomy

For a personalized visualization of organs the concept of a Magic Mirror is employed. The camera image is flipped horizontally and shown on the screen such that the user has the impression of standing in front of a real mirror. The system tracks all user movements using a depth camera and detects the pose of the user. Then, virtual elements are added to the image of the real scene. The augmentation context visualization [13] is used, such that the virtual objects are only seen through a circular focus window (see Figure 3.5). This leads to a better
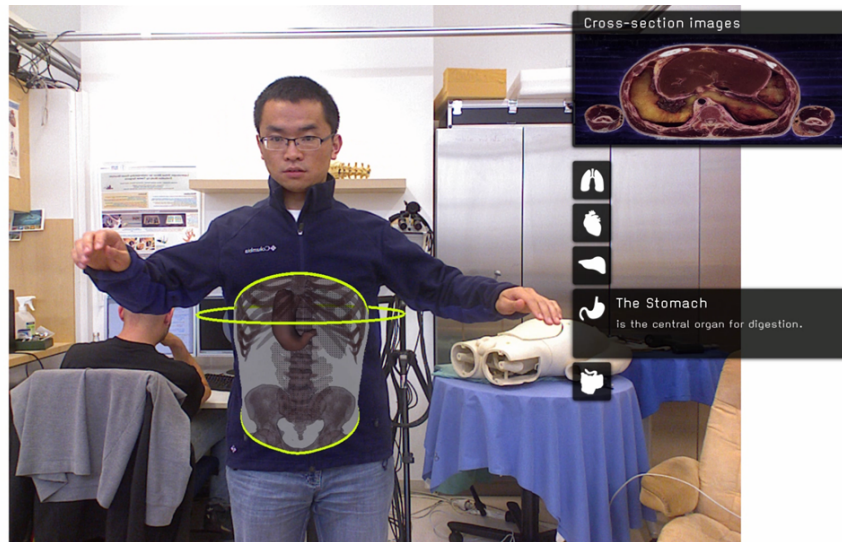
Figure 3.5 Magic Mirror Demo. The AR *in-situ* visualization of human anatomy in the prototype Magic Mirror demo.

perception of depth, compared to a simple augmentation of the whole CT. By using the Magic Mirror metaphor, the user is led to believe that he or she is able to look inside their own body. At the same time, different medical information (CT data and a fully segmented dataset of cross-sectional photographs of the human body) can be displayed [16]. For visualization of the bones a transfer function is used as bone can be distinguished easily in the CT volume based on the voxel intensities. For some applications where CT dataset of the user is available, it can directly augment the CT dataset onto their own body to improve the perception.

Presentation of additional medical information   In addition to the AR *in-situ* visualization of anatomy, the system still has to show more detailed information via text, images and 3D models. To achieve this, the system switches to a mode, where the whole screen is used to display additional information about anatomical structures, such as the 2D slices from the CT or photographic volume. Therefore, the system can switch to the 2D or the 3D models mode as needed, and both 3D view or 2D information can be displayed to improve the perception of the medical information.

Natural user interaction

All the body movements are tracked and can be directly employed for interaction. Arm and hand gestures are used to perform interaction in the Magic Mirror MR environment.

Relocating the see-through window   The *see-through* window plays an important role in creating the personalized *in-situ* visualization of the anatomy. As shown in Figure 3.6a, the window position is controlled by the hand position. A user specific offset vector is added to the left hand joint to calculate the window position. This interactive function motivates the user to learn where the interesting organ is. For public demos for children, we limited the window position to one dimension (see Figure 3.5) and the user moves up or down to change the window position to show different organs in the corresponding planes.
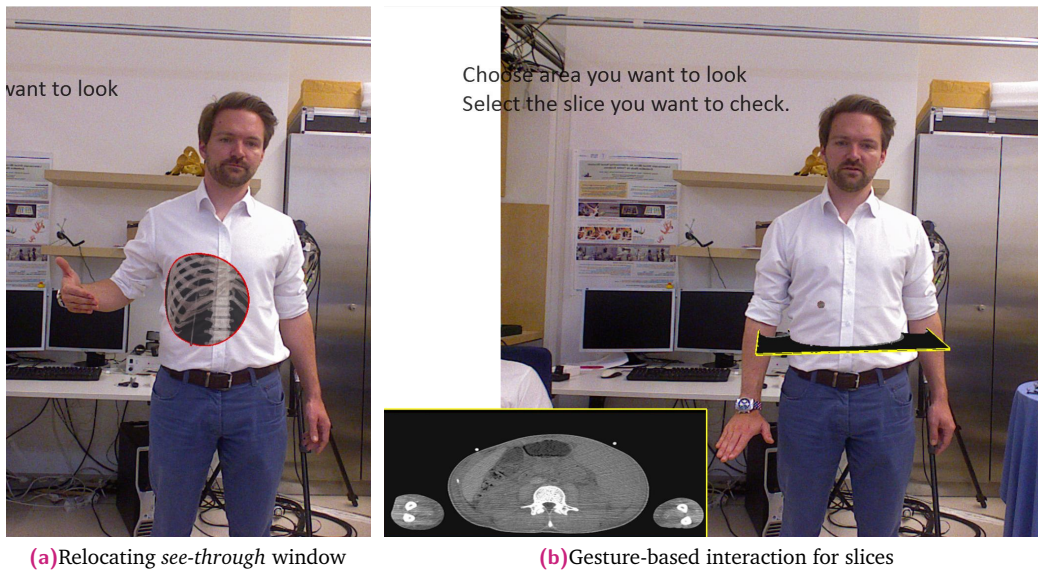
(a) Relocating *see-through* window   (b) Gesture-based interaction for slices

Figure 3.6 Arm and hand gestures can also be used to perform interaction with the Magic Mirror AR view. (a) The user can move the hand to relocate the *see-through* window. (b) When the system is in the transverse slice mode, the user could move their hand up or down to choose the interesting slice image.

Gesture-based interaction for slices   A volume can be seen as a stack of transverse slices starting from the top and going to the bottom. Usually medical volumes are checked slice by slice to get the desired information. The current slice is depicted on one corner of the monitor while a yellow circle or rectangle is augmented onto the user depicting the slice plane. The user can move their hand up or down to choose the interesting slice image (see Figure 3.5 and Figure 3.6b) when the system is in the transverse slice mode. It is similar for the sagittal slice mode. The system can easily switch slices between the CT and the photographic volume.

There is a trend towards competency-based education in medicine. Instead of defining a curriculum, learning outcomes are defined. Students have to fulfill these learning outcomes. The advantage of competency-based education is that all students will have the same competency in the end. A student who is less skilled has to take more time to learn than a student who already has good skills. One requirement for this is that the students have to be able to educate themselves until they reach the required competency. This framework can be made easily available to them so that they can use it for training and testing at any time.

3.1.3   Evaluation of the framework

The current system allows visualization of static anatomy on the user and offers a simple user interface to select CT or photographic slices. Demos of the prototype have been given on various public occasions. Medical doctors (MD) from physiotherapy, anatomy, radiology, sports medicine, children's medicine, and medical teacher and science centers are interested in using such a system in education of medical students and pupils. As an augmented reality anatomy learning application, system usability are very important prior to its translation in the classroom. We undertook one user study involving first year medical students to verify

the learning potential and acceptability of our technology as a compliment to atlases and textbooks in classroom.

Learning Scenario A list of questions were developed by the medical partners. Each question asks for the location of an anatomical structure. In the first step the student has to position the focus window at the correct location via the function "Relocating the see-through window". After positioning the window, the distance to the actual position of the structure is computed. If the distance exceeds a threshold it is considered as wrong. In this case the student has to reposition the focus window. When the correct position has been selected, the focus window is fixed and another mode is activated where a more fine grained selection is possible using "Gesture-based interaction for slices". In each question either one or more of the correct slices along the three directions have to be found.

Participants: 72 medical students from the Anatomische Anstalt der Ludwig-Maximilians-Universität München, having an age range between 21yrs and 28yrs, were enrolled in this study. The distribution of gender was 38 male and 34 female. Their medical experience was: 50 participants did not have any expertise, 8 had paramedic training, 5 with healthcare training, and 9 with some other degree (e.g. biology, chemistry, and therapy).

Analysis: Responses to questions were based on a four point Likert Scale: (1) *strongly disagree*, (2) *disagree*, (3) *agree*, and (4) *strongly agree*. For statistical analysis, the response categories were reduced to Yes/No: Yes (*agree + strongly agree*); No (*strongly disagree + disagree*).

The study procedure involved every participant to use the personalized Magic Mirror system for 15 minutes and try all available system functions. Participants were then asked to assess the learning potential of the technology by responding to the following questions:

Q1- is the system helpful for anatomy learning?

Q2- do you like to use the Magic Mirror?

Q3- is the presentation and visualization of the CT volume acceptable?

Q4- do you like the AR view of 3D human anatomy?

Q5- what other information would you want to see from the 3D rendering? (Open question)

Frequencies (F) and percentages (%) of response on the first four questions are given in Table 3.1. The questionnaire, focused on motivation and perception of the participants, showed statistically significant differences for all the questions. Percentages confirmed how the students accept this Magic Mirror system for anatomy learning.

The results of question 5 are depicted in Table 3.2. They indicate that the participants would eventually like the current Magic Mirror system to evolve to one that includes the visualization of complex anatomy models such as the vascular system, relations between different organs, extremities, etc.

Table 3.1 Frequencies and percentages of response of the medical students about the AR system.

	Q1		Q2		Q3		Q4	
	F	*%*	*F*	*%*	*F*	*%*	*F*	*%*
Yes	62	86.1	60	83.3	68	94.4	66	91.7
No	10	13.9	12	16.7	4	5.6	6	8.3

Table 3.2 Frequencies and percentages of new 3D anatomy expected by the 72 medical students

New 3D rendering functions	*F*	*%*
vascular system	19	26.4
relationship between different organs	13	18.1
extremities	10	13.9
landmarks	6	8.3
static AR view with their body	6	8.3
head/brain	5	6.9
muscle	5	6.9
others	5	6.9

3.1.4 Conclusion

This section presented a personalized AR Magic Mirror framework and a prototype for anatomy education. The prototype system creates a MR environment, which is certified useful for anatomy learning, and the majority of the medical students accepted the learning potential of the technology. The academic methodologies must be changed according to the advances of new technologies, but there is still a long way for this. Considering the benefits of the personalized and interactive AR system for motivation and perception of anatomy learning, new technologies can additionally be helpful to facilitate autonomous learning and secondarily to reduce laboratory material and instructor costs. Together with the anatomy community, we hope to initiate such discussions in integrating exciting user-specific and gaming concepts within curricula via new anatomy learning systems.

3.2 User specific registration

The main advantage of the Magic Mirror framework is to create a personalized MR environment, which builds a link between the user's own body and the medical information. A personalized perception of the medical information is the major objective of the Magic Mirror, and it includes two parts: (i) correct medical information for current user and (ii) accurate registration of the non-physical visual effect. Registration is the most important part to forge an AR view of the user's body and a satisfactory non-physical visual effect. The corresponding medical information is selected from the database, transformed to fit the user's body, and used to create a valid Magic Mirror effect (see Figure 3.4). Via the color and depth camera, the system should detect the personal information (e.g., gender, age, height, etc.) to select the best medical information for the current user. However, the detection of personal information is mainly based on machine learning to calculate a general acceptable result. Hence, an

interactive registration is introduced to improve the accuracy of the AR view. In this section, personal information detection and interactive registration are discussed.

3.2.1 Personal information detection

To improve accuracy of the system framework, personal information, including gender, age, body size and pose of the user, should be collected. It can be implemented via biometric recognition, skeleton tracking, and 3D reconstruction. Then the corresponding medical volume is selected, transformed, and augmented onto the user body taking the collected personal information as parameters.

Personal information to collect

If a new user is detected in the valid range of the system, the system goes to the stage of personal information detection and calculates all the parameters. The following list are the parameters we detect from each user.

1. Gender: Define the sex of the user.

2. Age: Classify into children, young, adult and elderly.

3. Height and length of limbs: Compute the height and limbs length of the user in millimeters.

4. 3D volume: Reconstruct the 3D volume of the user. The volume is the upper part of the user's body.

5. Weight: Estimate the weight of the user.

6. Circumference: After obtain the upper part of the 3D volume, measure the circumference of the bust, waist and hip information of users.

7. Shape: Give the body shape information of the user in the WHR (waist-hip ratio) index.

System design

The design of personal information collection is to develop a middleware, which works as an independent layer to collect the personal user information based on the color image, depth and skeleton stream from the Kinect sensor (see Figure 3.8). An independent middleware would not affect the basic interaction of the Magic Mirror MR system. The middleware does not involve the non-physical visual effect definition. Through the whole processing of the user recognition, the middleware provides all the personal information of the user as parameters to the Magic Mirror AR view. All these parameters would help the system generate a better MR environment.

Once the middleware detects a user within the valid range of the system, it starts to collect the personal information of the user and the whole procedure of the iteration is described as Figure 3.8. The first step of using the middleware software is the same as the Magic Mirror system. The first task is user detection, to check if there is a new user in the view of the system.
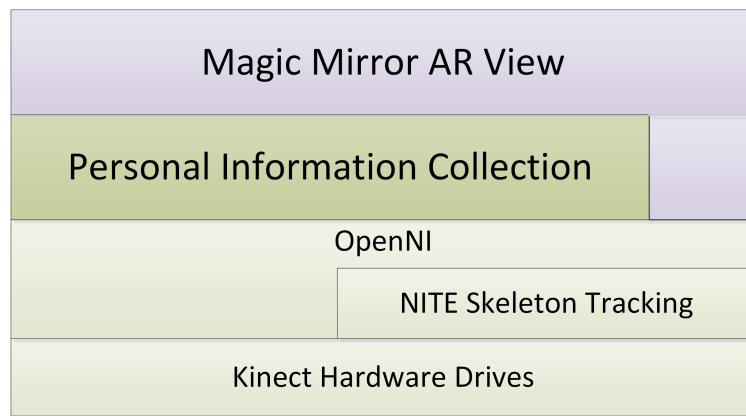
Figure 3.7 Personal information collection. It is developed as a middleware, which works as an independent layer to collect the personal user information.
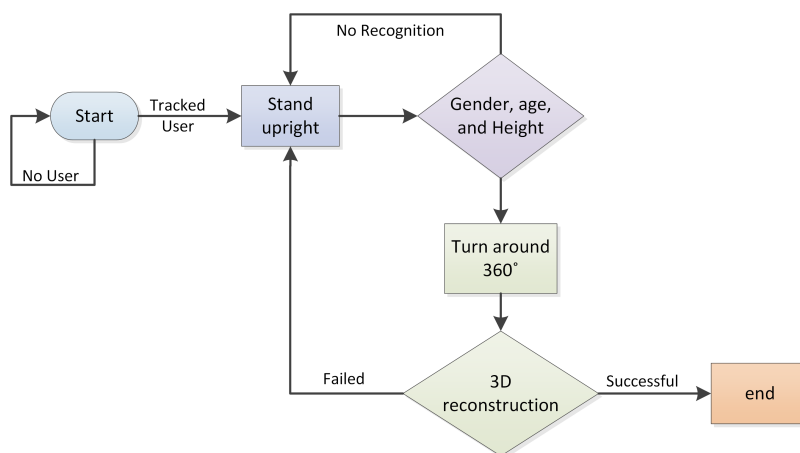
Figure 3.8 Workflow of the interaction with the middleware to collect user personal information.

Then, the user is asked to stand straight in front of the system. When the user stands upright, the system detects the gender, age, height and limbs length. In the second step, the user needs to turn in a circle in front of the sensor in order to scan the user's body completely. When the user finishes this movement, the system would complete the collection of the personal user information. If the reconstruction fails, the system depicts a hint to repeat the whole process.

Implementation

The color and depth cameras of the Kinect are the current sensors to observe the user and all the parameters of the user are detected via the color/depth images and the skeleton stream (see Figure 3.9). OpenBR framework is chosen to detect the gender and age from the captured color images. Meanwhile, the OpenNI and NITE libraries are adopted in this system for the height and limb length calculation. Besides the above parameters, weight, volume, circumference and body shape are all based on the 3D reconstruction of the user's upper body, which is done by the 3D reconstruction SDK and OpenCV[3] libraries. Combining all the following functions, the personal user information would be collected. The Magic

---

[3]http://opencv.org/
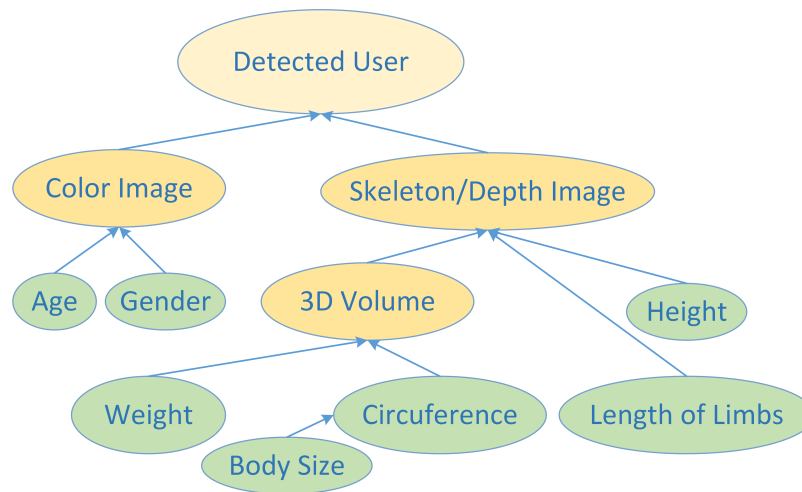
Figure 3.9 Personal information tree. All the parameters of the user personal information are detected via the color/depth images and the skeleton.

Mirror framework can refer to the user's information as personalized parameters to generate more useful and interesting AR applications.

Gender and age The detection, normalization, representation and extraction are done by the OpenBR framework. What the middleware does is to implement a thread for this work. When a user stands in front of the system, the system firstly estimates the rotation of the human body. When the user faces the camera, an image is captured. After the capturing, the image including the front face goes to the engine of the OpenBR framework, and then the gender and age are calculated. To get a more accurate result, a face area is selected based on the skeleton data.

Height and length of limbs From the basic understanding of the word human height, we all know that the height is the distance from the top of the head to the bottom of the feet. As the feet sometimes may not be seen and detected by the Kinect sensor, the height would be converted to the distance from the top of the head to the ground. Thanks to the help of the OpenNI and NITE SDK platform, the plane function of the floor could be easily assessed from the depth image and the user map is generated, defining the valid pixel belonging to the user's body in the depth image. The coordinate of the top point of the user can be calculated, based on the user's information and the depth image. For the length of limbs, the user could be asked to keep a cross pose to get a clear view of the body and limbs. The limb length is the distances between corresponding skeleton joints.

3D volume The ReconstructMe framework is used to implement the 3D volume reconstruction. One important task of the body reconstruction is to estimate the rotation angle of the user's body. If the total rotation angle of the user is less than $360°$, it means that the reconstruction is not finished and the system continues to reconstruct. The system tries to estimate the new position of the user, transforms the previous position to the new one and merges the data. After the transformation, it gets the next depth frame and repeats the reconstruction frame by frame until the sum of the angle changes is larger than 360 degree.
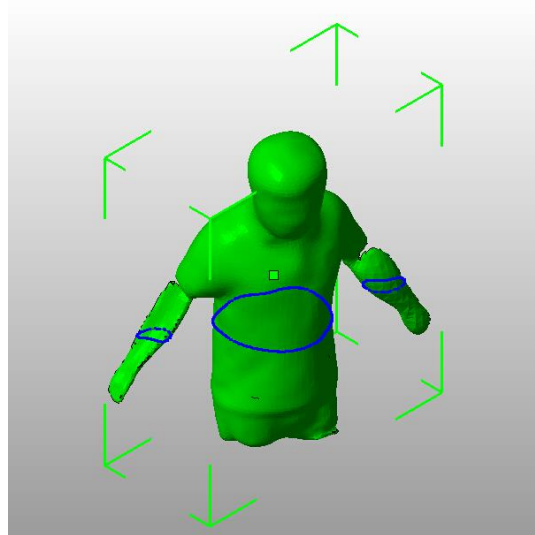
Figure 3.10 Circumference. The circumference of the circle in each layer in the 3D body volume gives us the information of the user's body shape.

Weight   In the proposed middleware, we provide a simple method to estimate the weight information of the user. It is a simple idea that the weight is related to the volume of the 3D reconstruction. After the 3D reconstruction, a watertight 3D mesh is created and the volume is calculated [166]. Due to the limitation of the Kinect sensor in the field of view, the 3D volume reconstructed cannot cover the whole body. However, we could just use the upper body of the user to estimate the full weight. As the result presented by [153], the upper body is about 68.2% of the entire body weight. No matter how the reconstructed result is, the upper body is the volume without the part below the hip. As the hip joint information is known, we just make a cut from the hip position and generate a new volume without the hips. As the density of human body is very near to the density of water $(1.0g/cm^3)$, the estimated weight of the user in the system would be as same as the volume.

$$Weight = (Volume * Density)/0.682 \tag{3.1}$$

Circumference and Body Shape   After the obtainment of the 3D body volume of the user, the last measurement is to get the information of the user's shape like the circumference of bust, waist and hip. From the reconstructed 3D volume, the body surface can be calculated easily. The circumference of the circle in each layer gives us the information of the user's body shape. As shown in Figure 3.10, it is a simple example of the user's body when we look at the volume in the middle layer from the cross-sectional view. The blue circle is the Z axis cut on the 3D volume and the length of the cut would refer to the circumference on that layer. The key point here for the body shape detection is how to find the right layer for the right circle we want. The bust, waist and hip are the most important parameters for the body shape detection within the system. The position of the bust, waist and hip can be estimated from the skeleton, we can refer to the correspondent layer of them and acquire the length of each parameters. Body shape information of the user refers to the index of health and the risk of developing serious health conditions. The Waist-Hip-Ratio (WHR)[115] is introduced to indicate the condition of obesity for the system's users. The "apple shaped" bodies would be more healthy

than the "pear shaped" as less weight around the waist makes the value of WHR smaller. The system would provide the value of Waist-Hip-Ratio index to the user and indicate the health condition of them.

3.2.2 Interactive registration

The skeleton output from Kinect also limits the precision of the Magic Mirror applications. Thus, the Magic Mirror augmentations would contain errors and users would easily distinguish anatomical offsets on their body resulting in a poor mixed reality environment and visualization. Alternatively, we considered projectors to display human anatomy and along with their existing limitations the same inaccuracies would exist [143]. Hence, an interactive solution is proposed to improve the registration as the Magic Mirror framework always takes the user movement as input. Our aim is then to propose a method to correct for such overlay inaccuracies that can be used in the general Magic Mirror framework.

Anatomical bone landmarks

Together with orthopedic surgeons, we have defined five bone landmarks that (i) users can easily identify and touch while standing in front of the Kinect, and (ii) that are accurately and easily identified in medical data such as CT [100]. These are: left and right acromion, left and right anterior superior iliac spine, and the manubrium (see Figure 3.11).
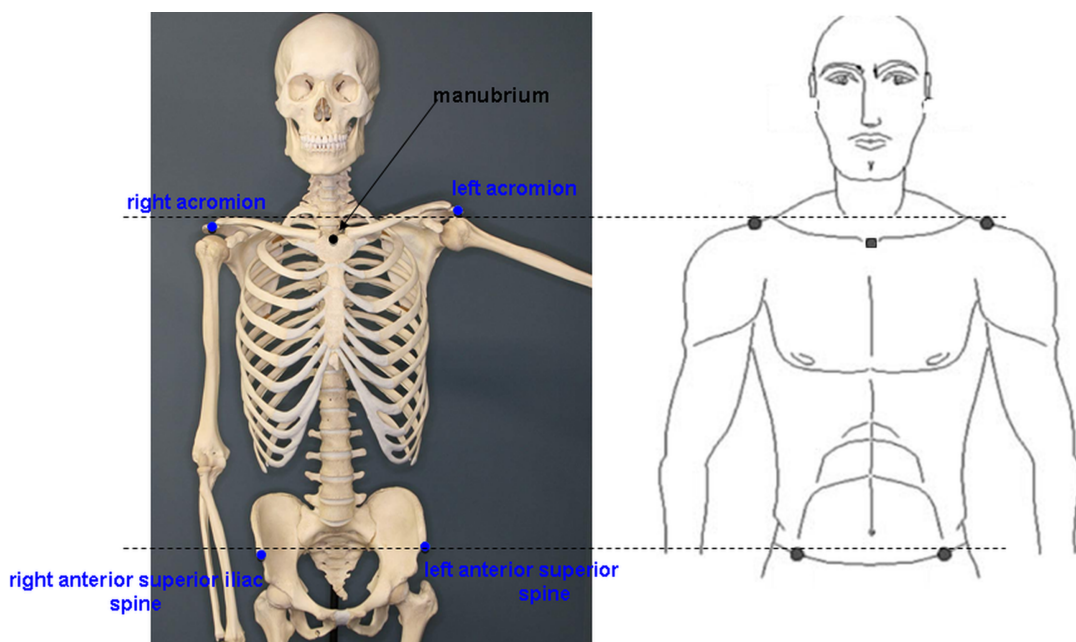
Figure 3.11 Anatomical bone landmarks. Selected anatomical points for Kinect skeleton improvement and subsequent CT warping and interpolation.

Prior to correctly using the Magic Mirror system, the users stand in front of the monitor and virtual marks are displayed near the five bone landmarks according to the raw skeleton stream. The users are asked to interactively adjust the positions of the five marks to fit their own bone positions. In addition, the exact locations in the VKH CT dataset of the five selected bone landmarks are known. A linear interpolation was executed to estimate the

torso point (i.e. a 6th landmark) in the CT volume to improve the overlay. Then the scale factors and transformation matrix were computed to render the anatomical image onto the user's body. These landmark positions allow the deformation and interpolation of the medical data correctly within the Magic Mirror and onto the human body, resulting in a more precise augmentation. A user study involving surgeons and anatomy experts confirmed our findings and will be presented in the following section.

Improvement of Kinect skeleton

We present a general method to interactively improve and correct the Kinect skeleton for anatomy education purposes. The method can also be applied to projectors or other sensors as well as for general augmented reality applications. A thorough validation of our method demonstrated improved precision of anatomical landmarks and opens the avenue to future improvements in medical education. Figure 3.12 shows a comparison between the traditional Kinect skeleton and its proposed improvement. The first row depicts visually the exactness of the new skeleton. The second row depicts the skeleton landmarks directly on CT data. We observe that the shoulder and anterior superior spine are inaccurate in the images. The last row depicts the selected bone landmark positions within CT as well. Transverse and sagittal CT slices of the visible human Korean are seen respectively in rows 2 and 3.

Normal solution   To improve the perception of the AR view, the medical data has to be registered to the current user. A simple method used by the Magic Mirror framework is that an orthopedic clinician helps to find the corresponding Kinect skeleton joints in the CT volume (see Figure 3.12a and Figure 3.12c). The points 1-5 are raw Kinect skeleton joints and slices 1-5 are estimated by the clinician according to the yellow line drawn on user's body using the raw Kinect skeleton joints information. So the scaling factor from CT volume to the current user is:

$$WidthScale = Avg(\frac{W_{12}}{W_{12}^{CT}}, \frac{W_{34}}{W_{34}^{CT}}) \tag{3.2}$$

$$HeightScale = Avg(\frac{H_{14}}{H_{14}^{CT}}, \frac{H_{23}}{H_{23}^{CT}}) \tag{3.3}$$

T_{K_CT} is the transformation from Kinect coordinate system to CT volume. T_{K_n} and T_{CT_n} is the transformation matrix from the original point in Kinect world and CT volume to the skeleton joint n, correspondingly.

$$T_{K\_CT} = T_{K\_5} * T_{CT\_5}.inv() \tag{3.4}$$

Proposed method   The goal of the proposed method is to create a more precise user-specific AR overlay environment. The positions of the anatomical bone landmarks can improve the deformation and interpolation of the medical data, resulting in a more precise Kinect skeleton and augmentation. The following scale factors were calculated for the improved system via the selection of bone landmarks.

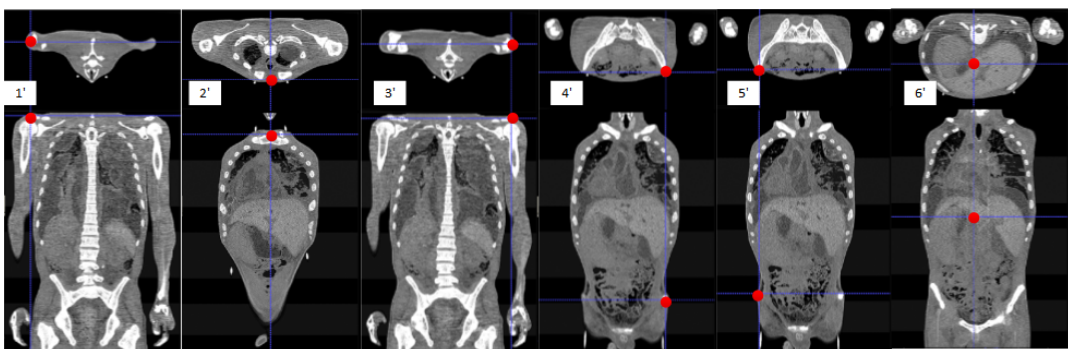$$WidthScale = Avg(\frac{W_{1'3'}}{W_{1'3'}^{CT}}, \frac{W_{4'5'}}{W_{4'5'}^{CT}}) \tag{3.5}$$
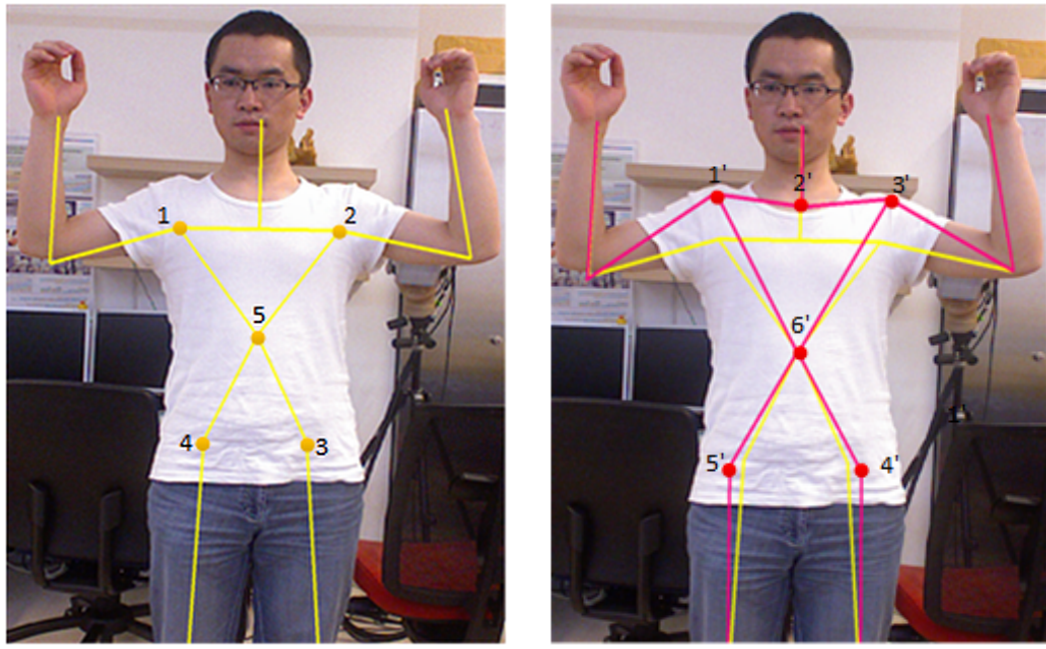
(a) Kinect skeleton

(b) Improvement of Kinect skeleton
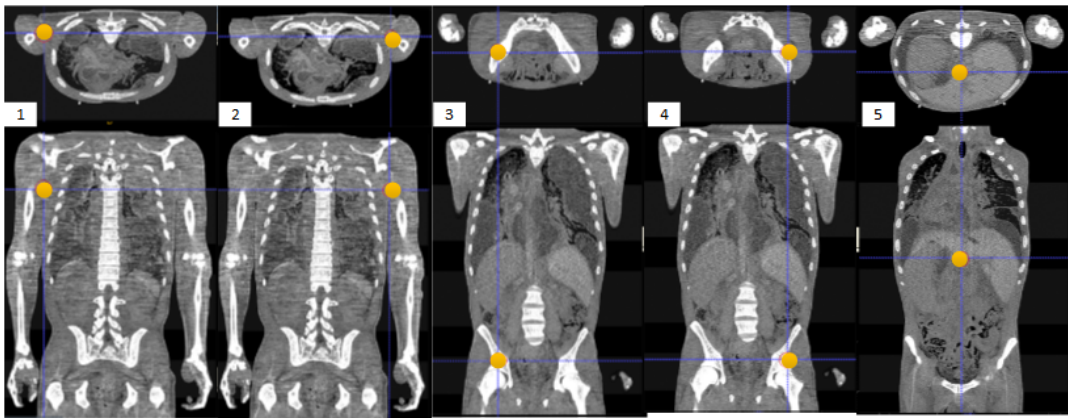
(c) Skeleton joints in CT volume

(d) Bone landmarks in CT volume

Figure 3.12 The inaccurate Kinect skeleton (a) The points 1-5 are raw Kinect skeleton joints. (b) The inaccurate Kinect skeleton, in yellow, compared to the improved Kinect skeleton positions in red. The points 1'-5' are the position of the five selected bones, which are suggested by orthopedic clinicians. (c) Slices 1-5 are estimates by orthopedic clinicians according to the yellow line drawn on the user's body using the Kinect skeleton joints information. (d) Slices 1'-5' are the selected five bones suggested by orthopedic clinicians and the slice 6', which is computed from slice1'-5', is the corresponding slice for joint 6'.

$$HeightScale = Avg(\frac{H_{1'5'}}{H_{1'5'}^{CT}}, \frac{H_{3'4'}}{H_{3'4'}^{CT}}, \frac{H_{2'4'}}{H_{2'4'}^{CT}}) \qquad (3.6)$$

$T_{K\_n'}$ and $T_{CT\_n'}$ is the transformation matrix from the original point in Kinect world and CT volume to the bone marker n', correspondingly. Using Eq.3.7, the position of the torso in the CT volume is calculated. The transformation of the torso is the linear interpolation of the five bone landmarks.

$$T_{CT\_6'} = Avg(T_{CT\_n'} * T_{K\_n'}.inv() * T_{K\_5}) \qquad (3.7)$$

$$T_{K\_CT'} = T_{K\_6'} * T_{CT\_6'}.inv() \qquad (3.8)$$

Evaluation

As an augmented reality anatomy learning application, both accuracy and system usability are very important prior to its translation into the classroom. We undertook one user study with particular users having expert anatomy knowledge to evaluate if this system is precise enough for anatomy learning.

Assessing the Magic Mirror system precision and usability Participants: Seven participants were included in this study (two surgeons and five final year undergraduate medical students). Analysis: a Likert scale was used which is a type of psychometric response scale often used in surveys and the most widely used scale in survey research. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement. The format of our 5-pt Likert was: (1) *strongly disagree*, (2) *disagree*, (3) *neither agree nor disagree*, (4) *agree*, (5) *strongly agree*.

To assess the precision of our personalized Magic Mirror we asked the participants to interact with the system platform which integrates user-specific anatomical landmark selection. Participants were asked to provide an estimated numerical offset, if any, on how far specific bone landmarks or organs were with respect to their own body. For this, they interacted with the Magic Mirror window, CT data, and used their own medical knowledge and expertise for judgment. CT data was displayed in an interface depicting both transverse and sagittal planes, and participants would quantify the offsets. If needed, a ruler was provided to assist them. The anatomical targets during evaluation were defined as: the anterior superior iliac spine, manubrium, heart, and liver. Results from this exercise are shown in Table 3.3, with offsets measured in centimeters. The user study showed the impact of the proposed method to increase precision for a better visualization of anatomy. The offsets of specific anatomical landmarks decreased significantly. Results from the user study show that the precision of the user-specific learning environment is on average 0.96cm.

The seven participants were then asked to judge the usability of the AR Magic Mirror system by responding to the following questions: (i) is the overlay accurate w.r.t human body (ii) is the user interface easy to use, (iii) is it fun to play, (iv) can it be used for medical education, and (v) would it have stronger impact for medical education learning? The Likert scale results for the first four questions are shown in Table 3.4.

Table 3.3 Precision (in cm) of magic mirror system based on anatomical offsets

Anatomy	Offset (Mean±STDev)	Improved Offset (Mean±STDev)
anterior superior iliac spine	6.83 ± 4.97	0.67 ± 0.52
manubrium	3.17 ± 1.72	0.67 ± 0.75
heart	1.83 ± 1.47	1.17 ± 1.60
liver	5.08 ± 2.42	1.33 ± 1.21

Table 3.4 Likert scale results regarding magic mirror usability

	Mean± STDev
is the augmented reality overlay accurate w.r.t human body	4.00 ± 0.89
is the user interface easy to use	3.67 ± 1.03
is it fun to play	4.50 ± 0.55
can it be used for medical education	4.17 ± 0.75

For the last question regarding the impact of our technology, there was a unanimous response that the AR Magic Mirror system should be considered as a potential platform to complement existing anatomy learning tools inside anatomy classrooms.

Conclusion

The precision of our method is visually demonstrated in Figure 3.13. We observe that the acromion and anterior superior iliac spine, using the traditional Kinect skeleton, is not positioned correctly within the CT data compared to the modified Kinect skeleton version (1-2 vs. 1'-2'; 3-4 vs. 3'-4'). The orthopedic surgeons participating in our study confirmed this. Results from a user study show the impact of interactively improving the Kinect skeleton to increase precision for a better visualization of anatomy. The offsets of specific anatomical landmarks decreased significantly. The following comments were also collected from the user study:

1. the precision improved; making the user touch anatomical landmarks is cool since this is the way it is done in clinic ...

2. two additional landmarks easily accessible are the sternum and bottom of rib cage...

3. make the Magic Mirror circle bigger for larger anatomy ...

4. voice command is a good idea but it is sensitive to the surroundings noise (i.e. multiple people talking in the room, etc.)

5. interactions between observers ...

6. could introduce the female CT visible human Korean ...

7. could use a healthy patient CT or other modality ...

8. could make the CT slices bigger on the screen ...

Figure 3.13 The Magic Mirror before (top) and after (bottom) the Kinect skeleton adjustment. Column 2 depicts the bottom of the rib cage being positioned correctly after skeleton correction.

This section presented a general method to collect the personal information and interactively improve and correct the Kinect skeleton for anatomy education purposes. A thorough validation of our method demonstrated improved precision of anatomical landmarks and opens the avenue to future improvements in medical education.

3.3 Interactive mixed reality and serious gaming

Knowledge is a process of understanding by connecting. Although we can learn by reading and watching, the learner always needs to encode the knowledge by themselves, rather than simply receive what is shown. This is one of the reasons that interacting with the mixed reality system can be a powerful learning tool - we do not only get the benefit of the AR view, but also receive immediate feedback. These interactions are often useful in creating the 'WoW' moment – the joining of medical knowledge with a learner's own body and movement. In Section 3.2, user's personal information is collected to improve the perception of the medical information. Here, we implement several applications by integrating exciting user-interaction and gaming concepts within the proposed Magic Mirror framework.

In this section, a new interaction with anatomy is firstly developed for learning in the MR Magic Mirror framework. It is demonstrated by two new modules using the organ atlas and an education game. Then, we focus on the muscles of one human arm to find out a solution for muscle learning. Finally, serious gaming for education and rehabilitation is discussed and a knee rehabilitation demo is developed.

3.3.1 Interaction for anatomy learning

Basic knowledge of human anatomy is also considered common knowledge that is taught at various levels of school education. To improve the conventional book-based methods for anatomical education, we have investigated the suitability of an MR system using a Magic Mirror metaphor for this task. With the existing platform, we develop an interactive educational application consisting of an organ atlas and a complementary organ learning game, leveraging the benefits of self-referential learning and motivation through game challenges. The organ atlas is a good way of imparting anatomical knowledge on the users, allowing them to explore their body without any instructor or supervisor. The education game is developed taking advantage of the new interaction method with the anatomy.

Motivation and objective

In the original Magic Mirror MR learning environment, the users can see themselves on the screen like in a mirror with medical information augmented onto their body and a feeling of "looking inside" the human body. The user can also relocate the see-through window and interact with slices as described in Section 3.1.2, but the user cannot directly interact with the organ anatomy. Here, we try to expand the existing framework to be used as an alternative and entertaining way to learn human anatomy. An educational application is developed to allow the user to explore their body and learn about spatial relations, visual appearance and function of the organs. As atlas and text are still the main learning methodologies, we also want to merge these traditional information into the MR learning environment.

The objective is to implement an interactive anatomy atlas that also focuses on the inner organs of the torso. The user can explore these organs and read additional information on each of the visualized organs. The work is separated into two different parts: general interaction methods performing all kinds of basic operations not specific to any application and the logic implementation of the actual application, and realizing the organ atlas application. The general interaction methods are designed to be one part of the Magic mirror AR view, as shown in Figure 3.3, and provide easy access to implement custom application logic, making it easier to experiment with new ideas based on the Magic Mirror framework.

Challenges of interaction   The main challenging issue is the interaction method with the organ model. We need a method to reliably detect which organ the user has selected while leveraging the mirror analogy as far as possible. The most intuitive and direct way would be to use a virtual hand technique to directly select the organs [60]. However, this is bound to fail for two reasons. First and most obviously, it is not possible for the user to put a hand into their body and directly touch the organ. Only the approximate region in front of the organ is reachable, but it is not accurate in many cases as it is not really possible to detect reliably which organ the user points to. Especially in the abdominal area many organs are fit very

tightly together in one place, making it very hard to distinguish between two organs. Secondly, the tracking capabilities of the current version of Kinect sensor are not robust. The only thing suitable for selection is the position of the hands. Untrained users usually use their fingers when asked to indicate a point. Neither orientation of the hand nor information about the fingers is available when the hand is put on their body. This makes it even harder to develop a suitable, intuitive interface.

General interaction methods

Since conventional point-and-click selection methods are only of limited use for self-referential selection in MR environments, we have developed new methods better suited for our scenario: (i) One is driven by a two-handed interaction method to allow the user to select the region of interest with one hand and using the other hand to select a specific organ; (ii) the other uses a single-handed approach where the user indicates the organ directly by touching the specific region of their torso [79].

Organ Explosion   Our approach to overcome the selection challenge is a technique that we have called "Organ Explosion" effect (see Figure 3.14). This technique is inspired by exploded view drawings known from engineering, where the goal is to depict how mechanical parts need to be assembled. Usually they are drawn with the innermost part at the center while the other parts are moved some fixed distance outwards to display all parts of the assembly that would otherwise be hidden. We use the same basic idea in our interaction technique by implementing a two-handed interaction method. Using the left hand, the user is able to set the focus height for the region they are interested in. Organs at approximately the same height are then projected outwards. Because we use a spherical projection with the center position at the height of the hand, but at a point behind the body, the organs are moved outwards creating the illusion of seeing them "fly out" in front of the body. Lines are drawn to indicate the original position of the organs. After separating the organs from each other, it is easy for the user to select the organ of interest using the other hand. This way we are able to separate the organs close together since organs at different angles to the projection point are projected into different regions. We scale the magnitude of the displacement down with the distance between hand and body to allow the user to see the organs in their original arrangement but still enable the user to "zoom in" for detail and selection. Figure 3.14b depicts how this technique looks in practice. Once we are able to separate the organs from each other, it is easy to perform selection by checking the intersection between the hand and the organ. Another functionality of the organ explosion effect is allowing the user to rotate and observe the 3D organ models and perceive the spatial relationship between them. The interaction methods we developed are not necessarily limited to self-referenced selection and manipulation. In general cases, they are suitable for scenarios where many possible items are densely packed and a user wants to select a single or multiple items in an augmented or mixed reality environment. The large number of possible modifications to the methods make them difficult to configure, but the general idea should be applicable to many other cases.

Interaction test for selection   Intersection detection between different 3D objects is a built-in functionality in many rendering libraries. A virtual predefined sphere or box can be attached to the control hand of the current user and defined as the hand object, and the selection function is triggered when the intersection between the hand object and the organ models occurs. For performance reasons, we only use bounding-box intersection tests and skip the narrow
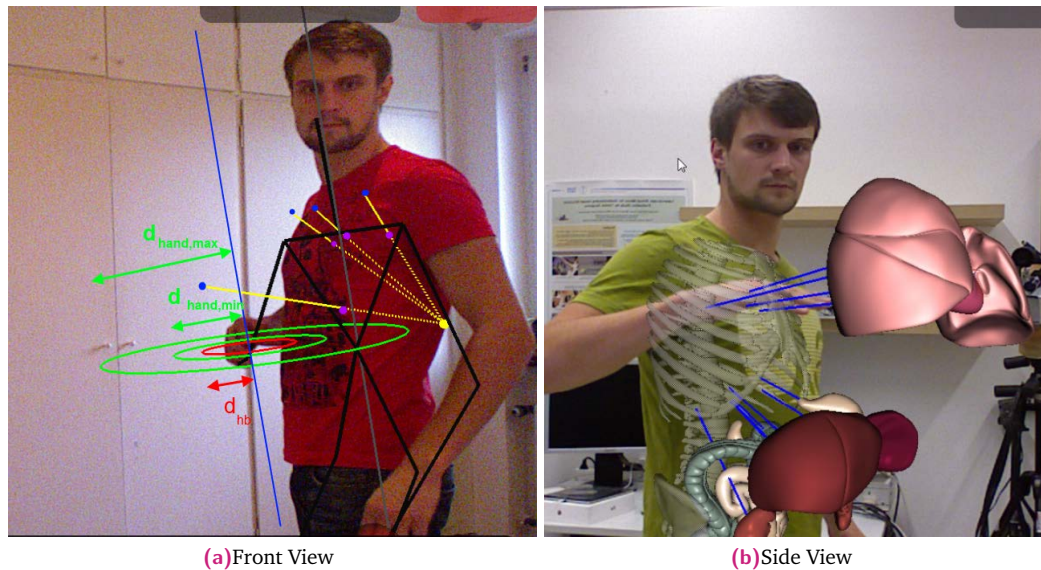
(a)Front View   (b)Side View

Figure 3.14 Left: Quantities involved in the calculation of the position offsets for the technique (organs not displayed). In this image, four organs are displaced. Purple dots are original organ positions, blue dots displaced organ positions and the Yellow dots signify the spherical projection. Right: Screenshot of the explosion technique.

phase, treating intersecting bounding-boxes as intersections between geometry. This obviously reduces the accuracy of bounding box tests. This coarse intersection check is better not only in performance, but also from a usability perspective. Intersection tests in our applications are used for checking intersections of the hands with organs. Due to the inaccurate tracking of the hands by Kinect, this actually works quite well when just verifying for "roughly" intersecting pairs.

Organ atlas application

An organ atlas is usually a collection of visual representations of organs or specific organ systems, annotated with names and physiological information. We reproduced a version of this concept using the Magic Mirror framework with personalized interaction. The main educational goal of this application is to allow the user to explore the organs of their body and learn about the shape, position, function and 3D spatial relations to the other organs. We approach this goal by augmenting all available organs onto a user's body and letting the user select the one they are interested in via an "Organ Explosion" effect. Once the user has selected an organ, the application displays textual information about the organ. This information includes the latin and/or greek names as well as a description of the basic function of the organ. During the selection process, the knowledge about the spatial position is also obtained. An example of what the application looks like in action and when finished is shown in Figure 3.15. The organ atlas offers an interesting new way to explore the human anatomy. However, the described methods offer much room for experiments and improvement. The organ learning application in its current prototype version already provides a fun experience, but it lacks the content to be distributed as a full game. This could easily be changed by adding more game modes for other organ systems, multi-player modes or other elements commonly found in computer games.

(a) In action   (b) Finished
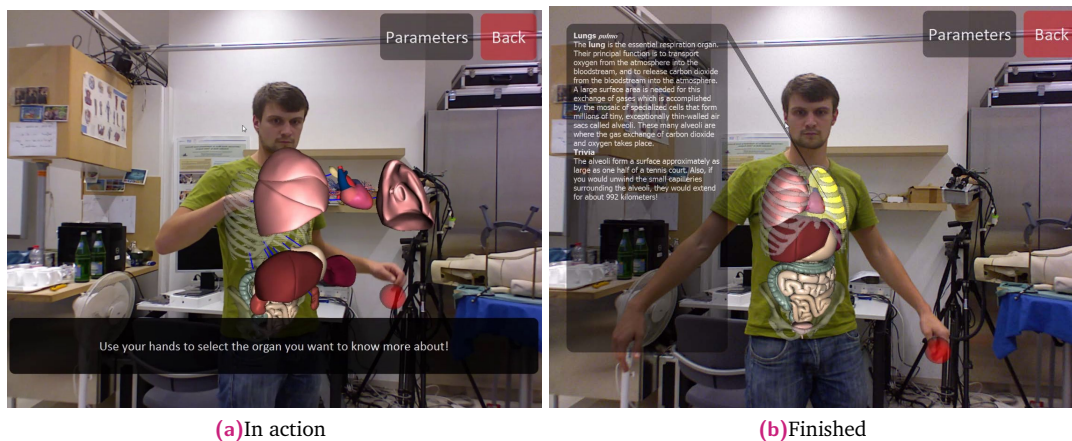
Figure 3.15   Organ Atlas implementation. Lungs are selected and the infobox shows textbook knowledge about them.

Organ game

A serious game is developed for general users to learn basic anatomy called "Whack an Organ". The game is based on the idea of the classic arcade games at carnivals, "Whack-a-Mole". We implement a similar game idea using the proposed interaction methods. It is a combination of Whack-a-Mole and classical quiz games. Questions regarding human anatomy are shown to the user, and the answers a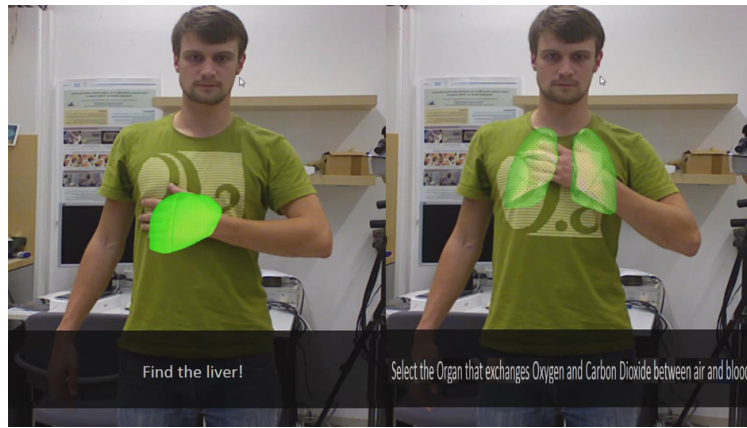re always one organ or organ system. This knowledge will serve as a basis for the game. This game works in a quiz-show like manner where the user has to answer questions by indicating the right positions of the organs. To answer the question, the user has to point to the location of the organ directly on their body. The game then decides if the location is correct and awards points if so. Questions can come from different question sets with varying difficulty, ranging from simple location questions: *where is the liver?*, to more complex knowledge questions: *which organ is infected if you have Hepatitis?*. Still, the questions have to have a concrete organ as an answer, limiting the possible range of anatomy-related questions. At the end of a question round, the achieved points are shown to provide some feedback on the performance of the player. We hope to provide an experience that is both entertaining and fun to persons of all ages.

Organ bounding box selection   A direct approach is to do the bounding box interaction test between all the organs and a virtual box, which is attached to the hand. Once an intersection has been detected, we use a timeout to allow the user to position the hand correctly. Once the timeout has run out, we collect all organs that are currently intersected as our selection set (see Figure 3.16a). We increase the length of the virtual box in negative Z direction (towards the back of the body) to be as long as the user's torso width. This is because some organs like the kidneys are located in the dorsal area of the torso and they still have to intersect with the box when the player touches the front.

Sphere intersection   Another solution is inspired by taking the Whack-a-Mole concept one step further. We define a fixed number of selection regions attached to the body of the user. Those regions are labeled by virtual spheres or boxes, which are attached to the front of the trunk and are separated far enough to allow for an unambiguous selection. We compute the closest sphere for every organ at the initialization in advance. These spheres act as a proxy for

(a)Bounding box   (b)Sphere interaction
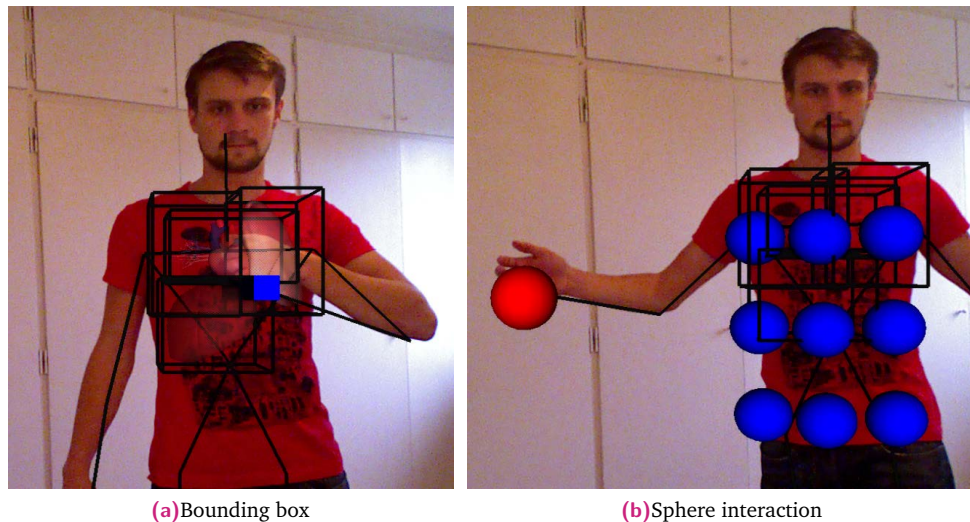
Figure 3.16 Debug visualization output of the two selection techniques. (a) Organ Bounding Box Selection: Blue box is intersector box of the left hand. (b) Sphere Selection: Blue spheres are the selection regions, red sphere is the intersector of the right hand.

the actual organs, so the user only has to select the sphere closest to the organ in question. We detect the selection of the spheres via bounding box intersection. Once we have detected the selection, we check if the correct organ(s) are associated with the selected sphere and determine the correctness of the answer. The number of spheres can be varied and several spheres can be associated with one organ.

Implementation   The implementation of the game is rather straightforward based on the application framework. We first load an appropriate question set from the hard drive. Question sets consist of a list of questions together with a set of correct answers for each question. The answers are given as a list of organ names and then we calculate the closest sphere if needed. After setting up the basic game state variables like the current question and counters for correct/wrong answers, we proceed to set up a default behavior for all organs. This behavior keeps an organ invisible for most of the time. Only when the user has selected an answer and the organ is in the set of correct answers for the current question, it is made visible for a short time (see Figure 3.17). At the time, it flashes either red or green, depending on whether the answer is right or wrong. Once a selection set is determined, the application checks whether it contains a correct answer. According to this, a short feedback is displayed and the counter for correct or wrong answers is modified accordingly. Afterwards the next question is presented if there are any questions left. One important thing for usability is to prevent the system from accepting the current position as an answer for the next question as well, because the hand is still intersecting and the timer has run out. We chose to solve this by setting a flag that prevents accepting new answers. This flag is only cleared when the user has moved the hand completely away from the body.

The organ learning game implemented as a proof of concept offers an engaging and fun way to learn about the human anatomy. The intuitive concept of indicating the organs on the players' own body should offer an increased learning intention. We hope to extend the system by providing different difficulty levels to make it more interesting for a larger target group.

Figure 3.17 Whack an Organ. This game works in a quiz-show like manner where the user has to answer questions by indicating the right positions of the organs.

Table 3.5 The selected motion of one arm and the corresponding main muscles

Motion	Muscles
Flexion	M.biceps brachii
	M.brachialis
Extension	M.triceps brachii
	M.anconeus
Pronation	M.pronator teres
	M.pronator quadratus
Supination	M.supinator

3.3.2 Interactive muscle learning

Muscle learning is a very important and difficult topic for all the medical students as it is a complex system and several muscles are active to perform one motion. Functions of the muscle are not easy to learn as it is very difficult to observe certain motions in humans. Bones of the skeleton system are rigid objects and are controlled by muscles via attachment points. From any outside tracking systems, the poses of the bone can be calculated. The main objective of the application is to enable the user to acquire knowledge about the muscles of the arm by directly connecting all the virtual information with their own arm movement. Based on the Magic Mirror framework, the non-physical visual effect is the activities of the muscles and the bones. A virtual model of an arm is rendered according to the real arm movement and the AR view is generated. After discussion with our medical partners, the main muscles of one arm related to four basic movements *flexion*, *extension*, *pronation*, and *supination* are defined. The learning targets are the main muscles, which expand and contract during the completion of the above movements (see Table 3.5).

Application design

The application contains an AR view, which shows the color image from the Kinect with the augmented virtual arm. This view acts as the Magic Mirror effect and helps the user to map all the activities of the bones and muscles onto their own arm. There is another VR view which is

also synchronized with the real arm motion. The virtual view concentrates on the upper or lower arm, showing the details of the muscle model. To preserve the connection between the user and the virtual arm, the AR-view is shown on the left side of the screen after the user is calibrated (see Figure 3.22). The learning procedure include 'muscle-oriented', the muscle is displayed one by one, and 'motion-oriented', the user performs the target motions and only the muscles related the current motion is shown. The system also allows the user to observe the muscle model from different viewpoints to learn the spatial information and the learning flow is flexibly scripted [125].

Dynamic muscle model   An anatomical model of a human male body from ANATOMIUM 3D[4] is employed as the virtual arm model. Since the application covers just one arm and quite a few simple motions, this muscle and skeleton model was simplified under the supervision of a senior medical student. Only the bones of the arm and the shoulder, and the muscles responsible for the motions, flexion, extension, pronation, and supination, are chosen, i.e. the *musculus biceps brachii, musculus triceps brachii, musculus brachialis, musculus anconeus, musculus pronator teres, musculus pronator quadratus and musculus supinator*. The *musculus coracobrachialis* is added to this collection according to the medical partner's suggestion, although it does not contribute to the motions mentioned above.

The bone models are used as they were, but after consulting with medical professionals, we decided to recreate the muscle models, since the existing ones did not resemble their real counterparts closely enough. To achieve a more realistic visual effect, the muscle model should be deformed during the selected motions. The texture of the 3D muscle model is realistic, especially in a close up view. But it is quite difficult to apply vertex weights to the mesh topology of the original muscle model. In this application , most muscles are modeled with a cylindrical base mesh and created using polygonal modeling techniques and subdivision algorithms according to images from the Sobotta Atlas of Human Anatomy (see Figure 3.18). Only the *supinator* and the *pronator quadratus* are modeled from a flat box, due to their shape. The biceps and triceps are both created as one contiguous model instead of separate parts for each head and detailed attention is given to the anchor points for all muscles to make sure they are at the correct positions.

Muscle deformation   To enable the animation of the arm model, the bones are weighted to different armatures that consists of one control object. The ulna and radius are weighted to separate control objects to behave correctly when the hand is rotated. To deform the muscles, the vertices of each model are weighted to several control objects, which can be transformed easily according to the bones (see Figure 3.19). The vertices follow the transformation of each related object to a special percentage, which is determined by their weighting. The muscles' animation for contraction and expansion was achieved in 3DS MAX[5] via stretchy bones, which automatically contract or expand, if their length is changed. Whenever a new user is detected, the neutral length of each muscle is recalculated after the skeleton is scaled. This is accomplished by a configuring class attached to each muscle's control object. It contains references to the control object's two anchor points and uses them to define its position and rotation. The position is set directly to that of the first anchor point, while the rotation is calculated via a relative rotation. The application measures the distance between the control

---

[4]http://www.anatomium.com/
[5]http://www.autodesk.com/products/3ds-max/overview

(a)Flexors

(b)Extersors   (c)Pronators   (d)Supinator
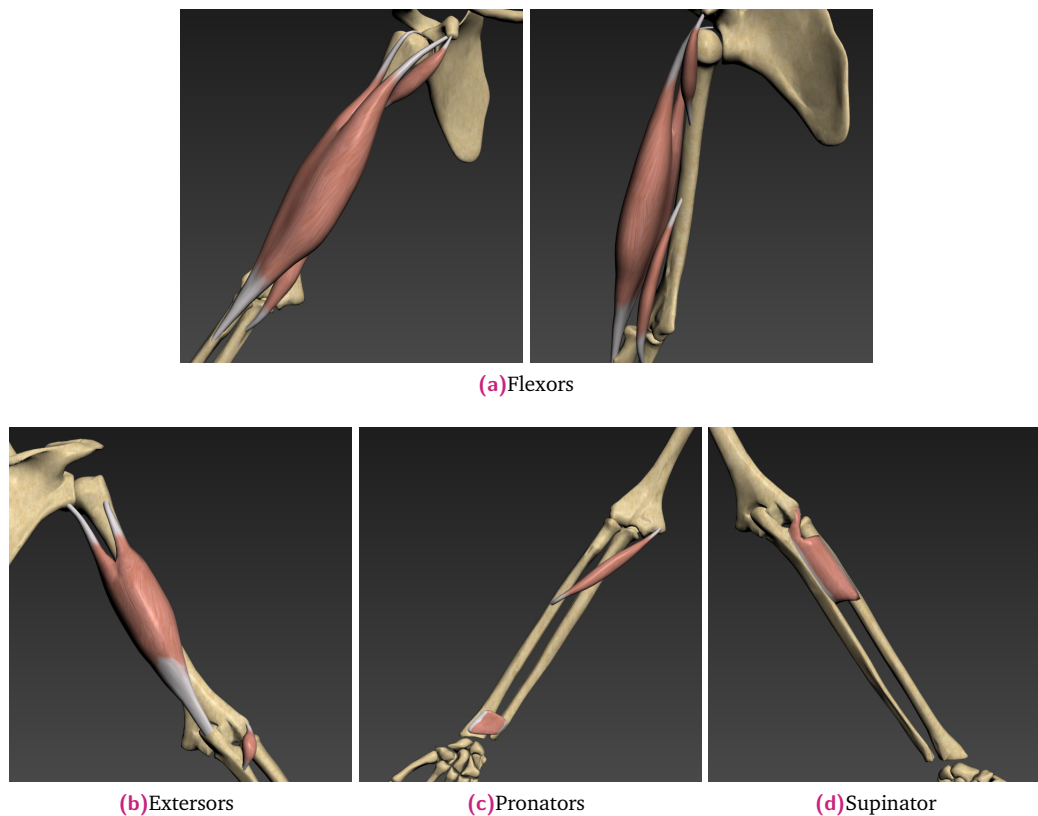
Figure 3.18 Dynamic Muscle Models. Most muscles were modeled with a cylindrical base mesh that was modified using polygonal modeling techniques and subdivision algorithms according to images from the Sobotta Atlas of Human Anatomy. Each picture shows the muscles involved in each motion.

object's anchor points and scales it along its local x-axis according to the ratio between that distance and its neutral length. Along the y- and z-axis the inverse of that ratio is used for scaling, multiplied by a factor to control the magnitude of expansion.

Learning functions

Virtual view   The details of the muscle are very important, so a virtual view is introduced to present a close up view of the arm model to show the attachment points and the spatial relationship. There are several solutions to place the virtual camera. One is that the camera rotates around a special object and generates the virtual view from different angles. The cameras are attached to a rotating helper object to create an orbiting movement around the arms. The close views are implemented with additional cameras and copies of the arm model. By putting these copies into separate layers it is ensured that these cameras only render their respective objects. The first copy takes part in just the flexion and extension, while the second one participates in the pronation and supination. The muscles, which are not responsible for the selected movement, are removed, to keep them from obscuring the relevant muscles or occlude them.

Self-control virtual view   A self-control virtual camera is introduced and its position and rotation can be controlled by the user. When the user's right hand is placed close to the left arm, the self-control camera model is tr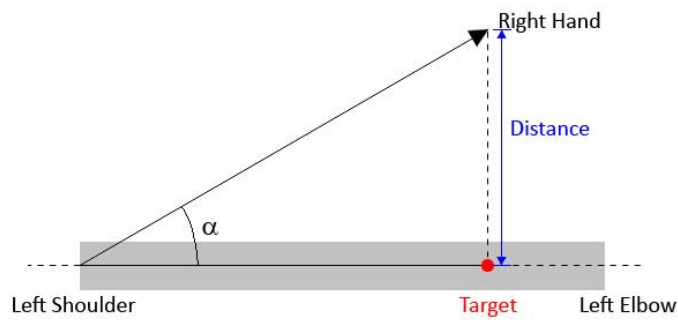iggered. The camera moves to the position of the hand and tries to focus on the desired area by calculating the closest point on the arm and

**Chapter 3**   Personalized Augmented Reality

(a)Arm Model   (b)Armature   (c)Armature with Muscles
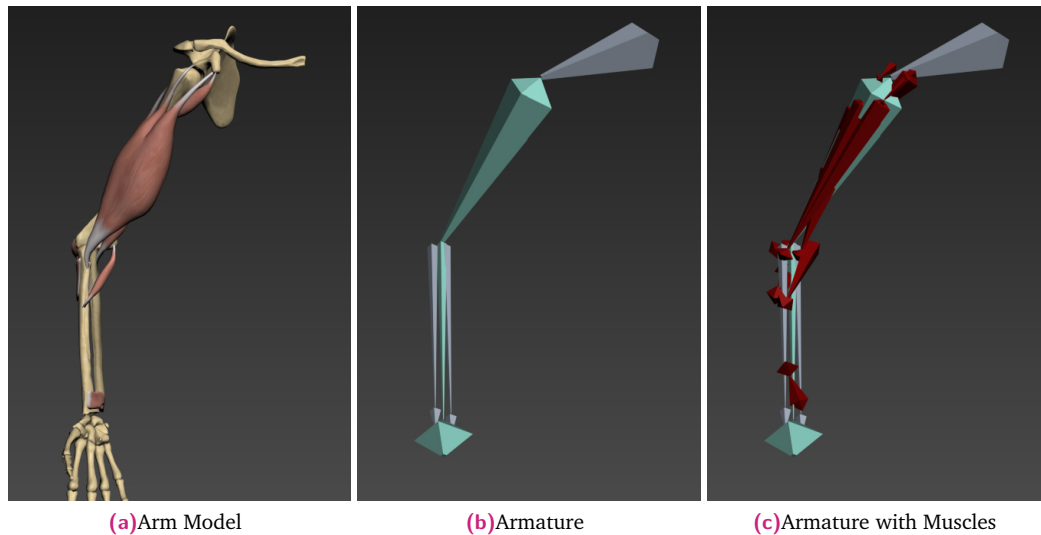
Figure 3.19 Dynamic Muscle Models. The bones are weighted to different armatures and the vertices of each model are weighted to several control objects in the armetures.

rotating towards it (see Figure 3.21). This point is computed in the user-controlled arm's local coordinate system and transformed into that of the arm currently observed. As shown in Figure 3.22, the user can naturally move the virtual camera to focus on different objects from different view angles. This function introduces more friendly interaction for muscle learning.

Interactive learning   A configuring class is defined, including methods to control the muscles' visibility, the highlighting of working muscles and of anchor points, and the labels that display the muscles' names and function. For highlighting it determines the angle between the upper and lower arm, as well as the rotation of the hand, relative to the frame before. This information is used to determine which motion is being performed, and in turn which muscles are active according to Table 3.5. Additional functions are created to toggle the visibility of all muscle-labels to the user interface and control the visibility of muscles by type. Similar to the muscles, the bones are controlled by a configuring class. These determine the color-coding of the bones and expose this functionality to the UI. In the muscle oriented model, one or more muscles are shown successively and the AR view and virtual view of the muscle is generated. The user can perceive the detail of the muscle, the start and end point and the spatial relationship with the bone (see Figure 3.23). In the motion oriented model, users are asked to perform special arm movements and only the muscle involved is highlighted (see Figure 3.20b). Then the functions of the muscle are learned. During the learning procedure, the user can always move the arm and see the muscle state and shape deformation. When a motion is performed, active muscles would be visually highlighted in order to make them more prominent.

Learning Sequence   Based on the AR and Virtual view, the system can generate a lot of learning material. We prefer a simple user interface and let the user focus on the learning instead of how to control this system. Hence, we introduce another function, "Learning Sequence". The sequence of events for the learning session is predefined via a configuring file.

(a)View 1　　　　　　　　　　　　　　(b)View 2

Figure 3.20 Virtual Camera concentrating on Flexion and Extension.

Figure 3.21 The camera moves to the position of the hand and tries to focus on the desired area by calculating the closest point on the arm and rotating towards it.
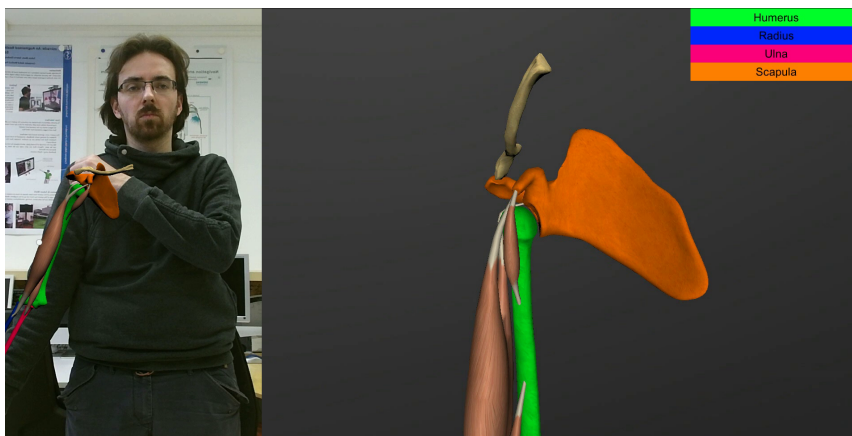
Figure 3.22 When the user's right hand is placed close to the left arm, the self-control camera model is triggered.
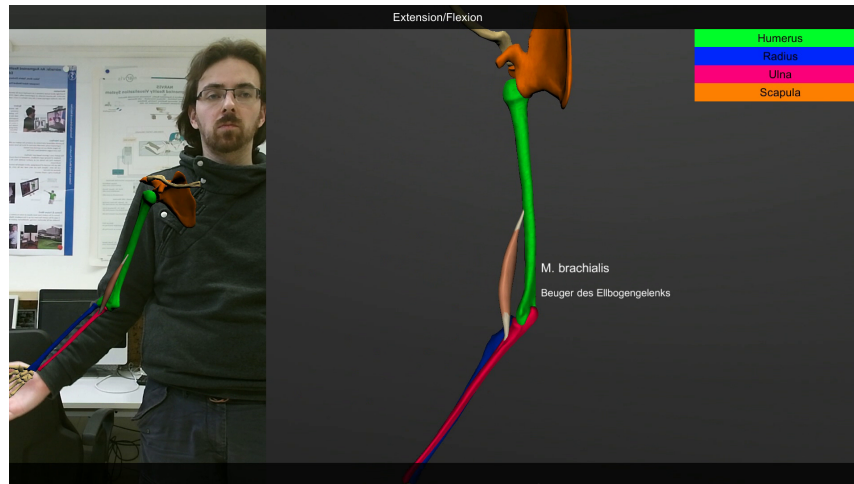
Figure 3.23 Muscle oriented model: one muscle is shown and the user can perceive the detail of the muscle, the start and end points, and the spatial relationship with the bone.

This is an easily editable text format with values separated by semicolons. Each line in this file contains the data for one event. The data for one event consists of an integer for the duration and the camera mode, boolean values to enable or disable the color-coding for the bones, the muscles' labels and highlighting of their anchor points, and a bitmask that determines the muscles visibility. Each bit in this eight-digit binary number controls visibility for one muscle. An optional seventh value can be used to display text at the bottom of the screen. This file is opened and read line after line by a simple parser at the start of the program, and the data is stored in an array of event structures. The sequence is played using a coroutine that takes one entry from the event-array and uses its values to set the relevant parameters and waits for the specified time before triggering the next event, until it reaches the end of the array. Simple functions can control the running sequence by stopping and restarting the coroutine or by modifying the array-index to skip or replay events.

To facilitate the learning session, the users run through a predefined sequence of events. The learning sequence can be generated according to the user's needs. Here is one example. It starts with the AR-view, showing the bones of the arm, identified by color, with the muscles hidden. It then switches to the first close view to introduce the muscles for flexion and extension individually. Next to each muscle is a label displaying its name and additional information about its function. After the muscles contributing to each motion have been introduced, they are shown together to give the user the opportunity to recap. This process is then repeated for pronation and supination. The whole sequence runs for approximately ten minutes and can be stopped, paused or resumed at any time. Additionally, it is possible to skip events or return to passed ones. Both the color-coding and the labels showing the muscle-information can also be displayed independently from the learning sequence.

Based on our discussion with medical students we found that they really liked the AR visualization on their own bodies, however they were not used to the natural user interface and gestures required for interaction. A learning curve was necessary for them to appease initial frustrations they had when first using our system. Most students who tried the application viewed it as an interesting and engaging source of information. The possibility of viewing the arm from every direction offered them a new insight that they could not get from static

images. The students confirm the effectiveness of acquiring knowledge with an AR-program. However, this application is limited by necessity, so the textbook still constitutes a more universal resource and it can function as a useful supplement.

3.3.3 Serious gaming for rehabilitation

Leg injuries are common injuries in both sports and everyday life. After some complicated injuries many patients can not use their leg for a long period. Following this lengthy healing progress, a period of rehabilitation is required to restore functionality and resilience of the injured leg. Various rehabilitation exercises are used by physiotherapists to increase the patient's ability to keep balance. These exercises include balancing on one leg and moving the other leg or arms, or even simple weight shifting exercises. Other exercises require the patient to stand on one leg, bend their knee slightly and balance with a straight back. Balance boards can also be used to increase the difficulty of the balancing exercises. These rehabilitation exercises are tailored towards every patient individually and are often supervised by a physiotherapist. The conventional exercises are typically repeated many times which makes them tedious and feel like a chore.

Motor rehabilitation consists in iterative repeating exercises to strengthen the affected body area. Since the rehabilitation must be performed over quite a long period it is common for patients to get disinterested, and consequently, they perform the exercises in a casual and incorrect manner. Usually, motor rehabilitation occurs by regular meetings between patient and physiotherapist with depending on the availability of both. We believe that it becomes crucial for the patient to perform exercises at home to speed up the rehabilitation process. However, at home both therapeutic instructions and corrections by the therapist are not verbally provided to the patient, which potentially leads the patient to perform the exercises incorrectly.

A serious game for balance rehabilitation is developed for leg stability and balance training. It is designed to entertain the user while doing the exercise as well as monitor and correct the patient. To deliver motivation, the exercise has to entertain the patient via the Magic Mirror effect. It will encourage the user to continue the exercise and accelerate the healing process. The game also takes over monitoring work from physiotherapists, so the patient can do exercises on their own. It offers self-monitoring of stances, such as "standing on one leg" or "standing straight" and displays messages to warn the patient if they fail to keep their balance or stance. The resulting game is tested by patients and reviewed by doctors and physiotherapists to assess its usefulness with regard to motivating the patient and its usefulness from the medical perspective [132].

Desired functions of the game

The first requirement for the game is to monitor the patient and to ensure that the right poses and movements are performed for the rehabilitation exercise. The poses contain standing with one leg while the corresponding knee slightly bends hovering over the foot and standing with two legs on a balance board while keeping the upper body straight. Another requirement is to motivate the patient to continue doing rehabilitation exercises. This is achieved by adding gaming components to the tedious exercise via the non-physical mirror visual effect. Varied mixed reality sense and different challenges are implemented to attract the user. A score is

Leg set on the ground   Leg held up in the air   threshold border lines

Left Hand Side   Front View   Right Hand Side   red leg inside
                                                                    border ->
                                                                    stretched

z axis   x axis
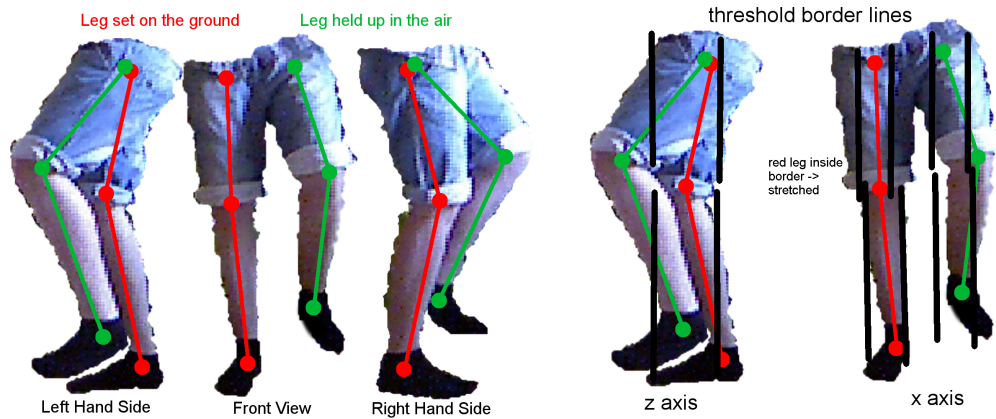
(a)The position of the lifted leg is changing bigger then the   (b)The lifted leg is not in the limited area.
other.

Figure 3.24 Stance detection. The system has to detect which leg the patient is standing on and the particular standing pose.

calculated based on the participant's movement. The user is awarded higher score if the pose is kept correctly, but the punishment by losing score is triggered after failing to keep a required pose.

To meet these requirements, the game is designed as a catch game based on the Magic Mirror framework. A mixed reality mirror view is generated and the corresponding targets are flying around the user, then they are asked to catch all the targets using the hands or lifting leg. During the game, the system also monitors the stance poses and displays sound and visual feedback when any wrong pose is detected. To monitor and correct the stance pose, *Stance Detection*, *Detection of unstable stances*, and *Stance Correction* are implemented. The game conceptions, such as *MR scenario*, *Difficulty Level*, and *Rewarding Score*, are designed to motivate the patient to continue the exercise.

Stance detection   To lift one leg and move around the arms while still remaining in balance is one of the common exercises for balance rehabilitation. The system needs to distinguish between the two legs and detect which one is lifted. The lifted leg can also be used as input for playing the game. During the exercise, the patient is required to stand with a slightly bent leg. The knee of the bent leg has to be above the toe of the corresponding foot. Hence, the system has to detect which leg the patient is standing on and the particular standing pose. The skeleton stream from the Kinect sensor includes the positions of all important joints of the user body in three dimensional space. As seen in Figure 3.24, the joints that are relevant to detect lifted legs are marked by green dots. There are two approaches combined to decide whether a leg is lifted up or not. The first one is to compare the position change of the leg joints. The leg which is left up would have larger movement than the other. The median position of the last 10 frames is compared with the current position. If the offset exceeds a distance threshold, the leg is chosen as the lifted one (see Figure 3.24a). The first approach cannot detect the lifted leg if the foot is held very still in the air. To supplement this case, a second approach is proposed. The program calculates a limited area for each leg according to the ground plane and the hip positions. If it detects that any leg joint is exceeding the limited area more than a threshold, it is considered lifted as well (see Figure 3.24b). The limited area and threshold

(a) Angle of the bent Knee.   (b) Player not standing with straight upper body.
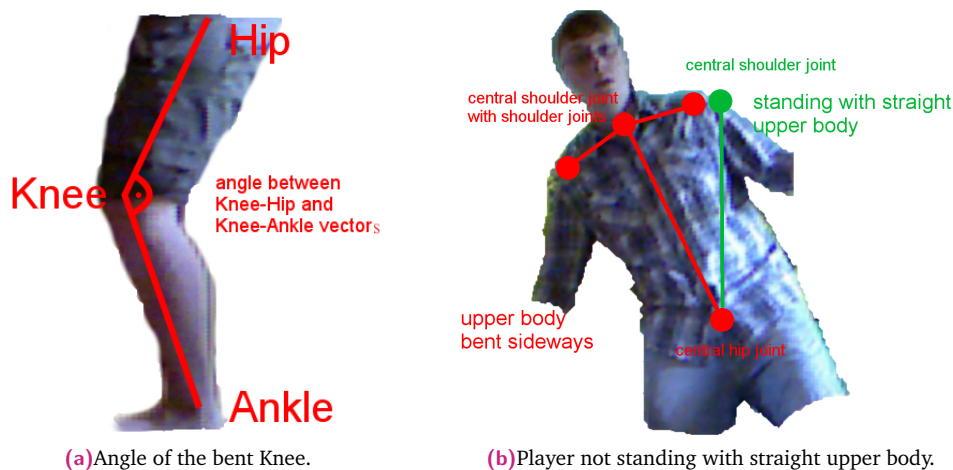
Figure 3.25 Detection of stance states. Angle of the bent Knee is calculated based on the skeletal joint position and the spine direction is calculated to check if the user stands straightly.

are adjusted to the size of the person. The limited area is marked by black lines. A stable solution is to combine both approaches to detect the lifted leg. To ensure that the knee is bent properly, the angle between the three dimensional "Knee-Ankle" and the "Knee-Hip" vectors is calculated (see Figure 3.25a). If this angle gets over $175°$, the leg can no longer be considered bent. Bending the leg too much also can be detected with this method.

Detection of unstable stances   The objective of the balance rehabilitation is that the patient can keep the upper body stable while standing with one leg or on a balance board. The system has to detect if the upper body is kept straight all the time. A large change of the shoulder position can be also used to check if the user has lost their balance. The spine direction also has to be calculated to check if it still remains as straight. It calculates the vector from the shoulder center (neck) to the center of the hips. For a person standing straight, the vector should be perpendicular to the ground plane. The angle between the vector and plane is calculated to estimate the state of the upper body (see Figure 3.25b). Normal movement should not instantly be detected as loss of balance, such as small movements to keep balance. Large, fast movements or twisting of the body however need to be detected. The person in Figure 3.25b is not standing straight and has lost his balance. The user then has to be notified and prompted to correct his stance and regain balance. The length of the vector between the center of the shoulders and the center between the hips is calculated and used to calculate appropriate thresholds. The game should be able to be played by patients in different states of their rehabilitation. For this, it has to offer a certain threshold of customization. The customization mainly includes restrictions to stances allowing an easier start for the patient. For patients who are further advanced in the rehabilitation process, it can offer a challenge setting.

Stance correction   This function provides feedback to the patient after detection of incorrect pose and unstable stance. The feedback can be printed on the screen as a normal message and special sound to attract the attention of the user. Instead of an error message, which discourages the user, the system tells the patient what to do to recover the correct pose. When a balance violation is first detected, one hint message is generated to recover the balance.

Figure 3.26 Rehabilitation gaming. A virtual background image is displayed in the Magic Mirror to generate a MR environment and a user tries to catch the moving objects around him.

Then the message is shown on the screen and a special sound effect is played to attract the attention of the user. These messages are displayed for a configurable amount of time. While one message is displayed, no other message can overwrite it. We think displaying multiple correction messages would confuse the patient. The new message is allowed to overwrite the existing message before the predefined time or when the correct pose is recovered. The system also has to detect the case when both legs remain on the ground and then the patient has to receive a message that asks them to lift one of their legs. If this rule for bent legs is violated, another message is displayed to tell the patient how to recover it.

MR scenarios   An MR scenario is designed to motivate the patient following the general rules for games. Based on information from the Kinect sensor, a segmentation of the patient's body in color image is generated. A virtual background image is displayed in the Magic Mirror to generate a MR environment (see Figure 3.26). To motivate the patient to continue playing the game, a wide variety of themes are offered. The background image, target object and its movement patterns are thematically designed to create a new experience for a different theme. Sound effect is also employed to enhance the experience of the players. The scenarios can either be unlocked after achieving a certain proficiency at the game, e.g. achieving a certain amount of score, or be available from the beginning to offer a wide variety of choices to the user.

Difficulty level   Games usually offer variety, challenges, and difficulty levels, and rewards the player if he or she overcomes the difficulties and challenges. Our game offers difficulty settings that are configurable independently for every limb. The difficulty influences the speed as well as the patterns of the moving targets. Higher difficulty means higher speed and more complex movement patterns. The player starts with an easy basic mode and continues with more difficult levels after achieving a special objective or a certain amount of score.

Rewarding score During the game, the player gets rewarded when the balance is kept continuously for a period of time. The system also decreases the score when a incorrect pose is detected. The score for each target is set according to the difficulty and the scenario. The easier the mode, the less score gets awarded. The final score gives the player a sense of accomplishment while also gaining a way to compare their performance in the game with previous gaming performances.

Conclusion

The mixed reality game is more immersed and interactive than normal exercises and thus offers enjoyment while doing rehabilitation exercises, motivating the patient. Playing an MR game helps the user to generate a sense of enjoyment when doing rehabilitation exercises. Having fun through exercises can motivate the patient to perform their exercises more frequently. The Magic Mirror concept also offers the users to observe themselves as looking into a mirror to perceive the wrong stances. The visual and sound feedback provides them instant hints to recover the correct pose. With the help of this system, the patient would be able to do exercises at home without being constantly monitored by a physiotherapist. All the data about the exercise can be collected and analyzed by the physiotherapist later via WiFi transmission for example.

Personalized User Interface 4

Human centered computing aims at adapting computers to human minds and habits, and an important research direction in this area is the design of natural user interfaces and ergonomic human-computer interaction techniques. Pointing gestures are fundamental to human behavior and are used consistently across cultures. In section 4.1, we first propose a method to accurately recover the *eye-rooted* pointing ray without tracking the eye gaze in an egocentric setting, enabling the user directly performing pointing gesture to interact with ambient media and objects without visual feedback. Then, section 4.2 introduces a novel personalized user interface, allowing the surgeon to personally perform touchless interaction with the various medical systems, and switch effortlessly among them, all of this without modifying the systems' software and hardware.

4.1 Pointing gesture recovery in an egocentric setting

Pointing gestures are fundamental to human behavior [97] and are used consistently across cultures [99]. The gestures begin at an early developmental stage [20] and let humans reference proximal objects as well as abstract concepts in the world. Today, pointing gestures are not only part of our gestural language but are inherently used for interaction [109]. As shown in Figure 1.2(a), there is a limited working zone when positioning a camera next to the monitor or on the ceiling to observe the user. As an alternative to the traditional setting of external sensors observing the user, an egocentric setting mounts a portable camera on the head or clothing of the user and perceives the interaction from an egocentric perspective [44, 90]. However, there is a significant angle between the ray starting from the camera and that from the eye when the user tries to point towards objects which are not reachable (see Figure 1.2(c)).

In this section we focus on how the pointing origin and geometry can be modeled in an egocentric setting, i.e. to understand the user-specific pointing gesture without external camera(s) or tracker(s) from the egocentric view. Regarding the ray's origin, pointing techniques can be classified into two groups: *hand-rooted*, where the ray originates at the user's hand, and *eye-rooted*, where the ray starts at the user's eyes. Whenever a *hand-rooted* technique is used, the objects that are along the pointing ray might differ from those that the user focuses on [5]. The more interesting scenario is the latter. Thus, we focus on the *eye-rooted* approach and demonstrate how to approximate and recover the pointing geometry within an egocentric setting.

The pointing gesture recovery in an egocentric setting would pave the way for performing real-time interaction with various digital objects and medias. With the similar viewpoint as the user, it allows the user to perform direct interaction with objects in the most natural way without relying in visual feedback. The pointing ray in the *eye-rooted* technique is defined

jointly by the user's eyes and fingertip positions. Forsberg et al. [50] specified that the choice of the origin point was an important consideration. This issue has been ignored in literature since: (i) the eye position is known in a virtual environment, and (ii) the user's dominant eye is customarily tracked using external cameras and trackers as the origin. Most often a visual feedback is also provided to improve the accuracy.

We propose a method to calculate a virtual eye center as the origin for the *eye-rooted* pointing ray technique without tracking the eye gaze in an egocentric setting. The user specific pointing geometry can then be recovered for interaction with ambient media and objects without visual feedback. The pointing ray starts from the virtual eye center and goes through the fingertip to the desired target information. During the initial calibration, the user is asked to perform pointing gestures towards several targets displayed within the field of view. A user specific virtual eye center is then estimated based on the detected fingertip in the coordinate system of the wearable device. Then the *eye-rooted* pointing ray is recovered when a user is performing a pointing gesture. We provide both a mathematical validation and a user study involving ten participants to assess the precision of our unique pointing gesture recovery.

4.1.1 Methodology

General Design

In a general hardware design, the egocentric setup should include a wearable sensor, which is put on the user's head and that contains two sensors, S_f and S_c. S_f is used to detect the fingertip and calculate its 3D position. S_c acquires the contextual information of the objects in front of the user. In general, S_f should provide a near-range depth map as the 3D position of the fingertip has to be accurately calculated. The contextual information is defined as the geometry and 3D position of the objects and medias perceived by the user. S_c can be a color camera, a depth sensor or a RGB-D sensor acquiring 3D position of given objects with known geometry or a valid depth map of the real world. As S_c and S_f are just different types of camera, the self-calibration of the wearable device can be performed using Zhang's method [167].

Pointing At Several Targets (PAST) calibration

Both human eyes perceive the visual information as a stereo camera and the images are processed by the brain to generate a 3D scene of the objects in the field of view. One always perceives a unified 3D scene no matter how we move our eyes. In other words, there could be a self-centered fixed origin point of the 3D scene in our brain, which we try to approximate as the virtual eye center. When performing pointing gestures, users always can directly position their fingertips somewhere between their eyes and the target generating a line of sight, which goes through the fingertip and the target. Using perspective-based pointing [126], the pointing geometry is modeled as a ray, which starts from the virtual eye center and goes through the fingertip towards the target, as shown in Figure 4.1. It is important to emphasize that the virtual eye center is user specific and its location is affected by the fact that users may be left or right eye dominant (see Figure 4.12). The PAST method accounts for this as its objective is to find the origin of the user specific pointing geometry. A calibration is required to calculate the origin and the result is the position of the virtual eye center E. As shown in Figure 4.1, the target T_i is presented in front of the user, who is then asked to perform a pointing gesture
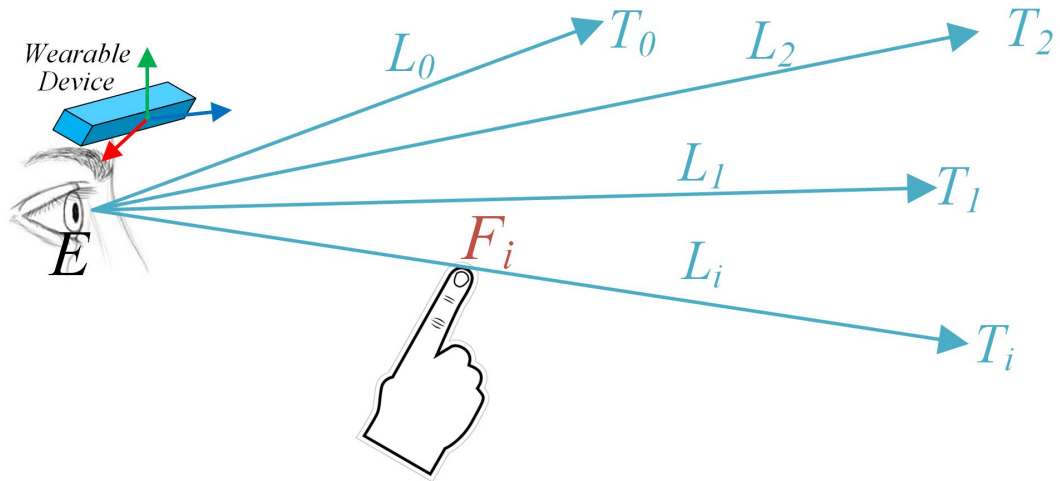
Figure 4.1 The PAST Calibration. The user is asked to point at several targets T_i using one finger. The pointing rays which pass through the fingertip F_i and T_i are collected. The intersection of these lines is E, the user specific virtual eye center.

towards it. The 3D positions of the fingertips F_i and T_i are collected through S_f and S_c, respectively. The pointing ray L_i ,which passes through F_i and T_i, would also go through E. After enough pointing rays are collected, E is defined as the best approximation of their intersection. (The PAST calibration procedure is similar to Single Point Active Alignment Method (SPAAM) [154], which was designed to calibrate optical see-through HMDs. Our method, however, does not require any display and our main objective is to estimate a virtual eye center for *eye-rooted* pointing gesture.) After calibration, the user specific virtual eye center E is estimated and the pointing rays can be recovered as L_i, going through E and F_i, when a pointing gesture is performed.

Target the user points at

As defined in section 4.1.1, a mesh surface M can be generated based on the contextual information which is observed via the sensor. After the PAST calibration, the user specific pointing geometry can be recovered as the pointing ray L_i. As shown in Figure 4.2, the intersection between L_i and M is T_i, the target which the user points at. No additional camera or marker is used to track the user's eye gaze or head and the user can directly interact with the media and objects in the real world using solely natural pointing gestures without any visual feedback.

4.1.2 Mathematical analysis of the PAST calibration

The pointing gesture is fundamental to human behavior and is also quite personal [97]. If the images in the user's eyes were accessible, then F_i , the fingertip detected in S_f, and E, the eye center calculated in the PAST calibration, would be perfect. In our proposed system, the user is asked to focus on the target and not on the fingertip. This allows the user to perform the pointing naturally. In addition, the fingertip detection is based on a noisy depth image. Thus, the fingertip used for pointing in the user's view is not always the same as the one detected by S_f. In this section, we first certify that there is a valid point E for the collected pointing rays,

even when the fingertip cannot be detected consistently. Then the accuracy of the recovered pointing ray is evaluated.

As shown in Figure 4.3, F_i and T_i are the fingertip and the target, which are detected by the wearable device. E is the virtual eye center calculated through the PAST calibration procedure and L_i is the pointing line estimated by the wearable device. If $\bar{F}_i$ and $\bar{T}_i$ are the fingertip and target in the user's own view, then a vector V_{fi} can be defined to present the offset between $\bar{F}_i$ and F_i, $F_i = \bar{F}_i + V_{fi}$, when the pointing is performed. During the calibration procedure, the user is asked to point towards a predefined target T_i, so $\bar{T}_i = T_i$. $\bar{E}$ is defined as the ideal virtual eye center and $\bar{L}_i$ is the ideal pointing line in the user's view.

Calibration

Based on our model of the pointing geometry $\bar{L}_i$ is starting from $\bar{E}$ going through $\bar{F}_i$ and ending at the desired target $\bar{T}_i$. We use Eq.4.1 to represent all the points along the pointing ray $\bar{L}_i$, and $\bar{E}$ is the intersection of all the pointing rays $\bar{L}_i$.

$$P_{\bar{L}_i} = \bar{T}_i + \bar{\lambda}(\bar{F}_i - \bar{T}_i), \bar{\lambda} \in \mathbf{R} \tag{4.1}$$

This means that there is a $\bar{\lambda} = \bar{\lambda}_i$ making both Eq.4.2 and Eq.4.3 true for every $\bar{L}_i$.

$$\bar{\lambda}_i = \frac{D_{\bar{E}2\bar{T}_i}}{D_{\bar{E}2\bar{T}_i} - D_{\bar{E}2\bar{F}_i}} \tag{4.2}$$

where $D_{\bar{E}2\bar{F}_i}$ is the distance from $\bar{E}$ to $\bar{F}_i$ and $D_{\bar{E}2\bar{T}_i}$ is the distance from $\bar{E}$ to $\bar{T}_i$.

$$\bar{E} = \bar{T}_i + \bar{\lambda}_i(\bar{F}_i - \bar{T}_i) \tag{4.3}$$

The pointing ray estimated by the wearable device is L_i, which goes through T_i and F_i. All the points along L_i can be represented as:

$$P_{L_i} = T_i + \lambda(F_i - T_i), \lambda \in \mathbf{R} \tag{4.4}$$
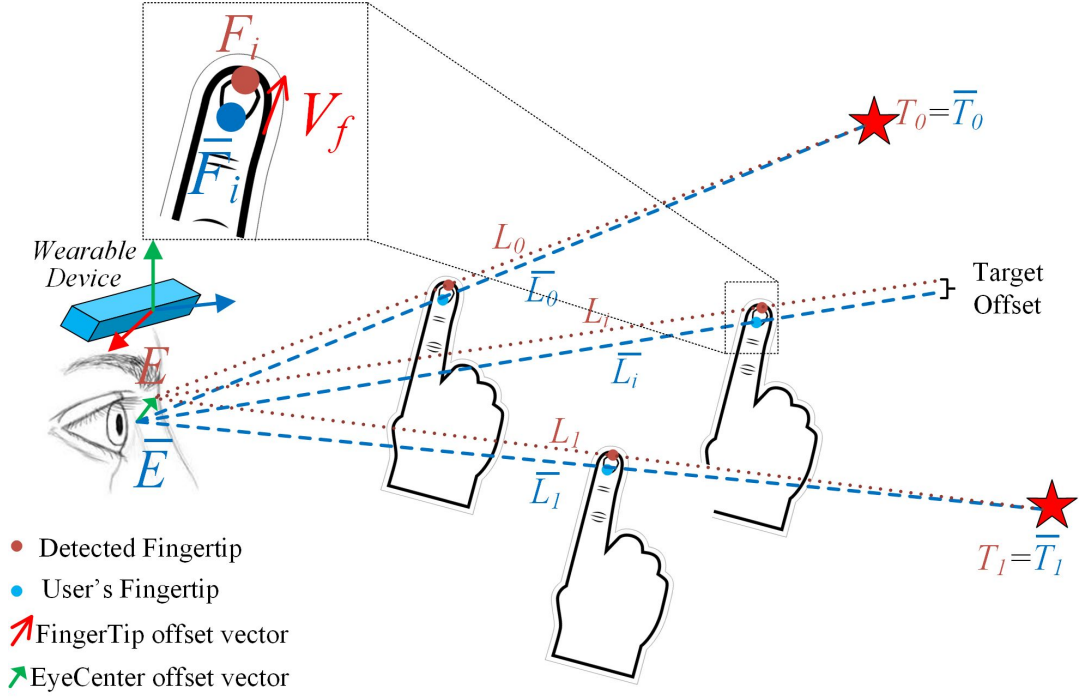
Figure 4.3 The pointing geometry during the calibration and recovery procedure. There is a vector V_{fi} from $\bar{F}_i$ to F_i when the pointing gesture is performed and a vector V_E, from the ideal virtual eye center $\bar{E}$ to E.

F_i in Eq. 4.4 can be replaced by $\bar{F}_i + V_{fi}$ and $T_i = \bar{T}_i$ during calibration. Hence,

$$P_{L_i} = \bar{T}_i + \lambda(\bar{F}_i - \bar{T}_i) + \lambda V_{fi}, \lambda \in \mathbf{R} \tag{4.5}$$

There is a point E_i on the line P_{L_i}, where $\lambda = \bar{\lambda}_i$ in Eq.4.5.

$$E_i = \bar{T}_i + \bar{\lambda}_i(\bar{F}_i - \bar{T}_i) + \bar{\lambda}_i V_{fi} \tag{4.6}$$

After substituting Eq.4.3 into Eq.4.6, there is always a point E_i along every L_i that can be represented as:

$$E_i = \bar{E} + \bar{\lambda}_i V_{fi} \tag{4.7}$$

As shown in Figure 4.4, $\bar{\lambda}_i$ in Eq.4.2 decreases slowly as $D_{\bar{E}2\bar{T}_i}$ increases in three different example cases. $\bar{\lambda}_i$ can be rewritten as:

$$\bar{\lambda}_i = \bar{\lambda}_c + \Delta\bar{\lambda}_i \tag{4.8}$$

where $\bar{\lambda}_c = 1.15$ is the average and $\Delta\bar{\lambda}_i \in [-0.1, 0.15]$, obtained from Figure 4.4. V_{fi} can be different every time as the depth map contains noise and the pose of the finger is different, hence it can also be represented as $V_{fc} + \Delta V_{fi}$, where V_{fc} is a stable offset and ΔV_{fi} is a small random deviation. Thus,

$$\bar{\lambda}_i V_{fi} \approx \bar{\lambda}_c V_{fc} + \Delta\bar{\lambda}_i \Delta V_{fc} \tag{4.9}$$
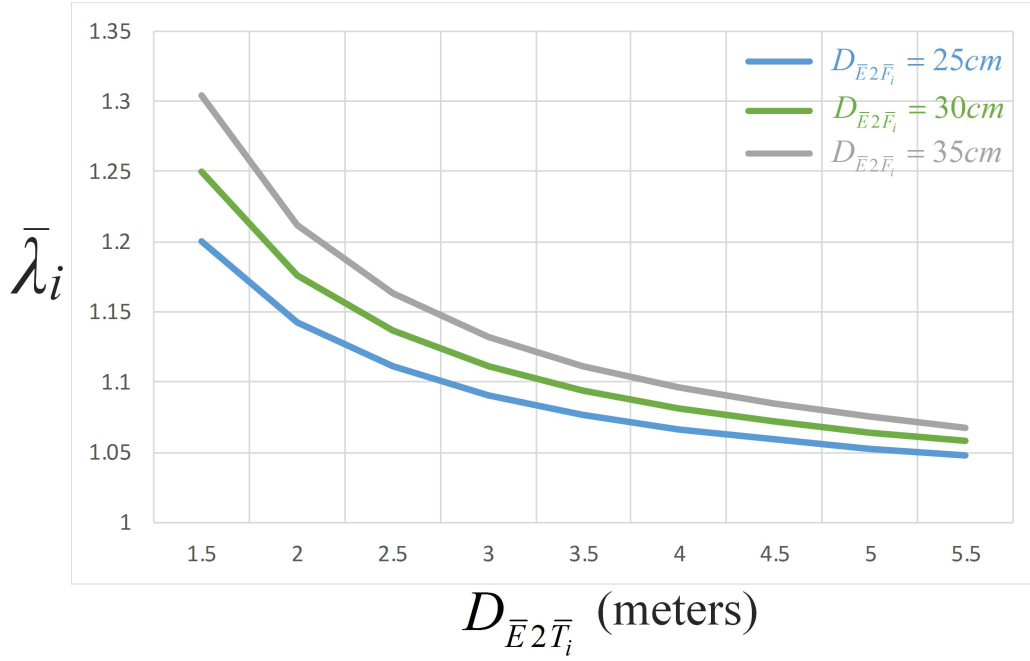
Figure 4.4 The curve of $\bar{\lambda}_i$. It decreases as $D_{\bar{E}2\bar{T}_i}$ increases when $D_{\bar{E}2\bar{F}_i}$ is fixed, and decreases as $D_{\bar{E}2\bar{F}_i}$ decreases.

where $\bar{\lambda}_c \Delta V_{fi} + \Delta \bar{\lambda}_i \Delta V_{fi}$ is very small. As $\bar{\lambda}_c V_{fc}$ is consistent, $\Delta \bar{\lambda}_i V_{fc}$ is the main contribution of the variation for $\bar{\lambda}_i V_{fi}$. There is a candidate intersection point E, which is very close to all the lines L_i.

$$E = \bar{E} + \bar{\lambda}_c V_{fc} \tag{4.10}$$

The offset between E and $\bar{E}$ is approximately $\bar{\lambda}_c V_{fc}$. Even when there is about 2 cm offset between F and $\bar{F}$, there exists a point E and its distance to any line L_m is not greater than 0.3 cm after substituting the numerical values in Eq.4.10. Hence, there exists a valid approximation for the intersection of the pointing line L_i making the PAST calibration feasible.

Pointing ray recovery

As shown in Figure 4.3, E is the calibration result and there is an offset $V_E \approx \bar{\lambda}_c V_{fc}$, between E and the ideal virtual eye center $\bar{E}$. When the user points towards a target after calibration, the position of the target calculated by the wearable device T_i, can be represented by E and F_i as:

$$T_i = E + \frac{D_{E2T_i}}{D_{E2F_i}} (F_i - E) \tag{4.11}$$

where D_{E2F_i} is the distance from E to F_i and D_{E2T_i} is the distance from E to T_i.
$\bar{T}_i$, the position of the ideal target in the user's view, can be expressed as:

$$\bar{T}_i \approx \bar{E} + \frac{D_{E2T_i}}{D_{E2F_i}} (\bar{F}_i - \bar{E}) \tag{4.12}$$

$\angle_E$, the angle error between the user's pointing lines to T_i and $\bar{T}_i$ can be calculated using Eq.4.13 since D_{E2T_i} is much larger than the offset between the targets T_i and $\bar{T}_i$.

$$\angle_E \approx \tan^{-1} \left(\frac{|T_i - \bar{T}_i|}{D_{E2T_i}} \right) \tag{4.13}$$
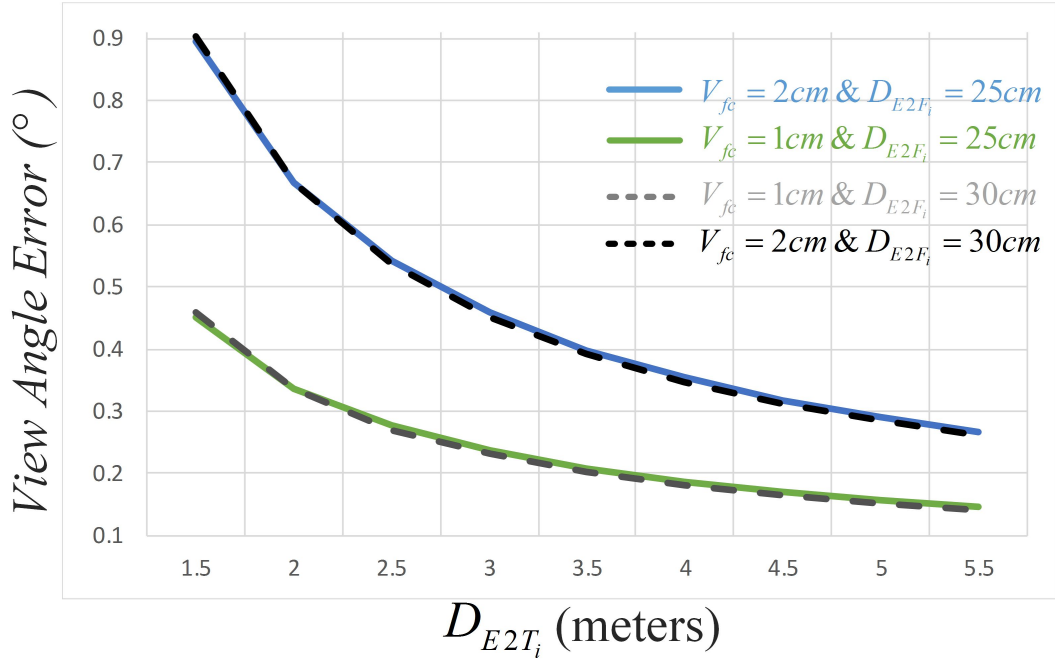
Figure 4.5 The curve of angle error. It decreases as D_{E2T} increases and V_{fc} decreases. D_{E2F} does not really affect the accuracy of the pointing gesture.

After substituting Eq.4.10, Eq.4.11 and Eq.4.12, we get :

$$\angle_E \approx \tan^{-1} \left(\frac{\lambda_c |V_{fc}|}{D_{E2T_i}} + \frac{|\lambda_i V_{fi} - \lambda_c V_{fc}|}{D_{E2F_i}} \right) \tag{4.14}$$

As $\angle_E$ is very small, we get after substituting Eq.4.9:

$$\angle_E \approx \frac{\lambda_c |V_{fc}|}{D_{E2T_i}} + \frac{|\Delta \bar{\lambda}_i V_{fc}|}{D_{E2F_i}} \tag{4.15}$$

As shown in Figure 4.5, the angle error decreases as V_{fc} decreases and D_{E2T_i} increases. D_{E2F_i} does not really affect the accuracy of the recovered pointing gesture. The accuracy of a recovered pointing line is below 0.9°in these four cases.

Based on the above feasibility analysis we can conclude that the PAST method can find a usable origin for the pointing ray which enables the wearable device to determine accurately where the user points to.

4.1.3 Implementation

In this section, we implemented a scenario with a wearable RGB-D sensor and two different displays. The 3D contextual information was acquired by tracking the display. During the PAST calibration process, the targets are shown on the display. The mesh, M, is then defined by the planar surface of the display. The display can be located at any angle and distance from the user.
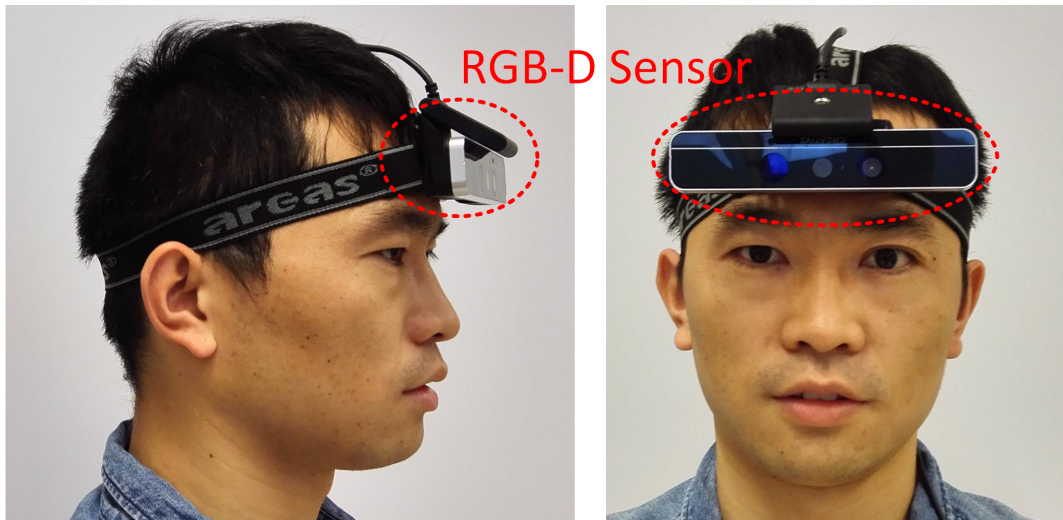
RGB-D Sensor

Figure 4.6 The hardware setup. The side and front view of the wearable device. The *RealSense* 3D sensor was strapped on the user's head using a bandana.

Hardware

An Intel *RealSense* 3D sensor was selected as the wearable RGB-D device. It contains a 640x480 depth sensor (Valid Range = 0.2-1.2meters) and a 1920x1080 color camera. The *RealSense* 3D sensor was strapped on the user's head using a bandana with an adjusted height and angle (see Figure 4.6). A 46″ TV display and a 24″ monitor presenting various information were positioned about the user as the contextual environment. The users could walk around and rotate their body or head freely while performing interactions with the display and monitor.

Fingertip detection

A real-time depth image of the hand or finger was generated via the near range depth sensor while the user performed pointing gestures. The depth sensor and the color camera were packaged in one device and were oriented in the same direction. When the color camera was adjusted to have a similar view as the user, there was no valid depth map of the full hand even when the user was asked to point upwards. Thus the fingertip detection technology developed by Betancourt et al. [12] was not employed. However, this limitation would be removed when a RGB-D sensor with the depth and color camera looking at different directions is developed commercially. In the meantime, with our current setup, the hand and finger blobs were directly extracted from the depth image using the *Intel RealSense* SDK, and the contours and convex hull were calculated using the method provided by Suzuki and Be [144]. The best peak was chosen as the detected fingertip position.

Contextual information

To simplify the task of acquiring the contextual information, the *ARTooKitPlus* [160] markers were employed to track the TV display and monitor. There were two implementations: (i) some markers were directly displayed on the screen without interfering information (see Figure 4.7-left). The markers were detected to calculate the planar function of the screen, and (ii) one physical printed marker was placed next to the display (see Figure 4.7-right). A calibration was then performed to calculate the transformation matrix from the printed marker to the screen. The printed marker was then detected to calculate the information of
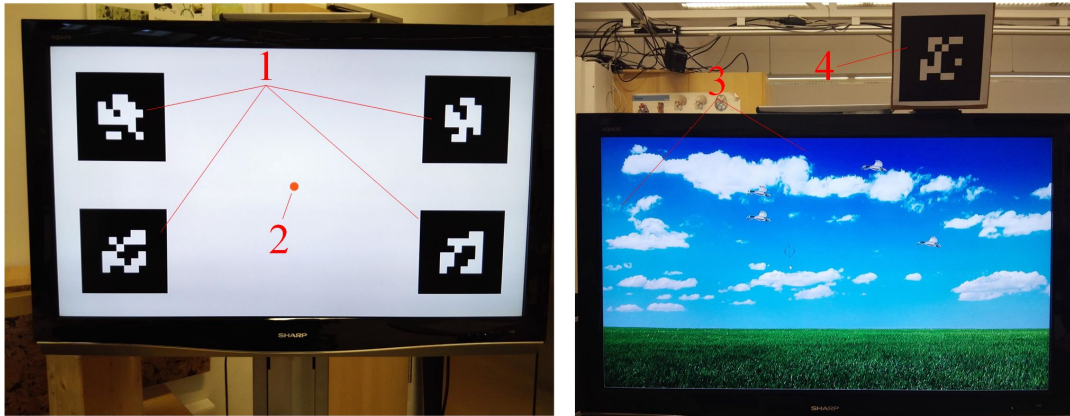
the screen. As shown in Figure 4.7, the screen was detected using the first implementation for the PAST calibration and evaluation, whereas the second implementation was employed during the *DuckHunt* demo (see Section 4.1.3) and the medical application.

DuckHunt Demo

Inspired by the Nintendo game, a *DuckHunt* demo[1] was developed using the pointing gesture to perform a dunk hunting task. Several flying ducks were shown on the screen and the user was asked to point to a duck to shoot it. After the completion of the PAST calibration, the user could directly perform a pointing gesture towards the screen to choose which duck to hunt without any visual feedback. The fire command was triggered when the user pushed the fingertip along the pointing line.

4.1.4 User study and results

To evaluate if a valid origin would be found by the PAST calibration and how accurate the recovered pointing ray was, we conducted a quantitative user study followed by qualitative evaluation to assess the proposed pointing technique in the egocentric setting. During the user study, several tasks were designed within different scenarios and the participants were asked to finish them and answer a questionnaire.

Scenario for experiments

The TV display and the monitor were placed around the participant. There were three basic tasks for the user to perform during the user study:

(A) *Calibration:* a red target was shown at 9 different positions on the TV display and the user was asked to perform pointing gestures towards the target using their pointing finger.

---

[1] https://youtu.be/Y4zO52oFDDg

(B) *Evaluation:* a blue target was shown at 9 different positions on the TV display or the monitor and the user was asked to perform pointing gestures towards the target using one finger. Data from 10 frames was collected and the average was calculated as the final result.

(C) *Evaluation with visual feedback:* a blue target was shown at 9 different positions on the TV display. The user performed pointing gestures using one finger towards the display and showcasing a green point as a visual feedback. The user was asked to match the green point and the blue target as close as possible. Data from 10 frames was collected and the average was used for accuracy analysis.

Study description

Participants   Ten people (6 Male and 4 Female, Age Range: 22-34) from different cultures were enrolled in the user study. Three of them grew up in Germany, three in China, two in Chile, one in France and one in India. Seven of them had a computer science degree. Nine were reported to be right-handed, while one was ambidextrous.

Study procedure   First, the participants put the wearable RGB-D device on their head and adjust the height and angle to make sure that the color camera has a full view of the TV display in front of them, while ensuring that the depth sensor had a valid depth map of their fingertip when performing the pointing gesture. Prior to beginning the user study, the participants were allowed to perform pointing gestures around the working environment and understood the volume in which their fingertip could be detected stably (this was limited by the field of view and pose of the depth camera). The user was asked to focus on the target of interest and not on their fingertip when performing a pointing gesture. After a simple training session, participants were asked to complete the following 10 steps for the evaluation.

(1) Perform task (A) three times from 2 meters away to complete the PAST calibration and evaluate if the calibration is stable. The first calibration result of the three is employed for all the remaining steps.

(2) Perform task (C) on the TV display from 2 meters away, verify the calibration result, and calculate the accuracy the user can achieve with the visual feedback.

(3) Perform task (B) on the TV display from 2 meters away to evaluate the accuracy of the recovered pointing ray.

(4) Perform task (B) on the TV display using the opposite hand from 2 meters away to evaluate if the calibration is independent of the hand.

(5) Perform task (B) on the TV display from 1.5,3 and 4 meters away, to evaluate the accuracy of the recovered pointing ray at different distances.

(6) Perform task (B) on the monitor from 2 meters away to evaluate if the calibration is independent of the display.

(7) Perform task (A) three times from 1.5, 3 and 4 meters away to evaluate if the calibration is independent of the target distance.
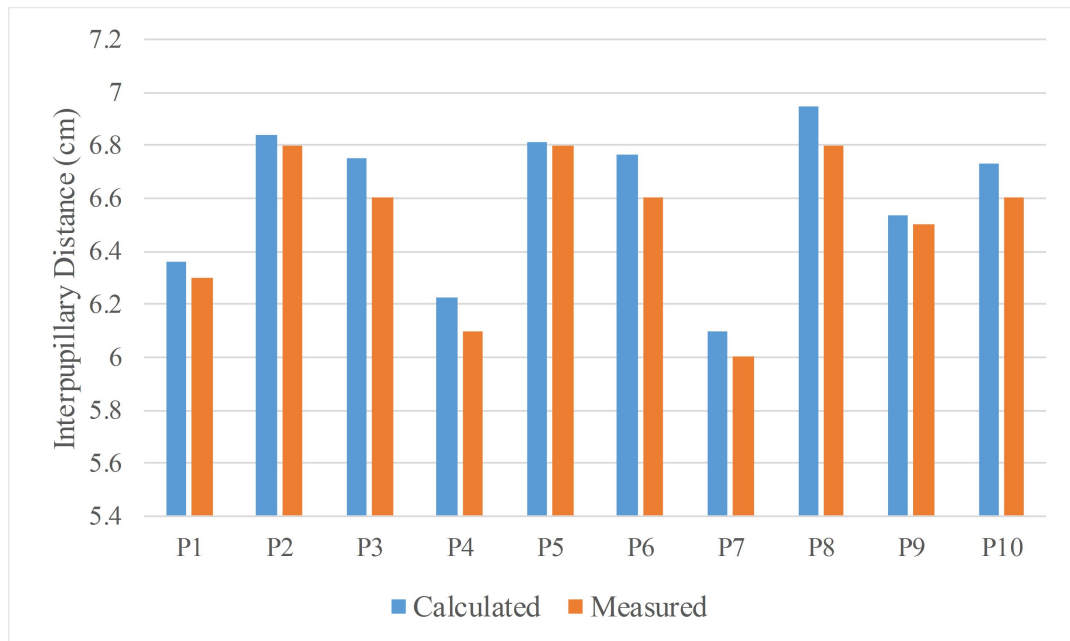
Figure 4.8 Interpupillary distances. The calculated interpupillary distances of the 10 participants were very close to the ground-truth obtained using a ruler.

(8) Perform task (A) with only the left or right eye open, from 2 meters away, to understand the relationship of the two eyes and the calculated virtual eye center.

(9) Play with the *DuckHunt* demo for 2 minutes using the first calibration result calculated in step (1).

(10) The interpupillary distance of the participant was measured with a ruler and the user completed a questionnaire to evaluate the pointing technique.

Results

PAST Calibration   The real interpupillary distances of the ten participants were measured in step (10). They varied from 6.0 to 6.8cm (see Figure 4.8). In step (8), the calibration resulted in the positions of the left and right eye in the coordinate frame of the head-worn sensor system since the participants only used one eye. The interpupillary distances were also subsequently calculated from the distance of the two eyes. As shown in Figure 4.8, the calculated distances were very close to the measured one using the ruler with an average difference of 0.10 ± 0.05cm. From this value, we conclude that the PAST method could accurately calculate the pointing ray origin based on the collected data. The average calibration results in step (1) were taken as the ground-truth to verify whether the PAST calibration was stable and independent of the scenario. For each participant, the offsets of the calibration results in step (1) and (7) relative to the ground-truth are shown in Figure 4.9. The offset in the XY plane was less than 0.5 cm and that along the Z-axis was less than 1cm in most cases. The offset along Z-axis was larger than the XY plane since the PAST method estimated the intersection of the pointing rays which oriented towards the Z-axis. The 3D distances from the virtual eye centers calculated at 1.5, 2, 3, and 4 meters away relative to the ground-truth were $0.60 \pm 0.35cm$, $0.56 \pm 0.41cm$, $0.90 \pm 0.73cm$, and $1.02 \pm 0.73cm$, respectively. The 3D distance calculated

from 1.5 meters was greater than the one from 2 meters away, since $\bar{\lambda}_i$ increases as the $D_{\bar{E}2\bar{F}_i}$ decreases (see section 4.1.2). The offsets at 3 meters and 4 meters away were larger since the marker tracking accuracy decreased with increasing distance. The four plots in Figure 4.9 show that the PAST method could provide a stable calibration result for different users and the calibration is independent of the contextual environment.

Accuracy of the pointing ray starting from E The angle error of the recovered pointing ray, $\angle_E$, is calculated according to Eq.4.15. The precision is the difference of $\angle_E$ in 10 consecutive frames while the participant performs pointing gestures to the same target. $\angle_E$ turns out to be independent of the rotation of the participant, the size of the target and the spatial relationship with the target and the fingertip. To get a clear distribution of $\angle_E$, the vector from the $\bar{T}_i$ to T_i is also calculated. As shown in Figure 4.10 and Figure 4.11, for every point in the plots, the distance to the origin is defined as the angle value in degrees(°), and the points are distributed according to the vector from $\bar{T}_i$ to T_i. The first PAST calibration result in step (1) is employed to evaluate the pointing ray in different scenarios, with and without visual feedback, from different distances, and towards a different display. The accuracy of the pointing ray with and without visual feedback is shown in Figure 4.10. The accuracy with visual feedback is $0.22 \pm 0.13°$ (see Figure 4.10(a)). The accuracy without any visual feedback from 2 meters away is $0.43 \pm 0.27°$ and $\angle_E$ is less than $1°$ as shown in Figure 4.10(b), which is consistent with the result presented in Figure 4.5. In addition, the precision of the pointing ray without visual feedback is $0.10 \pm 0.07°$. Comparing Figure 4.10(a) and Figure 4.10(b), the pointing ray starting from the virtual eye center E (the PAST calibration result), had the same level of accuracy as the case with visual feedback. Therefore, the pointing ray is accurately recovered in this scenario.

The pointing accuracies shown in Figure 4.11(a-d) are $0.69 \pm 0.52°$ from 1.5 meters away, $0.63 \pm 0.44°$ from 3 meters away, $0.85 \pm 0.94°$ from 4 meters away, and $0.70 \pm 0.57°$ using a $24''$ monitor, respectively. As expected in Figure 4.5, $\angle_E$ increased at 1.5 meters away in Figure 4.11(a), compared to Figure 4.10(b). However, Figure 4.11(b) and 4.11(c) show that $\angle_E$ did not decrease when the pointing was performed from 3 and 4 meters away since the accuracy of the marker tracking declined. Figure 4.11(d) demonstrates that the calibration was contextually independent and the pointing ray was still recovered when using the $24''$ monitor.

Figure 4.10 and Figure 4.11 show that $\angle_E$ is less than $1°$ in most cases in different scenarios. The overall pointing accuracy at 1.5, 2, 3, and 4 meters is $0.67 \pm 0.71°$. We conclude that using the PAST calibration result E, the virtual eye center can be set as the origin for the perspective-based pointing interaction and the pointing ray can be recovered accurately without any visual feedback.

Position of the virtual eye center   From step (1) and (8), we obtained the positions of the virtual eye center and the two eyes in the coordinate system of the wearable device. When all interpupillary distances were upscaled to 7 cm, the spatial relationship of the left eye, right eye, and the virtual eye center could be drawn in a 2D plane (see Figure 4.12). For four participants, the virtual eye center was very close to the left eye and these participants were determined as left-eye dominant using the method from WikiHow[2]. The position of the virtual

---

[2]WikiHow:http://www.wikihow.com/Determine-Your-Dominant-Eye
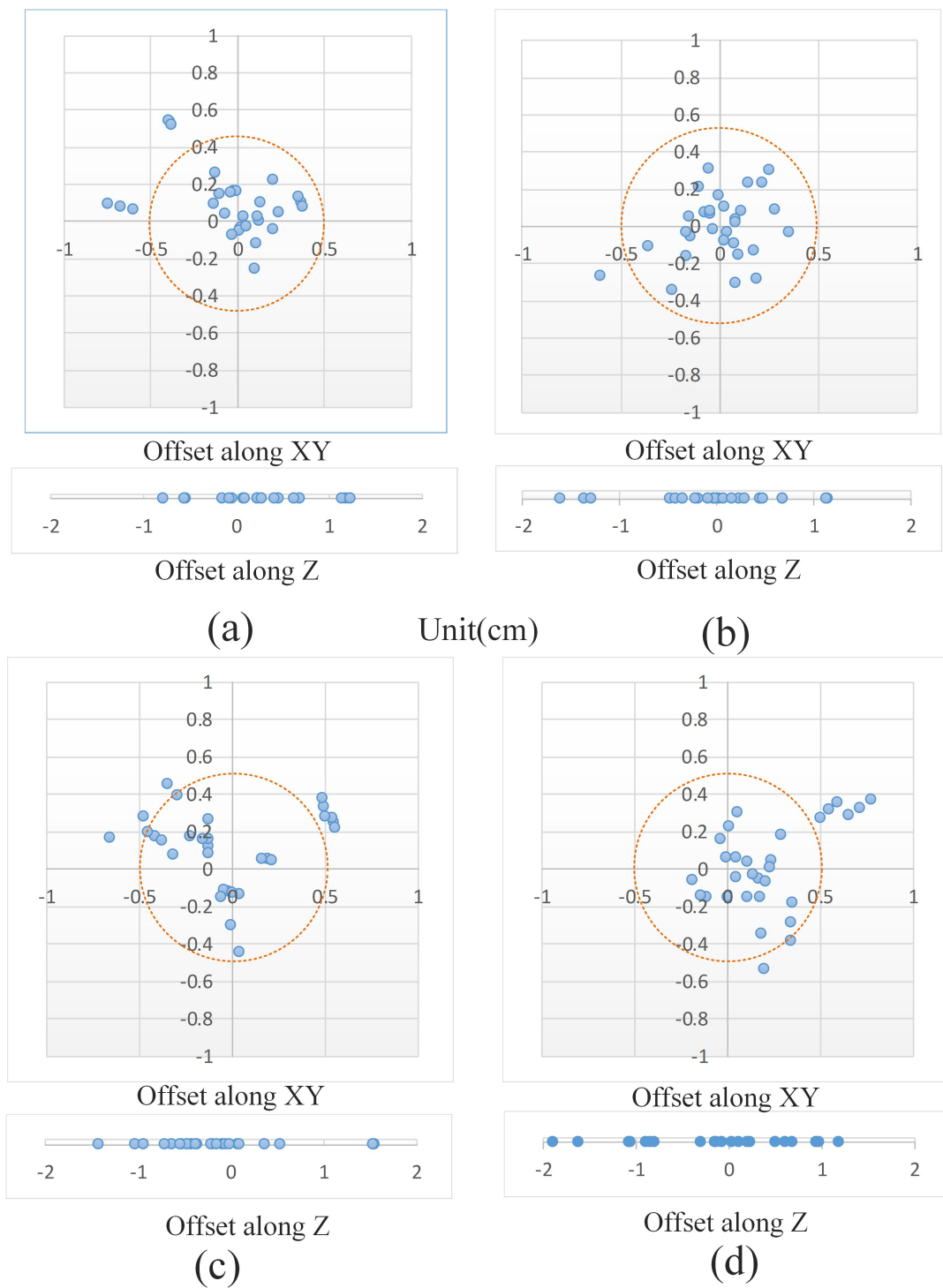
Figure 4.9 The calibration results in different scenarios. The results were compared to the reference one showing distribution of the offset along X,Y and Z-axis in the coordinate system of the wearable sensor. The XY plane was the one approximately parallel to the user's face. These four charts show the offset of the calibrations from different distances (a) 1.5 meters away (b) 2 meters away (c) 3 meters away (d) 4 meters away.

(a) Unit(degree) (b)
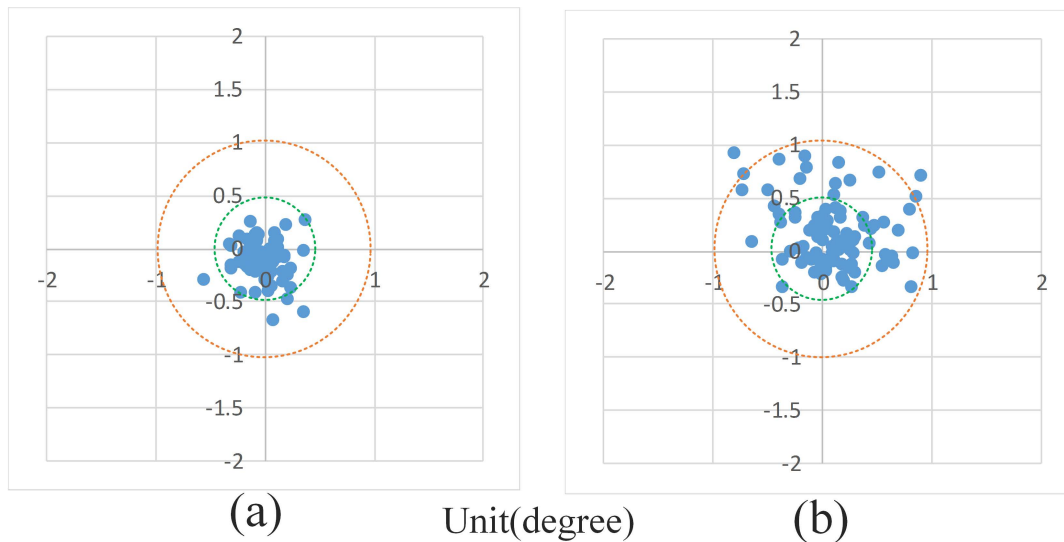
Figure 4.10 After calibration, the participants performed pointing gestures towards some targets to evaluate if the pointing geometry could be correctly recovered. (a) with feedback at 2 meters away, and (b) without feedback at 2 meters away.

eye center was very user specific and influenced by the dominant eye. The potential positions of the origin for the pointing ray are the two eyes and the middle of them [50]. These three points were ranked as close, middle, and far based on the distance to the virtual eye center E. The accuracy of the pointing rays starting at these three points is shown in Figure 4.13. It is obvious that the point close to E could be considered the best origin for the pointing ray. The PAST method can also be employed to estimate the origin when the eye positions are known. In step (4), the participants performed the pointing gesture using the opposite hand, different than the one used for calibration. In general, the results are good except for Participants 5 and 9 (see Figure 4.14). Finally, the PAST calibration was dependent on the fingertip detected from the depth sensor and influenced by the user's specific pointing habit. The hand pose sometimes was quite different when the participant performed pointing with the opposite hand, thus the virtual eye center may be different for each hand.

System usability To evaluate the proposed perspective-based pointing technique in an egocentric setting, a questionnaire was designed to obtain a qualitative evaluation. A Likert scale was used. The format of our 5-pt Likert was: (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, (5) strongly agree.

As shown in Table 4.1, all questions attained positive scores above 4.0. Q1 had the lowest score as some participants felt a little tired while continuously lifting their arm to perform pointing gestures. In the "DunkHunt" demo, the fire command was recognized when the user moved their fingers along the pointing line, which was a little different from a natural click hand gesture. Therefore, the participants required some time to get used to it. In general the participants liked the pointing interaction with an egocentric setup and could learn it easily and quickly.

Here we made a mistake while creating the questions in Table 4.1, the tense of which were not consistency. Hence, we want to declare that all the questions were asked after the users used the system for a while, and the users were clear that they were asked to answer the questions based on their experience.
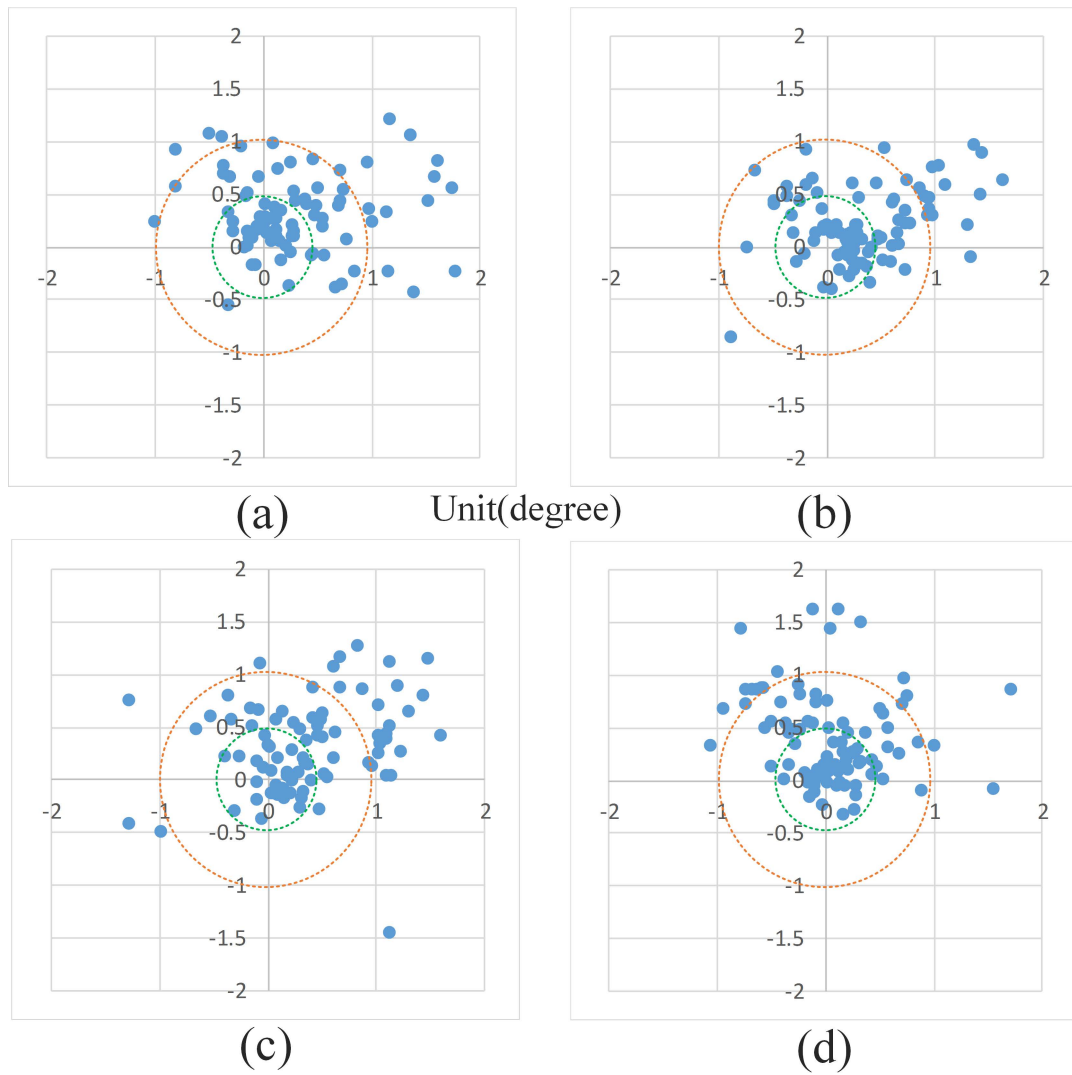
Figure 4.11 The participants performed pointing gestures towards some targets in different scenarios to evaluate if the calibration was contextually independent. (a) perform pointing from 1.5 meters away (b) 3 meters away (c) 4 meters away and (d) perform pointing towards the 24″ monitor.
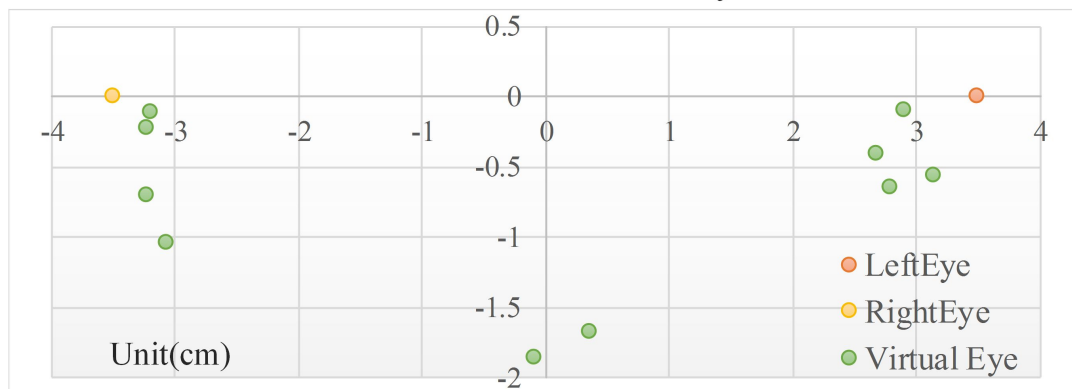
Figure 4.12 The spatial relationship of virtual eye center and the two eyes. The middle point of the two eyes was set as the original center and the calculated virtual eye center was shown in green.
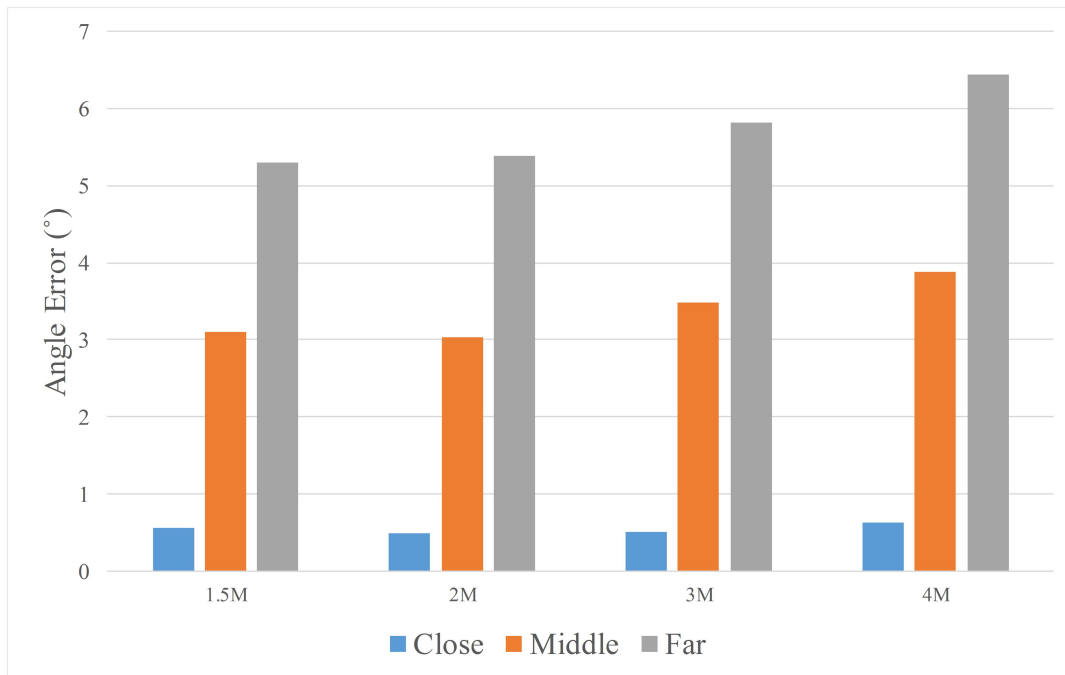
Figure 4.13 The accuracy of the pointing ray starting from the positions of the two eyes and the average of them. These three points were ranged as close, middle and far based on the distance to the virtual eye center E, calculated in the PAST calibration.
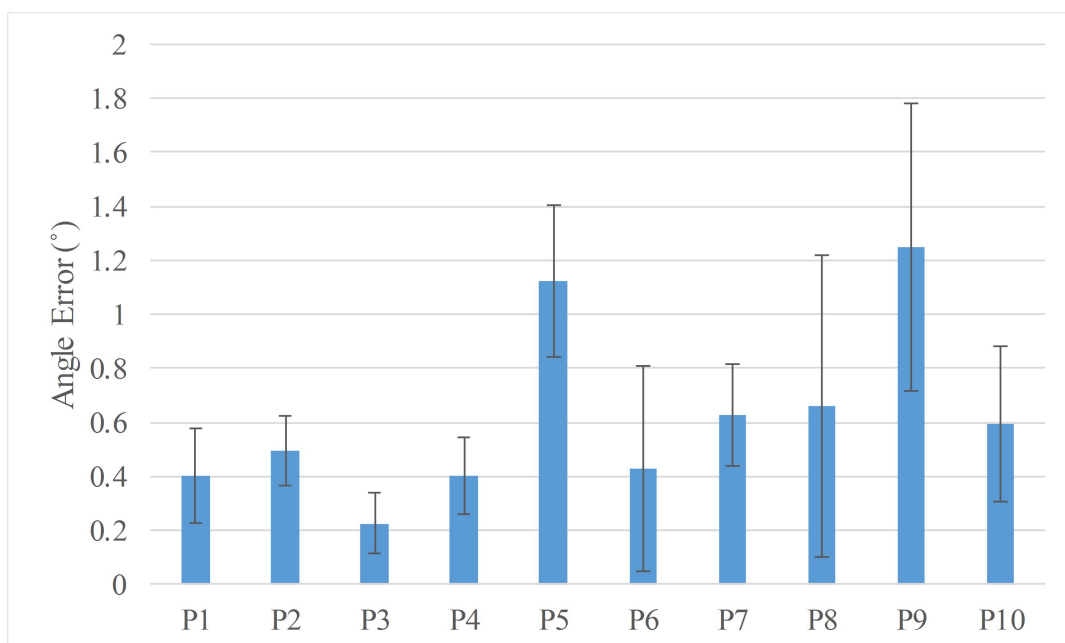
Figure 4.14 The accuracies of pointing with the opposite hand.

Table 4.1 The Likert scale results of the questionnaire

Questions	Scale
Q1: I think that I would like to use this system	4.0 ± 0.8
Q2: I found the system unnecessarily complex	4.7 ± 0.5
Q3: I thought the system was easy to use	4.2 ± 0.9
Q4: I don't think that I would need the support of a technical person to be able to use this system	4.1 ± 1.0
Q5: I think the system is consistent	4.3 ± 0.6
Q6: I would imagine that most people would learn to use this system very quickly	4.4 ± 0.5
Q7: I felt very confident using the system	4.3 ± 0.6
Q8: I didn't need to learn a lot of things before I could get going with this system	4.5 ± 0.5

4.1.5 Conclusion

This section presented the Pointing At Several Targets (PAST) method to calculate an origin for the *eye-rooted* pointing ray and recover the user specific pointing geometry in an egocentric setting without external cameras and trackers.

This work is different from the state-of-art techniques for hand gesture recognition and gesture based interaction. Its main difference lies in calculating the ray origin for the user's specific habit, recovering the pointing geometry accurately, and enabling direct interaction with real and virtual digital elements of interest within the user's environment without any visual feedback.

Mathematical analysis and experimental results demonstrate that the PAST calibration can estimate the origin point stably and recover the user specific pointing ray accurately. The calibration method is contextually independent and in most cases the pointing line is recovered with an accuracy of less than $1°$. The overall pointing accuracy at 1.5, 2, 3, and 4 meters is $0.67 \pm 0.71°$. The PAST solution would enable a whole series of applications with an egocentric setting in which the user can directly acquire or interact with different media and objects in real-time. It will allow users to naturally perform multi-control interaction with virtual or real elements placed within their environment.

In future work, one could improve the hand gestures command combined with the pointing gesture enabling multi-modal interaction. Moreover, this work offers the possibility of developing an intuitive method for selecting desired objects in a real scene thanks to the recovered pointing gestures.

4.2 Device- and System-independent Personal Touchless User Interface for Multi-Display

As the computing technologies develop, more and more information is digitized from the physical world, but the mainstream display technology is still 2D display. Researchers work on

how to enable the user to perceive and interact with the information naturally and effectively. As shown in Figure 1.1, in the modern day operating room the surgeon performs surgeries with the support of different medical systems that showcase patient information, physiological data, and medical images. Surgeons rely on these medical systems to capture, browse, and manipulate patient information, physiological data, and medical images, but the current control input of these systems are still mouse, keyboard, touch screen, and joystick. To retrieve the desired information, the user has to press either a physical/virtual button using their own finger or a visual cursor controlled by another device. Yet, one of the most important rules in the OR is to maintain a strict boundary between what is sterile and what is not [**OHara2014a**].

In this section, a novel user interface is developed that allows the surgeon to personally perform touchless interaction with the various medical systems, switch effortlessly among them, all of this without modifying the systems' software and hardware. To achieve this, a wearable RGB-D sensor is mounted on the surgeon's head for inside-out tracking of their finger with any of the medical systems' displays. Android devices with a special application are connected to the computers on which the medical systems are running, simulating a normal USB mouse and keyboard. When the surgeon performs interaction using pointing gestures, the desired cursor position in the targeted medical system display, and gestures, are transformed into general events and then sent to the corresponding Android device. Finally, the application running on the Android devices generates the corresponding mouse or keyboard events according to the targeted medical system. To simulate an operating room setting, our unique user interface was tested by 7 medical participants who performed several interactions with the visualization of CT, MRI, and Fluoroscopy images at varying distances from them. Results from the System Usability Scale and NASA-TLX workload index indicated a strong acceptance of our proposed user interface.

4.2.1 Design and methods

Based on the pointing gesture recovery in Section 4.1, we propose a solution to give surgeons direct control of the various medical systems' displays inside the OR using only a single user interface via personalized pointing gesture. The collaboration of the Android devices and software allow the combined system to act as a usual mouse or keyboard and facilitates working with any device that does not see a difference in its input/output.

Hardware setup

We introduce a wearable RGB-D sensor and a wearable computing device for each surgeon, and an Android device for each medical system. The proposed hardware setup with one surgeon and one medical system is shown in Figure 4.15. The wearable RGB-D sensor is connected to the wearable computing device to recover the pointing gesture, and the Android device is connected to the medical system simulating a normal USB mouse and keyboard. The wearable computing and Android devices exchange information via a WiFi router. To simplify the display tracking task, a 2D marker plane is attached to each display, which is connected to different medical system.
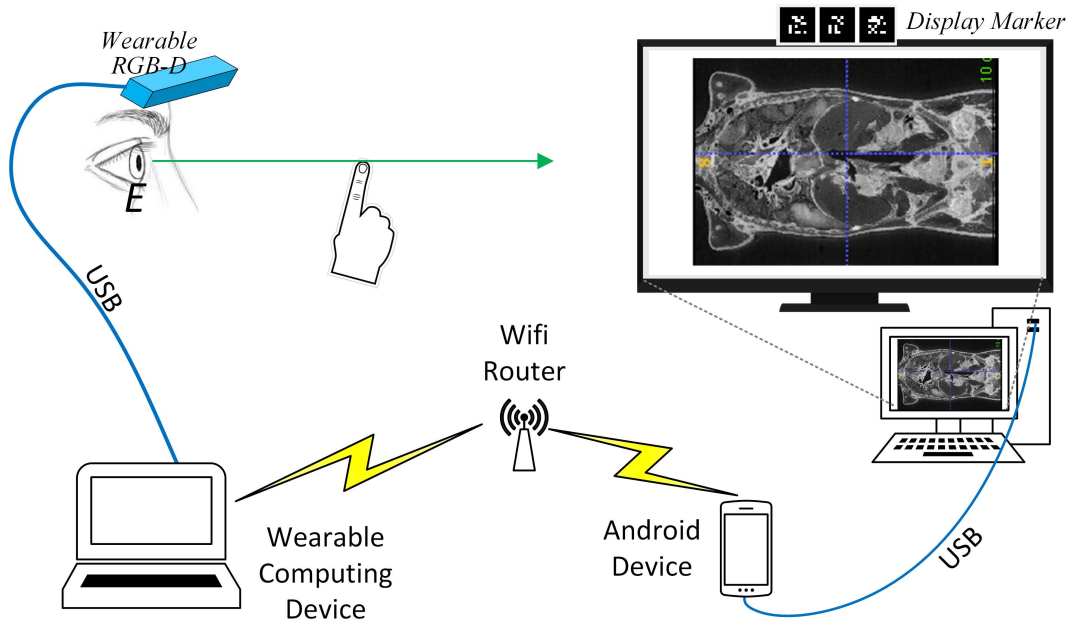
Hardware setup. The proposed system include a wearable RGB-D sensor, a wearable computing device, a medical system with one display, and an Android device.

Calibration

For the setup there are two parts that require calibration: i) the transformation between the display and its marker, and (ii) the user specific pointing gesture. The calibration methods are shown in Figure 4.16 and presented below.

Display with markers   Without modifying any software/hardware of the existing medical systems, this calibration is performed with a marker printed on an A4 paper, as shown in Figure 4.16. The A4 paper is affixed closely next to the top-left corner of the display. As the paper is very thin, the printed marker and the display are supposed to be coplanar. With the tracking information of the markers, the transformation matrix of the display relative to the display marker can be calculated. Since the size of the A4 paper and the display are known, the translation of the printed marker relative to the display screen can be calculated. This one-time calibration process is performed offline for each display in the operating room. Then a configuration file is created for each display marker to describe which display it is attached to and which medical system it is connected to this display.

User specific pointing gesture   In section 4.1.1, the objective is to calibrate and recover the user specific pointing geometry. The result of the calibration is the position of a virtual eye center E for the user. As shown in Figure 4.1, the target, T_i, is shown somewhere in front of the user, who is then asked to perform a natural pointing gesture towards T_i using one finger. The 3D positions of the fingertips F_i and T_i are collected through the wearable RGB-D sensor. The pointing line L_i is defined such that it passes through F_i and T_i. After sufficient pointing lines are collected, the intersection of the lines is calculated defining the user specific virtual eye-center E. After calibration the pointing geometry is recovered as L_i, going through E and F_i, when the pointing gesture is performed.
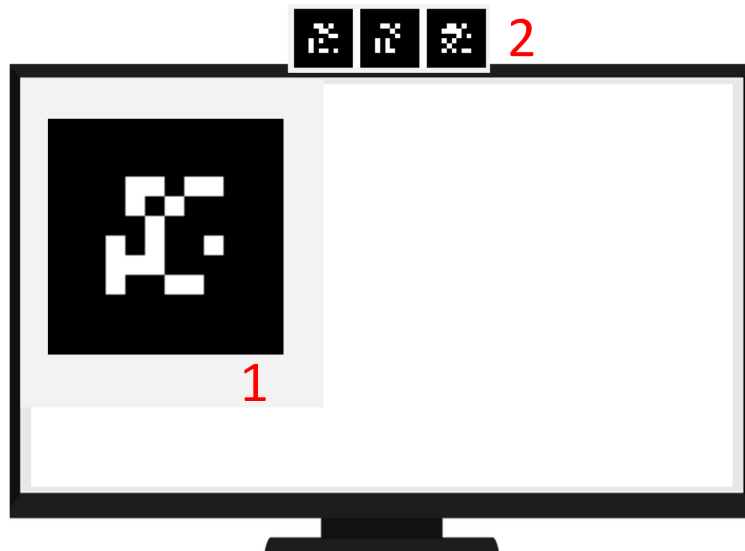
Figure 4.16 Calibration methods: (a) A printed marker (1) is placed next to the right-top corner of the display to calculate the transformation matrix between the display marker (2) and the display.

System framework

The system framework is shown in Figure 4.17. The left side shows the usual elements in the OR and the right side shows the data-flow of our user interface paradigm. When a surgeon performs interaction with a medical system, the display marker and fingertip can always be detected via the RGB-D sensor as it has a similar view as the surgeon. Through the following processing: *Tracking Info*, *Pointing Geometry*, and *Interaction Events*, mouse and keyboard events are generated to control the medical systems.

Tracking Info   All the images are perceived via the wearable RGB-D sensor. Therefore, this is an inside-out tracking and does not require an external reference. The fingertip is detected from the depth image using 3D blob extraction and the display marker position is calculated via the color image. The display's position can be calculated based on the calibration result in Section 4.2.1.

Pointing Geometry   Based on the calibration of "User specific pointing gesture" and the tracking information, the pointing geometry between the pointing direction and the display screen plane is recovered. The target position in real world and the desired cursor position on the screen can be calculated. We can also detect some predefined finger and hand gestures. Finally, all the information is transformed into general events and then sent to the corresponding Android device through UDP (User Datagram Protocol) connections.

Interaction Events   Two virtual devices are created in a Linux kernel to simulate USB mouse and keyboard when the Android device is connected to a computer. The two devices send normal initial signals to the host computer and are recognized as USB mouse and keyboard. The special application running on the Android system builds up a UDP connection with the computing devices to receive general events transformed from the wearable computing device and generates the corresponding mouse or keyboard events according to the current medical system. In our general design, all the interactions are implemented via the Android mouse or
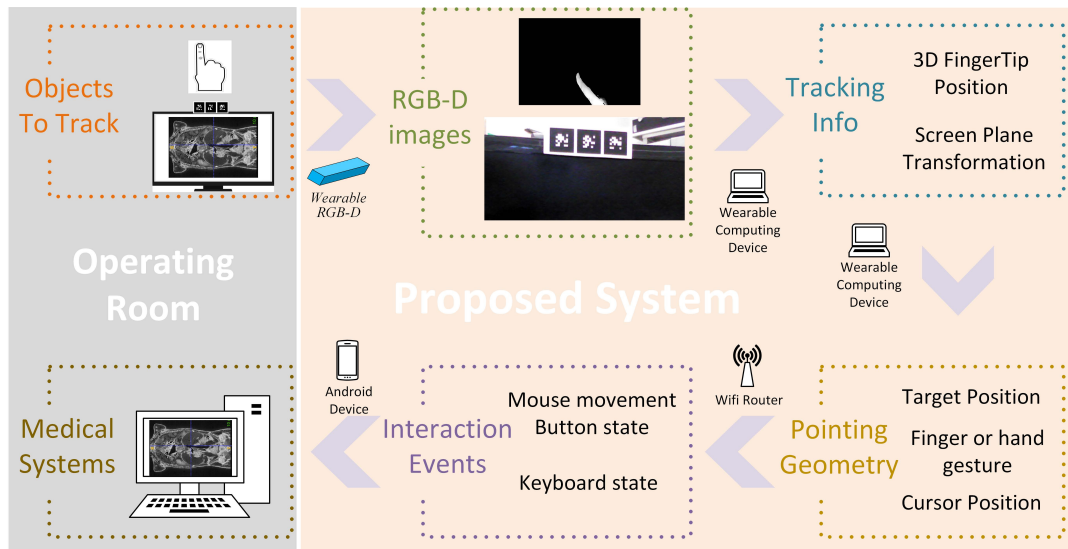
Figure 4.17 The framework of the proposed system. Tracking information of objects in the OR after processing is used to control the medical systems.

keyboard as the medical system is taken as a black box. However, the wearable computing device could directly communicate with the medical system when this is allowed.

Multi-user Multi-system Scenario There can be more than one surgeon and more than one medical system in the operating room scenario, and the proposed setup is shown in Figure 4.18. In the Pointing Geometry stage, a personal configuration file should be loaded to implement the multi-user case and personalized figure-pointing gestures. In the configuration file, each member of the surgical team is assigned a priority level according to their responsibility. When Android devices receive information from more than one wearable computing device at the same time, they implement the mouse and keyboard event according to the surgeon with higher priority level. This can solve the issues which arise when these systems are used in the context of an actual surgical procedure, including collaboration, engagement and disengagement. When there are more than one displays connected to a medical system, it still works well as the Android device knows which display is the desired one. Even when there is no display for a medical system, a display marker still can be positioned somewhere to simulate a virtual display, which is logically connected to the medical system. In this case, the surgeon can point at the display marker and perform interaction with the corresponding system.

Personalized Pointing Gesture Performing interactions requires two steps: i) choosing the target, and (ii) releasing the command. Through pointing gestures the surgeon can choose the target naturally via the pointing and immediately perform a finger or hand gesture to release the command. During the calibration procedure, the calibration of user pointing provides the characteristics of a user specific pointing geometry. In addition, since the pointing gesture is personal, pointing characteristics could be saved for each surgeon in order to recognize their gesture and map it to different commands.
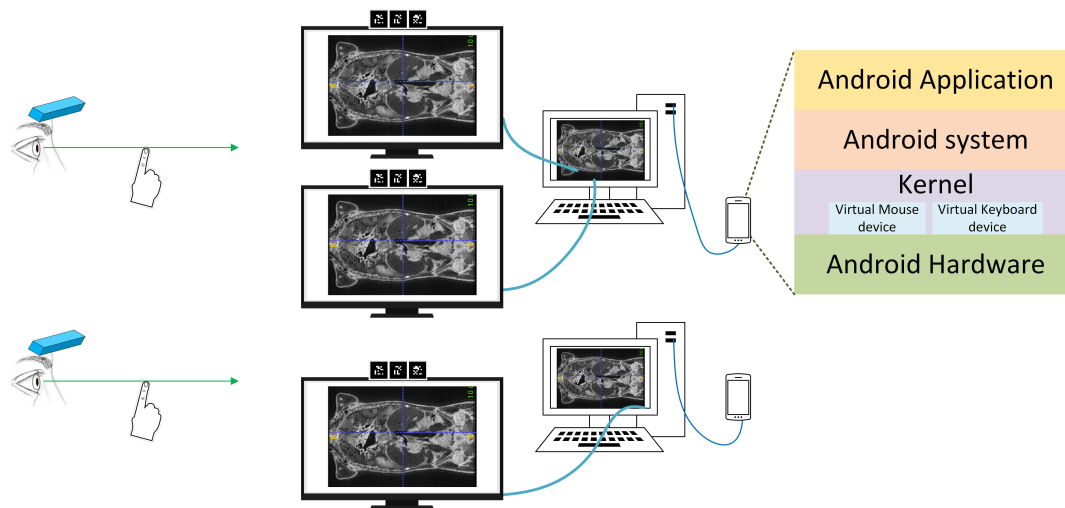
Figure 4.18 A setup with Multi-user Multi-system and the Android system structure. In this setup, there are two surgeons and two medical systems, one of which has two displays.

4.2.2 Evaluation & results

The main goal of the user study is to answer the following questions:

1. To what degree does the surgeon accept the interaction with multiple medical systems via the proposed user interface?

2. Does the surgeon find the system a viable alternative to the traditional indirect control of the medical systems and their displays in the modern day operation room?

Here we note that we are not evaluating the gesture accuracy, but we are assessing the Android-based communication with the different systems.

Implementation

Proposed system   During the user study, the Intel Realsense Creative camera was chosen as the wearable camera and Microsoft Surface 2 as the computing device, which was put in a backpack. The special Android application was developed and installed on two Nexus 7 devices to simulate the USB mouse and keyboard for each medical system.

Patient scenario   Anonymous patient is wheeled inside the emergency room following a car accident. A CT was taken after neck trauma to rule out fractures. As no fractures were found but instability due to disc ligament tear was suspected, an MRI was carried out that confirmed it. Anonymous patient then went to surgery where C-arm fluoroscopy images were taken intra-operatively to assist the surgeon during intervention.

Simulated medical systems   The operating room was simulated in our laboratory with three displays showing CT, MRI, and C-arm fluoroscopy images. *MicroDicom*[150] was used to visualize the C-arm fluoroscopy image in **display A**, *InVesalius*[107] was used to visualize the CT volume in **display B**, and MRI images were loaded directly in **display C**. The sizes of A, B, and C are 24 ″, 42″ and 22″ and were placed at different locations and distances from the
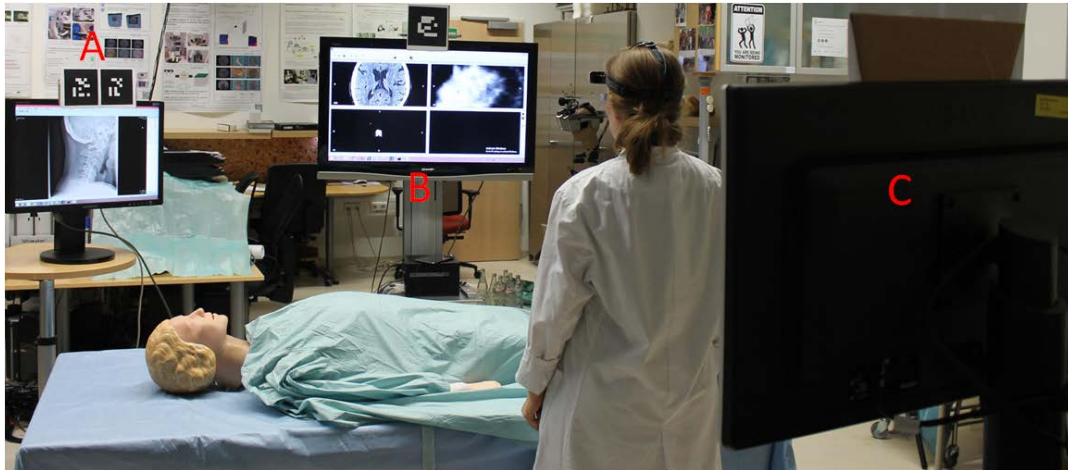
Figure 4.19 Simulated OR. There are three displays in the simulated scenario, where the C-arm fluoroscopy , CT, and MRI images are shown in display A, B, and C, respectively.

surgeon, as shown in Figure 4.19. The interactions to be performed with these medical images were defined similar as those performed inside the modern day operating rooms:

Fluoroscopy image: up-down flip , left-right flip

CT & MRI images: go through the images in axial, sagittal, coronal
zoom in/out in axial, sagittal, coronal

Mouse and keyboard events   Going through the images in axial, sagittal, and coronal in *InVesalius*[107] is done by clicking at the image window and rotating the mouse wheel. Zooming is triggered by pressing the right mouse button down and moving the mouse up or down. To flip images using *MicroDicom* [19], it is necessary to trigger it by pressing buttons 'Ctrl' with 'F' or 'H' at the same time. There is a medical system configuration file for each Android application, and it declares how to generate the mouse and keyboard events to perform desired function.

Proposed interaction- The intersection between the pointing ray and the screen plane is calculated as the cursor position, and a mouse click function is generated when the surgeon points to a new image window via his or her finger. Mouse wheel events are generated to go through the CT & MRI data when the surgeon is pointing towards the *InVesalius* image window and moving their finger left or right in the middle of the window. For zooming, the 'right mouse button pressed down' event is triggered when moving their finger up or down. The button event 'Ctrl+H' is generated for left-right flipping of the C-arm fluoroscopy image when the surgeon is pointing towards the *MicroDicom* and moving their finger left or right. Similarly, the 'Ctrl+F' button event is triggered to flip up-down the C-arm fluoroscopy images when surgeons move their finger up or down on the right side of the window.

User study design

In order to answer the above questions, a qualitative user study was conducted with 7 participants (3 expert surgeons and 4 final year medical students). First, we introduced the concepts, user interface system, types of interactions, and experimental setting to the
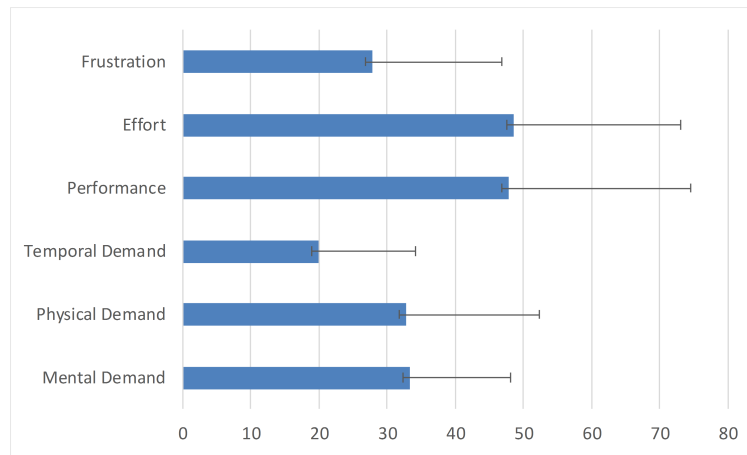
Figure 4.20 The results of NASA-TLX cognitive workload.

participants. Second, each participant put on the wearable RGB-D sensor and underwent a test period lasting 3 minutes to get comfortable with the initial calibration step and subsequent interactions. The distance of the displays to the participants are: **display A**-1.5 meters, **display B**-3 meters, and **display C**-1 meter.

The interaction sequence with the three different displays and data is as follows. Each participant was asked to visualize the CT images for neck fracture(s), browsing through the slices and magnifying specific regions within the slice in **display B**. Then, the participants were asked to either rotate their head or body to face **display C** to confirm the site of the ligament tear in the MRI images by browsing through the images and magnifying them. Finally, the participants looked at the C-arm fluoroscopy **display A** to flip the X-ray images correctly (Please check the demo video here).

Metrics

The following metrics were measured during the course of the experiments. First, we measured the total time in seconds for the participants to interact with the three displays. Then, following the user study, we had a short interview with each of the participants and they were asked to complete an adapted System Usability Scale (SUS) [19] questionnaire using a Five-point Likert scale (*1-Strongly disagree, 2-Disagree, 3-Neither agree nor disagree, 4-Agree, 5-Strongly agree*). Lastly, the participants completed the NASA-TLX [63], which is a multi-dimensional scale ranked on 100 designed to obtain cognitive workload estimates from one or more operators while they are performing a task or immediately afterwards. The total workload is divided into six subscales that are represented as: *Mental Demand, Physical Demand, Temporal Demand, Performance Effort, and Frustration*. The higher value from NASA-TLX means a higher cognitive workload.

Results

Although we did not have the same number of surgeons as students the measured average time to finish all the basic interactions with the three displays was 114 ± 13 seconds for the medical students and 115 ± 15 seconds for the experts.

---

https://youtu.be/FLKBZOG26XY

Table 4.2 The Likert scale results of the usability questionnaire.

Questions	Scale
Q1: The pointing gesture is intuitive for interaction.	4.1 ± 0.7
Q2: It is very smooth to switch between different medical system.	4.3 ± 0.8
Q3: It is not complicated to get used to the system.	4.1 ± 0.7
Q4: The system feels responsive.	3.4 ± 0.7
Q5: I think the system is consistent.	4.0 ± 0.6
Q6: I prefer the proposed system instead of indirect interaction.	4.4 ± 0.8
Q7: The system is easy to use.	3.9 ± 0.9
Q8: I think that I would like to use this system.	4.3 ± 0.6
Q9: I would imagine that most people would learn to use this system very quickly.	4.3 ± 1.0
Q10: I felt very confident using the system.	3.9 ± 0.5
Q11: I didn't need to learn a lot of things before I could get going with this system	4.7 ± 0.4

The results for the adapted SUS questionnaire are shown in Table 4.2. All responses were positive. The highest scores were credited to questions #6 and #11 indicating that the participants would prefer our user interface over the existing ones in the operating rooms (4.4 ± 0.8), and that such user interface did not require a significant amount of learning (4.7 ± 0.4). The lowest scores were attributed to questions #4 and #10 which indicate that participants did not feel confident (3.9 ± 0.5) which did not allow for a subsequent smooth interaction. Finally, the result of NASA-TLX cognitive workload is shown in Figure 4.20. The highest workload occurred for *Effort* and *Performance* with average scores of 48% and 49% respectively. The participants did not have high cognitive workload for the remaining subscales with values ranging between 20% and 33%.

4.2.3 Conclusion

Via a user study, we proposed and evaluated a new UI paradigm that gives the surgeon direct control of the interaction of different medical systems. There is no need to modify the software/hardware of the existing medical systems, and our solution can integrate pointing with personal gesture commands, while enabling switching smoothly among the different medical systems and operators. Our system can also be used for multi-user multi-system platforms. In future work, we plan to design a full user study and compare the proposed method to the traditional interfaces.

Here we made a mistake while creating the questions in Table 4.2, the tense of which were not consistency. Hence, we want to declare that all the questions were asked after the users used the system for a while, and the users were clear that they were asked to answer the questions based on their experience.

Framework for Collaborative Mixed Reality

<div style="text-align: right">5</div>

The magic mirror is introduced to create a MR environment, improving the user's perception during medical education and rehabilitation exercise in Chapter 3. During these procedures, there are some scenarios where more than one user are involved at the same time, such as a teacher with a student, a doctor with a patient, a nurse with a rehabilitation exerciser, and so on. There are communication issues when multiple users get visual feedback from the AR view and when performing actions in the same MR environment. From the real world to a MR environment, there are functional seams and cognitive seams for the user during the collaboration. The objectives of this chapter is to propose a framework for collaborative mixed reality and try to remove the functional and cognitive seams in the mentioned scenarios.

The pointing gesture in egocentric setting discussed in Chapter 4 is a natural interaction methodology and flexible for multi-user scenarios. It allows the user to perform interaction with ambient media and objects in the MR environment. Hence, a collaborative MR framework is introduced by integrating the MR magic mirror and the pointing gesture in egocentric setting. All the functions in the magic mirror framework are kept to guarantee the user's personalized perception of the medical information. The pointing gesture is introduced to allow a teacher or a doctor to communicate with the student or patient via the AR view. The normal communication is kept and new communication patterns are created in the MR environment. The framework is discussed in two scenarios, anatomy learning with teacher and rehabilitation exercise with nurse. Until now, we just have implemented a simple prototype of the proposed framework and a master student is working on this project, trying to implement the anatomy learning scenario and do a serious evaluation.

5.1 Framework design

Based on the magic mirror framework and the pointing gesture in Chapter 3 and Chaper 4, a new framework is introduced. We first discuss the system components of this framework and then the specific interactions with the shared AR view are presented.

5.1.1 System components

As shown in Figure 5.1, there is a shared AR view, two users, perception and interaction between the user and AR view, and communication between the users in the mixed reality environment. In the following we discuss the function of each component and how to create them.

Mixed Reality Environment

AR view

Perception
Interaction

Perception
Interaction
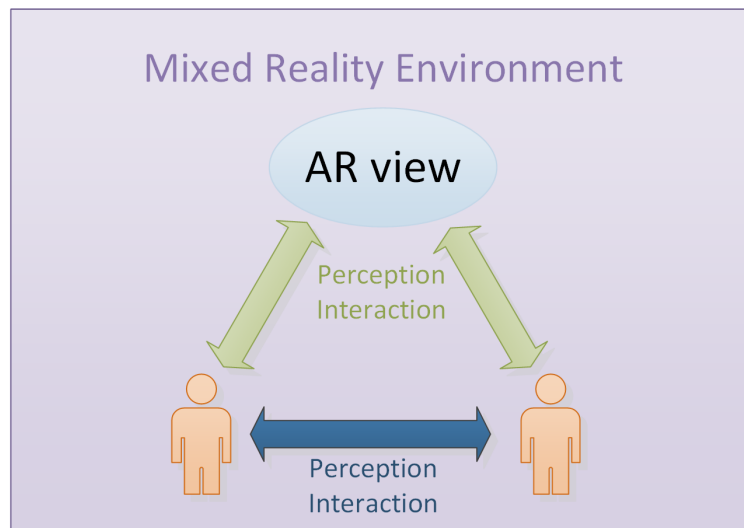
Figure 5.1   System components. There is a shared AR view, two users, perception and interaction between the user and AR view, and communication between the users in the mixed reality environment.

Shared AR View   The AR view is the most important part of the MR environment, combining the real world with the virtual world. The shared AR view is created via the magic mirror framework, and all the users can perceive the real world and virtual elements from the magic mirror effect. At the same time, the AR view also behaves as a bridge between two users.

Perception and Interaction   In the AR view, the virtual elements is still augmented onto the user's body based on the personal information to improve the knowledge understanding. The personalized perception from magic mirror is kept and the tracked user still can directly perform interaction with the AR view via the natural body gestures, as described in Chapter 3. For the user who is not tracked, mouse and keyboard is not an ideal solution to perform interaction with of the application naturally and the communication between the users cannot be extended smoothly. The method in section4.1 recovers the pointing gesture of the user with wearable device, enabling directly interaction with all the objects and medias in the MR environment. Hence, we introduce the pointing gesture in an egocentric setting to this framework as another interaction input.

Communication   In the MR environment, all the normal communication methodologies between the users should be kept, such as talking, body language, direct physical interaction and so on. In addition, there are new communication patterns created based on the natural interaction with the AR view.

5.1.2 Natural interactions with the AR view

As the users share the same AR view in a MR environment, the communication between the users also can be performed via the AR view. One user can directly interact with the AR view while communicating with the other in the real world. The communication in the real world can be enhanced by the interaction with the AR view. The interactions with the AR view can be classified into four categories.

1. *Interaction with magic mirror* The tracked user can directly interact with the magic mirror and generate a desired AR view for their communication with the other user in the real world.

2. *Interaction with static objects* GUI (graphic user interface) elements, such as button, menu, and so on, can be clicked via the pointing gesture. This kind of interaction can be used to adjust the parameters of the AR view.

3. *Interaction with virtual elements* The user with the wearable device also can directly control the virtual elements and modify the AR view via our recovered pointing gesture. E.g. a medical model can be re-scaled and moved to another place if needed.

4. *Interaction with human body* The user with the wearable device can naturally point to the other user in the mirror view. The system can calculate what is pointed at in the real world according to the mapping between color and depth image. Then, a specific AR view is generated in the desired location. For example, a teacher can point to the body of a student in the mirror view to tell them where is the heart by generating a AR view of the heart.

In this MR environment we try to remove the functional and cognitive seams as much as possible. The AR view is presented in a magic mirror concept to remove the cognitive seams as the mirror effect is very natural. The normal communication between the users without AR view is kept, and the interaction with the AR view is implemented via the personalized body or pointing gesture to remove functional seams.

5.2 Implementation

In this section, the general hardware is presented and then two example scenarios, medical teaching and rehabilitation exercise, are simply examined.

5.2.1 General hardware

As the framework integrates the magic mirror and the pointing gesture in an egocentric setting, the general hardware contains the hardware from both systems and is presented in Figure 5.2. There is one Kinect, one wearable RGB-D sensor, one display, a wearable computing device and computing device. Figure 5.2-right shows a display, a Microsoft Kinect and a computing device are used to create the magic mirror MR environment. Each user with one wearable RGB-D device and one wearable computer to perform the pointing gesture is shown in Figure 5.2-left. The pointing gestures are detected and recovered and commands are transferred to the magic mirror system via a Wifi router. The framework is also feasible in a scenario with more than one pointing systems and more than one magic mirror systems.
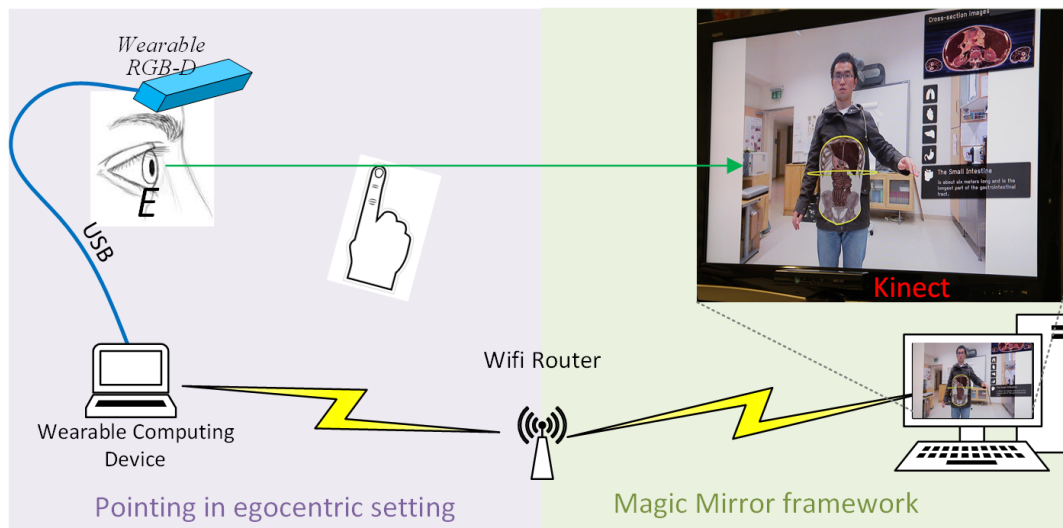
Figure 5.2 General hardware. Left side is the pointing gesture recovery in an egocentric setting and the right side is the magic mirror system. A pointing gesture is detected and understood by the wearable device and then transferred to the magic mirror system to control the AR view.

5.2.2 Example scenarios

In normal anatomy teaching or rehabilitation scenario, there is a teacher or doctor to teach and monitor the student/patient for the learning/exercise procedure. The implementation of these two scenarios are examined below.

In this framework, the teacher directly points to the MR view and controls it to initiate the learning procedure, and the nurse in charge of the rehabilitation exercise in the same way. In both scenarios, a teacher or a nurse can work with several students or patients as they can switch from one MR environment to another seamlessly. Similarly, several teachers or nurses also can be involved in one MR environment if needed.

Medical teaching   The demo scenario of medical teaching is presented in Figure 5.3. A teacher can teach a student about the anatomy knowledge in a collaborative MR environment. Based on the proposed framework, the following functions can be implemented.

1. The teacher can choose the best medical dataset for the current student and knowledge topic by clicking the predefined button or menu.

2. The student can generate a AR view of a special organ according to the teacher's question.

3. A specific anatomy model can be manually augmented to a special location the teacher prefers.

4. The teacher can control the AR view according to their teaching. The teacher can directly say: "Here is your stomach." and point to the student's stomach in the mirror view, and simultaneously the magic mirror system generates a AR view of the stomach.

(a) One teacher with one student.

(b) One teacher with several students.
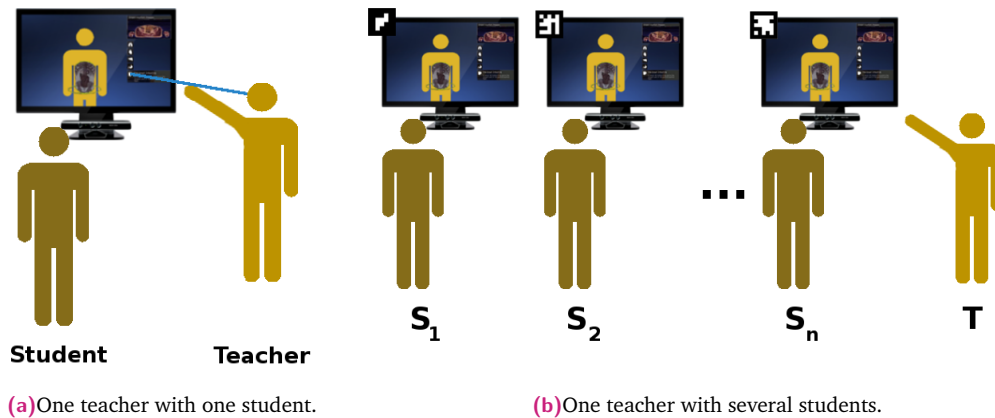
Figure 5.3 The demo scenario of medical teaching. The teacher interacts with the learning procedure using pointing gestures.

5. One teacher can take charge of several magic mirror systems and switch between them smoothly.

Rehabilitation exercise The tracking accuracy for patients might not be accurate as there is an increased possibility of not keeping a steady pose in front of the sensor due to their disability. As the medical teaching scenario, the nurse can take advantage of the framework for the rehabilitation exercise, such as adjusting the system parameters, creating specific AR view to perform the patient education, taking care of several patients and switching between them smoothly. The nurse can use one hand to touch or hold the patient's body and create a specific AR view for the current rehabilitator.

5.3 Conclusion

A framework which combines the personalized MR magic mirror and pointing gesture interaction in an egocentric setting is introduced. The communication in the real world can be enhanced by the interaction with the AR view. We investigated the framework with two scenarios, medical education and rehabilitation exercise.

Conclusion

<div style="text-align: right; font-size: 3em;">6</div>

6.1 Summary

In this thesis, a magic mirror framework is first presented to create a MR environment with an *in-situ* augmented reality view. The system can track the current user and finish the registration between the medical datasets and the user based on the personal information. It inherently allows the user to perform interactive registration and directly map the medical information onto their body. This opens interesting possibilities for personalized mixed reality for medical education and rehabilitation. Through the participation of 7 clinicians and 72 anatomy students, two user studies were designed to respectively assess the precision and acceptability of the magic mirror system for anatomy education in the classrooms of tomorrow. We also take the advantage of the interactive mixed reality to generate a personalized learning procedure. In addition, "Organ Explosion" and "Self-control virtual view" systems are designed and implemented for anatomy leaning. Serious games are also developed in the framework for health-care education and rehabilitation exercise.

Secondly, a method is proposed to calculate a virtual eye center as the origin for the *eye-rooted* pointing ray without tracking the eye gaze in an egocentric setting. The user specific pointing geometry can then be recovered for interaction with ambient media and objects without visual feedback. The pointing ray starts from the virtual eye center and goes through the fingertip to the desired target information. During the initial calibration, the user is asked to perform pointing gestures towards several targets displayed within the field of view. A user specific virtual eye center is then estimated based on the detected fingertip in the coordinate system of the wearable device. Then the *eye-rooted* pointing ray is recovered when the user is performing a pointing gesture. The proposed method is mathematically analyzed and demonstrated that the calibration is feasible with a measured accuracy of the recovered pointing ray below $0.9°$. The proposed method is also evaluated in a user study consisting of 10 participants performing pointing interaction with objects shown on the display while wearing a RGB-D sensor. The study shows that the calibration is independent of the environment and the pointing ray is recovered with an accuracy of $0.67\pm0.71°$ without any visual feedback. Based on the recovered pointing gesture, a novel user interface is introduced, allowing surgeons to personally perform touchless interaction with the various medical systems, and switch effortlessly among them, all of this without modifying the systems' software and hardware.

Finally, in conjunction with the proposed magic mirror framework and recovered pointing gesture, a multi-user collaborative MR framework is introduced. The human-human communication is enhanced via pointing gesture and MR framework.

6.2 Discussion & Outlook

6.2.1 Personalized AR

The magic mirror system does not require expensive hardware and can be set up with different display devices. Therefore, we can create a demo setup consisting only of a Kinect, a display device and a laptop. The system has been shown to the public at several occasions. It was shown during the open day of a hospital, in a school and during conferences. The feedback of the users, in particular from children has been very positive. While the MR *in-situ* visualization attracts the attention of people, it turned out that most users would spend much more time to understand where different organs are and what they look like. During the demos, the system has been used by many people and it has shown to be robust. The calibration sometimes fails, in particular when people are wearing thicker coats or jackets. But if the position and angle of the Kinect is set well, calibration problems are reduced largely.

It was originally suggested to us that our system would have much more of an impact for medical education if it were to be translated today in medical schools and anatomy classrooms. As such, with the help of our anatomy professor partners, we delivered the improved AR magic mirror to two anatomy classes within the Anatomy Department of the Ludwig Maximilian University (LMU) Medical School, Munich, Germany. The anatomy professors made it clear that the proposed system is exciting, but has to go beyond state-of-the-art technology to be truly useful for education. The user study participants stressed the importance of visualization of anatomy that can change dynamically resulting from the actions of the user moving the body. Furthermore, more advanced user interactions like the use of gaming elements would be required to make the use of the system for learning and rehabilitation tasks more interesting.

6.2.2 Personalized user interface

Pointing gesture recovery in an egocentric setting

In Section 4.1, we investigated whether our proposed method could recover a stable and precise pointing origin and corresponding pointing rays accurately, using the collected data via the wearable device. The interpupillary distances were calculated very accurately as the participants had a clear view of the finger when performing pointing with only one eye open (see Figure 4.8). From Figure 4.9, we find the proposed method is contextually independent and can be used to calculate the eye position in the wearable device coordinate system. As seen in Figure 4.12, the virtual eye center calculated by the PAST calibration is user specific. When the eye positions are known, this method can be used to determine the position of each eye, or the average of both eyes, as the best origin for the pointing ray.

As demonstrated by our user study, the calibration of the virtual eye center is completed stably by the participants. The pointing ray is recovered in different scenarios with an accuracy of less than $1°$ in most cases. The participants easily learned how to use this new interaction method.

Eye gaze tracking is an important technique in an egocentric setting, which allows users to navigate and explore holograms by looking at them. In this case, the eye gaze geometry

is employed to calculate where the user is looking at and the pointing gesture is used as a confirmation command. Compared to pointing techniques, there is only one available controlling point since the eyes can only focus on object of interest. In addition, it is very difficult to perform gestures using the eye-gaze, such as drawing a circle or line, and the visual feedback is necessary to combine the hand/finger movement for interaction, e.g. selecting a special area.

The PAST method can also be employed in other interesting scenarios such as: (a) **a wearable RGB-D sensor and an optical see through HMD**. A 3D virtual world can be presented using a HMD and contextual information of the virtual world in the HMD is known after calibration with the RGB-D sensor. The targets can be shown in the HMD during calibration and then the user can start interacting within a mixed reality world using pointing gestures. (b) **a wearable RGB-D sensor in a known environment**. The real world environment can be tracked through the color camera, thus contextual information of the entire real world is known. After calibration, the user can start interacting with elements in the real world using a pointing gesture without any visual feedback. (c) **a near range and far range depth sensors**. The user and the computer can directly work together to collect information from the real world, such as 3D scene understanding.

A special RGB-D camera, in which the depth sensor looks at a different direction compared to the color camera, should be developed to acquire the depth image of the full hand to improve the fingertip tracking. Positioning S_f and S_c at different directions also enables users to perform a pointing gesture more naturally. The PAST calibration can be separately performed for each finger to improve the accuracy, allowing the user to use multi-fingers to perform a multi-control interaction in a mixed reality setting. Given the recent increased interest in using speech to interact with mobile devices, it is also a logical next step to support a user pointing at an object while stating a command or asking a question about it.

Device and System Independent Personal Touchless User Interface for Multi-Display

Like most new technologies, the learning curve for the participants needs to be considered when interpreting the results of our user study. As each of the 7 medical participants did not have any prior knowledge with our user interface, it was not surprising to note that there was no significant difference in the time required to complete the interactions. It was also expected that participants did not feel confident using the user interface for the first time as seen from the adapted SUS results. This effect translated to the cognitive workload NASA-TLX with the highest scores to the following questions:*(Performance) How successful were you in accomplishing what you were asked to do?* and *(Effort) How hard did you have to work to accomplish your level of performance?*

Albeit these shortcomings, the vast majority of the results were positive. With consistent use of our user interface, participants agreed that the potential impact of the proposed solution compared to the traditional interfaces found in the operating room is clear. Our unique framework allows the surgeon to personally perform touchless interaction with the various medical systems at different locations and distances, switch effortlessly among them, all this without modifying the systems' software and hardware. Comparing with the traditional methods our proposed framework is very general and can be adapted to multi-users and

multi-displays. The objective of this section was to present the main user interface paradigm, and not evaluate the accuracy of the different gestures for displays A-B-C.

In 2014, the authors in [124] provided insights and guidelines when developing user interfaces for the operating room. A key consideration is the physical location of surgeons when they need to interact with different medical systems. The patient table can be a crowded environment, with surgeons often in close proximity to other members of the surgical team. Not only can this affect a system's approach to tracking but also impose constraints on the kinds of movement available for gesture design. In sterile practice, hand movements must be restricted to the area extending forward from the surgeon's torso between the hips and chest level. Moreover, the patient table itself hides the lower half of the surgeon's body. Also, sterility restrictions in the operating theater make touchless interaction an interesting solution to give the surgeon direct control over the medical systems. Our proposed user interface addresses these concerns.

Currently, finger tracking is not perfect and the marker tracking is challenged since the wearable RGB-D sensor works at a rate of 30 frames per second. In the future, with the advancement of algorithms and hardware these current limitations should be rectified. Lastly, with the volumetric acquisition of scans, the data is increasingly visualized as a 3D volume. Hence, the volumes would be better exploited using a full 3D interaction technique. The tracking of hands and gestures in 3D space opens up a much richer set of possibilities on how surgeons manipulate and interact with images through the full six degrees of freedom [124].

6.3 Framework for Collaborative Mixed Reality

Collaboration between people in MR environment becomes more and more common. In normal anatomy teaching or rehabilitation scenario, there is a teacher or doctor to teach and monitor the student/patient for the learning and exercise. When we try to introduce the new technique for these scenarios, one question is to make sure that it supports the collaborative MR. From the real world to a MR environment, there are functional seams and cognitive seams for the user during the collaboration. In the proposed framework, the normal interaction methods are kept in the MR environment and the human-human communication is enhanced via interaction with AR view. We proposed such a framework for only two scenarios. The next step is to develop some applications in medical teaching and rehabilitation exercise and perform user studies to evaluate the framework.

Bibliography

[1] L Ross A Gilroy, B Macpherson. "Atlas of Anatomy". In: *Atlas Anat.* 2012, p. 11 (cit. on p. 10).

[2] A Ammann, T Kiss, A Klebba, and H K Matthies. "The next generation of patient education: multilingual Dental Explorer 3D." In: *Int. J. Comput. Dent.* 13.1 (2010), pp. 43–55 (cit. on p. 13).

[3] Leonard a. Annetta, James Minogue, Shawn Y. Holmes, and Meng-Tzu Cheng. "Investigating the impact of video games on high school students' engagement and learning about genetics". In: *Comput. Educ.* 53.1 (2009), pp. 74–85 (cit. on p. 25).

[4] David Anton, Alfredo Goni, Arantza Illarramendi, Juan Jose Torres-Unda, and Jesus Seco. "KiReS: A Kinect-based telerehabilitation system". In: *2013 IEEE 15th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2013.* 2013, pp. 444–448 (cit. on p. 15).

[5] F. Argelaguet, C. Andujar, and R. Trueba. "Overcoming eye-hand visibility mismatch in 3D pointing selection". In: *Proc. 2008 ACM Symp. Virtual Real. Softw. Technol.* (2008), pp. 43–46 (cit. on pp. 28, 30, 67).

[6] Ronald Azuma and Ronald Azuma. "A survey of augmented reality". In: *Presence Teleoperators Virtual Environ.* 6.4 (1997), pp. 355–385 (cit. on p. 18).

[7] Jorge Bacca, Silvia Baldiris, Ramon Fabregat, and Sabine Graf. "Augmented Reality Trends in Education: A Systematic Review of Research and Applications". In: *Educ. Technol. Soc.* 17.4 (2014), pp. 133–149 (cit. on pp. 12, 22).

[8] K. M. Baird and W. Barfield. "Evaluating the effectiveness of augmented reality displays for a manual assembly task". In: *Virtual Real.* 4.4 (1999), pp. 250–259 (cit. on p. 24).

[9] Amartya Banerjee, Jesse Burstyn, Audrey Girouard, and Roel Vertegaal. "MultiPoint: Comparing laser and manual pointing as remote input in large display interactions". In: *Int. J. Hum. Comput. Stud.* 70.10 (2012), pp. 690–702 (cit. on pp. 28, 30).

[10] Cagatay Basdogan, Mert Sedef, Matthias Harders, and Stefan Wesarg. "VR-based simulators for training in minimally invasive surgery". In: *IEEE Comput. Graph. Appl.* 27.2 (2007), pp. 54–66 (cit. on p. 12).

[11] Eva Beranova and Catherine Sykes. *A systematic review of computer-based softwares for educating patients with coronary heart disease.* 2007 (cit. on pp. 13, 14).

[12] Alejandro Betancourt, Pietro Morerio, Lucio Marcenaro, et al. "Towards a unified framework for hand-based methods in First Person Vision". English. In: *2015 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2015.* IEEE, 2015, pp. 1–6 (cit. on p. 74).

[13] Christoph Bichlmeier, Felix Wimmer, Sandro Michael Heining, and Nassir Navab. "Contextual anatomic mimesis: Hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality". In: *2007 6th IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR* (2007), pp. 129–138 (cit. on p. 36).

[14] Mark Billinghurst and Hirokazu Kato. "Collaborative Mixed Reality". In: *Proc. ISMR '99.* 1999, pp. 261–284 (cit. on p. 4).

[15] Tobias Blum, Valerie Kleeberger, Christoph Bichlmeier, and Nassir Navab. "Mirracle: An augmented reality magic mirror system for anatomy education". In: *Proc. - IEEE Virtual Real.* (2012), pp. 115–116 (cit. on p. 22).

[16] Tobias Blum, Valerie Kleeberger, Christoph Bichlmeier, and Nassir Navab. "Mirracle: Augmented reality in-situ visualization of human anatomy using a magic mirror". In: *Proc. - IEEE Virtual Real.* (2012), pp. 169–170 (cit. on pp. 1, 31, 32, 37).

[17] Tobias Blum, Ralf Stauder, Ekkehard Euler, and Nassir Navab. "Superman-like X-ray vision: Towards brain-computer interfaces for medical augmented reality". In: *ISMAR 2012 - 11th IEEE Int. Symp. Mix. Augment. Real. 2012, Sci. Technol. Pap.* 2012, pp. 271–272 (cit. on p. 23).

[18] N. Alberto Borghese, Renato Mainetti, Michele Pirovano, and Pier Luca Lanzi. "An intelligent game engine for the at-home rehabilitation of stroke patients". In: *SeGAH 2013 - IEEE 2nd Int. Conf. Serious Games Appl. Heal. B. Proc.* IEEE, 2013, pp. 1–8 (cit. on p. 15).

[19] John Brooke. "SUS - A quick and dirty usability scale". In: *Usability Eval. Ind.* 189.194 (1996), pp. 4–7 (cit. on p. 90).

[20] Jeremy I M Carpendale and Ailidh B. Carpendale. "The development of pointing: From personal directedness to interpersonal direction". In: *Hum. Dev.* 53.3 (2010), pp. 110–126 (cit. on pp. 3, 27, 67).

[21] Kuo En Chang, Chia Tzu Chang, Huei Tse Hou, et al. "Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum". In: *Comput. Educ.* 71.July 2015 (2014), pp. 185–197 (cit. on p. 22).

[22] Meng-Tzu Cheng and Len Annetta. "Students' learning outcomes and learning experiences through playing a Serious Educational Game". In: *J. Biol. Educ.* July 2015 (2012), pp. 1–11 (cit. on p. 25).

[23] Meng-Tzu Cheng, Jhih-Hao Chen, Sheng-Ju Chu, and Shin-Yen Chen. "The use of serious games in science education: a review of selected empirical research from 2002 to 2013". In: *J. Comput. Educ.* 2.3 (2015), pp. 353–375 (cit. on p. 25).

[24] Chien-huan Chien, Chien-hsu Chen, and Tay-sheng Jeng. "An Interactive Augmented Reality System for Learning Anatomy Structure". In: *Proc. Int. MultiConference Eng. Comput. Sci.* I (2010) (cit. on p. 23).

[25] Sangwoo Cho, Jeonghun Ku, Yun Kyung Cho, et al. "Development of virtual reality proprioceptive rehabilitation system for stroke patients". In: *Comput. Methods Programs Biomed.* 113.1 (2014), pp. 258–265 (cit. on p. 15).

[26] Gertjan Cleeren, Marc Quirynen, Onur Ozcelik, and Wim Teughels. "Role of 3D animation in periodontal patient education: A randomized controlled trial". In: *J. Clin. Periodontol.* 41.1 (2014), pp. 38–45 (cit. on p. 13).

[27] Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, et al. "Mime". In: *Proc. 26th Annu. ACM Symp. User interface Softw. Technol. - UIST '13.* New York, New York, USA: ACM Press, 2013, pp. 227–236 (cit. on pp. 29, 30).

[28] Thomas M. Connolly, Elizabeth a. Boyle, Ewan MacArthur, Thomas Hainey, and James M. Boyle. "A systematic literature review of empirical evidence on computer games and serious games". In: *Comput. Educ.* 59.2 (2012), pp. 661–686 (cit. on p. 25).

[29] Francesca Cordella, Francesco Di Corato, Loredana Zollo, Bruno Siciliano, and Patrick Van Der Smagt. "Patient performance evaluation using Kinect and Monte Carlo-based finger tracking". In: *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics.* 2012, pp. 1967–1972 (cit. on p. 15).

[30] Vincent Crocher, Pilwon Hur, and Na Jin Seo. "Low-cost virtual rehabilitation games: House of quality to meet patient expectations". In: *2013 Int. Conf. Virtual Rehabil. ICVR 2013.* 2013, pp. 94–100 (cit. on p. 15).

[31] Dragos Datcu, Marina Cidota, Heide Lukosch, and Stephan Lukosch. "On the usability of augmented reality for information exchange in teams from the security domain". In: *Proc. - 2014 IEEE Jt. Intell. Secur. Informatics Conf. JISIC 2014*. 2014, pp. 160–167 (cit. on p. 24).

[32] Larry Davis, Felix G. Hamza-Lup, Jason Daly, et al. "Application of Augmented Reality to Visualizing Anatomical Airways". In: *Proc. SPIE* 4711 (2002), pp. 400–405 (cit. on p. 22).

[33] Teófilo E. De Campos, Walterio W Mayol Cuevas, and David W. Murray. "Directing the attention of a wearable camera by pointing gestures". English. In: *Brazilian Symp. Comput. Graph. Image Process.* IEEE, 2006, pp. 179–186 (cit. on pp. 29, 30).

[34] Chris Dede. "Immersive Interfaces for Engagement and Learning". In: *Science (80-.)*. 323.5910 (2009), pp. 66–69 (cit. on p. 18).

[35] Ángela Di Serio, María Blanca Ibáñez, and Carlos Delgado Kloos. "Impact of an augmented reality system on students' motivation for a visual art course". In: *Comput. Educ.* 68 (2013), pp. 585–596 (cit. on p. 22).

[36] a Dix, J Finlay, G D Abowd, and R Beale. *Human-Computer Interaction*. Vol. Third. December. 2004, p. 834 (cit. on p. 26).

[37] Matt Dunleavy, Chris Dede, and Rebecca Mitchell. "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning". In: *J. Sci. Educ. Technol.* 18.1 (2009), pp. 7–22. arXiv: `arXiv:1002.2562v1` (cit. on p. 18).

[38] Richard Dunnill. "Education – An Anatomy of the Discipline". In: *Widening Particip. Lifelong Learn.* 15.4 (2013), pp. 92–95 (cit. on pp. 2, 10).

[39] L. C. Ebert, G. Hatch, M. J. Thali, and S. Ross. "Invisible touch-control of a DICOM viewer with finger gestures using the kinect depth camera". English. In: *J. Forensic Radiol. Imaging* 1.1 (2013), pp. 10–14 (cit. on p. 16).

[40] L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross. "You Can't Touch This: Touch-free Navigation Through Radiological Images". In: *Surg. Innov.* 19.3 (2012), pp. 301–307 (cit. on p. 17).

[41] Jun Ehara and Hideo Saito. "Texture overlay for virtual clothing based on PCA of silhouettes". In: *Proc. - ISMAR 2006 Fifth IEEE ACM Int. Symp. Mix. Augment. Real.* (2007), pp. 139–142 (cit. on p. 21).

[42] P. Eisert, P. Fechteler, and J. Rurainsky. "3-D tracking of shoes for virtual mirror applications". In: *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR* (2008), pp. 1–6 (cit. on p. 21).

[43] P Fallavollita, L Wang, S Weidert, and N Navab. "Augmented Reality in Orthopaedic Interventions and Education". In: *Comput. Radiol. . . .* (2016) (cit. on p. 23).

[44] Alireza Fathi, Ali Farhadi, and James M. Rehg. "Understanding egocentric activities". In: *Proc. IEEE Int. Conf. Comput. Vis.* IEEE, 2011, pp. 407–414 (cit. on pp. 4, 67).

[45] Angel Fenol and Simi Mathew. "Compliance to recall visits by patients with periodontitis - Is the practitioner responsible?" In: *J. Indian Soc. Periodontol.* 14.2 (2010), pp. 106–108 (cit. on p. 13).

[46] Marco Feuerstein, Thomas Mussack, Sandro M. Heining, and Nassir Navab. "Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection". In: *IEEE Trans. Med. Imaging* 27.3 (2008), pp. 355–369 (cit. on p. 24).

[47] Mark Fiala. "Magic Mirror system with hand-held and wearable augmentations". In: *Proc. - IEEE Virtual Real.* (2007), pp. 251–254 (cit. on p. 21).

[48] Gabor Fichtinger, Anton Deguet, Ken Masamune, et al. "Image overlay guidance for needle insertion in CT scanner". In: *IEEE Trans. Biomed. Eng.* 52.8 (2005), pp. 1415–1424 (cit. on p. 24).

[49] Abraham Flexner. *Medical Education in the United States and Canada*. 1943 (cit. on p. 9).

[50] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. "Aperture based selection for immersive virtual environments". In: *Proc. 9th Annu. ACM Symp. user interface Softw. Technol.* (1996), pp. 95–96 (cit. on pp. 28, 68, 80).

[51] Jason R. Frank (Ed.) *The CanMEDS 2005 physician competency framework: Better standards, better physicians, better care.* 2005, pp. 1–40 (cit. on pp. 11, 22).

[52] Alana Da Gama and Pascal Fallavollita. "Motor Rehabilitation Using Kinect : A Systematic Review". In: *GAMES Heal. J. Res. Dev. Clin. Appl.* 4.2 (2015), p. 13 (cit. on p. 15).

[53] Alana Da Gama, Thiago Chaves, Lucas Figueiredo, and Veronica Teichrieb. "Guidance and movement correction based on therapeutics movements for motor rehabilitation support systems". In: *Proc. - 2012 14th Symp. Virtual Augment. Reality, SVR 2012*. 2012, pp. 191–200 (cit. on p. 15).

[54] José García-Rodríguez and Juan Manuel García-Chamizo. "Surveillance and human-computer interaction applications of self-growing models". In: *Appl. Soft Comput. J.* 11.7 (2011), pp. 4413–4431 (cit. on p. 29).

[55] S Gauglitz, B Nuernberger, M Turk, and T Höllerer. "In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation". In: *Proc. ACM Symp. Virtual Real. Softw. Technol. VRST*. New York, New York, USA: ACM Press, 2014, pp. 197–205 (cit. on p. 24).

[56] General Medical Council. "Tomorrow ' s Doctors - Outcomes and standards for undergraduate medical education". In: *2009* 1 (2009), pp. 1–108 (cit. on p. 22).

[57] Michael Gonzales and Laurel Riek. "A Shared Interface to Improve Oncologist-Patient Communication". In: *Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthc.* (2012) (cit. on p. 14).

[58] C Grätzel, T Fong, S Grange, and C Baur. "A non-contact mouse for surgeon-computer interaction." In: *Technol. Health Care* 12.3 (2004), pp. 245–57 (cit. on p. 16).

[59] Jerome Grosjean and Sabine Coquillart. *The magic mirror: A metaphor for assisting the exploration of virtual worlds*. 1999 (cit. on pp. 21, 31).

[60] Taejin Ha and W Woo. "An Empirical Evaluation of Virtual Hand Techniques for 3D Object Manipulation in a Tangible Augmented Reality Environment". In: *IEEE Symp. 3D User Interfaces*. 2010, pp. 91–98 (cit. on p. 51).

[61] Taejin Ha, Steven Feiner, and Woontack Woo. "WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception". English. In: *2014 IEEE Int. Symp. Mix. Augment. Real.* Figure 1. IEEE, 2014, pp. 219–228 (cit. on pp. 29, 30).

[62] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. "OmniTouch". In: *Proc. 24th Annu. ACM Symp. User interface Softw. Technol. - UIST '11*. New York, New York, USA: ACM Press, 2011, p. 441 (cit. on pp. 4, 28–30).

[63] S. G. Hart. "Nasa-Task Load Index (NASA-TLX); 20 Years Later". en. In: *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 50.9 (2006), pp. 904–908 (cit. on p. 90).

[64] A. Hilsmann and P. Eisert. "Optical flow based tracking and retexturing of garments". In: *2008 15th IEEE Int. Conf. Image Process.* (2008) (cit. on p. 21).

[65] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. "Exploring MARS: Developing indoor and outdoor user interfaces to a mobile augmented reality system". In: *Comput. Graph.* 23.6 (1999), pp. 779–785 (cit. on p. 24).

[66] John Hsu, Jie Huang, Vicki Fung, et al. "Health information technology and physician-patient interactions: Impact of computers on communication during outpatient primary care visits". In: *J. Am. Med. Informatics Assoc.* 12.4 (2005), pp. 474–480 (cit. on p. 13).

[67] Chi-Chiang Huang, Rong-Hao Liang, Liwei Chan, and Bing-Yu Chen. "Dart-it". In: *ACM SIGGRAPH 2014 Emerg. Technol. - SIGGRAPH '14*. New York, New York, USA: ACM Press, 2014, pp. 1–1 (cit. on pp. 28, 30).

[68] Johannes Huber, Andreas Ihrig, Wolfgang Herzog, et al. "Patients' view of their preoperative education for radical prostatectomy: Does it change after surgery?" In: *J. Cancer Educ.* 27.2 (2012), pp. 377–382 (cit. on p. 13).

[69] Antje Hübler, Christian Hansen, Oliver Beuing, Martin Skalej, and Bernhard Preim. "Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining". In: (2014), pp. 0–3 (cit. on p. 16).

[70] Andreas Ihrig, Wolfgang Herzog, Christian G. Huber, et al. "Multimedia support in preoperative patient education for radical prostatectomy: The physicians' point of view". In: *Patient Educ. Couns.* 87.2 (2012), pp. 239–242 (cit. on p. 14).

[71] A Ionescu. "A mouse in the O.R." In: *Ambidextrous, Stanford Univ. J. 4* 4 (2006), pp. 30 –32 (cit. on p. 16).

[72] Alejandro Jaimes, Nicu Sebe, and Daniel Gatica-Perez. "Human-centered computing". In: *Proc. 14th Annu. ACM Int. Conf. Multimed. - Multimed. '06* (2006), p. 855 (cit. on p. 9).

[73] S Jalaliniya and T Pederson. "Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach". In: *Pervasive Comput. IEEE* 14.3 (2015), pp. 22–31 (cit. on p. 17).

[74] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Büthe, and Thomas Pederson. "Touchless interaction with medical images using hand & foot gestures". In: *Proc. 2013 ACM Conf. Pervasive ubiquitous Comput. Adjun. Publ. - UbiComp '13 Adjun.* New York, New York, USA: ACM Press, 2013, pp. 1265–1274 (cit. on p. 17).

[75] Youngkyoon Jang, Seung-Tak Noh, Hyung Jin Chang, Tae-Kyun Kim, and Woontack Woo. "3D Finger CAPE: Clicking Action and Position Estimation under Self-Occlusions in Egocentric Viewpoint". English. In: *IEEE Trans. Vis. Comput. Graph.* 21.4 (2015), pp. 501–510 (cit. on pp. 29, 30).

[76] H. B. Jimison, P. P. Sher, R. Appleyard, and Y. LeVernois. "The Use of Multimedia in the Informed Consent Process". en. In: *J. Am. Med. Informatics Assoc.* 5.3 (1998), pp. 245–256 (cit. on p. 14).

[77] L Johnson, S Adams Becker, V Estrada, and A Freeman. *NMC Horizon Report: 2014 Higher Education*. 2014, pp. 55–57 (cit. on p. 22).

[78] Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. "Exploring the potential for touchless interaction in image-guided interventional radiology". In: *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. 2011, pp. 3323–3332 (cit. on p. 3).

[79] Jokob Weiss. "An Interactive Anatomy Education Game using Augmented Reality". Bachelor. Technische Universität München, 2013 (cit. on p. 52).

[80] Carmen Juan, Francesca Beatrice, and Juan Cano. "An Augmented Reality system for learning the interior of the human body". In: *Proc. - 8th IEEE Int. Conf. Adv. Learn. Technol. ICALT 2008* (2008), pp. 186–188 (cit. on p. 23).

[81] Carolien Kamphuis, Esther Barsom, Marlies Schijven, and Noor Christoph. "Augmented reality in medical education?" In: *Perspect. Med. Educ.* (2014), pp. 300–311 (cit. on p. 23).

[82] HA Karimi. *Telegeoinformatics. Location-based computing and services*. 2005, p. 319 (cit. on p. 18).

[83] Moritz Kassner, William Patera, and Andreas Bulling. "Pupil". In: *Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Adjun. Publ. - UbiComp '14 Adjun.* (2014), pp. 1151–1160. arXiv: `arXiv:1405.0006v1` (cit. on pp. 29, 30).

[84] Joshua C. Klontz, Brendan F. Klare, Scott Klum, Anil K. Jain, and Mark J. Burge. "Open source biometric recognition". In: *2013 IEEE Sixth Int. Conf. Biometrics Theory, Appl. Syst.* 2013, pp. 1–8 (cit. on p. 21).

[85] Ernst Kruijff, J. Edward Swan, and Steven Feiner. "Perceptual issues in augmented reality revisited". In: *9th IEEE Int. Symp. Mix. Augment. Real. 2010 Sci. Technol. ISMAR 2010 - Proc.* IEEE, 2010, pp. 3–12 (cit. on pp. 1, 18, 21).

[86]   Arun Kulshreshth, Chris Zorn, and Joseph J. Laviola. "Poster: Real-time markerless kinect based finger tracking and hand gesture recognition for HCI". In: *IEEE Symp. 3D User Interface 2013, 3DUI 2013 - Proc.* IEEE, 2013, pp. 187–188 (cit. on p. 27).

[87]   Vito Lampret. "A sharp double inequality for sums of powers". In: *J. Inequalities Appl.* 2011.2 (2011), pp. 1–20 (cit. on p. 18).

[88]   Sören Lenman, Lars Bretzner, and Björn Thuresson. "Using marking menus to develop command sets for computer vision based hand gesture interfaces". In: *Proc. Second Nord. Conf. Human-computer Interact. - Nord. '02*. 2002, p. 239 (cit. on p. 27).

[89]   T. M. Lewis, R. Aggarwal, N. Rajaretnam, T. P. Grantcharov, and A. Darzi. "Training in surgical oncology - The role of VR simulation". In: *Surg. Oncol.* 20.3 (2011), pp. 134–139 (cit. on p. 22).

[90]   Yin Li, Zhefan Ye, and James M. Rehg. "Delving into egocentric actions". In: *2015 IEEE Conf. Comput. Vis. Pattern Recognit.* 1. 2015, pp. 287–295 (cit. on pp. 4, 67).

[91]   Jiandong Liang and Mark Green. "JDCAD: A highly interactive 3D modeling system". In: *Comput. Graph.* 18.4 (1994), pp. 499–506 (cit. on pp. 28, 30).

[92]   Jianfeng Lu, Zhigeng Pan, Hai Lin, Mingmin Zhang, and Jiaoying Shi. "Virtual learning environment for medical education based on VRML and VTK". In: *Comput. Graph.* 29.2 (2005), pp. 283–288 (cit. on p. 12).

[93]   KM Ludmerer. "The Internal Challenges to Medical Education". In: *Trans. Am. Clin. . . .* 114.314 (2003), pp. 241–253 (cit. on p. 10).

[94]   Yuan Ping Luh, Jeng Bang Wang, Jin Wan Chang, Shun Ya Chang, and Chih Hsing Chu. *Augmented reality-based design customization of footwear for children*. 2013 (cit. on p. 21).

[95]   S Lukosch and M Billinghurst. "Collaboration in Mediated and Augmented Reality (CiMAR) Summary". English. In: *Mix. . . .* IEEE, 2015, pp. 1–2 (cit. on p. 24).

[96]   Stephan Lukosch, Mark Billinghurst, Leila Alem, and Kiyoshi Kiyokawa. "Collaboration in Augmented Reality". In: *Comput. Support. Coop. Work CSCW An Int. J.* 24.6 (2015), pp. 515–525 (cit. on p. 24).

[97]   Danielle Matthews, Tanya Behne, Elena Lieven, and Michael Tomasello. "Origins of the human pointing gesture: A training study". In: *Dev. Sci.* 15.6 (2012), pp. 817–829 (cit. on pp. 3, 27, 67, 69).

[98]   John C. McLachlan, John Bligh, Paul Bradley, and Judy Searle. *Teaching anatomy without cadavers*. 2004 (cit. on p. 11).

[99]   David McNeill and David McNeill. "Catchments and contexts: non-modular factors in speech and gesture production". In: *Lang. Gesture* (2000), pp. 312–328 (cit. on pp. 3, 27, 67).

[100] Ma Meng, Pascal Fallavollita, Tobias Blum, et al. "Kinect for interactive AR anatomy learning". In: *2013 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2013*. Adelaide, Australia, 2013, pp. 277–278 (cit. on pp. 22, 45).

[101] Ma Meng, Kevin Merckx, Pascal Fallavollita, et al. "[POSTER] Natural User Interface for Ambient Objects". English. In: *2015 IEEE Int. Symp. Mix. Augment. Real.* Fukuoka, Japan: IEEE, 2015, pp. 76–79 (cit. on p. 29).

[102] Alma S Merians, Howard Poizner, Rares Boian, Grigore Burdea, and Sergei Adamovich. "Sensorimotor Training in a Virtual Reality Environment: Does It Improve Functional Recovery Poststroke?" In: *Neurorehabil Neural Repair* 20.2 (2006), pp. 252–267 (cit. on p. 2).

[103] Cheryl D. Metcalf, Rebecca Robinson, Adam J. Malpass, et al. "Markerless motion capture and measurement of hand kinematics: Validation and application to home-based upper limb rehabilitation". English. In: *IEEE Trans. Biomed. Eng.* 60.8 (2013), pp. 2184–2192 (cit. on p. 15).

[104] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. "Augmented Reality: A class of displays on the reality-virtuality continuum". In: *SPIE (International Soc. Opt. Photonics)* 2351.Telemanipulator and Telepresence Technologies (1994), pp. 282–292 (cit. on p. 17).

[105] Leslie M. Miller, Ching I. Chang, Shu Wang, Margaret E. Beier, and Yvonne Klisch. "Learning and motivational impacts of a multimedia science game". In: *Comput. Educ.* 57.1 (2011), pp. 1425–1433 (cit. on p. 25).

[106] Pranav Mistry, Pattie Maes, and Liyan Chang. "WUW - wear Ur world". In: *Proc. 27th Int. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA '09* (2009), p. 4111 (cit. on pp. 4, 28–30).

[107] Thiago Franco Moraes, Paulo Henrique Junqueira Amorim, Fábio S Azevedo, and Jorge Vicente Lopes Silva. "InVesalius - An open-source imaging application". In: *Comput. Vis. Med. Image Process.* (2012), pp. 405–408 (cit. on pp. 88, 89).

[108] Michael Moyle and Andy Cockburn. "Gesture navigation: an alternative 'back' for the future". In: *CHI '02 CHI '02 Ext. Abstr. Hum. factors Comput. Syst.* 2002, pp. 822–823 (cit. on p. 27).

[109] Suranga Nanayakkara, Roy Shilkrot, Kian Peen Yeo, and Pattie Maes. "EyeRing". In: *Proc. 4th Augment. Hum. Int. Conf. - AH '13* (2013), pp. 13–20 (cit. on pp. 4, 27, 67).

[110] Nassir Navab, Sandro Michael Heining, and Joerg Traub. "Camera augmented mobile C-arm (CAMC): Calibration, accuracy study, and clinical applications". In: *IEEE Trans. Med. Imaging* 29.7 (2010), pp. 1412–1423 (cit. on p. 23).

[111] Nassir Navab, Tobias Blum, Lejing Wang, Asli Okur, and Thomas Wendler. "First deployments of augmented reality in operating rooms". In: *Computer (Long. Beach. Calif)*. 45.7 (2012), pp. 48–55 (cit. on pp. 1, 23, 32).

[112] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, et al. "KinectFusion: Real-time dense surface mapping and tracking". In: *2011 10th IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2011* (2011), pp. 127–136 (cit. on p. 19).

[113] Tao Ni, Amy K. Karlson, and Daniel Wigdor. "AnatOnMe". In: *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. CHI '11 (2011), p. 3333 (cit. on pp. 14, 22).

[114] Kai Nickel and Rainer Stiefelhagen. "Pointing Gesture Recognition based on 3D-Tracking of Face , Hands and Head Orientation Categories and Subject Descriptors". In: *Proc. 5th Int. Conf. Multimodal interfaces* (2003), pp. 140–146 (cit. on pp. 28, 30).

[115] C Nishida, Gt Ko, and S Kumanyika. "Body fat distribution and noncommunicable diseases in populations: overview of the 2008 WHO Expert Consultation on Waist Circumference and Waist–Hip Ratio". In: *Eur. J. Clin. Nutr.* 64.1 (2009), pp. 2–5. arXiv: `/www.nutrociencia.com.br/upload{\_}files/artigos{\_}download/Distribui{\%}C3{\%}A7{\%}C3{\%}A3o{\%}20de{\%}20gordura{\%}20corporal.pdf [http:]` (cit. on p. 44).

[116] Donald a. Norman. "Natural user interfaces are not natural". In: *Interactions* 17.3 (2010), p. 6 (cit. on p. 17).

[117] Michael North, Marta Omedes-Pujol, and Courtney Williamson. *Investigation of lewis acid versus lewis base catalysis in asymmetric cyanohydrin synthesis*. Tech. rep. 37. 2010, pp. 11367–11375 (cit. on p. 18).

[118] Antti Nurminen, Juha Järvi, and Matti Lehtonen. "A Mixed Reality Interface for Real Time Tracked Public Transportation". In: *10th ITS Eur. Congr.* June (2014) (cit. on p. 18).

[119] Kenton O'Hara, Gerardo Gonzalez, Graeme Penney, et al. "Interactional order and constructed ways of seeing with touchless imaging systems in surgery". In: *Comput. Support. Coop. Work CSCW An Int. J.* 23.3 (2014), pp. 299–337 (cit. on p. 3).

[120] L. M L Ong, J. C J M de Haes, A. M. Hoos, and F. B. Lammes. "Doctor-patient communication: A review of the literature". In: *Soc. Sci. Med.* 40.7 (1995), pp. 903–918 (cit. on p. 13).

[121] Noritaka Osawa, K Asai, and YY Sugimoto. "Immersive graph navigation using direct manipulation and gestures". In: *Proc. ACM Symp. . . .* (2000), pp. 147–152 (cit. on p. 27).

[122] Marina Papastergiou. "Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation". In: *Comput. Educ.* 52.1 (2009), pp. 1–12 (cit. on p. 25).

[123] Jin Seo Park, Min Suk Chung, Sung Bae Hwang, Byeong Seok Shin, and Hyung Seon Park. "Visible Korean human: Its techniques and applications". In: *Clin. Anat.* 19.3 (2006), pp. 216–224 (cit. on p. 35).

[124] T Pederson, L E Janlert, and D Surie. "Towards a model for egocentric interaction with physical and virtual objects". In: *Proc. 6th Nord. Conf. Human-Computer Interact. Extending Boundaries*. New York, New York, USA: ACM Press, 2010, pp. 755–758 (cit. on pp. 17, 102).

[125] Philipp Jutzi. "Muscle Learning through Interactive Augmented Reality". Bachelor. Technische Universität München, 2015 (cit. on p. 57).

[126] Jeffrey S Pierce, Andrew S Forsberg, Matthew J Conway, et al. "Image plane interaction techniques in 3D immersive environments". In: *ACM i3D* (1997), 39–ff (cit. on pp. 28, 30, 68).

[127] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. "As if being there". In: *Proc. ACM 2012 Conf. Comput. Support. Coop. Work - CSCW '12* 5 (2012), p. 1267 (cit. on p. 24).

[128] BA Ponce, JK Jennings, and TB Clay. "Telementoring: Use of augmented reality in orthopaedic education". In: *J. Bone . . .* 96.10 (2014), e84 (cit. on p. 23).

[129] Jagdish L. Raheja, Ankit Chaudhary, and Kunal Singal. "Tracking of fingertips and centers of palm using KINECT". In: *Proc. - CIMSim 2011 3rd Int. Conf. Comput. Intell. Model. Simul.* 2011, pp. 248–252 (cit. on p. 27).

[130] Siddharth S. Rautaray and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey". In: *Artif. Intell. Rev.* 43.1 (2012), pp. 1–54 (cit. on p. 27).

[131] Michael J. Reale, Shaun Canavan, Lijun Yin, Kaoning Hu, and Terry Hung. "A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing". In: *IEEE Trans. Multimed.* Vol. 13. 3. 2011, pp. 474–486 (cit. on p. 28).

[132] Florian Reichhold. "Rehabilitation Game using Kinect for leg stability and balance training". PhD thesis. Technische Universität München, 2014 (cit. on p. 62).

[133] Guillermo M. Rosa and María L. Elizondo. "Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report". English. In: *Imaging Sci. Dent.* 44.2 (2014), pp. 155–160 (cit. on p. 16).

[134] Guilherme Cesar Soares Ruppert, Leonardo Oliveira Reis, Paulo Henrique Junqueira Amorim, Thiago Franco de Moraes, and Jorge Vicente Lopes da Silva. "Touchless gesture user interface for interactive image visualization in urological surgery". In: *World J. Urol.* 30.5 (2012), pp. 687–691 (cit. on p. 27).

[135] Farrokh Saidi. "American medical education 100 years after the flexner report". In: *Arch. Iran. Med.* 10.1 (2007), pp. 131–133 (cit. on p. 10).

[136] Jaime Sánchez and Ruby Olivares. "Problem solving and collaboration using mobile serious games". In: *Comput. Educ.* 57.3 (2011), pp. 1943–1952 (cit. on p. 25).

[137] Björn W. Schuller, Ian Dunwell, Felix Weninger, and Lucas Paletta. "Serious gaming for behavior change: The state of play". In: *IEEE Pervasive Comput.* 12.3 (2013), pp. 48–55 (cit. on p. 25).

[138] John Schulman, Alex Lee, Jonathan Ho, and Pieter Abbeel. "Tracking deformable objects with point clouds". In: *Proc. - IEEE Int. Conf. Robot. Autom.* 2013, pp. 1130–1137 (cit. on p. 19).

[139] Loren Arthur Schwarz, Ali Bigdelou, and Nassir Navab. "Learning gestures for customizable human-computer interaction in the operating room". In: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 6891 LNCS.PART 1 (2011), pp. 129–136 (cit. on p. 17).

[140] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, et al. "Real-time human pose recognition in parts from single depth images". In: *Stud. Comput. Intell.* 411 (2013), pp. 119–135 (cit. on p. 20).

[141] George Stetten, Aaron Cois, Wilson Chang, et al. "C-mode real-time tomographic reflection for a matrix array ultrasound sonic flashlight". In: *Acad. Radiol.* 12.5 (2005), pp. 535–543 (cit. on p. 23).

[142] Matt Strickland, Jamie Tremaine, Greg Brigley, and Calvin Law. "Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field". In: *Can. J. Surg.* 56.3 (2013), E1–6 (cit. on p. 16).

[143] Jing Sun. "Motion Learning with Biomechanics Principles". In: (2013) (cit. on p. 45).

[144] Satoshi Suzuki and KeiichiA Be. *Topological structural analysis of digitized binary images by border following*. 1985 (cit. on p. 74).

[145] Colin Edward Swindells. "That one there! Pointing to establish device identity". PhD thesis. 2002, p. 151 (cit. on p. 27).

[146] Z. Szalavari, D. Schmalstieg, A. Fuhrmann, and M. Gervautz. ""Studierstube": An environment for collaboration in augmented reality". In: *Virtual Real.* 3.1 (1998), pp. 37–48 (cit. on p. 24).

[147] Palin-Palokas T., Nordblad A., and Remes-Lyly T. *Video as a medium of oral health education for children with mental handicaps*. 1997 (cit. on p. 14).

[148] Hong Z. Tan. "Perceptual user interfaces: haptic interfaces". In: *Commun. ACM* 43.3 (2000), pp. 40–41 (cit. on p. 1).

[149] Justin H Tan, Cherng Chao, Mazen Zawaideh, Anne C Roberts, and Thomas B Kinney. "Informatics in Radiology: developing a touchless user interface for intraoperative image control during interventional radiology procedures." en. In: *Radiographics* 33.2 (2013), E61–70 (cit. on p. 16).

[150] T Tangcharoen, A Bell, S Hegde, et al. *Detection of coronary artery anomalies in infants and young children with congenital heart disease by using MR imaging*. 2011 (cit. on p. 88).

[151] Anthony S. Thijssen and Marlies P. Schijven. "Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees?" In: *Am. J. Surg.* 199.4 (2010), pp. 529–541 (cit. on p. 22).

[152] Rhys Gethin Thomas, Nigel William John, and John Michael Delieu. "Augmented reality for anatomical education." In: *J. Vis. Commun. Med.* 33.1 (2010), pp. 6–15 (cit. on p. 23).

[153] a. Tozeren. *Human Body Dynamics*. 2000, p. 335 (cit. on p. 44).

[154] M. Tuceryan and N. Navab. "Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR". In: *Proc. IEEE ACM Int. Symp. Augment. Real. (ISAR 2000)* FEBRUARY 2000 (2000), pp. 149–158 (cit. on p. 69).

[155] K I Ustinova, J Perkins, W A Leonard, and C J Hausbeck. "Virtual reality game-based therapy for treatment of postural and". In: *Brain Inj.* 28.4 (2014), pp. 486–495 (cit. on p. 2).

[156] Julien Valentin, Philip Torr, Vibhav Vineet, et al. "SemanticPaint". In: *ACM Trans. Graph.* 34.5 (2015), pp. 1–17 (cit. on p. 29).

[157] Koen W. Van Dongen, Gunnar Ahlberg, Luigi Bonavina, et al. "European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills". In: *Surg. Endosc. Other Interv. Tech.* Vol. 25. 1. University Medical Centre, Utrecht, The Netherlands. 2011, pp. 166–171 (cit. on p. 22).

[158] Daniel Vogel and Ravin Balakrishnan. "Distant freehand pointing and clicking on very large, high resolution displays". In: *Proc. 18th Annu. ACM Symp. User Interface Softw. Technol.* (2005), pp. 33–42 (cit. on p. 28).

[159] Juan Wachs, Helman Stern, Yael Edan, et al. "Real-time hand gesture interface for browsing medical images". In: *Int. J. Intell. Comput. Med. Sci. Image Process.* 2.1 (2008), pp. 15–25 (cit. on p. 16).

[160] Daniel Wagner and Dieter Schmalstieg. "ARToolKitPlus for Pose Tracking on Mobile Devices". In: *Proc. 12th Comput. Vis. Winter Work. CVWW07*. 2007, pp. 139–146 (cit. on p. 74).

[161] H Waitzkin. "Doctor-patient communication. Clinical implications of social scientific research." In: *JAMA* 252.17 (1984), pp. 2441–6 (cit. on p. 13).

[162] Xiangyu Wang and Phillip S. Dunston. "Comparative effectiveness of mixed reality-based virtual environments in collaborative design". In: *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 41.3 (2011), pp. 284–296 (cit. on p. 24).

[163] Thomas Wendler, Alexander Hartl, Tobias Lasser, et al. "Towards intra-operative 3D nuclear imaging: reconstruction of 3D radioactive distributions using tracked gamma probes." In: *Med. Image Comput. Comput. Assist. Interv.* 10 (2007), pp. 909–917 (cit. on p. 24).

[164] Lauren Wilcox, Dan Morris, Desney Tan, and Justin Gatewood. "Designing patient-centric information displays for hospitals". In: *Proc. 28th Int. Conf. Hum. factors Comput. Syst. - CHI '10* (2010), p. 2123 (cit. on p. 13).

[165] H Woodford and C Price. "EMG biofeedback for the recovery of motor function after stroke." In: *Cochrane database Syst. Rev.* 2 (2007), p. CD004585 (cit. on p. 15).

[166] Yisong Qiao. "Improving User Recognition for an Augmented Reality Magic Mirror". Master. Technische Universität München, 2015 (cit. on p. 44).

[167] Zhengyou Zhang. "A flexible new technique for camera calibration". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.11 (2000), pp. 1330–1334. arXiv: `arXiv:1011.1669v3` (cit. on p. 68).

[168] Zhengyou Zhang. "Microsoft kinect sensor and its effect". In: *IEEE Multimed.* 19.2 (2012), pp. 4–10 (cit. on p. 18).

[169] Qian-yi Zhou and Vladlen Koltun. "Depth Camera Tracking with Contour Cues". In: *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2015), pp. 632–638 (cit. on p. 20).

[170] Ren Zhou, Meng Jingjing, and Yuan Junsong. "Depth camera based hand gesture recognition and its applications in Human-Computer-Interaction". In: *2011 8th Int. Conf. Information, Commun. Signal Process.* (2011), 5 pp.–5 pp. (Cit. on p. 27).

List of Figures

List of Tables

List of my Publications

[1] M. Ma, P. Fallavollita, S. Habert, S. Weidert, and N. Navab. Device- and system-independent personal touchless user interface for operating rooms : One personal UI to control all displays in an operating room. *Int. J. Comput. Assist. Radiol. Surg.*, mar 2016.

[2] M. Ma, P. Fallavollita, I. Seelbach, A. M. Von Der Heide, E. Euler, J. Waschke, and N. Navab. Personalized augmented reality for anatomy education. *Clin. Anat.*, dec 2015.

[3] M. Ma, K. Merckx, P. Fallavollita, and N. Navab. Natural User Interface for Ambient Objects. In *2015 IEEE Int. Symp. Mix. Augment. Real.*, pages 76–79. IEEE, sep 2015.

[4] M. Meng, P. Fallavollita, T. Blum, U. Eck, C. Sandor, S. Weidert, J. Waschke, and N. Navab. Kinect for interactive AR anatomy learning. In *2013 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2013*, pages 277–278, Adelaide, Australia, oct 2013.

[5] S. Habert, M. Ma, W. Kehl, W. Xiang, F. Tombari, P. Fallavollita, and N. Navab. Augmenting Mobile C-arm Fluoroscopes via Stereo-RGBD Sensors for Multimodal Visualization. In *Mix. Augment. Real. (ISMAR), 2015 IEEE Int. Symp.*, pages 72–75. IEEE, sep 2015.

[6] W. Xiang, S. Habert, M. Meng, H. Chun-Hao, P. Fallavollita, and N. Navab. RGB-D/C-arm Calibration and Application in Medical Augmented Reality. In *Mix. Augment. Real. (ISMAR), 2015 IEEE Int. Symp.*, pages 100–103. IEEE, sep 2015.

[7] P. Stefan, P. Wucherer, Y. Oyamada, M. Ma, A. Schoch, M. Kanegae, N. Shimizu, T. Kodera, S. Cahier, M. Weigl, M. Sugimoto, P. Fallavollita, H. Saito, and N. Navab. An AR edutainment system supporting bone anatomy learning. In *Proc. - IEEE Virtual Real.*, pages 113–114. IEEE, mar 2014.

[8] L. GUO, Y. TANG, Y. DOU, Y. LEI, M. MA, and J. ZHOU. Window Memory Layout Scheme for Alternate Row-Wise/Column-Wise Matrix Access, dec 2013.

[9] G. Lei, Y. Dou, W. Wan, F. Xia, R. Li, M. Ma, and D. Zou. CPU-GPU hybrid accelerating the Zuker algorithm for RNA secondary structure prediction applications. *BMC Genomics*, 13(Suppl 1):S14, jan 2012.

[10] M. Meng, D. Yong, and D. Lin. Data Access Characteristics of OLTP on PostgreSQL. *J. Converg. Inf. Technol.*, 6(10), 2011.

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature,and acknowledgment of collaborative research and discussions. The work was done under the guidance of Professor Nassir Navab and Professor Pascal Fallavollita, at Fakultät für Informatik, Technische Universität München.

Garching bei München, November 3, 2016

Meng Ma