Dissertation

Human-Computer Interaction for Medical Education and Training

Tobias Blum

TECHNISCHE UNIVERSITÄT MÜNCHEN

Chair for Computer-Aided Medical Procedures & Augmented Reality
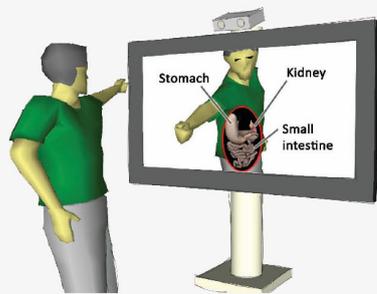
Human-Computer Interaction for Medical Education and Training

Tobias Blum

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: Unif.-Prof. B. Brügge, Ph.D.

Prüfer der Dissertation:

1. Univ.-Prof. Dr. N. Navab
2. Prof. T.M. Peters, Ph.D.
   University of Western Ontario, Kanada

Die Dissertation wurde am 11.07.2012 bei der Technischen Universität München eingereicht

und durch die Fakultät für Informatik am 02.04.2013 angenommen.

Abstract

Due to the increasing interest in computer-based medical education and training, human-computer interaction is becoming an important research topic in this domain. Medical education and training is a complex domain where interests of medical students, hospitals and well-being of the patients have to be carefully considered. To optimally support medical education, future systems need to make use of latest visualization techniques, allow for intuitive interaction and use appropriate domain models such that computer systems can interpret the interaction with the user and provide optimal feedback.

An important strategy for improving human-computer interaction in medical education is to develop systems that have a model of the knowledge domain. For this purpose methods for generating statistical models of medical workflow are discussed and examples of their use in medical training for laparoscopic surgery and ultrasound examination are presented. In addition, this thesis presents the first comprehensive literature review on computer-based ultrasound simulators and teaching concepts that are enabled by these simulators. Furthermore, this thesis describes the implementation of an augmented reality (AR) ultrasound simulator and novel concepts for teaching ultrasound using AR.

For education of anatomy an AR magic mirror system is introduced, which creates the illusion that the user can look into her body and explore her anatomy in conjunction with its corresponding medical imaging data. A new metaphor for touch-free gesture-based interaction is introduced and its implementation within the AR magic mirror system is evaluated and discussed. This thesis provides therefore a complete chain of medical workflow modeling, simulation and AR visualization, and integrates them all into novel computer-based interactive teaching and training systems, which have been fully implemented and partially evaluated together with many clinical partners.

Keywords: Medical Education, Medical Training, Augmented Reality, Human-Computer Interaction, Medical Workflow Models

Zusammenfassung

Aufgrund zunehmender Verbreitung computerbasierter medizinischer Lehr- und Trainingssysteme ist Forschung über die Mensch-Maschine-Interaktion in diesem Bereich ein Thema von wachsender Bedeutung. Medizinische Lehre ist ein komplexes Gebiet in dem Interessen von Studenten, Krankenhäusern und das Wohlergehen der Patienten berücksichtigt werden müssen. Um die Lehre optimal zu unterstützen müssen sich zukünftige Systeme fortschrittlicher Visualisierungsmethoden bedienen, eine intuitive Interaktion erlauben und Modelle des Wissensgebiets verwenden, so dass Computersysteme die Interaktion mit dem Nutzer interpretieren und Feedback geben können.

Eine wichtige Strategie um die Mensch-Maschine Interaktion in der Medizinlehre zu verbessern ist es, Systeme zu entwickeln die ein Modell des Wissensgebiets haben. Zu diesem Zweck werden Methoden zur Erstellung statistischer Workflowmodelle medizinischer Abläufe präsentiert. Deren Nutzung in der Lehre für laparoskopische Operationen und Ultraschalluntersuchungen wird diskutiert. Des Weiteren wird in dieser Arbeit die erste ausführliche Auswertung von Literatur zum Thema computerbasierter Ultraschallsimulatoren und deren Nutzung durchgeführt. Die Implementierung eines Augmented Reality (AR) Ultraschallsimulators wird dargestellt und neue Trainingskonzepte für Ultraschall unter Nutzung von AR und statistischer Workflowmodelle werden eingeführt.

Zur Anatomielehre wird ein AR System vorgestellt, bei dem der Nutzer vor einem Spiegel steht und die Illusion erzeugt wird, dass er in den eigenen Körper blicken kann. Die Nutzerin kann dabei anatomische Strukturen und medizinische Schichtbilder am eigenen Körper sehen. Eine neue Metapher für berührungslose gestenbasierte Interaktion wird vorgestellt und anhand des AR Spiegels evaluiert. Anhand der beiden Applikationsbeispiele, des AR Spiegels und des Ultraschallsimulators, wird die Integration von medizinischen Workflowmodellen, medizinischer Simulation und AR Visualisierung und deren Nutzen dargestellt. Beide Systeme wurden in Kooperation mit verschiedenen Ärzten entwickelt und teilweise evaluiert.

Schlagwörter: Medizinische Lehre, Medizinisch Ausbildung, Augmented Reality, Mensch-Maschine Interaktion, Modelle medizinischer Arbeitsabläufe

Acknowledgments

Contents

1. Introduction

Working as a medical doctor (MD) is one of the most complex, demanding and responsible jobs. The competences and skills of an MD have a tremendous impact on the wellbeing and lives of patients. To master their job, MDs need profound theoretical knowledge about medicine and related topics such as physics, chemistry, psychology, and nowadays even computer science and they have to keep their knowledge up-to-date. At the same time, they need practical experience and for some medical professions such as surgery even excellent motor skills are required. In addition to theoretical knowledge and practical experience in medical topics an MD must also be able to communicate with patients, be aware of legal and regulatory rules and consider economic issues.

It is a huge challenge to educate and train future MDs on all these aspects. While in most other professions it is acceptable that a young professional does not master all aspects of her job, this is not the case for medical professionals. Even a small mistake can have an immense impact on a patient. Another problem is that a medical student needs practical experience with real patients, while it is not acceptable to let a student with improper knowledge and experience train on real patients. Furthermore, medical education is also subject to cost pressure and the duration of medical educations must be kept within an acceptable range.

All these aspects show how important and difficult it is to educate MDs in a way that is optimal for students and patients. In order to overcome some of these problems many different training and education methods have been explored and are used in medical education. One technique that is already applied for a long time is the use of animals and human cadavers. This allows a certain degree of realism in training without the risk of harming patients. Drawings, physical models of anatomy and simulators have been used already several centuries ago. A detailed discussion of illustrations of anatomy can be found in [Bichlmeier, 2010] and the early use of simulation in medical education has been discussed in large detail by [Owen, 2012]. Although medical education and training has a long history, in the last century the number of novel tools to support medical education has strongly increased. Methods to support medical education include simulated patients, where an actor simulates a real patient, patient phantoms and educational videos. A more detailed discussion of different methods that are used in medical education and training will be provided in section 1.2.

Over the last decades, computers have become a central part of medical education and training. The Internet has revolutionized the way we gather information, not only in the field of medicine. Before the Internet has become widely available, a student could only rely on textbooks or lecture scripts. Gathering additional information or an alternative representation of the same information was time consuming. Today the Internet provides fast access to all kind of information. A bad script of a lecture can be compensated by online resources. There is a wide range of images and animations, which are easier to understand than textual descriptions. Online encyclopedias such as Wikipedia provide an easy way to obtain basic knowledge, web pages targeted at MDs provide very detailed expert knowledge and online forums enable discussions with other experts on many topics.

Nowadays, multimedia learning content is often used for medical education. Interactive

3D animations provide a better presentation of medical knowledge than simple images or videos do. They allow the student to control the learning tempo. While not yet successful in all medical areas, computer-based simulators are widely used in some areas such as training of anesthesia and minimally invasive surgery. Many modern phantoms are control by a computer and simulate e.g. pulse and blood pressure. And serious gaming could have to potential to make learning more interesting for students.

In recent years, there has been an increasing demand for individualized education with standardized outcome [Cox et al., 2006, Cooke et al., 2010, Leung, 2002]. Today the curriculum is standardized. However, as different students have different previous knowledge and are talented in different areas, not all students will have the same levels of skills and knowledge after they finished their education. Therefore a standardization of the skills and knowledge a student should achieve instead of a standardization of the curriculum is demanded. To be able to achieve this, each student should receive an individualized education, which allows him to reach these standards. Computer-based training and education will be very important to realize this, as teachers do not have enough time to offer an individualized education to all of their students. Educational multimedia material, which fosters self-educated learning and simulators that adapt to the level of knowledge of the students will play an important role.

Developing systems that optimally support the individual medical student will be a big challenge for the future of medical education. While today, human-computer interaction (HCI) in medical education and training mainly consists of input via mouse and keyboard and output on the screen, future systems have to go beyond that. Today, already many simulators use e.g. haptic devices for input and some research projects use augmented reality for visualization. Building better systems will require both, developing more intuitive input methods and better ways to present medical information. Systems that help students in learning will require new HCI paradigms where computer systems have a complete representation and understanding of the medical domain. A computer system has to understand what a student is doing, has to adapt to the skill level of the student, and has to provide feedback to the student.

While all these topics are extremely relevant for MDs, medical knowledge is also an important part of the general education. Not everyone needs to understand medical details, but it is important to know the basics of anatomy and issues that are related to personal health. Teaching medical knowledge to someone who is not a medical specialist is another challenge where computers and multimedia presentation of knowledge can be of great value as they enable more intuitive presentation of knowledge.

1.1. Organization & Contribution

In this thesis several issues and new methods that are related to HCI for medical education and training, both for medical professionals and non-professionals, will be discussed and presented. In the remainder of this chapter, the history of medical education and training will be presented and current and future developments will be discussed. Furthermore, related topics such as augmented reality and serious gaming will be discussed. In chapter 2, a detailed literature review of one specific domain will be done, which is the simulation of ultrasound (US). In the context of US training, different advantages of computer-based simulation are identified and will be discussed in detail. The implementation of an US simulator using augmented reality (AR) technology and several learning concepts will be presented in chapter 3. In chapter 4, methods for modeling of medical workflows are introduced and examples of their use in medical training and education will be provided. In chapter 5, the concept of an AR magic mirror

Figure 1.1.: Photo of Abraham Flexner taken around 1895. (Source: wikipedia.org)

for education of anatomy will be discussed and in chapter 6 conclusions will be drawn and possible future research direction will be sketched. In appendix A, two issues related to user interfaces when using head-mounted displays are discussed. Although these issues are not directly related to medical education, they are relevant for all systems using head-mounted displays for augmented reality, including also systems for medical education.

1.2. History and State of the Art

In this section we will first examine the history and state of the art in medical education and training. Then we will turn our attention to topics that are related to computer-based education, such as patient phantoms, training centers and serious games. Afterwards human-computer interaction in medical education will be discussed, where special attention will be given to the use of AR.

1.2.1. Medical Education and Training

While education and training of medical practitioners has already been done long before, we will start in the year 1910 and focus on North America where the history is documented best. In this year Abraham Flexner (see figure 1.1), a researcher from the Carnegie Foundation for the Advancement of Teaching, published the so called Flexner report [Flexner, 1910]. At this time, the work of Flexner was the most comprehensive report on medical education so far.

While conducting his study, Flexner visited all 155 medical schools that existed 1909 in the United States and Canada. At that time only few medical schools offered a high-level education and most medical schools did not have a well-defined curriculum. While in France and Germany medical education was state-regulated [Bates, 2008] in the United Kingdom and North America there was no standardization of medical education. Flexner took some of the high-level medical schools as positive examples and identified four major problems of the medical education in most of the other medical schools at that time [Irby et al., 2010]:

- Lack of standardization: There were not accepted academic standards and medical

schools were no accredited. Therefore, there was a wide variety of different curricula. The curriculum was largely based on the wisdom of teachers and results from science were usually not considered in education.

- Lack of integration: The education consisted mainly of lectures. Students had very few laboratory or clinical experience. An integration of scientific knowledge and practical experience was lacking.

- Lack of inquiry: While first research laboratories in Germany started to experimentally examine diseases, only few American universities established scientific medicine. Instead, medical education relied on memorization of existing knowledge. Students were not taught to use inquiry and research to advance the practice of medicine.

- Identity formation: As students had few contacts with practitioners and researchers they had no role models to adapt to.

Based on the example of some medical schools that offered good education, he proposed several features that should become standard for medical education. The education should be done at universities and include clinical experience. Teaching should be done by physicians who also do research and the students should spend time in laboratories. His report had a huge impact and most of his recommendations were implemented and shaped the modern medical education, where patient care, teaching and research are combined.

However, over the course of the 20th century medical education has partially moved away from the ideals of Flexner [Cox et al., 2006]. Due to the enormous pressure to publish and due to economic reasons, academic hospitals focused more and more on research and care for paying patients, instead of teaching. Furthermore, an MD who wants to be successful in research has to spend much time on the research and has only little time for teaching. Vice versa, MDs who are engaged in teaching have problems finding enough time to do cutting-edge research.

Another problem is that today the research is very specific and is focusing on small subtopics [Ludmerer, 2003]. Often these subtopics do not directly relate to topics that are relevant for teaching. Today, much research is done e.g. in the molecular area, which is not directly linked to patient care. On the other hand, only very little research is done on some topics that are highly relevant for teaching. An example is gross anatomy where most things are already known and only very few interesting research topics exist. Therefore, many university hospitals have no researchers dedicated to gross anatomy, whereas it is a crucial part of education [Ludmerer, 2003]. Using computer-based education could be one way to overcome this problem. Experts can develop computer-based training material and students can use it even if their university hospital does not have an expert in this domain. While this is similar to using an educational textbook that has been written by an expert, computer-based education can be much more powerful, in particular for teaching a subject such as gross anatomy, where 3D visualization of great benefit.

In the last years, calls for a new way of medical education emerged [Cox et al., 2006, Cooke et al., 2010]. It is widely believed that in the future not the curriculum, but learning outcomes should be standardized. And to some extend this is already reflected in the current medical education [Leung, 2002]. At the same time the education should be individualized. It is not reasonable to offer the same education and training to every student. For one student who is less skilled in a certain area, the standard curriculum might not be sufficient to achieve a reasonable skill level. The same student might be very talented in another area and match the desired learning outcomes easily. Individualization of the education is important as MDs have

to acquire skills in a large number of areas and therefore education must be as efficient and effective as possible. Introducing this new educational paradigm will be a huge challenge for the next decades. Computer-based education and training can help to address many of the problems that have been discussed before. Computer systems that adapt to the knowledge and skills of students can help to provide an individualized education. At the same time they can help students to achieve standardized outcomes.

1.2.2. Simulated Patients and Patient Phantoms

One area that is related to computer-based education is simulated patients and patient phantoms. Simulated patients are often referred to as standardized patients. This is a term that has been phrased in 1963 [Wallace, 1997] and it denotes an actor who has been trained to take on the characteristics of a real patient. Experts define cases for the simulated patients. This allows students to train on real humans, including communication with the patient. Simulated patients can also be used to train situations that occur rarely in reality. Simulated patients are widely used in medical education [Lane et al., 2001, Kassebaum and Eaglen, 1999, Rosen, 2008] and have shown to be an efficient way to teach medical skills [Greenberg et al., 1999]. Drawbacks of incorporating simulated patients into the training of students are the high efforts and costs. At least one well trained actor is required and unless the actor itself is a medical expert, an additional expert should be present to provide feedback to students.

Wax anatomical models can be seen as predecessor of modern patient phantoms. Such wax models have been popular for anatomy education and in museums since the mid-1850s [Bates, 2008]. It took however more than a century until models of the human anatomy were not only used for education of anatomy but also for simulating medical procedures. Only by the 1970s, first mannequin simulators were introduced for training of mouth to mouth ventilation and cardiac compression [Cooper and Taqueti, 2004, Bradley, 2006, Rosen, 2008]. The first successful simulator was Rescusi Anne, which was introduced 1971 [Tjomsland and Baskett, 2002]. A photo of the simulator can be seen in figure 1.2. Early mannequins did not use computers and similar simulators are still used today for training tasks such as mouth ventilation.

1.2.3. Computer-Based High Fidelity Simulators

While in other areas such as aviation, spaceflight and nuclear power plants computer-based simulation is already used for decades [Rosen, 2008] it took a long time until the first medical simulators were successful. The first computer-controlled mannequin was already shown in 1969, but it was not successful. It was too expensive and there was no market for training other than for the standard model for training, which was the apprenticeship model [Cooper and Taqueti, 2004]. The first successful computer-based mannequins have only been developed in the 1980s. One example is an anesthesia simulator using a mannequin, which was developed in 1988 by [Gaba and DeAnda, 1988]. The simulator is shown in figure 1.3. Today anesthesia is still one of the areas with the most intense use of computer-controlled mannequins.

Over the last decades, technological advances have enabled a large range of different simulators. The most important technologies that are used in modern simulators are haptic interface devices and 3D visualization. Haptic interface devices (see [Salisbury et al., 2004] for an introduction), as seen in figure 1.4 allow interactions between the user and virtual objects. These

Figure 1.2.: The Rescusi Anne mannequin that is used for training mouth to mouth ventilation [Tjomsland and Baskett, 2002].
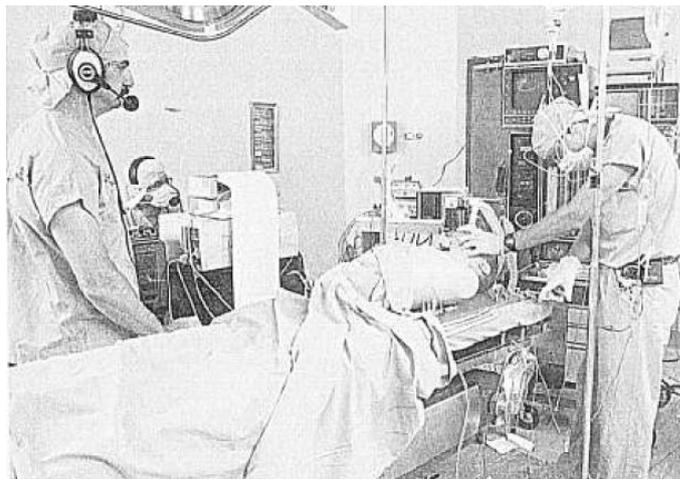
Figure 1.3.: One of the first simulators for anesthesia using computers [Gaba and DeAnda, 1988].
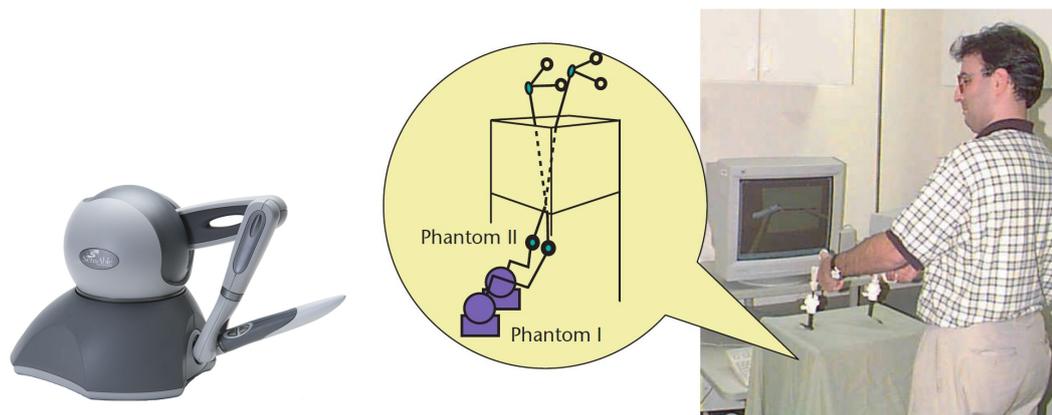
(a) Example of a haptic interface de- (b) Example of a laparoscopic simulator using a haptic device
vice. Phantom Omni from Sensable [Basdogan et al., 2004].
(Wilmington, United States).

Figure 1.4.: Examples of haptic interface devices.

devices can read the position of a robot end-effector and apply force to the end-effector. Haptic rendering algorithms are used to simulate forces that occur during the interaction between the user and the virtual object.

Visualization of medical data is a topic of large interest already for several decades. Compared to other areas where computer graphics are used, the visualization of medical data is very demanding. The visualization must correctly represent the reality and medical datasets are very large. However, as the computer gaming industry has developed high performance graphics cards that can also be used for medical visualization, today a very high level of quality can be achieved. Figure 1.5 shows two exemplary images of early and recent medical computer graphics.

While developing systems for medical education and training is still challenging, advances in technology have led to the development of a large range of different task-specific simulators. The areas where simulation is most successful and where it is in daily use at many universities are training for laparoscopic surgeries, cardiovascular disease simulators, multimedia computer systems and anesthesia [Issenberg et al., 1999]. One sub-area of medical simulation and training are simulators of ultrasound. This area will be discussed in great detail in chapters 2 and 3.

In the context of computer-based simulation often the terms virtual reality (VR) and augmented reality (AR) are used. Both can be seen as parts of the reality-virtuality continuum [Milgram and Kishino, 1994], which is shown in figure 1.6. In AR, the real world is enriched by virtual information. In VR on the other side, the user is fully immersed into a virtual environment. While simulators that use a haptic device and visualization on a screen are sometimes denoted as AR, we stick to the definition of Milgram and refer to such simulators as virtual reality or augmented virtuality (AV). Today, most systems for medical training and simulation are VR or AV systems. AR systems will be discussed in more detail in subsection 1.3.1 and an US simulator using AR will be presented in chapter 3.

While there are still few evaluations of simulator systems, many people see the future of medical training in simulation. [Vozenilek et al., 2004] reformulated the well-known "see one, do one, teach one" paradigm of medical education in "see one, simulate many, do one competently, and teach everyone". One argument which is often used in favor of computer-based

(a) Early visualization of a human brain [Sunguroff and Greenberg, 1978].
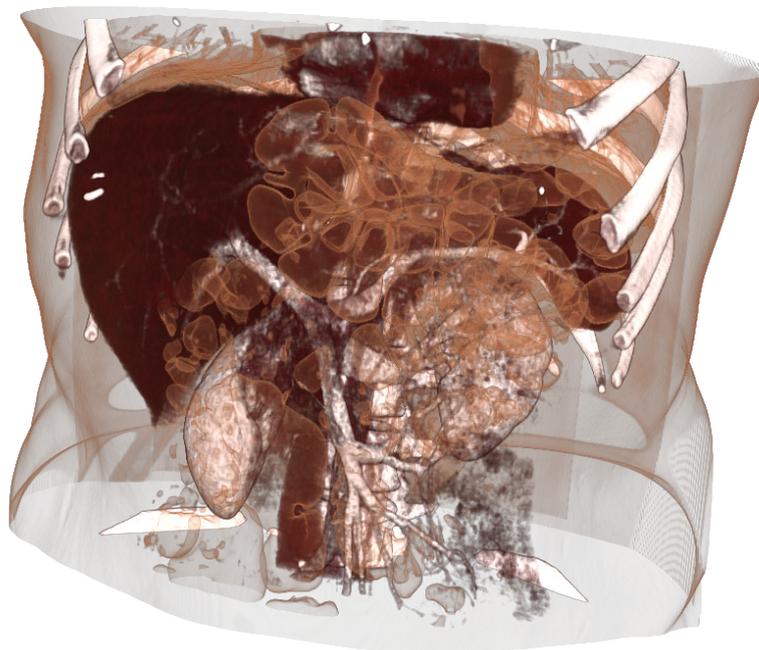
(b) Recent visualization using emission-absorption volume rendering [Kutter, 2010].

Figure 1.5.: Examples of early and recent visualization methods for medical volumes.
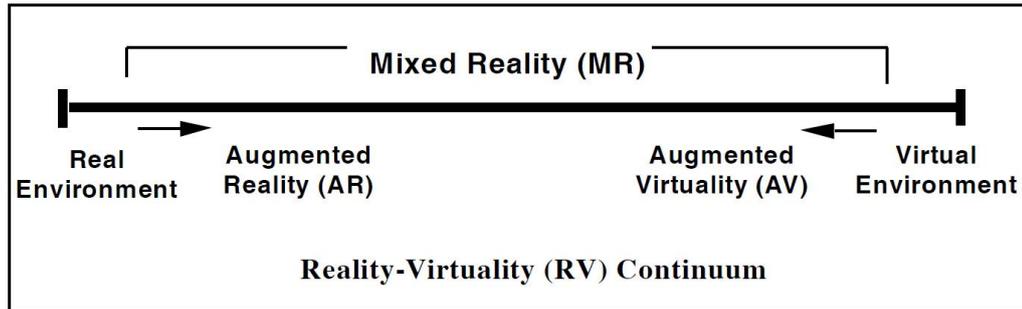
Figure 1.6.: The reality-virtuality continuum [Milgram and Kishino, 1994].
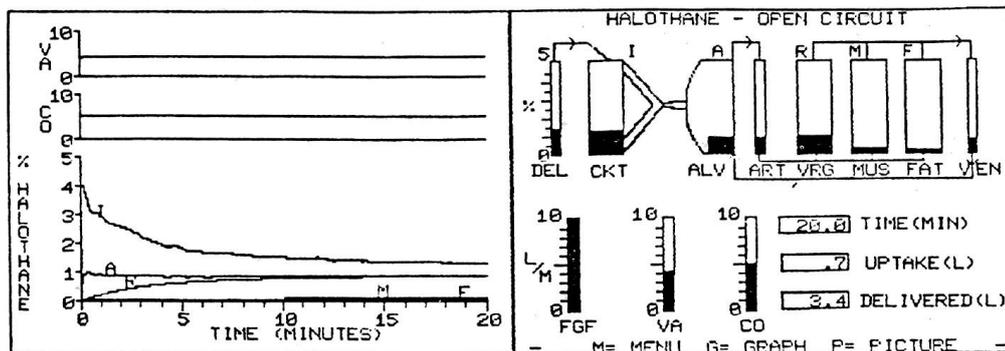
Figure 1.7.: Screenshot from an early screen-based anesthesia simulator [Philip, 1986].

high-fidelity simulators is the ethical imperative to not harm patients [Ziv et al., 2003].

1.2.4. E-learning

Screen-based medical training and education, often referred to as E-learning, is the area where today computers have the highest impact on medical education. E-learning includes a wide range of different learning tools. It ranges from websites that are used in self-managed learning, over web forums where details can be discussed to medical simulation software.

The first area where E-learning methods emerged is screen-based simulators. The first screen-based simulations have been developed in the 1960s. But it took until the 1980s to develop simulators that were used regularly in practice [Lane et al., 2001]. The first example of a successful simulator is Gas Man, an anesthesia screen-based simulator, which did not use a mannequin [Philip, 1986]. Instead, the system was operated by game paddles and relevant information was shown on a computer screen as can be seen in figure 1.7. The software was initially shown in 1986 and is sold until today. Such simulators allow students to do self-managed learning and they can offer different difficulty levels, a range of different cases and some of them can provide immediate feedback to the student [Lane et al., 2001]. A good overview on other screen-bases simulation systems can be found in [Rosen, 2008]. While such simulators can be made easily available to students, one drawback is that they are limited to visualizing information on the screen. There is only a low level of realism and immersion. Such systems can only be seen as an extension to books and lecture scripts but not as replacement of practical experience.

Besides simulation, there are also systems that use computers to present medical knowledge using text, images, video and animations. Furthermore there are decision-making sys-

tems and case databases. The growing availability of the Internet greatly influenced the use of E-learning tools [Ruiz et al., 2006]. While initially, mainly simulators running on workstation were used, the Internet allows the use of many more E-learning tools such as case databases and multimedia material. In the beginning, most E-learning tools were distributed via CD-ROM or only installed on certain workstations. Today the majority of E-learning content is distributed via the Internet. For example 83% of the virtual humans are distributed via the Internet [Huang et al., 2007].

By 1998 71.2% of the university hospitals have been using software simulations or models of biological processes and 33.6% used software for problem solving or decision-making in basic science courses [Moberg and Whitcomb, 1999]. Several studies showed that E-learning can be an effective training tool. [Leong et al., 2003] showed that a case database is more effective than paper cases, [Kamin et al., 2003] showed the effectiveness of virtual patients and [Triola et al., 2006] showed that virtual patients lead to comparable learning outcomes as simulated patients.

Another development that impacts the medical education is that the Internet allows us to access information anytime and anywhere. This in particular influences the continuing education as it allows just-in-time learning. An MD can use laptops or smartphones to access relevant information when it is required. This integration of learning with practice is often referred to as convergence [Choules, 2007]. A survey by [Casebeer et al., 2002] has shown that the main motivation for using the Internet for physicians is to solve particular patient problems, which indicates that the integration of learning with practice is already happening.

Beyond information on medical topics and simulators that are offered via the Internet, in recent years collaborative platforms (also known as Web 2.0) such as wikis and blogs are gaining increasing attention. While not all students are familiar with these technologies [Sandars and Schroter, 2007] and there might be some negative side effects [Boulos et al., 2006], such as low quality of the content, first experiences are encouraging. For example [Berger et al., 2007] describes the implementation and use of a wiki that is maintained by medical students and that is used regularly and is considered useful by the students.

1.2.5. Training Centers

One development that is closely related to the development of novel methods for training is the increasing popularity of training centers. These are big centers that are specialized on training MDs, largely MDs who have already graduated, on novel procedures and technologies. To understand why such training centers have emerged in the last two decades, it is useful to have a look at existing centers and their business models.

In table 1.1 a list of some of the most important centers in Europe is shown. It can be seen that currently all centers offer training on minimally invasive surgery (MIS). The use of MIS is growing, however it is still a relative new method and many surgeons are not very confident in using it. At universities it is usually possible to teach the use of MIS as many researchers at university hospitals are interested in new methods and usually there are MDs mastering MIS. At normal hospitals it is more difficult to train surgeons on new technologies. A method such as laparoscopic surgery is too complex to just learn it by trying it on a patient or by reading books. And a normal hospital does usually not have the budget to set up an own training program. Therefore many MDs visit training centers that are specialized on MIS. Further subjects where courses are offered by many training centers are medical imaging and emergency medicine.

To understand why some of these centers do research and why many of them are funding

	IRCAD	IMSat	AMTS	ESI	CCMID
Location	Strasbourg, France	Dundee, Scotland	Lucerne, Switzerland	Norderstedt, Germany	Cáceres, Spain
Focus	MIS	MIS, imaging	MIS, emergency medicine	MIS	MIS
Research	yes	yes	no	no	cooperations
Funding	public / private	public / private	private / public	private	public / private

Table 1.1.: A list of training centers in Europe.

by a private public partnership we will have a look at the business model of such centers in figure 1.8. The main customer segments for a training center are hospitals, industry, research and government. The value proposition of a training center to hospitals has been discussed before. They receive better training for new technologies and therefore hospitals are paying attendance fees for their MDs.

Interestingly, training centers also offer a value proposition to the industry, mainly to manufacturers of devices that are used in new medical procedures. Today, technology has become very complex. When introducing a new medical technology, MDs have to be trained. Otherwise they will not be able to use the new technology. In particular for new products this has become critical. MDs will only buy a product if they understand the benefits and know how to use it. If it takes too long until a new technology is adopted by the users, competitors have time to develop and market similar products. Therefore the value proposition to the industry is to make MDs familiar with their new technologies. In fact, many of the training courses are sponsored and in some training centers even organized by industry.

Another stakeholder for such centers are researchers and start-up companies. Some training centers have research groups attached or they are cooperating with researchers. The value proposition for researchers is that they can get in contact with MDs and that they can use the infrastructure of the center. Often, training centers have an infrastructure e.g. to train surgeries on animals. Such an infrastructure is very valuable for studies.

Many centers are financed by public private partnerships. The main reason for this is that government wants to support the other stakeholders of the center, namely industry, hospitals and research.

In particular as training centers often combine training, research and industry they have become very important in the field of education and training. Expensive systems can only be purchased and maintained by some university hospitals and training centers. Expensive training systems using technology such as VR and AR are mainly interesting for training centers where systems are used frequently. And often research and development on novel technologies for medical training is done at training centers or in cooperation with them.

1.2.6. Serious Games

The idea of serious games is to use computer game elements to make an useful activity more interesting. While the last sections gave an overview on different aspects that are related to training and education of medical experts, the use of serious games is currently mainly focusing on non-professionals. The area where the largest body of
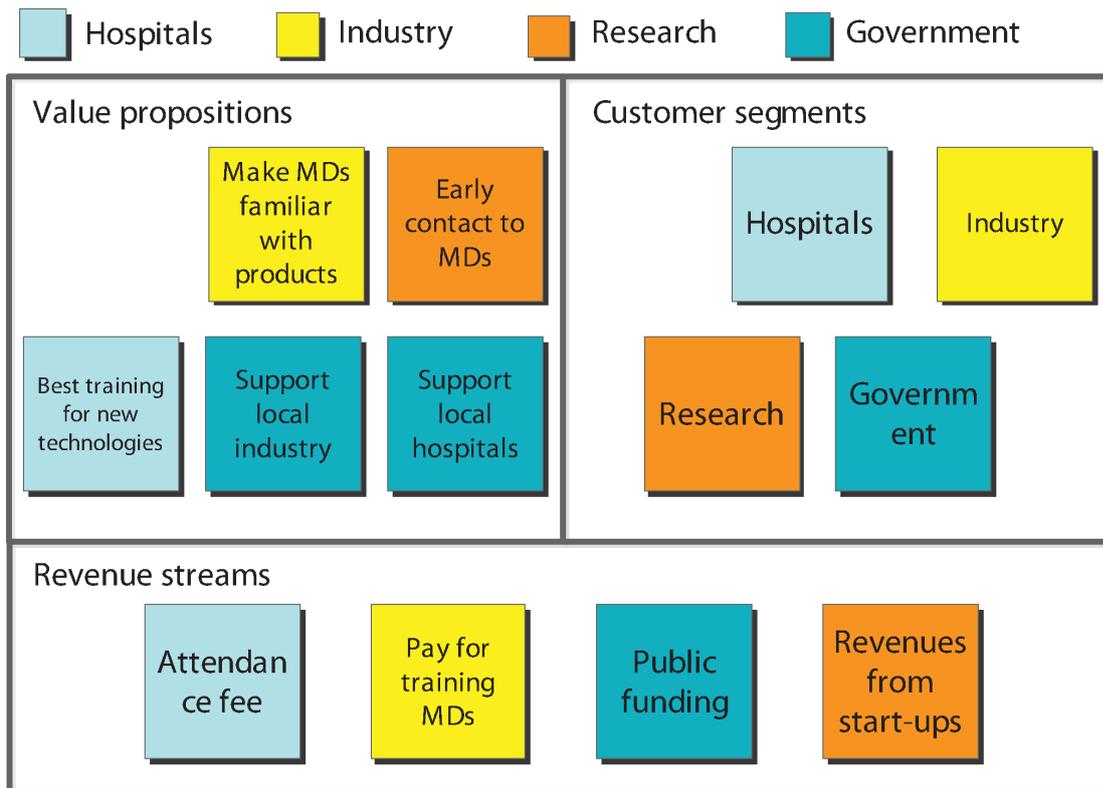
Figure 1.8.: The business model of medical training centers.

work exists is the use of gaming elements in motor rehabilitation, mainly for stroke patients. While the first systems that used VR technology for motor rehabilitation did not include any gaming elements [Jack et al., 2001, Holden et al., 1999, Luo et al., 2005, Piron et al., 2005], many newer system use gaming elements to make exercises more motivating [Flynn et al., 2007, Crosbie et al., 2008, Alankus et al., 2010, Standen et al., 2010]. The games typically use standard elements known from normal computer games, such as flying a spaceship [Standen et al., 2010] as shown in figure 1.9 or playing baseball [Alankus et al., 2010]. Other examples of rehabilitation systems, are games that are designed for balance rehabilitation[Betker et al., 2007, Lange et al., 2010, Fitzgerald et al., 2010] or multiple scleroses [Notelaers et al., 2010] rehabilitation. Serious games often use very advanced HCI methods, utilizing e.g. methods to track the movement of the user and gesture-based input.

Although in rehabilitation games, the gaming elements are mainly used for motivation, there are also some systems that use games for training or to provide knowledge on a topic. One example is the Videodope game, which educates the player about misuse of drugs [Gamberini et al., 2007]. Another example is a game for education of children on chronic pediatric diseases [Lieberman, 2001].

While serious games could also be used to teach medical professionals about certain procedures or diseases, there are few games in this area. Previously, a system using gaming elements has been shown to teach emergency medicine [Vidani et al., 2010] and one system for training of triage [Kizakevich et al., 2006] has been developed. As serious games are a topic where currently a lot of research is done, an increasing use for education of medical professionals can be expected. Here it should be noted that the use of serious gaming elements does not necessarily involve elements such as flying a spaceship or playing baseball. Elements such as

Figure 1.9.: A serious game for motor rehabilitation where the patient can control a spaceship by hand movements [Standen et al., 2010].

giving performance scores and comparing scores of different users on a leaderboard can also be motivating without turning the whole learning exercise into a game.

1.3. Human-Computer Interfaces in Medical Training and Education

In the last section the increasing importance of computer-based solutions for medical training has been discussed and an overview covering different aspects of medical education and training has been given. All computer-based systems for medical education and training use some sort of human-computer interface. The earliest computer-based systems already used a very sophisticated interface, utilizing a mannequin of a human. By using a mannequin, the user does not directly interaction with a computer, and often users are not even aware that they are interacting with a computer system. Instead the illusion of interacting with a human is created. Screen-based systems are much less immersive as they do not mimic a real situation. On the other hand screen-based systems are inexpensive and can be made widely available as they use a standardized user interface, which can be used with any PC. In particular, since today many systems can be used through the Internet such systems can be made available to a large audience at low costs.

In the last decades, a huge range of new interfaces emerged. Ranging from speech input over touch displays to touch-less gesture-based interaction methods. While the use of keyboard and mouse has been the primary interaction method for decades, multi-touch displays are about to replace them in many application areas, such as mobile phones and casual gaming. One interface that is important for medical simulation are haptic devices, which allow simulating the interaction between medical instruments and a human. New visualization methods allow building more realistic simulation systems. In particular VR and AR allow new ways to improve
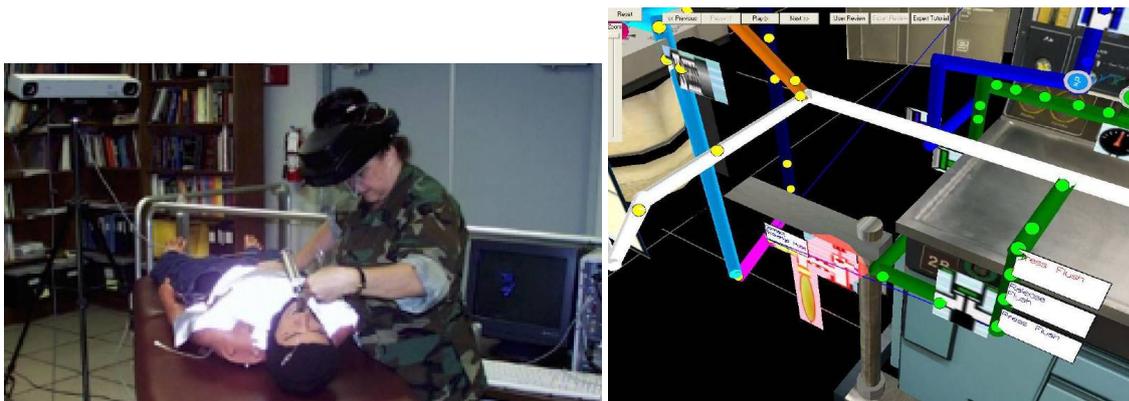
(a) An HMD-based AR simulator visualizing anatomical airways onto a real person [Davis et al., 2002].

(b) Augmentation of internal processes onto an anesthesia simulator [Quarles et al., 2008].

Figure 1.10.: Two examples of AR systems for medical training and education.

computer-based teaching, as will be discussed later on the example of an US simulator. In this context the use of different display technologies such as head-mounted displays plays an important role.

One issue that is also closely related to HCI are systems that can interpret human actions and react appropriately. A computer-system for education has to provide feedback to the user. Feedback might be given as performance measure. When using advanced visualization methods such as VR or AR, also new ways of providing feedback have to be investigated, to optimally educate medical students.

1.3.1. Augmented Reality for Medical Training and Education

As augmented reality systems play an important role in this thesis, we will now discuss the use of AR for medical education. AR systems enrich the real world by virtual objects. See [Azuma et al., 2001] for a general overview on AR, [Sauer et al., 2008] for a discussion of medical AR and [Sielhorst et al., 2008] for an overview on display devices for medical AR. The first AR system was shown in 1968 by Sutherland [Sutherland, 1968] and used a head-mounted display (HMD) to visualize 3D information. In the 1990s, first AR systems for medical use have been developed, including systems using a HMD to augment US images [Bajura et al., 1992], a monitor-based augmentation of magnetic resonance (MR) images for neurosurgery [Lorensen et al., 1993] and augmentation of virtual objects into the image of an operating microscope [Edwards et al., 2000].

AR is also a valuable tool for medical education and training. Unlike VR simulators it can show information in the real world and therefore a higher degree immersion of the user can be achieved. Several AR systems for training have been proposed. [Davis et al., 2002] presented a system that uses a HMD to augmented anatomical airways onto a real person. A monitor-based AR system to train obstetric forceps delivery was proposed by [Lapeer et al., 2004]. [Quarles et al., 2008] presented a system that augments a simulator of an anesthesia machine in order to show the internal state of the machine. The system also allows teachers and students to perform an after action review. Two of the systems can be seen in figure 1.10. The use of AR in training and education will be discussed later in more detail on the example of an AR ultrasound simulator in chapter 3 and an AR magic mirror system in chapter 5.

1.4. Summary & Discussion

As discussed throughout this chapter, there are several important trends and developments in medical training and education. There is a demand for standardized outcomes and individualized education. For university hospitals it is very challenging to provide education at the highest level due to problems such as increasing specialization of researchers. In particular, the demand for standardized outcomes and individualized education poses problems as this requires additional resources. On the other hand there are different developments and technologies that can help to provide a better education. Training centers can provide high-level education on very specific topics and the Internet and E-learning tools enable self-educated learning. While computer-based simulators are not used for all medical disciplines, yet, they are established in some of them. For the future development, computer-based systems have to provide more intelligent human-computer interaction. They have to understand the medical domain and analyze the interaction with the user in order to allow individualized training and enable students to optimally achieve standardized outcomes. Furthermore, advanced visualization is important to allow knowledge transfer to the student.

2. Literature Review on Computer-based Ultrasound Simulators

2.1. Introduction

After the general field of medical training and education has been introduced and discussed in chapter 1, we will investigate one specific domain in large detail. In this chapter, a literature review on existing computer-based US simulators is done and HCI and training concepts in this area are discussed. In chapter 3 an implementation of an AR simulator for US will be presented.

In this review, we will categorize different simulators according to

1. the method that is used for simulation of the US image in subsection 2.2.1,

2. the user interface, which consists of the input device, haptic simulation and output device, in subsection 2.2.2,

3. and the medical application domain in subsection 2.2.3.

In section 2.3 we will discuss training concepts that can be realized using computer-based US simulators and advantages over traditional training. A brief overview on commercial US simulators is provided in section 2.4 and we summarize existing evidence on the learning effect when using computer-based US simulators in section 2.5.

2.1.1. Basics of Ultrasound

Depending on the frequency, sound pressure is classified into three different categories. Sound waves that can be heard by humans are called acoustic. Sound with a frequency lower than 20Hz can usually not be heard by humans and is called infrasound. The highest frequency that can be heard by humans is around 20kHz and sound above this frequency is called ultrasound. Medical ultrasound is usually in the range between 2MHz and 10MHz.

Medical US devices use a probe that can emit and detect ultrasound waves. The sound waves travel through the body and fractions of the sound are reflected whenever the wave encounters a surface between two materials having different acoustical impedance. By detecting the reflected sound, such surfaces can be imaged. Over the course of the last decades, US has become indispensable for a wide range of diagnostic, therapeutic and surgical applications. US has several advantages over other imaging modalities. It does not involve ionizing radiation and devices are inexpensive compared to other imaging modalities. Small and portable US devices have become available and can be used bedside, during interventions or in ambulances and helicopters. In emergency medicine US has become a valuable tool for providing a first diagnosis, as it requires less preparation time than other modalities such as computed tomography (CT).

(a)   (b)
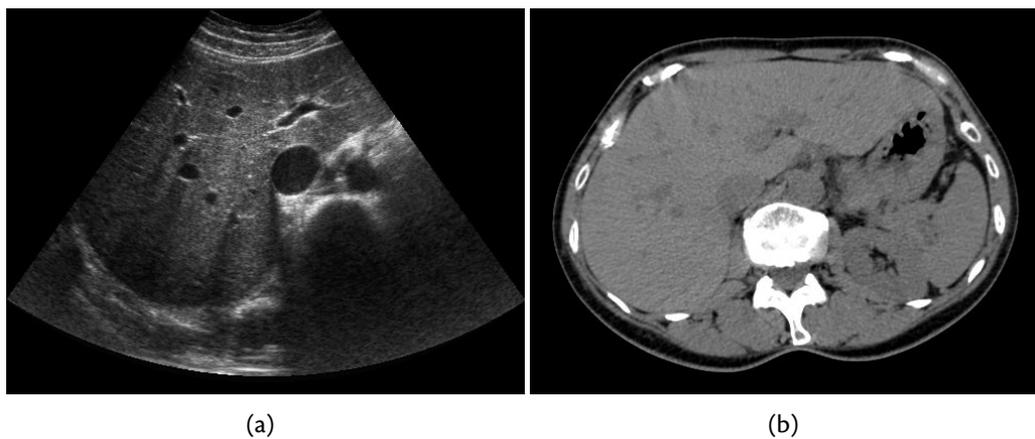
Figure 2.1.: On the left an US image of the liver is shown and on the right a CT image of the same area (images taken from [Wein, 2007]).

Due to these advantages and due to improvements in image quality, US is today used in many diagnostic areas such as cardiology [Cheitlin et al., 2003, Warnes et al., 2008, Cardiac, 2004], bedside cardiology [Beaulieu, 2007, Coletta et al., 2006], screening for fetal anomalies [Whitworth et al., 2010, Pathak and Lees, 2009], dermatology [Schmid-Wendtner and Burgdorf, 2005], diagnosis of the abdomen [Beckh et al., 2002], detection of metastases [Voit et al., 2001], preoperative diagnosis of congenital heart defects [Tworetzky et al., 1999] and diagnosis of esophageal cancer [Kelly et al., 2001]. Furthermore, it is used in emergency medicine [Costantino et al., 2005, Kendall et al., 2007, Counselman et al., 2003] for image-guided biopsies [Memel et al., 1996, Kliewer et al., 1999], intraoperative monitoring [Mahmood et al., 2008] and many other applications.

2.1.2. Motivation for Computer-Based Ultrasound Simulation

While the use of US has many advantages, there are also some fundamental drawbacks. Most other medical imaging modalities show intensity values of material. US only shows interfaces between material with different acoustical impedance, which is more difficult to interpret. In figure 2.1 the difference between an US and a CT image can be seen. The image quality of US is poor. It suffers from low dynamics, low spatial resolution and a low signal to noise ratio. Furthermore, there are many artifacts present in US images, some of them depending on the viewing direction, which makes it very difficult for a novice to interpret US images. A radiologist can interpret every CT or MR volume due to standardized cut planes and viewing modes. For US it is difficult to do a diagnosis only based on an image without knowing the exact position of the probe relative to the patient or the amount of pressure the examiner used. This makes it difficult to teach the correct use of US and it requires a high amount of hands on training. Inter-observer and intra-observer repeatability is low, which has been shown for thyroid volumetry [Andermann et al., 2007, Schlögl et al., 2001, Brauer et al., 2005] and prostate volumetry [Tong et al., 1998]. Even for routine procedures, such as trauma ultrasound, sensitivity is low [Stengel et al., 2001].

The requirement for an intensive training of US is well known. This has led to a number of recommendations on the minimum amount of training novices should receive. Before being able to perform echocardiography independently, a minimum of 150 transthoracic plus

150 Doppler examinations is recommended and for fetal echocardiography a minimum of 25 procedures should be performed [Ones and Creager, 2003]. Other guidelines recommend a minimum of 480 examinations for echocardiography [Ehler and Carney, 2001], 150 for bedside echocardiography [Seward et al., 2002], 300 for critical care [Neri et al., 2007] and 20 for sentinel node biopsy [Tafra, 2001]. However, even such high numbers of cases might not be sufficient to use US confidently. [Hertzberg et al., 2000] showed that radiologists even after 200 cases, which are recommended by the American Institute of Ultrasound in Medicine and the American College of Radiology, still had problems depicting anatomical landmarks and performing examinations. Furthermore, a study by [Moore et al., 2004] showed that in emergency medicine only a low number of hospitals follows these recommendations.

2.2. Classification of Ultrasound Simulators

Over the last fifteen years several computer-based US simulators have been developed. Such simulators have several important advantages compared to classical training. These advantages will be discussed later in detail in section 2.3. In this section we will classify different simulator systems and discuss medical application areas.

Before doing a classification of computer-based simulators, the main difference to traditional ultrasound phantoms is discussed. Today training is often done following the "see one, do one, teach one" paradigm, where students observe experts performing US procedures and later learn how to do them on real patients under supervision of a teacher. For invasive procedures, such as US-guided needle biopsies, often physical phantoms are used. These phantoms mimic relevant physical properties of real tissue and are used with real ultrasound probes. They are made of different materials such as foam, gelatin or fiber [Bude and Adler, 1995, Smith Jr et al., 1998, Liu et al., 2010] and many different phantoms of whole human bodies and single organs exist. To build very inexpensive phantoms sometimes water, meat [Brown et al., 2008] or food [Wu, 2001] is used. We will not discuss such traditional simulation methods in more detail but turn our attention towards computer-based simulation methods.

Computer-based simulators do not use a real ultrasound probe, but simulate the ultrasound image in the computer. As user interface some systems use phantoms of an ultrasound probe and a patient. However these phantoms do only simulate the outer appearance and sometimes the haptics of a patient but not internal structures. Other systems use a virtual representation of the patient on a computer screen and a mouse to control the ultrasound probe. Below we classify existing system into three categories: The method to simulate ultrasound images, the user interface and the medical application. Table 2.1 summarizes key aspects of the most relevant research systems.

References	US image simulation method	User interface	Medical application	Training concepts
[Alterovitz et al., 2003b, Alterovitz et al., 2003a]	2D ultrasound	keyboard + mouse	prostate brachytherapy	
[Ehricke, 1998]	3D ultrasound	keyboard + mouse	no specific	co-registered images, different cases
[Ni et al., 2008, Ni et al., 2009]	3D ultrasound	haptic device	needle biopsy	visual model
[Sclaveranoa et al., 2009, d'Aulignac et al., 2005, Troccaz et al., 2000]	3D ultrasound	haptic device	deep venous thromboses, TRUS	
[Goksel and Salcudean, 2010, Goksel and Salcudean, 2009]	3D ultrasound	haptic device	prostate brachytherapy / needle insertion	visual model
[Abolmaesumi et al., 2004, Tahmasebi et al., 2007, Tahmasebi et al., 2008]	3D ultrasound	haptic device	no specific	visual model, co-registered images
[Heer et al., 2004]	3D ultrasound	phantom + tracked probe	transvaginal	
[Stallkamp and Wapler, 1998]	3D ultrasound	phantom + tracked probe	various	different cases
[Weidenbach et al., 2007, Berlage et al., 1996, Weidenbach et al., 2005, Weidenbach et al., 2000, Trochim, 2002, Weidenbach et al., 2004, Berlage, 2008, Weidenbach et al., 2009]	3D ultrasound + computer model	phantom + tracked probe, stereo monitor	TTE / TEE	visual model, difficulty levels, record and replay, different cases, statistical models

Forest et al. [Hostettler et al., 2005, Forest et al., 2007]	CT-based	haptic device	hepatic biopsy, thermal ablation, obstetrics	visual model, difficulty levels
[Vidal et al., 2008, Vidal et al., 2005, ap Cenydd et al., 2009]	CT-based	haptic device, semi-transparent mirror	needle biopsy	visual model
[Magee et al., 2007, Zhu et al., 2006, Magee and Kessel, 2005, Zhu et al., 2007]	CT-based	phantom + tracked probe	needle biopsy	visual model, record and replay, performance metrics
[Bommersheim et al., 2005]	Visible Human	keyboard + mouse	endoscopic	visual model, co-registered images
[Bürger et al., 2008, Abkai et al., 2007]	computer model	haptic device	IVUS	
[Köhn et al., 2004, Reis et al., 2006]	computer model	keyboard + mouse	echocardiography	different cases
[Arkhurst et al., 2001, Hacker et al., 2002, Arkhurst, 2005]	no simulation	keyboard + mouse	pediatric	visual model, co-registered images

Table 2.1.: Summary of key aspects of the most relevant research systems.

2.2.1. Simulation of Ultrasound Images

2.2.1.1. Interpolative Simulation Methods

Most simulators allow the user to move the US probe and therefore they have to generate US images from different viewpoints. The most common method to simulate 2D images is interpolation from 3D US volumes [Weidenbach et al., 2007, Sclaveranoa et al., 2009,   Aiger and Cohen-Or, 1998,   Stallkamp and Wapler, 1998, Ehricke, 1998,   Abolmaesumi et al., 2004,   Heer et al., 2004,   Terkamp et al., 2003, Maul et al., 2006, Ni et al., 2008, Markov-Vetter et al., 2009, Arkhurst et al., 2001]. An example can be seen in figure 2.2(c). Using state of the art hardware, reslicing of volumes is easy to implement and can be done in real-time on the graphics processing unit (GPU). Resliced images are very realistic as long as the probe in the simulator is placed at the same position and orientation as the real probe that was used to acquire the 3D volume. The main drawback is that, as soon as the user takes other views, view dependent effects are not shown correctly. There are different solutions to this problem. It is possible to acquire several 3D volumes from different viewpoints and switch between them depending on the position of the probe. While this method can provide multiple different viewpoints, it is not effective for small variations of the pose, as a high number of volumes would have to be acquired. Another approach is to add view dependent effects to the resliced image. Effects such as gain, depth gain compensation and focus can be added to the 2D slice [Aiger and Cohen-Or, 1998]. [Ni et al., 2009] have added shadow by ray-casting and finding boundaries. While such methods can make the images more realistic, they cannot simulate all view dependent effects, as there is no underlying model of the US physics. In particular when a 3D US volume already contains artifacts and shadows, it is hardly possible to remove them and replace them by new artifacts.

Interpolative methods can also be used to simulate additional objects, such as needles. [Zhu and Salcudean, 2010] used images of a real needle, which have been acquired in a water tank from different directions. During simulation, the appearance of the needle is interpolated from the prerecorded images.

While for most applications it is necessary to simulate US images from different viewpoints, for some applications it is adequate to only use one slice. [Alterovitz et al., 2003b] presented a method to simulate needle insertion and radioactive seed implantation, where the deformation caused by the needle is simulated, but no movement of the probe is considered.

For some procedures, such as cardiac US, motion is highly important. For prenatal heart diagnostic, [Wüstemann et al., 2008] acquired 4D US volumes. At runtime, the system interpolates the 2D slices from the 4D volume running in a loop. Solutions for recording 4D US are discussed e.g. by [Reis et al., 2003].

2.2.1.2. Generative Simulation Methods

For non-real-time applications, such as transducer design, generative methods have been developed that perform a physics-based simulation using other images, such as CT, as input. As these methods do not run in real-time, different simplified methods have been proposed for training simulators. These methods try simulating a subset of the phenomena involved in the real image formation and usually apply simplifications to them. Effects that are simulated in many methods are absorption, reflection and noise. For some of these effects, ray-based methods are used, which simulate the propagation of sound inside the human body by casting rays through a 3D volume. For applications that do not require real-time simulation, often wave
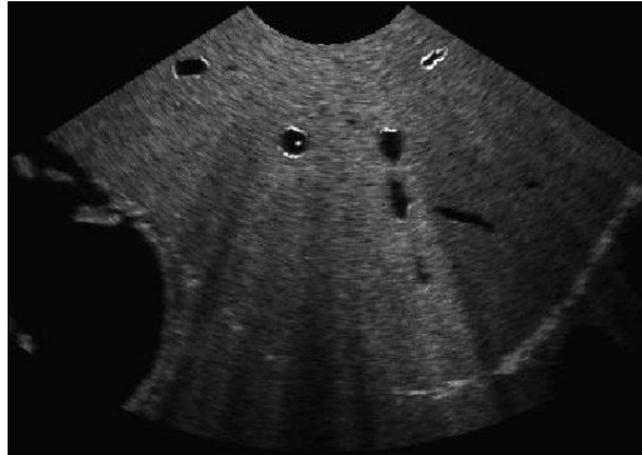
(a) Simulation from segmented Visible Human dataset [Bommersheim et al., 2005].

(b) Image generated from segmented CT volume and texture database [Magee et al., 2007].

(c) 2D slice interpolated from 3D US volume [Wüstemann et al., 2008].

Figure 2.2.: US images simulated with different methods.

propagation is used instead [Jensen, 1996]. Generative simulation methods can be classified into image-based and model-based methods.

Image-Based   Many generative simulation methods use another image, such as a CT or MR volume as input. In CT, a segmentation into air, bone and soft tissue can be obtained using thresholds on the intensities in the CT image. Using such a segmentation, the most basic effects such as absorption and reflection can be simulated as done by [Hostettler et al., 2005] and [Vidal et al., 2008]. A correlation between Hounsfield units and the acoustic impedance was assumed by [Reichl et al., 2008] and [Shams et al., 2008] who used this assumption to simulate absorption, reflection and transmission.

One way to allow more realistic simulation is using a finer segmentation and assigning tissue properties to every voxel. Such a segmentation can currently not be done automatically. [Imani et al., 2002] simulated US from the Visible Human Dataset (VHD) [Spitzer et al., 1996]. They used a segmentation, which is available for this dataset, and enriched this data by tissue characteristics such as acoustic impedance, absorption and scattering coefficients. They generate the images by creating beams where the attenuation is simulated based on the distance from the transducer and the tissue characteristics. [Bommersheim et al., 2005] took a similar approach by assigning tissue properties to different parts of the segmented VHD. They simulate attenuation along the US rays and add Rayleigh noise. An exemplary image of this method can be seen in figure 2.2(a). However, computing an image takes four to five seconds. [Bürger et al., 2008] use segmented CT and MR images, where object properties are assigned manually. They simulate ray propagation, beam forming and backscattering.

An alternative method to ray-based simulation is the use of textures. [Zhu et al., 2006] manually segmented CT data and assigned labels to each voxel. They simulate US images by texturing a 2D slice with textures obtained from real US images. Additional shadow effects are added by 2D ray-casting and radial blur is added in post-processing. An example of a simulation using textures is shown in figure 2.2(b).

Model-Based   While image-based methods have the advantage that large numbers of CT and MR images are available, which can be used to simulate many different patients, they have some limitations. In particular for US of small and moving anatomy, such as the heart, CT does not provide enough information. One way to overcome this is modeling of the anatomy. Sun and McKenzie have built a model of the heart that models the movement of the valves [Sun and McKenzie, 2008]. The US image is created by extracting a 2D slice from the model and texturizing it. For simulation of intravascular ultrasound (IVUS), [Abkai et al., 2007] used functional descriptions of a flexible tissue model of the vessel system. They use a model derived from wave-equations given by the Rayleigh integration method and additional filters to simulate IVUS images. Other work simulated the pumping motion of the heart, operation of the valves and the blood stream [Berlage et al., 1996]. Instead of building a full model of the heart, [Bürger et al., 2008] use 3D CT or MR volumes and animate the heart by forward free form deformation. For US simulation of the abdominal area [Ni et al., 2009] modeled respiratory motion.

One challenge for model-based simulation is to offer a range of different cases. Most models are based on images of one patient and do therefore only provide simulation of this specific patient. In order to allow simulation of different echocardiographic cases, with different pathologies, [Köhn et al., 2004] combined different ontologies to model the heart. They used an ontology that describes the anatomical structure of the healthy heart, and another one describing

different findings and defining the impact of a finding on the structure of the heart. Geometric models of the heart have been constructed manually from MR. In later work [Reis et al., 2006] a parametric representation of the heart was estimated from meshes that have been extracted from MR. The ultrasound image is generated by simulating attenuation, scattering, reflection and speckle based on the parameterized geometric model, which is instantiated using the ontologies. While this method is very appealing for generating an unlimited number of cases for different pathologies, a lot of work has to go into modeling, defining the ontologies and confirming that they are correct.

2.2.2. User Interface

In this section, we will discuss all issues related to user interfaces and user interaction. This includes the methods for tracking of the US probe, but also computation of haptic feedback and the output device. Some of the systems that are discussed can be seen in figure 2.3.

2.2.2.1. Input Devices

Many systems use a physical phantom of a patient, made of plastic or foam, and a phantom of an US probe, which is tracked. The most common choice for tracking is the use of electromagnetic (EM) tracking [Weidenbach et al., 2005, Stallkamp and Wapler, 1998, Heer et al., 2004, Terkamp et al., 2003, Magee and Kessel, 2005, Sun and McKenzie, 2008]. While for real US applications EM tracking can cause problems if magnetic material is present, it is usually possible to avoid magnetic material in simulators. In particular for applications where the US transducer is inserted into the patient, as in transesophageal echocardiography [Weidenbach et al., 2007], EM systems are the preferred way of tracking, as they do not suffer from the line of sight problem as optical tracking systems do. [Markov-Vetter et al., 2009] used optical tracking to estimate the pose of a phantom of a newborn and an US probe. One advantage of optical tracking systems compared to EM systems is that the tracking volume is larger. This is important if additional objects have to be tracked, such as a HMD for AR.

An alternative to physical phantoms are haptic devices, as used by [d'Aulignac et al., 2005, Alterovitz et al., 2003b, Abolmaesumi et al., 2004, Bürger et al., 2008, Ni et al., 2008, Goksel and Salcudean, 2010, Forest et al., 2007, Vidal et al., 2008, Reichl et al., 2008]. One advantage of haptic feedback devices is that they can simulate different tissue such as bone and soft tissue for different patients. Also for procedures such as needle insertion they are superior to a phantom, as inserting the needle into a physical phantom will destroy the phantom when used often. Another advantage is that the amount of pressure that has been used by the examiner can be measured by the haptic feedback device. The information about the applied force can be used to simulate the deformation of the image. In some applications such as diagnosis of thromboses or carotid stenosis it is important to apply a certain amount of pressure. For such applications it is crucial to estimate the exact force that is applied in order to allow realistic simulation and provide feedback to a trainee if too much or too little pressure was applied.

Building a system using haptic devices is much more complex than using a phantom and a tracked probe, because it requires simulation of haptic feedback, which we will discuss in the next section. Haptic devices do only have a very limited working range and combining them with a physical phantom is difficult. Therefore often only the haptic device without a phantom of a patient is used, which requires more abstraction from the user. Devices with 6 degrees of

(a) Haptic feedback device and standard monitor [d'Aulignac et al., 2005].

(b) Haptic device with stereoscopic overlay [Vidal et al., 2005].
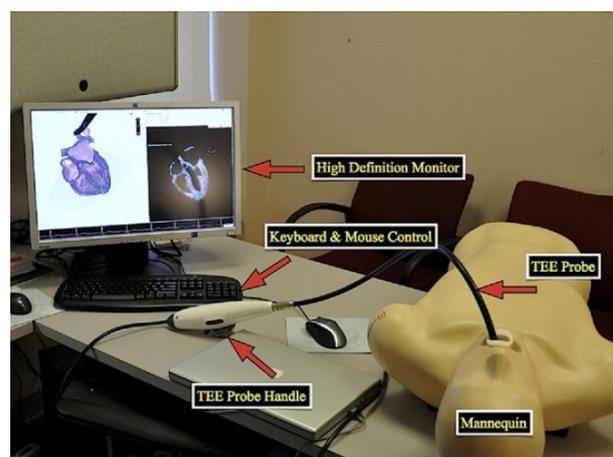
(c) Transesophageal simulator with phantom [Bose et al., 2009].

Figure 2.3.: Different user interface setups of US simulators.

`pie.med.utoronto.ca/TEE`	Illustration of TEE standard views. This system does not simulate ultrasound images but it shows prerecorded videos. However, it is a very good example of using visual models.
`www.teachivus.com`	Structures in IVUS images can be measured and compared to expert measurements. There is no simulation of ultrasound images, but only fixed 2D slices.
`www.ct2tee.agh.edu.pl`	TEE simulator showing different standard planes. The images are simulated from CT.

Table 2.2.: Websites with interactive content related to ultrasound training (last accessed: Feb. 3, 2012).

freedom (DOF) are very expensive. While 3-DOF devices are relative inexpensive, they cannot simulate any force for rotation and the maximum amount of force these devices can produce it limited. Another problem is that the working range of haptic devices is limited. A smaller haptic device, such as the Sensable PHANTOM Omni (Sensable, Wilmington, United States), has a working volume of only 16*12*7cm [1]. The Sensable PHANTOM Premium 1.5, which is one of the devices with the largest working volumes, has a working volume of 38*27*19cm [2]. Although the larger devices offer a working volume which is reasonable to simulate placing the probe at one location on the patient, the working volume is not large enough to simulate a procedure where the probe has to be placed on different locations.

With the aim of building a low-cost simulator for developing countries [ap Cenydd et al., 2009] used controllers from the Nintendo Wii console, which contain accelerometers and a simple infrared tracking system, to control probe and patient position.

There are some systems that do not mimic a real US probe but use keyboard and mouse to navigate a virtual probe [Ehricke, 1998, Arkhurst et al., 2001, Kempny and Piórkowski, 2010]. While this has the huge drawback that training of hand-eye coordination is not possible, such systems are inexpensive and students can use them on their own PC or even over the Internet. Such systems are better suited as complement to standard training or in addition to using a simulator having a physical probe. Depending on how complex the simulation of the US images is, such systems can also be provided online, which makes them easily accessible for many people. For instance, Kempny and Piórkowski [Kempny and Piórkowski, 2010] showed a TOE simulator that is accessible over the Internet. Table 2.2 provides some examples of websites offering interactive content related to ultrasound simulation. These online resources can also be used to see some of the teaching concepts that will be discussed throughout this chapter.

2.2.2.2. Haptic Simulation and Image Deformation

One crucial, but very complex issue for realistic simulation is to provide realistic haptic feedback when using haptic devices. Another related issue is the deformation of the US images as a result of pressure applied to the patient.

---

[1] `http://www.sensable.com/haptic-phantom-omni.htm`, accessed May 3rd, 2012
[2] `http://www.sensable.com/documents/documents/Premium_1.5_6DOF.pdf`, accessed May 3rd, 2012

One of the standard methods to simulate haptic feedback is the use of proxy-based methods, which have been used for a US simulator by [Vidal et al., 2008]. While these methods are fast to compute, they do usually not simulate image deformation. Common methods for simulating tissue deformation are the mass spring model and the finite element method (FEM). [Alterovitz et al., 2003b, Alterovitz et al., 2003a] and [Goksel and Salcudean, 2009] used the FEM to simulate soft tissue deformation due to the force applied through the probe. D'Aulignac et al. used mass spring models [d'Aulignac et al., 2005] with deformation parameters obtained experimentally from a real thigh using force sensors. And [Magee et al., 2007] used a localized mass-spring model to simulate tissue deformation during needle insertion.

Another possibility to simulate image deformation for specific applications is to record the deformations during a real US procedure. Troccaz et al. presented such a method [Troccaz et al., 2000], where US images with and without pressure have been obtained and the images are interpolated between both volumes during runtime. Several points were labeled in both images and based on these points a model of the deformation using splines was built. During simulation, this model can be used to interpolate different amounts of pressure applied to the patient. The advantage of this method is that it does not require a physics-based simulation of the deformation, which would be computationally very expensive. In addition, effects that can currently not be simulated with other methods are captured by this method. However, only deformations that have been recorded can be simulated and such methods will only work as long as the amount of force and the direction of the force that is applied in the simulator is similar to the recorded force. Furthermore, the process of acquiring the images is very time-consuming.

Several systems use haptic devices to simulate the insertion of a needle. These systems need to simulate the forces between the needle and the tissue as well as needle bending. [Alterovitz et al., 2003b, Alterovitz et al., 2003a] proposed a method taking into account the cutting at the needle tip, membrane puncture and friction. [Ni et al., 2008] simulated needle insertion, taking into account pre-puncture forces, friction forces and cutting forces. [Magee et al., 2007] simulated the bending of the needle and [Vidal et al., 2008] used a model that combines a proxy-based approach when the needle is outside the patient body and a method using the CT intensities and in vitro measurements when the needle is inside the body.

Realistic real-time simulation of haptics and image deformation is still a challenging problem. In particular when the simulated US images have to be computed at the same time and when a 3D visualization has to be generated. Furthermore, the simulation of the haptic feedback has to run with very high frame rates of about 1 kHz. As the visualization cannot be run at such high frame rates, methods have to be implemented to synchronize visualization and haptic rendering.

2.2.2.3. Output Devices

The requirements for the output device depend on the type of the input device. All systems have to present at least the 2D US slice to the user. The natural way to do this is to use a standard monitor as done by a real US device. Systems that do not use a physical phantom of the patient have to visualize relative poses of the probe and the patient on a screen. This can be done by showing a 3D scene on a monitor, including a virtual representation of the patient and the probe. The drawback of using standard monitors for this is the reduced immersion, since the user has to look at the virtual probe, which is not co-located in space with the haptic device that is used to control the probe. [Vidal et al., 2005] used a semi-transparent mirror

and shutter glasses to augment a visualization of the patient into the workspace of the haptic devices to solve this problem. This system can be seen in figure 2.3(b).

One advantage of computer-based US simulators is that information which cannot be seen during a real US procedure can be presented to the user. It is possible to show bones and organs in order to improve the understanding of spatial relations. Furthermore, simulators can show guidance or additional information related to organs. The different ways to show such information can be categorized according to the virtuality continuum [Milgram and Kishino, 1994], into augmented reality and augmented virtuality (see figure 1.6). Most US simulators that provide such information use AV, where a virtual representation of the scene is augmented by real information. For US simulators, the real information is the location of the tracked US probe. Usually such an AV scene is shown on a standard monitor, but also stereo monitors can be used as done by [Berlage et al., 1996] for showing a virtual representation of the heart and the US plane. In AR systems, the reality prevails and only some virtual information is augmented into the real scene. For instance, the system by Vidal et al. [Vidal et al., 2005] that has been mentioned before is an AR system. The system allows an AR visualization of bones and other key organs inside the patient. Guirlinger has presented a simple AR simulator for prenatal diagnosis [Guirlinger, 2007]. He uses a webcam to track markers that are attached to phantoms of a probe and a patient. The tracking data is used to overlay the geometry of the US plane onto the camera image.

2.2.3. Medical Applications

Many simulators were developed for cardiology. US simulators have been used in transthoracic echocardiography (TTE) to get a better understanding of the heart anatomy and to practice standard slices by [Berlage et al., 1996, Weidenbach et al., 2000, Weidenbach et al., 2005, Weidenbach et al., 2009] and [Sun and McKenzie, 2008].

While in most US procedures the probe is placed on the skin of the patient, there are also some applications in which the transducer is inserted into the patient. While the insertion of the probe is usually not invasive, training on patients is still problematic, as it is very uncomfortable for the patient. Due to the conflict between providing optimal treatment to patients and training of medical staff, there is an ethical imperative to perform simulator-based training [Ziv et al., 2003]. One example is echocardiography which can be performed transesophageal (TOE), through the esophagus, which is complicated as the examiner cannot see the head of the transducer. [Weidenbach et al., 2007], [Bose et al., 2009] and [Kempny and Piórkowski, 2010] presented simulators for TOE.

Another challenging field for training on real patients is prenatal diagnosis. Many findings only occur rarely, and therefore it is difficult to practice them on real patients. In particular for anomalies, it is delicate to use patients for training, as this is a stressful situation for them. Forest et al. developed a simulator for obstetrics [Forest et al., 2007] and [Maul et al., 2004] showed a simulator for prenatal diagnosis. Later they extended the same system to allow fetal echocardiography, where the 2D images are extracted from 4D US volumes to allow simulation of a beating heart [Wüstemann et al., 2008]. Simulators for neonatal cranial US were developed by [Markov-Vetter et al., 2009] and by [Arkhurst et al., 2001].

US simulators are important in particular for invasive procedures such as ultrasound-guided needle insertion where training on humans is problematic. Another example is radioactive seed implantation for prostate brachytherapy [Alterovitz et al., 2003b]. Simulation of US-guided needle insertion has been shown for biopsies [Ni et al., 2008, Magee and Kessel, 2005,
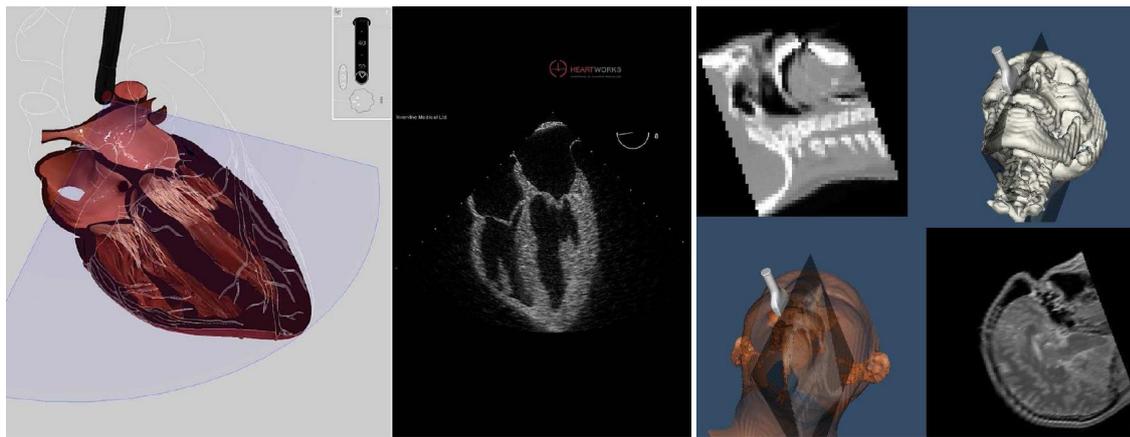
(a) Colocation of US slice and a model of the heart [Bose et al., 2009].

(b) Colocation of US image and CT/MR data from Visible Human [Tahmasebi et al., 2007].

Figure 2.4.: Different visualization methods that have been used in US simulators

Vidal et al., 2008] and hepatic biopsies [Forest et al., 2007]. Simulators for transrectal US-guided biopsy of the prostate have been proposed by [Sclaveranoa et al., 2009] and [Goksel and Salcudean, 2010]. Another invasive procedure that has been simulated is radiofrequency thermal ablation [Forest et al., 2007].

Simulators for gynecology were developed for ultrasound mammography [Marquardt et al., 2007] and a system for simulation of transvaginal US diagnosis has been developed by [Heer et al., 2004]. Simulators for radiology include systems which simulate IVUS [Abkai et al., 2007] and deep venous thromboses examination of the lower limb [Troccaz et al., 2000, d'Aulignac et al., 2005]. [Terkamp et al., 2003] showed a simulator that covers different pathologies in the abdominal area. Furthermore, a system for training longitudinal endoscopic ultrasound of the gastrointestinal tract [Bommersheim et al., 2005] has been developed.

2.3. Training Concepts

The main advantage of computer-based simulators is not to replace traditional training, but to offer new ways of training a student. In order to understand the advantages US simulators can offer, we will first discuss the need for a mental model. This discussion is based on work by [Trochim, 2002] and [Berlage, 2008]. Building up a mental model of complex relations and situations is a central part of learning. Models direct our actions, and therefore only adequate models can lead to adequate actions. Such mental models are constructed by interaction. We have a mental model that controls our actions and we can observe the outcome of our actions. Therefore, we are also able to check our mental model, reject or correct it. While for easy tasks, trial and error might be enough to build up a correct mental model, complex tasks such as the use of US require a transfer of knowledge from an expert to the trainee. One common problem for all learning tasks is how to transfer a mental model of an expert to a trainee. The most common method to do this is writing and reading books. However, it is difficult to use books to transfer a mental model including knowledge about anatomy, spatial relations and US physics. The main advantage of US simulators is that they provide new ways for a student

to build up a mental model.

In order to be able to assess the current state of US simulators it is also helpful to have a look at experiences with other simulators. [Issenberg et al., 2005] conducted a literature review and synthesized existing evidence from 109 studies regarding the features of high-fidelity medical simulators that lead to most effective learning. The most important feature of a simulator is to provide feedback, which conforms to the concept of a mental model that has to be updated. Other important features are, in order of their importance, repetitive practice, curriculum integration, range of difficulty levels, multiple learning strategies, capturing clinical variation, controlled environment and individualized learning. We will come back to some of these points when discussing the different training concepts that can be realized using an US simulator.

2.3.1. Visual Models

One of the main problems for beginners is the mental mapping between the 2D US plane and the 3D anatomical structures. Many novices mix transducer rotation with angular movements and fail to combine a number of consecutive images during an US sweep into a 3D structure [Berlage et al., 1996]. In echocardiography, the most difficult aspects to learn are the relationship between 2D images and 3D heart anatomy and the adjustment of standard planes [Weidenbach et al., 2005]. In particular for procedures that do not use a hand held probe, such as endoscopic or intravascular US, it is very difficult for a trainee to understand the spatial relations, as there is a lack of context and usually a very small field of view.

Visual models that show these relations in a 3D scene are a natural way to provide this information and help a student to build a mental model of the anatomy. Such a model might be very realistic, but it can also be abstract. Some examples of such 3D scenes are shown in figure 2.4. In order to transfer the mental model of an expert to a trainee, the expert might decide that less important information is completely hidden and more important information shown in more detail. Many simulators use 3D scenes to provide such a visual model to a trainee by showing the US scan plane co-located with additional 3D data such as the skin of the patient, bones, or other relevant organs. Some systems simply show the surface of the patient/phantom [Stallkamp and Wapler, 1998, Aiger and Cohen-Or, 1998, Ehricke, 1998]. Others show more detailed models of the anatomy [Goksel and Salcudean, 2010, Zhu et al., 2007] and allow changing visualization parameters [Weidenbach et al., 2000, Ni et al., 2008]. A visual model can also be enhanced by textual information. [Arkhurst et al., 2001] and [Bose et al., 2009] segmented and labeled anatomy. Later [Arkhurst, 2005] used a more advanced model where segmented organs were connected to a knowledge base that includes relations between organs.

2.3.2. Co-registered Images from other Modalities

Understanding an US image can be very difficult for a trainee due to bad image quality and artifacts. The mental model of using US must also include a mapping between the real structures that are seen in the image and their appearance in the US image. For other imaging modalities, such as CT or photographic slices, it is much easier to understand what is seen in an image. In US it is very difficult to build up a mental model about the appearance of structures in the image due to bad image quality and as only surfaces are shown in the image. One way to aid in building up a mental model is showing an US slice and the corresponding slice from another modality. [Ehricke, 1998] registered US volumes to CT/MR/photographic

slices from the Visible Human Dataset. Using a mouse, US slices and corresponding views from the VHD can be taken. Co-registered MR slices have been used by [Arkhurst et al., 2001]. [Tahmasebi et al., 2007] proposed a similar system. They use a haptic feedback device to move the US probe, while also showing corresponding slices from the VHD, which can be seen in figure 2.4(b). Co-registered images could also be presented printed or on a screen without using a simulator. However, it has been shown that simultaneous training of cognitive and motor skills enables faster learning than training of only one aspect [Kahol et al., 2009].

2.3.3. Feedback

The training concepts *Visual Models* and *Co-registered Images from other Modalities* can help to correct mental models only up to a certain degree. By showing additional spatial information, they can help building up an adequate model about the anatomy and spatial relations. However, they cannot provide any feedback on more complex actions such as how to reach the correct view of a structure or whether a student carried out a procedure correctly. One way to provide such feedback is to let a novice perform a task and provide record and replay capabilities. This can be used to identify and discuss errors done by the trainee. Many simulators provide a virtual view where the US plane and 3D anatomy is shown. During a discussion the expert can use this visual model to better explain his mental model. Systems with record and replay capabilities have been shown by [Berlage et al., 1996, Zhu et al., 2007]. Feedback can also be provided by showing how an expert performs the same procedure. This feature has been implemented by Aiger and Cohen-Or [Aiger and Cohen-Or, 1998]. More advanced methods to provide feedback will be discussed later in chapters 3 and 4.

2.3.4. Different Difficulty Levels

Another aspect that has been found very important for medical simulators is to provide a range of difficulty levels [Issenberg et al., 2005] for users with different skills. This is also important to provide an individualized education, where the level of knowledge of the student is considered. An example of this has been shown in [Weidenbach et al., 2000] where different visualization aides can be used in an echocardiographic training scenario. The system can show the relative position of the probe and the anatomy and it can visualize additional outlines of the target organs on the scan plane to aid beginners. For more advanced users these aides can be switched off. Similar, [Forest et al., 2007] offers different difficulty levels, where a transparency mode is used for less experienced users to improve understanding of spatial relations. An edutainment game offering different stages with varying difficulty for teaching US skills has been presented by [Chan et al., 2010].

2.3.5. Range of Different Cases

One reason for using computer-based simulators is the lack of appropriate cases in traditional training. A study by [Costantino et al., 2003] showed that the number of scans performed during the residency year is more important than the number of didactic hours. When using simulators it is possible to make a wide range of cases available to students, including rare cases. For instance, there are fetal abnormalities that only occur in one of 200,000 to 400,000 pregnancies [Lee et al., 1999]. Even if a patient with a rare abnormality would be available, training of a high number of trainees is delicate for prenatal abnormalities, as the mother is subject to

high mental stress in such a situation. Due to the lack of training cases, detection rates are low. The reported values range from 14.2% [Lys et al., 1989] to 30.3% [Queisser-Luft et al., 2002] and 44.5% [Levi et al., 1995]. Another application where appropriate cases are not easily available for training is diagnosis of trauma patients, where US is e.g. used for diagnosis of internal bleedings.

The UltraSim system [Aiger and Cohen-Or, 1998] offers a range of different cases including patient history, patient medical information and diagnostic data. A case database has also been built by [Ehricke, 1998]. The Sonotrainer system has a range of cases, each consisting of a 3D US volume, an assignment and answers to the assignment [Maul et al., 2006]. Different modules with certain types of cases are available for prenatal diagnostic, gynecology and breast diagnosis so that e.g. a student can start with normal findings, continue with common malformations and then advance to uncommon malformations. A parametric representation of the heart combined with an ontology has been proposed [Reis et al., 2006], such that an arbitrary number of different cases can be created.

2.3.6. Standardization of Training

Today, there is a low level of standardization in US training. US simulators could be used in a highly standardized learning curriculum, by providing a defined set of cases. In medical simulation, there is also a high interest in methods for automatic assessment of skills, which can also lead to higher standardization of education outcomes. Standardization of training and assessment of students is a crucial issue in order to guarantee high quality in US diagnostic. And as discussed in chapter 1, there is a demand for standardized outcomes. One problem when training on patients is that every patient is different and therefore equal conditions for training and assessment cannot be guaranteed. Simulators allow different students to train on exactly the same anatomy. Therefore, it is possible to provide every student with a standardized training. Also for assessing students, simulators have advantages, as objective criteria for evaluation can be defined. [Monsky et al., 2002] used a simulator for training of residents. To assess them before taking overnight calls, trainees had to perform measurements that have been compared to measurements of experts that were performed on the same patient data.

2.3.7. Invasive Procedures

For most applications US is less invasive than other imaging modalities such as CT, which uses ionizing radiation. Therefore, for many applications it is possible to train on patients or healthy subjects. However, there are also some invasive or semi-invasive applications. Procedures such as transesophageal, transvaginal or transrectal US can be uncomfortable and painful for the patient, which makes it difficult to practice them. This is even more problematic for some US-guided procedures that are invasive and where training on healthy patients can potentially harm the patient. Examples for this are needle biopsies or radioactive seed implantation. For these applications, simulators have huge advantages regarding patient comfort and safety.

2.4. Commercial Systems

Among the systems that have been discussed before there are three commercially available systems:

- The UltraSim system by MedSim (UltraSim, Ft. Lauderdale, United States), which has been presented in [Aiger and Cohen-Or, 1998].

- The Schallware system (Schallware, Berlin, Germany), which has been discussed in [Wüstemann et al., 2008, Maul et al., 2006, Maul et al., 2004, Marquardt et al., 2007, Staboulidou et al., 2010] and which can be seen in figures 2.2(c) and 2.3(a).

- The HeartWorks system by Inventive Medical (Inventive Medical, London, United Kingdom), which was shown in [Bose et al., 2009] and can be seen in figures 2.3(c) and 2.4(a).

In addition, there are other commercial systems that have not been described in the scientific literature. A summary of these systems is shown in Table 2.3.

We now briefly discuss these systems regarding the classification from section 2.2 and regarding the training concepts. The Vimedix and the HeartWorks systems are made for teaching of TTE and TEE. Both use model-based generative image simulation, since imaging of the heart requires a very detailed simulation, which can be achieved by using computer models. The UltraSim, ScanTrainer and Schallware systems use interpolative methods. One system, the SonoMom, shows prerecorded 2D ultrasound images, depending on the position of the probe in the simulator. This has the disadvantage that only the position but not the orientation of the probe is considered, however this solution does not require expensive hardware to track the orientation of the probe. While many research systems use image-based generative simulation, none of the commercial systems uses this technique. One possible reason for this is that these methods require very powerful hardware and most of the related research has only been published within the last years.

For the user interface, all commercial systems to date use standard monitors and no system uses technologies such as stereo monitors or augmented reality. Only the ScanTrainer uses a haptic device to control the probe. All other systems use phantoms of patients and ultrasound probes. Most commercial systems target applications where training on healthy patients is difficult. Both the Vimedix and the Heartworks systems provide training for TTE and TEE. Training for obstetrics can be done using the SonoMom or the ScanTrainer systems. Both the Schallware and the UltraSim systems provide different modules for a range of different medical applications.

Unlike the majority of research systems, most commercial systems provide case databases for training, often with different difficulty levels. Also the training concept of visual models is used by several systems. Both systems for training of TTE and TEE are based on computer models and use these models for visualizing the relative pose of the probe and the ultrasound plane with respect to the anatomy.

System	US image simulation method	User interface	Medical application	Training concepts
MedSim UltraSim	3D ultrasound	phantom + tracked probe	various	expert replay, different cases
SimuLab SonoMom & Trauma-Man F.A.S.T. module	2D ultrasound	phantom + tracked probe	obstetrics / FAST	expert examples, different cases
HeartWorks	computer model	phantom + tracked probe	TTE / TEE	visual model
MedaPhor ScanTrainer	3D ultrasound	haptic device	obstetrics	visual model, performance metrics, different cases
CAE VIMEDIX Ultrasound Simulator	computer model	phantom + tracked probe	TEE, TTE, FAST	visual model, different cases
Schallware	3D / 4D ultrasound	phantom + tracked probe	various	different cases

Table 2.3.: Summary of key aspects of commercial systems.

2.5. Evaluation

Several aspects of a simulator have to be evaluated. Face validity judges the degree of resemblance between simulator and reality and can be evaluated by questionnaires or by performing the same procedure on real patients and on a simulator. Content validity shows how appropriate a system is for teaching and to which extend it covers the subject matter of the real activity. Content validity is usually assessed by questionnaires. Construct validity is the ability of a simulator to discriminate between subjects of different ability or experience. This can be done e.g. by letting subjects of different experience perform the same task and measure criteria such as time or correctness of a diagnosis. Another way to evaluate a simulator is to use pre- and post-tests to see if using the simulator improves the performance of trainees. A simulator can be compared to traditional training methods by separating the subjects into two groups, one receiving simulator and one traditional training, and comparing the performances of both groups in a post-test.

An overview on different studies that have been published can be seen in table 2.4. Below the results of the different studies are discussed in detail. A summary of the most important results is provided in the next section.

Weidenbach et al. conducted a series of evaluations on transthoracic and transesophageal US. In [Weidenbach et al., 2000] they compared a group receiving standard training using illustrations, atlases and text to a group using a simulator to train standard views in TOE. The group receiving simulator training could adjust the standard views faster and more confident. In a subsequent evaluation [Weidenbach et al., 2005], 25 students participated in a training course where the simulator was involved for explanation and hands-on training. Participants had to fill out a questionnaire that showed high content validity and medium face validity. In a later study, the same simulator was integrated into a TOE course for anesthetists [Weidenbach et al., 2007]. This study reported positive feedback for face and content validity. A later study by [Weidenbach et al., 2009] found average ratings of 4.18 on a scale between 1 (worst) and 5 (best) on different questions on face validity and 4.49 on content validity. Construct validity was shown by presenting several cases to medical doctors with different skill levels. Experts had a mean performance grade of 0.98, intermediates of 0.69 and beginners 0.44, where 1.0 was the highest performance grade that could be achieved. The differences between the groups were significant.

The commercially available UltraSim system has been evaluated for different applications. [Knudson and Sisley, 2000] integrated the simulator into a training course for the focused abdominal sonography for trauma (FAST) protocol. Using a pre-test and a post-test the learning effect of using the simulator was compared to using phantoms or real patients. The results of both groups in the post-test did not show a significant difference. Training of FAST using the UltraSim system was also evaluated by [Salen et al., 2001]. Their course involved training on normal phantoms, peritoneal dialysis phantoms and the UltraSim system. 20 emergency medicine residents and 10 physicians rated how helpful the different training methods are. While 85% of the participants found the computer simulator somewhat helpful or very helpful, the results for the normal phantoms and the peritoneal dialysis phantoms were better, whereas the difference was not significant. The residents, who did not have previous FAST experience, were assigned to two groups, one receiving training with the simulator and one receiving training with the peritoneal dialysis phantom. A FAST interpretation post-test did show similar results for both groups without significant difference.

Face validity was shown by [Terkamp et al., 2003] for a system that simulates abdominal

US. Eleven physicians had to scan real patients and simulated patients with a range of different pathologies. On the real patients, 75% of the pathological findings have been identified whereas on the simulators 71% have been identified. The participants were a little bit more confident with their findings on real patients (68% compared to 64%) and the time required for the scanning was 10.57 min on real patients and 9.59 min on the simulator. None of the differences is statistically significant. However, the handling of the ultrasound machine (74%) was judged significantly better than the handling of the simulator (61%).

A series of evaluations was performed during development and use of the Sonotrainer system, a simulator for prenatal examinations. To show face validity, ten ultrasound simulations of normal scans and scans from abnormal fetuses were shown to seven specialists who had to diagnose the cases. The sensitivity for this task was 86% and the specificity 100% [Maul et al., 2004]. Face validity and content validity have been evaluated by a questionnaire involving 24 physicians. They rated image quality good (80%) to moderate (20%) and the training effect good (94%) to moderate (4%). Additionally, the performance of a group of 24 physicians that only received textbook training on measurement of nuchal translucency thickness (NT) and crown-rump length (CRL) was compared to a group of 21 physicians who received additional simulator training [Wüstemann et al., 2002]. After the training, both groups had to take measurements on real patients. Their results were compared results of experts, and the group that received additional simulator training performed significantly better. The mean absolute deviation for the measurement of NT was 0.31 and 1.48 for CRL for the group receiving simulator training. The group that only received theoretical training had a mean absolute deviation of 0.62 for NT and 3.27 for CRL.

In another study by [Maul et al., 2006], different cases were shown to nine experts and pathologies were identified with a sensitivity of 83% and a specificity of 100%. The image quality was rated good or medium in 97% of the cases. The system was integrated in a training module that was used to train 162 gynecologists. While 83% of them found the training module useful, only about 50% were satisfied with the image quality. After this study, the system was used in more than 50 courses to teach measurement of the nuchal translucency thickness where almost 2000 gynecologists participated. Questionnaires from a subset of the participants (N=153) showed that more than 90% found the whole module helpful, more than 80% found the integration of the simulator into the training course useful and more than 80% were satisfied with the image quality. In another large-scale study over a period of 19 month, 1266 gynecologists participated in training courses that made use of US simulators. They had to fill out questionnaires before and after the training course and the number of correct answers was used to show a training effect. Before the course, 51.4% of the questions were answered correctly and after the course, 75.3% were correct.

[Magee and Kessel, 2005] identified a set of metrics that can be used to differentiate between experts and trainees for US-guided needle insertion. Ten novices performed two training sessions with a gap of one week. Most of the students showed an improvement of the metrics. However, as variance was high and sample size low, only for few metrics there was a statistical significance. Furthermore, their system did not use simulated US, but only CT slices. A later version of the simulator, which used simulated US, was evaluated by 60 experts using a questionnaire [Zhu et al., 2007]. On a scale between 1 (worst) and 5 (best), questions related to face validity received average ratings between 3.41 and 3.80, with the only exception of tactile feedback where the score was 2.61. Questions related to content validity got average scores between 3.62 and 4.37. A subsequent study by [Magee et al., 2007] compared performance of consultant radiologists, radiology registrars and students on a version of the same simulator

enriched by simulation of soft tissue deformation. Ten different metrics were measured for each group. Significant differences could be found for eight metrics between consultants and students, for four metrics between consultants and registrars and for one metric between registrars and students.

A simulator for US-guided biopsy was also evaluated by [Ni et al., 2009]. Similar to Zhu et al. [Magee et al., 2007] they defined a set of metrics on the relations between needle tip, scan plane, entry plane and target. Four experts and 12 novices performed the biopsy six times each. For four out of six metrics a significant difference between experts and trainees could be shown. They also performed a study on the learning curve when using the simulator. Eight novices had to perform ten biopsies, where the time and the metrics were compared. Results for time and one metric have been presented, and both showed an improvement, however results for the other five metrics were not discussed.

Another simulator for US-guided needle insertion was presented by [Vidal et al., 2008]. The simulator was presented to inexperienced trainees who rated several aspects of the system using a questionnaire. Several questions related to content validity were answered with average ratings between 5.4 and 6.35 on a scale between 1 (worst) and 7 (best) and questions related to face validity were rated with an average between 2.58 and 2.84 on a scale between 1 and 5. As already observed for other evaluations, the realism of the force feedback received the lowest scores.

The use of a TOE simulator for training was evaluated by [Bose et al., 2010]. Two groups, each consisting of seven residents without TOE experience did written pre- and post-tests. In between, one group received simulator training and the other group traditional training. The results for the simulator group in the post-test were significantly higher than for the group that received traditional training.

2.5.1. Summary and Discussion of Evaluation

Multiple studies have investigated content and face validity by carrying out questionnaires with experts and trainees. Content validity was rated high to very high and face validity was rated medium to high. Details are discussed below.

The face validity for the realism of the simulated US image [Weidenbach et al., 2009, Weidenbach et al., 2007, Maul et al., 2004, Maul et al., 2006, Zhu et al., 2007, Weidenbach et al., 2009] and for the realism of the whole US procedures [Weidenbach et al., 2005, Weidenbach et al., 2009, Terkamp et al., 2003, Zhu et al., 2007, Vidal et al., 2008] has been rated medium to high in different questionnaires. Further questionnaires on face validity for the handling of the probe showed high ratings for using physical dummies [Weidenbach et al., 2007], but low ratings for using haptic feedback devices [Vidal et al., 2008, Zhu et al., 2007]. Face validity can also be shown by letting a user perform the same procedure on a simulator and on real patients and comparing the results. Two such studies showed medium [Heer et al., 2004] to high [Terkamp et al., 2003] correlations between real and simulated diagnoses. High specificity and sensitivity for performing a diagnosis on a simulator using US from real patients has been reported by [Maul et al., 2004, Maul et al., 2006].

Reference	US image simulation method	Medical application	Population	Aspect	Results
[Weidenbach et al., 2000]	3D ultrasound	TTE	no information	comparison to standard training	better performance for simulator group
[Weidenbach et al., 2005]	3D ultrasound	TTE	25 students	face validity	40% realistic, 40% neutral
				content validity	$> 90\%$ agree or strongly agree
[Weidenbach et al., 2007]	3D ultrasound	TOE	25 experts	face validity	~75% realistic or very realistic
				content validity	~93% agree or strongly agree
			31 novices	content validity	~98% agree or strongly agree
[Weidenbach et al., 2009]	3D ultrasound	pediatric TTE	12 experts, 16 intermediates, 15 beginners	face validity	different questions were rated with an average of 4.18 on a scale between 1 (worst) and 5 (best)
				content validity	different questions were rated with an average of 4.49 on a scale between 1 (worst) and 5 (best)
				construct validity	significant difference between experts, intermediates and beginners
[Knudson and Sisley, 2000]	3D ultrasound	FAST	74 residents	comparison to phantom and patient-based training	no significant difference
[Salen et al., 2001]	3D ultrasound	FAST	20 residents	comparison to phantom-based training	no significant difference
[Heer et al., 2004]	3D ultrasound	gynecology	25 experts	face validity	same diagnosis and measurement as in reality for 2 of 3 cases

				construct validity	measurement comparable to measurements of experts after simulator training
[Terkamp et al., 2003]	3D ultrasound	abdomen	24 students		
[Maul et al., 2004]	3D ultrasound	prenatal	11 residents	face validity	percentage of pathologies found on real patients not significantly different to simulated patients
			45 physicians	additional simulator training to textbook training	significantly better in doing measurements
			24 physicians	face validity	80% rated image quality good, 20% moderate
				content validity	94% rated training effect good, 4% moderate
			7 experts	face validity	86% sensitivity and 100% specificity for identifying abnormalities
[Maul et al., 2006]	3D ultrasound	prenatal	9 experts	face validity	83% sensitivity and 100% specificity for identifying abnormalities
					image quality rated good in 87% of the cases, medium in 10% and bad in 3%
			162 gynecologists	face validity	50% were satisfied with image quality
				content validity	85% agreed that the training module including the simulator is helpful
			153 gynecologists	content validity	≥90% agreed that the training module including the simulator is helpful
					≥80% agreed that the simulator is helpful
				face validity	≥80% were satisfied with the image quality
[Staboulidou et al., 2010]	3D ultrasound	prenatal	1266 gynecologists	pre-/post-test	48.6% questions answered correctly before, and 75.3% after training course

[Magee et al., 2007]	CT-based	needle placement	60 experts	face validity	different questions were rated with an average of 3.38 on a scale between 1 (worst) and 5 (best)
				content validity	different questions were rated with an average of 4.02 on a scale between 1 (worst) and 5 (best)
[Zhu et al., 2007]	CT-based	needle placement	8 students, 8 registrars, 8 consultants	construct validity	can discriminate between consultants and students and to some degree consultants and registrars
[Vidal et al., 2008]	CT-based	needle placement	20 trainees	content validity	different questions were rated with an average of 4.51 on a scale between 1 (worst) and 7 (best)
			19 trainees	face validity	different questions were rated with an average of 2.67 on a scale between 1 (worst) and 5 (best)
[Ni et al., 2009]	3D ultra-sound	needle placement	4 experts, 12 novices	construct validity	can discriminate between experts and novices for 4 of 6 metrics
[Bose et al., 2010]	computer model	TOE	14 residents	comparison to traditional training	median score of 64 for simulator training compared to 29 for traditional in a written post-test

Table 2.4.: Summary of different evaluations. Some of the results presented here summarize several questions that have been answered by the subjects.

While face validity should be taken into account when developing a simulator, it is not necessarily required to build a simulator that leads to a good training effect [Feinstein and Cannon, 2002]. For some domains, it even might be beneficial to employ a more abstract representation of the reality, which allows putting more emphasis on the important aspects, as discussed before in the context of mental modeling. Therefore content validity of simulators is even more important than face validity. High to very high ratings were given in different questionnaires regarding whether simulators are useful for training [Weidenbach et al., 2007, Weidenbach et al., 2009, Heer et al., 2004, Maul et al., 2004, Maul et al., 2006, Zhu et al., 2007, Vidal et al., 2008]. In particular training of spatial relations, transducer steering and standard planes has been rated very high [Weidenbach et al., 2005, Weidenbach et al., 2007]. Questions regarding content validity of haptic feedback simulation have only been rated medium [Vidal et al., 2008].

Different studies reported medium to high levels of construct validity, which means that a simulator can discriminate between experts and non-experts [Ni et al., 2009, Magee et al., 2007, Weidenbach et al., 2009].

While values for content and construct validity are very high, there are still few studies that investigate how well simulators perform compared to traditional training. Some studies have shown that simulator training improves performance, but most studies did not compared it to training using real patients. It has been shown that additional simulator training improves performance compared to standard training [Maul et al., 2004]. Two studies reported that students receiving simulator training instead of a standard lecture using images and drawings showed better results [Weidenbach et al., 2000, Bose et al., 2010]. In a pre-/post-test, it was shown that simulator-based training leads to comparable results as training on real patients [Knudson and Sisley, 2000] and another study reported that no significant difference to non-computer phantoms could be found [Salen et al., 2001]. While these results are promising, further studies are required to establish well-founded evidence about the effect of simulator-based training of using US. In particular comparisons of long term training effects between traditional and simulator-based training are lacking.

2.6. Related Applications

There are some applications that are related to using US simulators for education, which can also benefit from using similar technology. In this section some of them will be discussed briefly. One of the main advantages of US simulators is that they can help to build up a mental model of an US procedure e.g. by showing 3D anatomy co-registered with the US slice. Other areas where similar methods are useful are teleconsultance and virtual re-examination. Unlike for other 3D or 4D imaging modalities it is usually not possible to do a diagnosis only based on the US image, as it is difficult to recover the spatial relations without holding the probe in the own hands. [Berlage, 1997] showed a teleconsultance system for echocardiography. The system uses the same method as a previously developed US simulator [Berlage et al., 1996]. The US slice is shown co-registered with a virtual model of the heart, which allows understanding the spatial relations. Using this additional communication channel can facilitate a common understanding in teleconsultance. One case where the teleconsultance system has been used was presented by [Wick et al., 1999]. They compared using the teleconsultance system to performing the same procedure with the remote expert in the same room and found that there was a similar information density.

Similarly, in virtual re-examination a 3D US dataset is obtained from the patient.

An MD can re-examine the same patient using a patient-based simulator as discussed by [Tahmasebi et al., 2008]. A similar system for telemedicine has been presented by [Heer et al., 2001]. They acquired 3D volumes using transvaginal US, transmitted the data to a remote computer, and performed re-examination on a phantom using a tracked transvaginal US probe. Measurements performed on the simulator were comparable to measurements done on the real patient. Such methods can be used e.g. in rural areas where not always an MD skilled in the use of US is present. A less skilled MD or a nurse could perform the ultrasound scan and the re-examination is performed remotely by an experienced sonographer.

Another application where US simulators can be used is repeatability measurement. It is very important to know how big errors are when the size of certain structures is measured using US. It is important to know the inter- and intra-observer repeatability in order to assess whether US can replace other imaging modalities such as CT for a measurement task. Using a simulator for repeatability measurements has some advantages. The exact size can be estimated from the 3D volume that is used for simulation, whereas for measurements on real patients the real size is unknown. It can be guaranteed that every subject has exactly the same conditions, whereas a real patient could move. As the corresponding 3D volume is available, it is possible to analyze the cause of measurement errors. When the volume of a structure should be measured and e.g. most participants used slices that are not perpendicular to each other, this problem could be identified and analyzed in a simulator.

A study on repeatability of nuchal translucency thickness measurement using simulated US has been performed by [Newey et al., 2003]. Using a simulator allowed them to do measurements on images with uniformly distributed NT thickness and angle and showed a significant correlation between repeatability and measured thickness, gain and measured thickness and between gain and repeatability coefficient.

Some of the US simulators that have been discussed use only one dataset. However, most of the systems are patient-based and allow a range of different cases. One future step could be to develop patient-specific simulators, where data from a new patient can be integrated very fast. The main use of patient-specific simulators is to do a dry run of a procedure, before it is carried out. This can be helpful for novice surgeons or for very complex cases or it could also be used in pre-operative planning. Such a patient-specific surgical simulation is discussed by [Soler and Marescaux, 2008]. While patient-specific simulators are very interesting, several problems arise. Any post-processing has to be done automatically as usually there is not much time between acquisition of a 3D volume of the patient and the use of the simulator. More important, simulation has to be very realistic. When the simulation does not resemble the real procedure very well, patient-specific training is of no use, and might even threaten the patient if a surgeon draws wrong conclusions from the simulation.

2.7. Technical Issues

Although the main technical issues for US simulators are appropriate generation of images and haptic simulation, there are some more problems that have to be solved. In this section, we will discuss some of the additional technical issues.

2.7.1. Registration

Registration refers to the problem of having two images that have to be aligned to each other. The images may be of same or different dimensionality, come from the same or dif-

ferent imaging modalities, may be from the same or different patients and may have total or only partial overlap. For an overview on different registration methods for medical images see [Hajnal et al., 2001].

Many simulators can simulate different patients but use only one physical phantom. To enable this, patient data must be registered to the phantom. The patient data usually consists of a 3D US or CT volume. One method to register the patient data to the phantom is to obtain a CT scan of the phantom and register the surface of the phantom with the surface of the patient. [Magee and Kessel, 2005] discuss one method of registering two point clouds representing the surface of a phantom and the surface of a patient for use in an US simulator. They use the iterative closest point (ICP) algorithm for initial alignment and estimate a quadratic mapping using RanSaC for an exact registration.

Some systems use two different 3D volumes such a CT and US. Other systems use a combination of volumes and parametric models. In both cases a registration is required. [Pieper et al., 1997] used 3D US volumes and a model of the heart. They used several landmarks to register both. Similarly, [Ni et al., 2009] registered US and CT volumes based on manually selected landmarks. They refined their registration using mutual information.

Another issue where registration is required is stitching of large US volumes. 3D US volumes are usually obtained from one viewpoint and therefore the volume is only of small size. Registering multiple volumes can increase the overall volume from which 2D slices can be extracted. To solve this problem, we have to register data from the same patient and the same modality, but with little overlap. Methods for registering multiple 3D volumes have been used for US simulation by [Ni et al., 2009] and [Aiger and Cohen-Or, 2000]. Further strategies for registering 3D US volumes, taking into account viewing-angle dependent effects in US images, have been discussed by [Wachinger et al., 2008, Wachinger and Navab, 2009].

2.7.2. Calibration

When 3D US volumes are used for the simulation they have to be recorded before. Some commercial US systems have a moving transducer array and allow recording 3D volumes. But also normal 2D systems can be used in combination with a tracking system to obtain 3D US volumes. One crucial issue for reconstructing a 3D volume from 2D images is accurate calibration of the US probe. The extrinsic and intrinsic parameters have to be estimated. Tracking systems provide the pose of one point. When using an optical tracking system this is usually the pose of one optical marker. When using magnetic tracking, this is the pose of the head of the magnetic tracker. The extrinsic parameters describe the transformation between the tracked point and the center of the US image. The intrinsic parameters describe characteristics of the US plane, such as field of view and depth. Different methods of calibrating a probe are discussed by [Mercier et al., 2005]. The next step for obtaining 3D volumes is the reconstruction from a set of 2D slices with known pose. A review of existing methods can be found in [Solberg et al., 2007].

Intrinsic and extrinsic parameters have also to be known for the virtual probe. However, here an accurate calibration is less important as for the acquisition of 3D volumes. While for reconstruction of 3D volumes a small error will already result in a wrong reconstruction, small errors are not as critical for a training simulator. As calibration is a one-time procedure, this can be done manually by defining the intrinsic parameters based on the characteristics of the US probe that should be simulated. The extrinsic parameters can be set by manually changing rotation and translation or by performing a tool tip calibra-

tion where the tip of an instrument is fixed and the tool is rotated around the fixed point [Magee and Kessel, 2005, Tuceryan et al., 1995].

2.7.3. Reslicing

Reslicing a 3D US volume and generating a 2D image in real-time has been a major technical issue for early simulators. Solutions for this are detailed e.g. in [Weidenbach et al., 2000, Troccaz et al., 2000]. Today reslicing is usually done by the GPU. As GPUs provide functionality for interpolating 2D slices from 3D volumes and modern GPUs are fast enough to handle big 3D volumes, reslicing is not a major problem anymore, as long as no image deformation has to be considered at the same time.

2.8. Discussion

While long term training effects have not been investigated yet, the first studies on computer-based ultrasound simulation are promising, and simulators are already used in regular training programs [Maul et al., 2006]. However, there are still many technological, educational and organizational issues that have to be addressed.

In order to see whether computer-based ultrasound simulators will be helpful it is important to compare them to traditional simulation using physical phantoms. In terms of image quality, computer-based simulators do not provide as realistic images as physical phantoms do. While interpolative simulation methods can provide very realistic images, they are not able to create very realistic images from every viewpoint. Generative methods do not provide highly realistic images as they have to use simplified models of ultrasound physics to run in real time on current hardware. Another aspect where traditional systems are better than computer-based systems is haptics and deformation. Evaluations of systems using haptic feedback devices have shown that current methods are not very realistic. Furthermore, such systems do only provide haptics for the ultrasound probe. Feeling anatomical landmarks with the hands is not possible. Some computer-based simulators use physical phantoms, which can be very realistic. Only few systems simulate the deformation of the ultrasound image, which is very important for some application.

While there are some drawbacks, computer-based simulators offer many advantages over traditional systems. One important advantage, which is provided by most commercial systems, is a set of different cases. A physical phantom can only represent one specific anatomy or pathology, while a computer-based simulator can provide many different cases and can by extended by a software update. Another advantage that is implemented in several systems is the use of visual models. This is in particular important for applications such as echocardiography where a very detailed mental model is required.

In the long term we believe that a big advantage of computer-based simulators will be that they can unburden the teachers. When using traditional simulators, always an experienced medical doctor must be present to provide feedback. Visual models can be used to help a student building a mental model where otherwise a teacher would have to teach the student. Advanced training systems can even provide feedback to a student. While the simulator is used, the computer can analyze how a student is performing. For simulators of minimally invasive surgery there has already been a lot of research on how to analyze the performance of a student with the goal of doing skills assessment and for providing feedback [Reiley et al., 2011].

Investigating the use of such methods will be very important to make full use of the possibilities of computer-based ultrasound training. Advanced methods to provide feedback and metrics to evaluate the skill of a student have a big potential, as they allow autonomous and competency-based training. The use of such methods for an US simulator will be discussed later in more detail in chapter 4.

Current systems still suffer from several technological shortcomings. Both interpolative and generative image generation methods do not produce highly realistic images. Complex artifacts such as multiple reflections and diffraction cannot be simulated in real-time. Additional research is required to develop more realistic simulation of ultrasound images in real-time. Another issue is the simulation of Doppler ultrasound, which is difficult both for interpolative and generative methods. First attempts to simulate Doppler ultrasound have been done by Khoshniat et al. [Khoshniat et al., 2005] and Hirji [Hirji, 2006], but they have not been used in a simulator for training. One problem for all medical simulators is haptics and deformation. Face validity for haptic simulation has been reported to be low. Haptic simulation is important for invasive procedures such as ultrasound-guided biopsy. This is a very promising application and developing better methods for haptic simulation is very important in order to build meaningful simulators for invasive procedures.

While solving those educational and technological problems, organizational and financial issues have to be taken into account as well. It should be investigated whether simulators are only efficient when used under supervision of an experienced sonographer, or whether it would allow students to perform training without supervision. We believe that simulators cannot replace a teacher, but they can reduce the time experts have to be present. Moreover, the prices of simulators have to be considered. They range from system that use haptic devices and tracking systems, which often cost more than a physical phantom, to very inexpensive systems that can be run on a standard PC without additional hardware. Expensive systems will only be reasonable for skills labs and training centers that are used frequently, or when simulators are integrated into training courses that are offered at multiple hospitals. On the other hand, systems that do not require additional hardware could even be offered for free. Also combinations should be taken into account. Skills labs could offer access to a simulator using a physical phantom and a probe, while additionally students could recapitulate and refresh their knowledge on a software version of the same simulator running on their own PC. Introducing such novel teaching methods that are only possible using computer-based simulation could help reducing the costs, while increasing the quality of training.

3. An Augmented Reality Ultrasound Simulator

In this chapter, an US simulator, which has been developed in the course of this thesis, is presented. As has been discussed in the last chapter, there is a wide range of existing research systems and even some commercially available US simulators. The main motivation for building the system that is presented in this thesis was to investigate new methods of using computer-based US simulators for education. Most previous systems focused on achieving a realistic simulation of US procedures, such that a simulator resembles a US procedure on a real patient very closely. Achieving a realistic simulation is very important and has also been a goal in the development of this system. However the main motivation of this work was investigation methods such as AR visualization that go beyond what could be done on a real patient.

On the next pages, the hardware and software architecture of the system will be discussed and different training concepts, mainly related to AR visualization and providing feedback will be presented. The contributions of this chapter are new concepts of using an US simulator for training. However, the main contributions will be presented in chapter 4, where the generation of medical workflow models will be explained and examples of using workflow models in the US simulator will be presented.

First, in section 3.1, the method for generating US images will be explained. In sections 3.2 and 3.3 two setups of the system will be described. One setup is located at the hospital of the Technical University Munich (Klinikum Rechts der Isar, Technische Universität München) and uses a flexible phantom and a webcam for AR visualization. The other one is located at the hospital of the University Munich (Klinikum Innenstadt, Ludwig-Maximilians-Universität München) and uses a HMD for advanced AR visualization. Training concepts that can be realized in this simulator setup will be described.

3.1. Simulation of US Images

As already discussed in chapter 2, there are different methods to simulate US images. The method that is used in the majority of the simulators is to reslice 3D US volumes. As we want to use advanced visualization methods, we have decided not to use 3D US volumes, as high-quality visualization from US volumes is not possible due to noise and artifacts. Instead we use a generative approach, where US images are simulated from CT volumes.

The simulation of the US images has not been developed in the course of this thesis, but is based on previous work. [Wein et al., 2007] developed a method to simulate US images from CT volumes with the application of registering CT with 3D freehand US. Such a registration can be used to register intra-operative US with a preoperative CT [Wein et al., 2008]. Methods for registration of medical images usually require a similarity measure and an optimizer, which optimizes over the similarity measure. When registering CT with US, a major problem is that the same structures can have a very different appearance in both imaging modalities. Therefore the definition of a reasonable similarity measure is not trivial. Even methods such as mutual
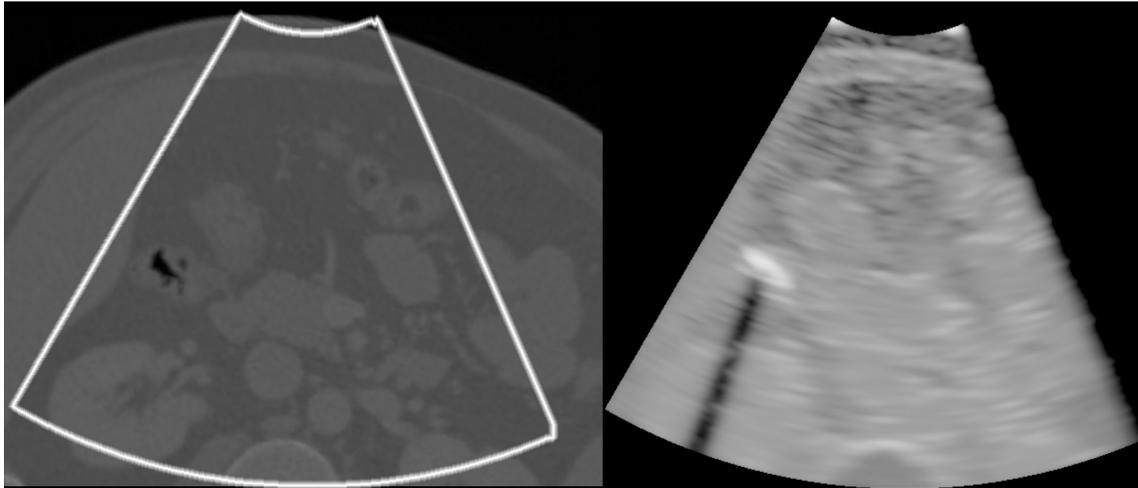
Figure 3.1.: On the left the source slice from the CT volume is shown with the geometry of the US image overlaid in white. The right image shows the corresponding simulated US image.

information, which provide a good similarity measure between different modalities such as CT and MR, have problems due to view-dependent artifacts in US images. Wein et al. proposed a new similarity measure that simulates an US image from the CT volume and compares this simulated image to a real US image.

For the simulation a correlation between Hounsfield units in CT images and acoustic impedance is assumed. Rays are cast through the CT volume and reflection and transmission are computed based on differences in Hounsfield values. A method for registering CT volumes with US images based on the simulated US images was proposed and evaluated in a study involving 25 patients [Wein et al., 2008]. The method could correctly register 76% of the cases.

Based on this work, [Shams et al., 2008] developed a more realistic simulation, with the motivation to use it for training. They extended the simulation model by Lambertian scattering which depends on the angle of incidence and the effect of beam width. Additionally, a precomputed image of speckle patterns was added. A GPU-based implementation was shown later by [Kutter et al., 2009]. The simulation problem was formulated as a ray-casting problem, which allows efficient parallelization on the GPU. This implementation allows real-time simulation of US images and is used in the system that is presented in the two following subsections. An example of an image simulated with this method is shown in figure 3.1.

3.2. Setup Using an External Camera

The setup described in this section uses a patient phantom and an optical tracking system. An external camera is used for simple AR visualization. The system is located at the IFL, which is an interdisciplinary research laboratory located at the university hospital of the Technical University Munich (Klinikum Rechts der Isar, Technische Universität München). The setup was implemented in cooperation with the department of surgery. A photo of the setup is shown in figure 3.2 and an illustration of all components and transformations can be seen in figure 3.3. In this section alternatives for the different components are discussed and the solution that was chosen for our system is described.
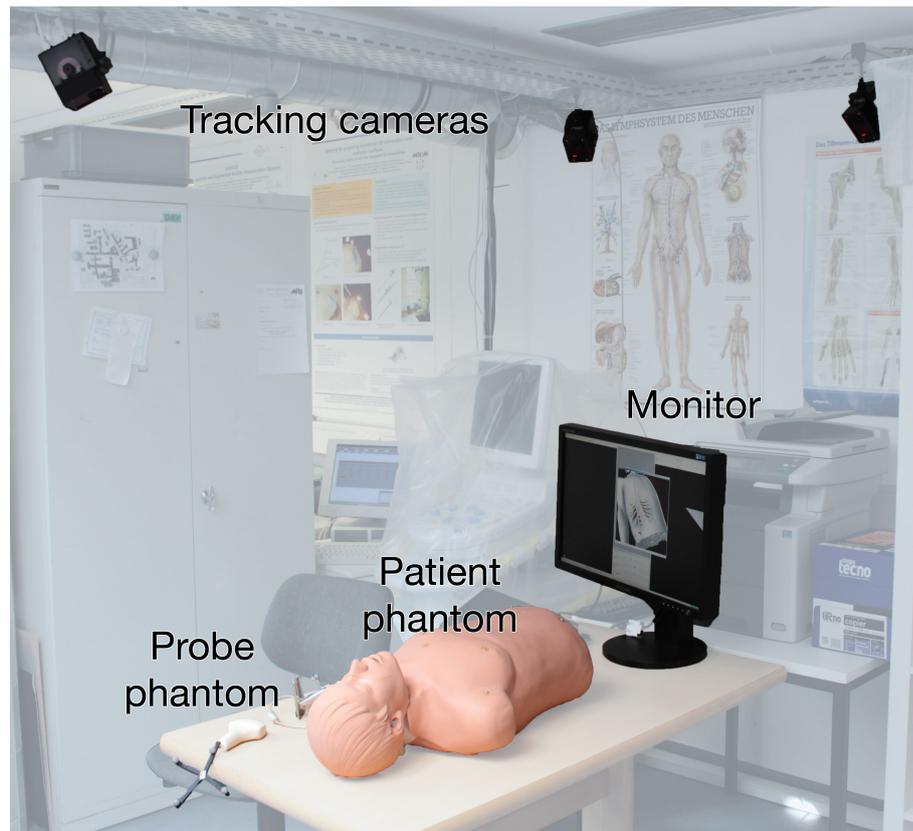
Figure 3.2.: Photo of the setup at the IFL at the university hospital Rechts der Isar, Munich.
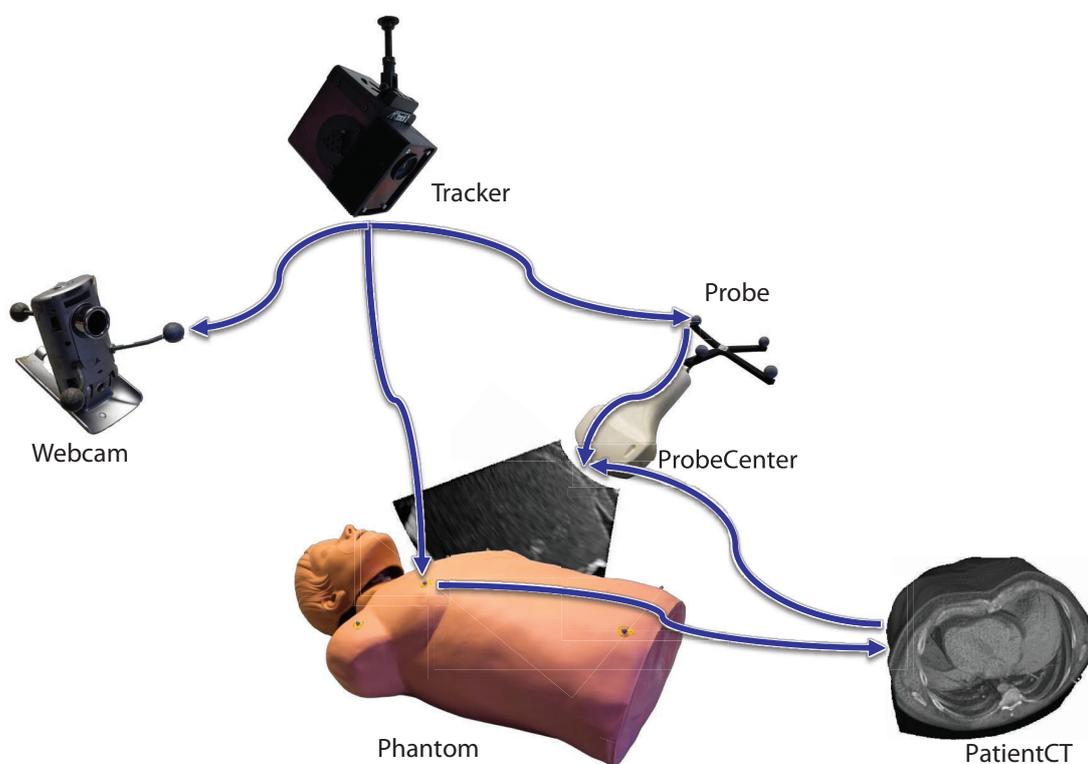
Figure 3.3.: Illustration of the US simulator setup using a webcam. The blue arrows denote the transformations that are used.

Figure 3.4.: The Simulaids Brad phantom is used in this setup. Optical infrared markers are put onto the phantom.

3.2.1. Input Device

Different solutions for input devices have been discussed before in section 2.2.2.1. We have decided to use a physical phantom with a tracked probe. The use of keyboard and mouse is no option as only a low immersion can be achieved. For haptic devices, evaluations of other US simulator systems have shown that the realism of haptic feedback has been rated low [Zhu et al., 2007, Vidal et al., 2008].

In order to build an inexpensive phantom we initially used a wetsuit as used for surfing or diving. This wetsuit was filled with foam. While the general shape of this phantom resembled the shape of a human well, it did not look very realistic and as it was only filled with foam, it did not resemble the haptic feedback of the bones. After initial experiences with the wetsuit we decided to use a standard phantom for medical training instead.

As already discussed in chapter 1, phantoms or mannequins are commonly used in medical training. A wide range of phantoms for different purposes is available. Most of these phantoms are partially or completely rigid, which is not desired for an US simulator, as a realistic haptic feedback is important. Most phantoms are designed for training of certain medical specialties such as pediatrics or nursing. They allow e.g. to do injections, smear tests or catheterization and most of them are very expensive. Other phantoms are specially designed to produce realistic images when used with imaging modalities such as CT or US. As these phantoms are also very expensive and as we use a computer-based simulation of US images, such phantoms are not reasonable. For training of reanimation there are inexpensive phantoms that are made of flexible material. We have chosen the Simulaids Brad reanimation phantom (Simulaids, Saugerties, United States). It is made of vinyl plastic over polyurethane foam. Anatomical landmarks such as the ribcage are modeled. The Brad phantom is shown in figure 3.4

A phantom probe is used instead of a real probe. The phantom was formed from a real probe using epoxy resin. A photo of the phantom probe is shown in figure 3.5

Figure 3.5.: A phantom of a probe formed from a real probe. Optical infrared markers are attached to the phantom for tracking.

3.2.1.1. Tracking

When using a phantom, it is necessary to estimate the pose of the phantom and the probe. In table 3.1 important properties of different tracking technologies are summarized. Below the different options for tracking are briefly discussed. A more detailed discussion of tracking methods, in particular optical and magnetic tracking, can be found in [Birkfellner et al., 2008].

The most robust, accurate and precise solution is mechanical tracking, where a mechanical link between the tracking device and the tracked target exists. Examples of this are robotic arms or haptic devices. These devices usually have sensors measuring the movement of each joint. When using a haptic device to simulate the US probe, this is the natural choice. The main problem is that the mechanical link limits the working volume. As discussed before, haptic devices are problematic, when the probe has to be placed on different sides of a patient, as their working volumes are relative small. In particular when using a physical patient phantom, as in our setup, haptic devices do not allow moving the probe to any position of the phantom.

Inertial tracking uses sensors such as gyroscopes, compasses and accelerometers to estimate the position and movement. While gyroscopes and compasses provide absolute orientation, an accelerometer can only be used to compute relative position changes. This results in two problems. First, in order to obtain absolute position values, the starting position must be known. This problem is known as initialization. This could be achieved by placing the sensor at a known position. The second problem is drift. While tracking an object, small errors accumulate. Therefore the precision of inertial tracking is low and it is not suited for an US simulator.

Most optical tracking systems consist of two or more infrared cameras with a built-in infrared flash. Markers that reflect infrared light very well are attached to the tracked objects. Instead of using a flash and passive markers, active markers can be used, which emit infrared light. As active markers require a power source, it is usually more convenient to use passive markers. The main drawback of optical tracking systems is that they suffer from the line-of-sight problem. To get a robust estimate of the pose of a tracked object, at least two tracking

Tracking methods	Properties
Mechanical	Highly robust, accurate & precise
	No line-of-sight problem
	Limited working volume
	Requires mechanical link
Optical	Accurate & precise
	Line-of-sight problem
	Big working volume
Magnetic	Precise but only medium accuracy
	No line-of-sight problem
	Medium sized working volume
	Problems with ferromagnetic material
Inertial	Low precision and accuracy
	Drift
	Initialization problem

Table 3.1.: Important properties of different tracking methods.

cameras should be able to see the target. As the user or other objects can obstruct the view of the tracking cameras, this cannot be guaranteed at all time. This problem can be reduced by using multiple tracking cameras or by placing them in a way such that occlusion of markers is unlikely. Using optical tracking, a big tracking volume can be achieved, which can be further extended by using additional cameras. Therefore optical tracking systems can be used for an US simulator.

Magnetic tracking systems generate a magnetic field. A sensor is attached to each tracked device. This sensor can measure its own pose within the magnetic field. Compared to optical and mechanical tracking the accuracy is lower. When ferromagnetic materials are present, the magnetic field gets distorted, which reduces accuracy. In a simulator setup the use of ferromagnetic material can usually be avoided and magnetic tracking delivers reasonable accuracy for a simulator. While for simulators magnetic tracking is an option, for generating 3D volumes from freehand US the accuracy is not sufficient. So, the commercial Schallware simulator uses magnetic tracking while optical tracking is used for recording 3D volumes. Another drawback of magnetic tracking is that the sensors that are attached to the tracked objects have to transmit their measurements to the magnetic field generator. This requires either a wired connection or a wireless connection in combination with a power source for each tracked device. As US probes are connected to the US machine via a wired connection, it is unproblematic to use a wired connection for the magnetic tracking. Therefore magnetic tracking is also a reasonable choice for a simulator system.

In both of our setups we use the optical tracking system ARTtrack2 (Advanced Realtime Tracking GmdB, Weilheim, Germany). The setup that is described in this section could also be realized with magnetic tracking. However one reason for using optical tracking is that we do not only track the relative pose of the probe to the phantom, but also an additional webcam. When multiple objects have to be tracked, the use of magnetic tracking is more complicated as every tracked device needs a wired connection to the magnetic field generator or a power source for a wireless connection. Another reason for using optical tracking is the bigger working volume. This is more important for the setup that is discussed in the next section where a HMD is used. The HMD requires a larger tracking volume as can be realized with magnetic

tracking. In our simulator, optical markers are attached to the phantom, the probe and a webcam. In both setups we use four tracking cameras that are mounted on the ceiling. The cameras use an infrared light-emitting diode (LED) flash with a wavelength of 880nm and have a maximum frame rate of 60 fps. The working volume highly depends on the number and setup of the cameras. In this setup the tracking volume has a size of approximately 90*200*130cm (w*d*h).

Passive infrared markers are attached to the probe and the webcam. Markers that are visible in CT and at the same time reflect infrared light are placed on the phantom. These markers are used for registration as will be described in section 3.2.2. We will denote the pose of object y in the coordinate system of object x as $_xT^y$. The inverse of $_xT^y$ is denoted as $\left(_xT^y\right)^{-1}$. The optical tracking system estimates the transformations $_{Tracker}T^{Probe}$, $_{Tracker}T^{Webcam}$ and $_{Tracker}T^{Phantom}$. The ARTrack2 system uses a dedicated computer, which acts as a server. The computer which runs the US simulator connects to the server via network and receives the tracking data.

3.2.2. Dataset & Registration

For the simulation a CT dataset is required. As any CT dataset with reasonable quality can be used, a large number of different cases can be simulated. However, we also use one special dataset, the Visible Korean Human (VKH) [Park et al., 2005]. The dataset consists of CT, MR and photographic slices of a male cadaver. The photographic slices have been obtained by freezing a corpse and consecutively milling away and photographing single slices. The resolution of the CT is 505 x 276 x 1718 and the resolution of the anatomical images is 3040 x 2008 x 8590. In the anatomical volumes, 938 anatomical structures are segmented. While the CT dataset of the VKH has a lower image quality as can be achieved using modern CT machines, there are two advantages of using the VKH dataset. The anatomic images are in color and contain more information than CT slices. As discussed before in section 2.3.2, showing co-registered simulated US and anatomical slices can help a student to understand the US image better. And as all important structures are segmented, this segmentation can be used to highlight any structure. As will be explained in section 3.2.8.3 the segmentation can also be used to provide feedback when training how to take measurements using US.

Every CT dataset that is used for simulation has to be registered to the phantom. This registration consists of two steps. First, the CT dataset that is used for simulation is registered to a CT of the phantom. This step is required as the shape of the CT dataset differs from the shape of the phantom. Second, the CT of the phantom is registered to the tracked target. This step has to be done, as the optical tracking system provides the pose of one point. In case of the ARTtrack2 system it provides the pose of one of the passive markers that are attached to the phantom. In order to place the CT at the correct pose, we need to know the pose of the corresponding point in the CT volume.

To register the CT dataset that is used for simulation with the CT of the phantom a surface-based registration is used, similar to the method used in [Magee and Kessel, 2005]. A surface-based registration is the only option, as the phantom does not contain any organs that could be used for an intensity-based registration. From both CT datasets surface points are extracted. A rough manual initialization is performed and then a registration using the ICP method is done. The result is the transformation $_{PhantomCT}T^{PatientCT}$.

The registration between the CT of the phantom and the point that is tracked by the optical tracking system is done using special markers. Instead of using standard passive optical

infrared markers, Beekley CT spots (Beekley Corporation, Bristol, United States) are placed on the phantom. These spots are shown very bright in CT. Infrared reflective tape is glued onto the CT spots such that they are also visible to the optical tracking system. From the optical tracking system the position of all CT spots in the coordinate system of the tracking system is obtained. By a one time manual segmentation, the position of the CT spots in the CT image is obtained. Point correspondences are computed automatically using a graph matching algorithm [Gold and Rangarajan, 1996], and a least-squares registration between the corresponding points is computed, resulting in the transformation $_{Phantom}T^{PhantomCT}$. The transformation $_{Phantom}T^{PatientCT}$ is then computed as $_{Phantom}T^{PhantomCT} *_{PhantomCT} T^{PatientCT}$. For one dataset and as long as the marker positions on the phantom are not changed, all these transformations are static and have to be computed only once.

3.2.3. Software Framework

The simulator is using the CAMPAR software framework. This framework has been developed at the Chair for Computer Aided Medical Procedures & Augmented Reality [Sielhorst, 2008]. Initially it has been developed for HMD-based medical AR [Traub et al., 2006], but it has also been used for laparoscopic AR [Feuerstein et al., 2008] and monitor-based AR [Wendler et al., 2006, Navab et al., 2010].

CAMPAR uses a micro-kernel concept where only basic functionality is implemented in the core of the framework. For all devices such as tracking systems and cameras, plug-ins are loaded dynamically during start up. Also applications such as the US simulator are implemented as plug-in. CAMPAR allows synchronized real-time AR visualization. All input devices, such as tracking systems, are synchronized by using a ring buffer with time stamps. For visualization it uses OpenGL (Open Graphics Library) [1] or an OpenGL-based volume renderer, which has been presented in [Kutter et al., 2008]. The volume renderer is implemented on the GPU and is part of a GPU library called GPUVIS, which includes sub-libraries for visualization, general purpose GPU (GPGPU) programming and the ultrasound simulation that has been described in section 3.1. CAMPAR allows either using an OpenGL visualization for HMD-based AR or the use of a graphical user interface using Qt [2], a cross-platform open source framework for developing graphical user interfaces. For the user interface of the US simulator Qt is used.

The software architecture of the US simulator is illustrated in figure 3.6. The CAMPAR framework uses the camplib, a library that provides basic functionalities for segmentation, registration, visualization and related mathematical functions. It also relies on several third party libraries. The US simulator is implemented as a plugin for CAMPAR. It uses several other CAMPAR plugins e.g. to access the optical tracking system or the webcam. The US simulator also uses the visualization and US image simulation components of the GPUVIS library.

3.2.4. Ultrasound Simulation

To generate the US image, first the CT slice that corresponds to the current relative pose between the patient phantom and the US probe has to be extracted. To extract this slice we need this transformation in the coordinate system of the CT dataset. This is computed by

$$_{PatientCT}T^{ProbeCenter} = \left(_{Phantom}T^{PatientCT}\right)^{-1} * \left(_{Tracker}T^{Phantom}\right)^{-1} *_{Tracker} T^{Probe} *_{Probe}$$

---

[1]`www.opengl.org`
[2]`qt.nokia.com`
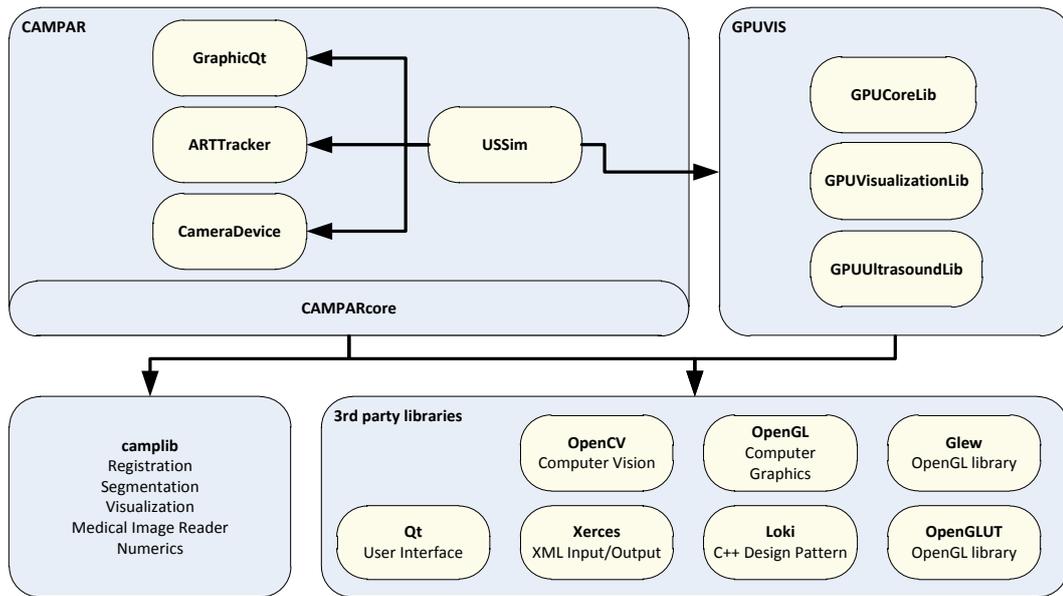
Figure 3.6.: The software architecture of the US simulator, which is implemented within the CAMPAR framework.

$T^{ProbeCenter}$, where $_{Probe}T^{ProbeCenter}$ is set manually for the phantom of a probe. As an exemplary application for the simulator we have chosen the focused abdominal sonography for trauma (FAST) protocol, which is used to scan for internal fluids due to internal bleedings. This is a procedure with a fixed workflow that is standardized in most hospitals. The FAST protocol is usually performed for trauma patients in emergency medicine as it allows to detect internal bleedings very fast.

3.2.5. Visualization & User Interface

The user interface offers three different kinds of views. A *slice view*, a *virtual view* and an *augmented reality view*. All views are shown in windows with customizable position and size. Different views can be shown at the same time. A screenshot of the user interface is shown in figure 3.7.

3.2.5.1. Slice View

The *slice view* shows the 2D slice corresponding to the current pose of the US probe. The *slice view* can show the original CT slice that is used for simulation, the simulated US image or the corresponding image from the photographic dataset. Multiple *slice views* can be used to display e.g. the US image and the corresponding photographic image at the same time. In figure 3.7 three *slice view* windows are used at the same time, showing the corresponding CT, US and photographic images. As discussed before in section 2.3.2 the use of such co-registered images is useful to better understand the US image.
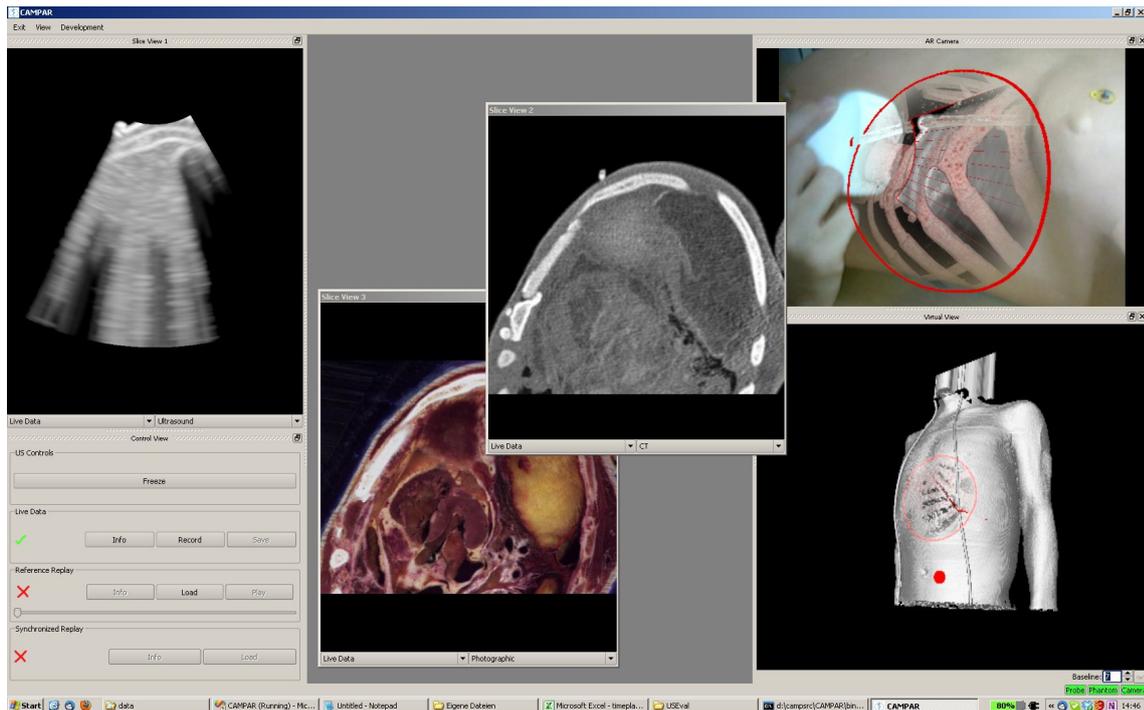
Figure 3.7.: Screenshot of the user interface showing three *slice views*, the *virtual view* and the *augmented view*.

3.2.5.2. Virtual View

In the *virtual view* a volume rendering of the CT is shown. The visualization uses direct volume rendering with pre-integrated transfer functions to visualize the CT volume. Additionally, the pose of the tracked probe and the US image plane are shown. For a better visualization of the pose of the image plane within the CT volume a focus and context (F+C) technique is used [Krüger et al., 2006] where the bones are only shown through a focus window. The advantages of F+C visualization techniques will be discussed later in the context of HMDs in section 3.3.2 where it is more relevant. The view can be rotated, zoomed and panned using the mouse. An example of the *virtual view* can be seen in the lower right of figure 3.7.

3.2.5.3. Augmented View

The *augmented view* allows visualizing the US slice into the video image of a tracked external camera. A volume rendering of the CT volume is augmented into the camera image. To be able to augment virtual objects into the camera image, we have to perform a calibration of the intrinsic and extrinsic camera parameters. The calibration uses a checkerboard and is based on the method proposed by [Tsai and Lenz, 1988]. Similar to the *virtual view*, the US image plane is augmented using a video-based F+C visualization. An example of the *augmented view* is shown in the upper right of figure 3.7.

The user interface allows customizing all views. Via an XML-configuration file it can be configured whether the *virtual view* and the *augmented view* are visible at start-up. Additional *slice views* can be opened. The user interface is realized in Qt via the QDockWidget class.
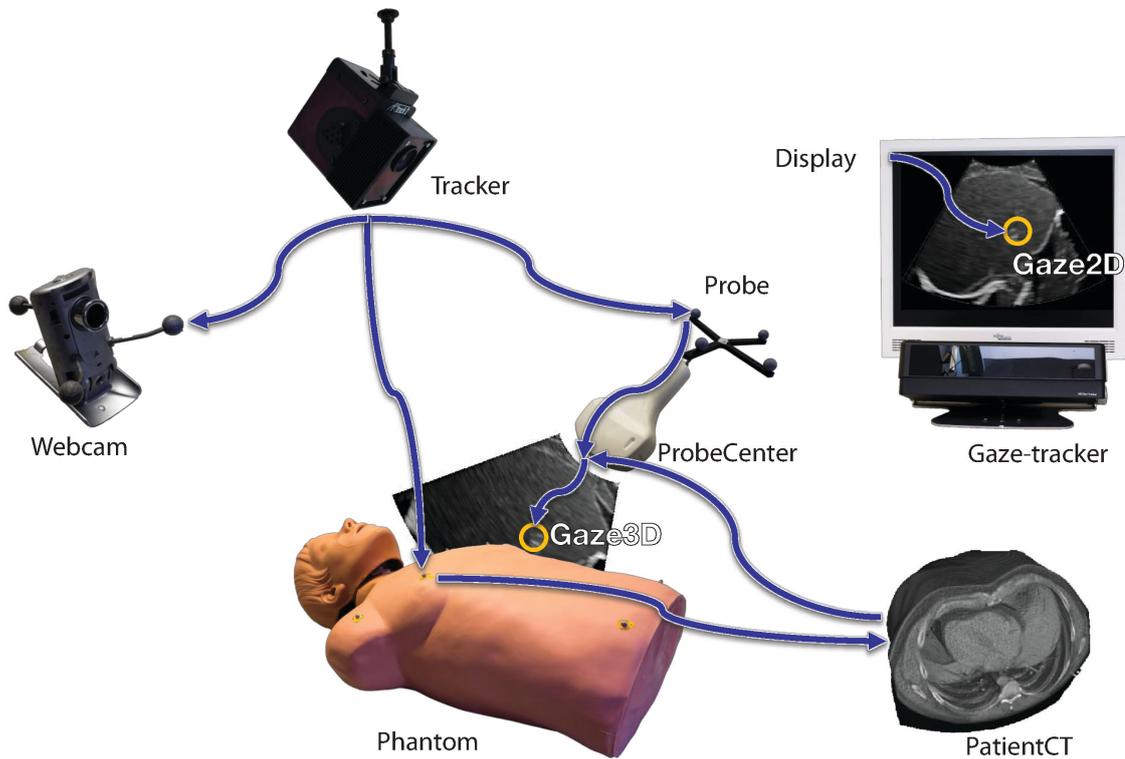
Figure 3.8.: Illustration of the US simulator setup with an gaze-tracker.

3.2.6. Output Device

As output device we are either using a standard PC monitor or a stereo monitor. As stereo monitor we are using a Zalman ZM-M220W monitor (Zalman Tech Co., Seoul, Korea). The monitor has a size of 22" and uses horizontal polarization to display stereo images. The user has to wear polarized glasses, where one eye only sees even pixel rows and the other eye only odd pixel rows. The monitor has a resolution of 1680 x 1050 and a viewing angle of $90°$ in horizontal direction and $11°$ in vertical direction. When using the monitor to display stereo images the vertical resolution per eye is halved. The virtual view can be displayed in stereo. To generate a stereo image the raycaster renders the image twice, from slightly different positions. To display the stereo image OpenGL quad-buffering is used.

3.2.7. Gaze-tracking

For analyzing how experts and students use US, it is interesting to exactly know where the examiner is looking. To be able to estimate this, we integrated an gaze-tracking device. We are using the Tobii X60 gaze-tracking device (Tobii Technology, Danderyd, Sweden). The X60 is a stand-alone gaze-tracker, which can be placed below or above any planar surface to track the gaze of the user on this surface. It tracks with 60 fps, has a spatial resolution of $0.2°$ and an accuracy of $0.5°$. The tracking volume has a size of 44x22x30cm. The user can move her head within this volume. For each user the gaze-tracking device must be calibrated once, before using it.

The gaze-tracking device provides us with the 2D gaze position on the screen in pixels. This is illustrated in figure 3.8. The gaze-tracker provides the transformation $_{Display}T^{Gaze2D}$ in pixels. As we know the size and position of the US image on the screen, we can compute the

gaze position on the simulated US image. For the simulation of the US image, characteristics of the US probe such as the field of view and the inner and outer radius of the US image plane have to be provided. From these characteristics the relation between the physical size in the CT volume and pixels in the simulated US images can be estimated. Therefore we can compute where on the US image the user is looking in the coordinate space of the US probe $_{ProbeCenter}T^{Gaze3D}$. The position of the gaze within the 3D CT volume can be computed as $_{PatientCT}T^{Gaze3D} =_{PatientCT} T^{ProbeCenter} *_{ProbeCenter} T^{Gaze3D}$.

The gaze tracking data has to be processed as humans switch between fixations and saccades. Saccades are very quick and short movements of the eyes and have to be filtered out. To do this the standard method for filtering saccades is applied, which uses thresholds on the minimum duration and the maximum movement of the eye during a fixation [Duchowski, 2007].

3.2.8. Teaching Concepts

In this section several basic teaching concepts are discussed. More advanced methods that use a workflow model of a procedure will be introduced in chapter 4.

3.2.8.1. Co-registered Images

As already discussed in section 2.3.2, presenting co-registered slices from different modalities helps students to get a better mental model. In this system in addition to the US image, slices from the CT volume and photographic images can be shown. Furthermore, all 938 anatomical structures that have been segmented for the VKH can be highlighted in the US, CT or photographic image. A screenshot of this is shown in figure 3.11

3.2.8.2. Simple Shapes

This teaching concept is inspired by the book [Harness and Wisher, 2001], which provides an introduction to the basics of US and clinical applications of US. In this book the basics of US imaging are explained using cross-sections of simple shapes such as cubes, balls or cones. Understanding how cross-section e.g. of a cone look like is important for every novice. AR is a powerful tool to implement a similar concept in a computer-based simulator. It is not required to build real objects, but virtual objects can be used. Both the US image plane and the virtual object are seen in the *AR view*. First, very simple shapes such as a box, a ball and a tube are shown. Here a student can learn how 2D cross-sections of 3D objects look like. In the space around the objects everything has the same echogenicity, so it is isoechoic and will be shown as gray area in the US image. The objects are hyperechoic, so they are more echogenic and their boundaries are imaged brighter. Three simple shapes are shown in figure 3.9.

In addition to the simple objects also more complex objects such as the box shown in figure 3.10 can be used. While the simple shapes consist only of one material, this box consists of multiple materials with different acoustic impedance. Some parts are hyperechoic and some hypoechoic. The task of the student is to figure out which parts are hyperechoic and which hypoechoic. Different choices are presented to the student in 2D images. This is shown in figure 3.10(b). The student can move the probe and has to figure out which of the 2D images corresponds to the 3D structure. Here it should be noted that the brightness in the image does not only depend on the echogenicity of a material, but also on the echogenicity of materials the sound waves have passed before.
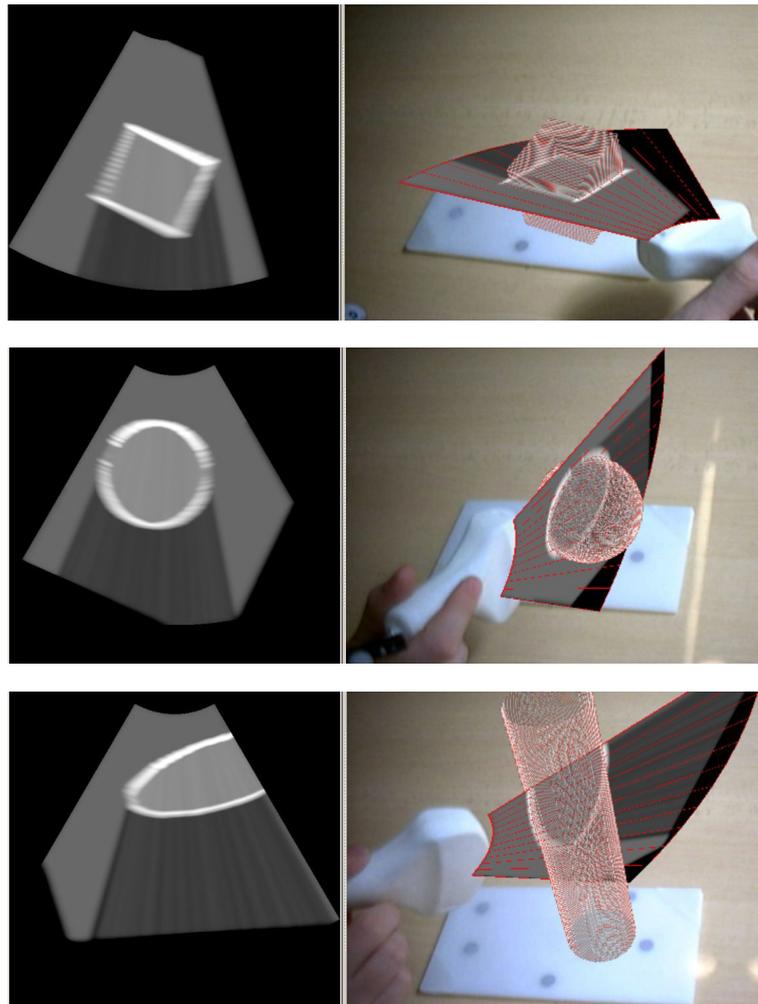
Figure 3.9.: AR view and simulated US image of a box, a ball and a cylinder. The objects are hyperechoic while the surrounding is isoechoic.

3.2.8.3. Measurement

As already discussed in section 2.1.2, measurements using US have low inter- and intra-observer repeatability. To allow practicing measurements, the simulator allows to freeze the images and to measure a structure by clicking on points in the US image, as also done on real US machines. As discussed in section 2.3 it is important to provide feedback so that a student can correct the own mental model. To provide feedback the US simulator uses the segmentation of the VKH dataset. While the student performs a measurement the pose of the probe is recorded. After a student measured a structure, the segmentation is used to highlight the structure. For this the *slice views* are used, so structures can be highlighted in the US, CT or photographic image. This can be seen in figure 3.11. The previously recorded poses of the probe are used to provide a replay. So the student can see where she did measurement errors. Additionally, a replay of an expert performing a measurement on the same dataset can be shown.
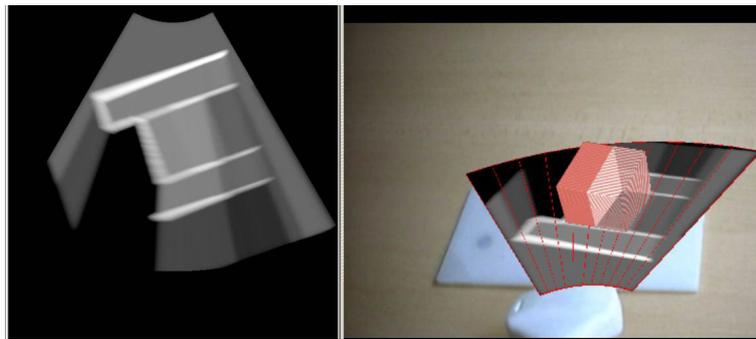
(a) US simulation and AR view of a more complex object.
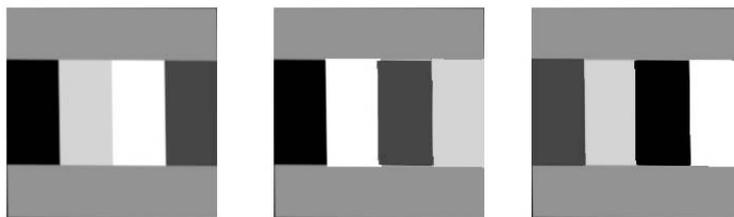
(b) The student has to choose the correct underlying object from these three choices.

Figure 3.10.: A more complex object with different materials.
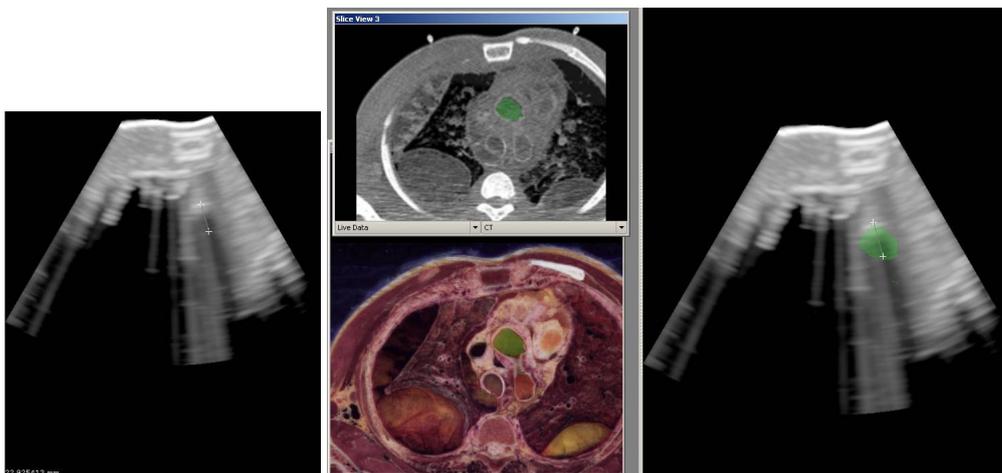
(a) For a measurement the user clicks on two points, shown by crosses. The measured distance is shown on the lower right.

(b) The measured structure is highlighted in green using the segmentation of the VKH

Figure 3.11.: Feedback for measurements.

3.2.8.4. Gaze Visualization

As described before in section 3.2.7 the simulator can record gaze-tracking information and extract fixations. Such information can be used for a scanpath-based feedforward training. The idea is to record the gaze patterns of an expert and show them to a trainee so that the trainee can learn from these expert examples. Feedforward training has been done previously by [Sadasivan et al., 2005] for training visual inspection of aircrafts. They recorded the gaze of experts in a virtual environment and showed it to trainees. Their study showed that a group that received feedforward training was more effective in detecting defects than a control group. In a study by [Nalanagula et al., 2006] feedforward training using gaze-tracking was done for visual inspection of circuit boards. The gaze of experts was shown to novices during inspection, which increased the number of defects they detected. In another study MDs had to detect pulmonary nodules in chest X-ray images, while the gaze of another examiner was shown to them [Litchfield et al., 2008]. When seeing the gaze of another person the radiographers performed better than a control group. In particular the performance of novices increased.

The same concept can also be applied to education of US. Using the gaze-tracking data from the US simulator a replay of the gaze and a visualization of the fixations was realized. This can be seen in figure 3.12. The blue balls visualize fixations of one user, where the size of the balls represents the duration of the fixation. They are connected via lines that represent their temporal order. The gaze of another user on the 2D US image plane is shown by a green ball. Such a visualization of gaze is interesting, in particular in combination with after action reviews as will be discussed later.

3.3. Setup Using a HMD

Based on the first setup a more advanced setup using a HMD for AR visualization has been built. This setup is located at the Narvis-lab at the hospital of the University Munich (Klinikum Innenstadt, Ludwig-Maximilians-Universität München). Most of the software components from the first setup are also used in the HMD-based system. This section will explain the differences of the HMD-based setup. A photo of the system can be seen in figure 3.13.

3.3.1. Input Device & Tracking

To be able to correctly augment images into the HMD we have to know the relative pose between the HMD, the probe and the phantom. As discussed in section 3.2.1.1 there are several possibilities for tracking. In a setup with a HMD the tracking volume has to be big enough such that the HMD and the phantom can be tracked at the same time. As optical tracking systems provide a big working volume they are a good choice for a HMD-based system. We are using the same type of optical tracking cameras as in the first setup, which are ARTtrack2 cameras. Again, we use a four camera setup to achieve a big tracking volume and avoid line-of-sight problems. In addition to this outside-in tracking system, a second optical inside-out tracking camera is mounted onto the HMD. Both tracking systems track a common target such that we can bring both into the same coordinate system. More details on the setup of the tracking system, the calibration and advantages of using a combination of inside-out and outside-in tracking can be found in [Sielhorst, 2008].
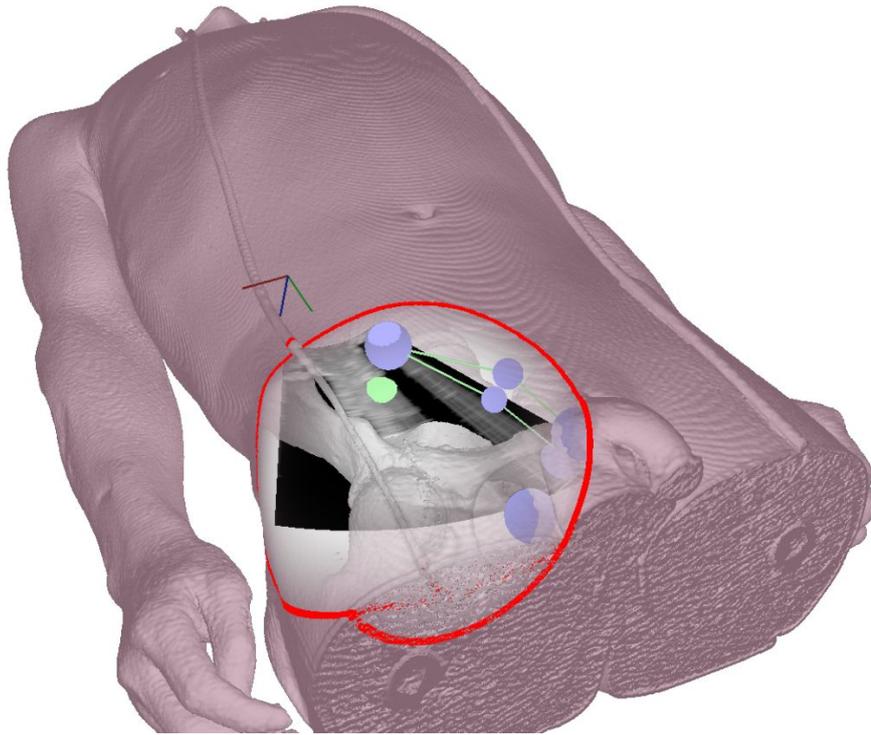
Figure 3.12.: Visualization of the gaze. The blue circles represent fixations of one recorded US procedure. The green circle is drawn onto the 2D US plane and represents the current gaze position of another replay.
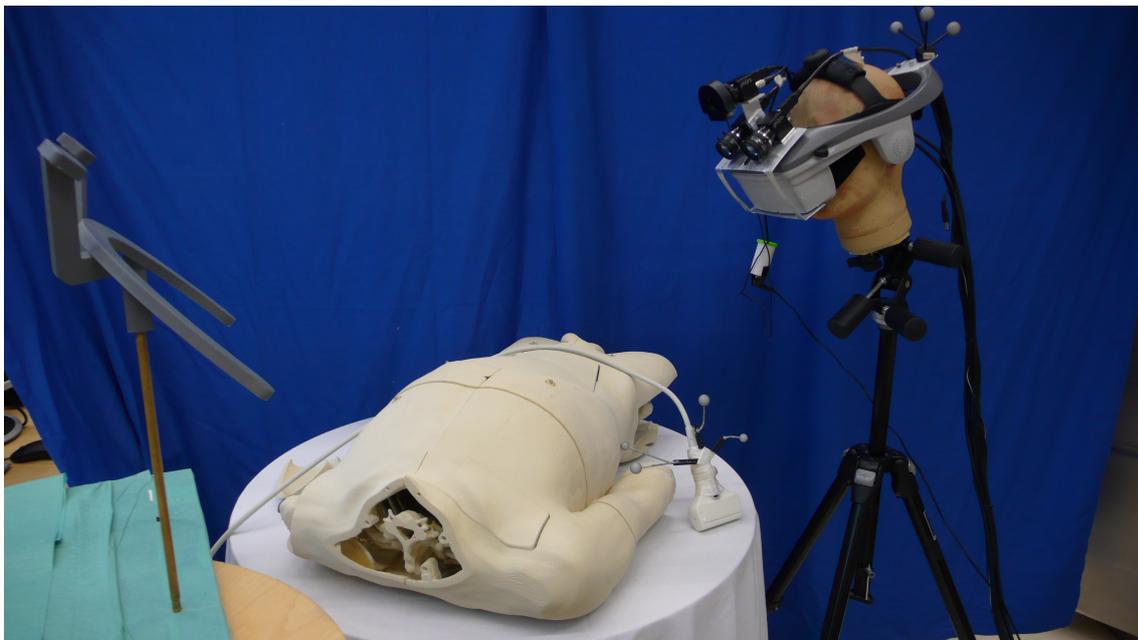
Figure 3.13.: Setup of the US simulator using a HMD and a phantom that was built using a rapid prototyping printer.
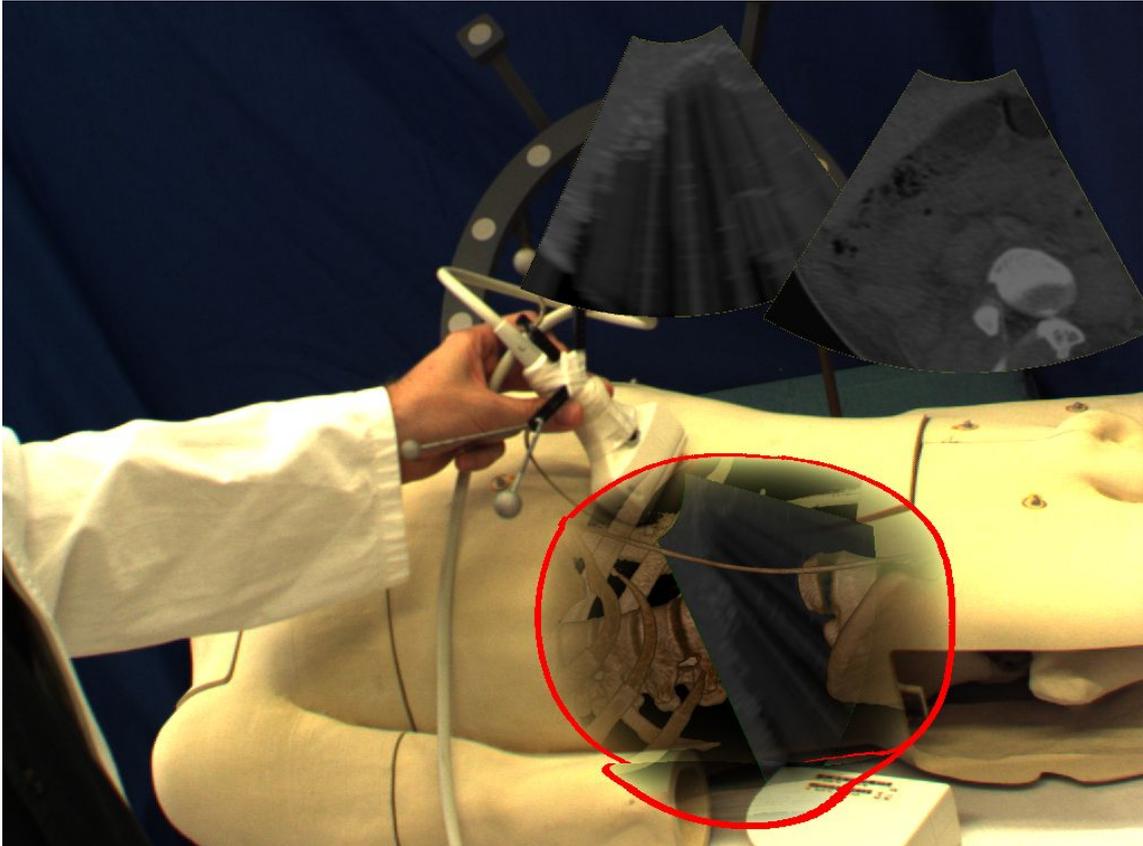
Figure 3.14.: Augmented view as seen through the HMD. The US image plane is shown co-located with a volume visualization of the bones. Above the phantom two virtual monitors are placed, which show the simulated US image and the corresponding CT slice.

In this setup, we use another phantom than in the first setup. This phantom is based on the VKH dataset and it was printed using rapid prototyping techniques [Bichlmeier et al., 2008]. This has both advantages and disadvantages over the phantom used in the first setup. As it was printed using rapid prototyping techniques the phantom is very rigid and does not feel like a real human. On the other side it is an exact copy of the VKH dataset, and therefore the dataset exactly fits to the phantom.

3.3.2. Software, Visualization & User Interface

This setup uses the same CAMPAR software framework as the first setup. Instead of using Qt for a window-based user interface it draws the images for the left and right eye using OpenGL. As already done for the AR visualization using a webcam in the first setup, the CT dataset is augmented onto the phantom using a focus and context visualization [Bichlmeier et al., 2007b]. Most AR systems augment the whole virtual dataset onto the real object. This, however, leads to problems in depth perception as the virtual objects occlude the real objects. As occlusion is one depth cue that is used by humans to estimate the relative distance of objects, anatomy which is inside the patient is perceived as being in front of the patient. Using focus and context visualization the virtual objects are visualized through a focus window, which leads to a better perception of depth. The F+C visualization can be seen in figure 3.14.

The user interface for a HMD has to be different than a user interface for a monitor-based system as the user is standing next to the phantom and no mouse or keyboard can be used. Instead of using multiple windows that are placed on a desktop to show different information, all information must be located in the real space. Similar to the *augmented view* of the webcam in the first setup, the US image plane is augmented into the view of the user. Instead of having a second window, as for the *slice view*, in the HMD-setup the slices can be placed at a fixed point in the real world. This is shown in the upper part of figure 3.14. Again, US, CT and photographic slices can be shown.

3.3.3. Output Device

For AR there is a wide range of display technologies that can be used. A simple AR visualization has already been used in the first setup of the US simulator, where the image is only shown on the screen. For some medical applications such as endoscopy [de Lange et al., 2010, Szpala et al., 2005] or laparoscopy [Feuerstein et al., 2008, Nakamoto et al., 2002] the use of monitor-based AR is reasonable as monitors are usually used for these procedures. However for other applications a direct augmentation of virtual objects into the real world is preferred. There are several technologies that allow to augment objects directly into the real world, including HMDs, projection and semi-transparent mirrors. Below, the most important output devices for AR are discussed. A more detailed discussion of display devices for medical AR can be found in [Sielhorst et al., 2008].

There are two types of HMDs. Optical see-through (OST) and video see-through (VST) HMDs. OST-HMDs use semi-transparent displays that are located in front of the eyes of the user. The user can see the real world through the semi-transparent display, while virtual images can be overlaid. VST-HMDs have two monitors instead of the semi-transparent displays. Two cameras are attached to the HMD and the camera images, enriched by virtual objects, are shown on the monitors.

OST-HMDs are smaller and have less weight. On the other hand, they have some drawbacks. The first problem is that they have to be calibrated for each user, as the relative position of the eyes to the display must be known in order to allow a correct overlay. A calibration procedure for this has been proposed by [Tuceryan et al., 2002]. Furthermore they have to be calibrated each time the user puts the HMD on, as the transformation between the eye and the display changes slightly every time the user puts the HMD on. If the HMD is moved while it is used it has to be re-calibrated. The second problem is that OST-HMDs that are available today only allow displaying semi-transparent virtual objects. Because of this, no virtual object can completely occlude real objects. This is problematic, in particular when working with advanced visualization techniques as the focus and context visualization. The third problem is that the real part of the image is seen instantly, while virtual objects are shown with a delay. It takes some time until the tracking data arrives. Then the visualization has to be drawn by the computer and the image has to be sent to the HMD. Due to this delay, if the user moves the head, virtual objects will lag behind. Using a VST-HMD a delay can be added to the camera image so that the real and the virtual objects are shown synchronous. While this introduces a delay to the whole image, this is preferable to a mismatch between real and virtual images. Another choice are HMDs that use laser projection into the eye of the user. However these devices are limited to drawing line graphics and are therefore not suited for visualization of medical image volume. Previously OST-HMDs have been used for AR in the domains of needle biopsy [Rosenthal et al., 2001], cranofacial surgery [Salb et al., 2002], endoscopy

[Bockholt et al., 2003], neurosurgery [Eggers et al., 2005] and dental surgery [Katić et al., 2010]. Research on AR using VST-HMDs has been done in the application domains of ultrasound [Bajura et al., 1992, Sauer et al., 2001b], neurosurgery [Sauer et al., 2001a], needle biopsy [Wacker et al., 2005], spine surgery [Bichlmeier et al., 2007a] and training of forceps delivery [Sielhorst et al., 2004].

Another method is the use of projectors to project virtual objects onto the surface of the patient. An AR system using such a technology for neurosurgery has been discussed in [Glossop et al., 2003]. While this technology does not require the user to wear a HMD it has several drawbacks. The projection will only work well on white surfaces and the surface geometry has to be known. Both limits the usefulness of this method. Furthermore the projection is only correct for objects that are directly on the projected surface. When showing objects that are below the surface, as the plane of the US image, additional head tracking would be required to create an image with a correct perspective for the user.

Semi-transparent mirrors are another technique that has been used previously. A half-silvered mirror is used such that the user can see the real world while additional virtual objects from a display are overlaid. Such systems have been used in the domains of neurosurgery [Liao et al., 2004], needle insertion [Fischer et al., 2005] and ultrasound [Stetten et al., 2003]. While it has been shown that this method works for US, the drawback is that the virtual objects cannot fully occlude real objects and virtual objects can only be augmented in a very limited space around the mirror. Due to this limited space for augmentations it is difficult to display additional information, which is desired for a simulator for training.

For some medical applications operating microscopes are used. For digital microscopes augmented objects can be added to the microscope image as has been done by [Edwards et al., 2000] and [Birkfellner et al., 2002]. However, for US the use of operating microscopes is only of minor relevance.

As we are highly interested in advanced visualization we have chosen to use a video see-through HMD. This is the technology that is suited best to visualize virtual objects that can occlude real objects in high quality. The HMD that is used for this simulator is an nVisor SX (NVIS, Reston, United States). It is a VST-HMD with a horizontal field of view (FOV) of $48°$ and a vertical FOV of $36°$. The resolution is 1280 by 1024. Two FLEA color cameras (Point Grey, Richmond, Canada) are mounted onto the HMD. They use an ICX204AK sensor (Sony, Tokyo, Japan) with a resolution of 1024 by 768. In addition to the four optical tracking cameras, this system uses a infrared tracking camera mounted onto the HMD. For this purpose a FLEA-BW camera with a Sony ICX424AL chip with a resolution of 640 by 480 is used. The setup is described in more detail in [Sielhorst, 2008].

When using a HMD the user is totally immersed into the real world. On the other hand HMDs still have some drawbacks. The use of HMDs often leads to simulator sickness and perception of depth is often not correct. Another problem is developing appropriate user interfaces when using HMDs, as traditional user interfaces such as mouse and keyboard are not well suited. In appendix A.1 two studies on related problems are presented.

3.3.4. Teaching Concepts

All of the teaching concepts that have been discussed for the first version of the simulator in section 3.2.8 can also be realized in the HMD-based system. However the HMD has the additional advantage that the information can be shown in-situ, on the phantom. A student can use the US probe on the phantom and at the same time see virtual information. This can

help to understand spatial relations better.

In this chapter a method to provide feedback to a user by showing a segmentation of a target structure has been presented in section 3.2.8.3. In this example a very simply workflow was used where the size one structure was measured. Many procedures where US is used have more complex workflows. In the next chapter it will be discussed how an US simulator can be used to provide feedback on more complex workflows. First the generation of workflow models will be discussed and later examples of using these models in the US simulator are given.

4. Workflow Modeling

Most medical procedures have a workflow that reflects how a procedure is performed. Such a workflow is interesting for medical training and education for several reasons. First, a medical student has to learn and understand the workflow of procedures. Finding new ways to model and present such workflows to students is therefore of importance. Second, standard workflows can be used for assessment of students by comparing the performance of a student to the standard workflow. Third, models of workflows can be used to provide feedback. As already discussed before, providing feedback is one of the most crucial elements of simulators. In this chapter, different methods of constructing statistical workflow models and using them for training and education will be discussed. The methods that are described in this chapter have been developed for two different applications areas. In addition to education and training, also the use for surgical workflow analysis and context-aware operating rooms has been considered, and the motivation and results for this application area will be addressed briefly.

In section 4.1.1 applications of workflow modeling, not only for educational purposes but also for context-aware operating rooms (OR), are discussed in more detail. In section 4.2 the data is discussed, where in addition to data from the US simulator, laparoscopic cholecystectomy will be introduced as an additional medical procedure. Then the use of dynamic time warping (DTW) for modeling workflows is discussed in section 4.3 and the use of DTW-based models for medical education in section 4.4. The use of hidden Markov models (HMM) for creating workflow models is discussed in section 4.5 and the use of HMM-based models for education in section 4.6. While for this thesis mainly the use for training and education is relevant, results for context-aware ORs will be discussed briefly at the end of the chapter in section 4.7.

The work presented in this chapter is partially based on previous work. The use of DTW to synchronize pose data in a simulator for forceps delivery has been investigated before by the author of this thesis [Blum, 2005]. The use of DTW to model the workflow of laparoscopic cholecystectomies, not for educational purpose but for context-aware ORs, has been discussed before by [Ahmadi et al., 2006]. Most of the data for the laparoscopic cholecystectomy, which is used in this work, was recorded by [Ahmadi et al., 2006]. The use of DTW and HMM to model workflows to enable context-aware ORs has been discussed before in [Blum, 2007] by the author of this thesis. The methods presented in sections 4.3 and 4.5 are based on [Blum, 2007], where construction of these models and methods for context-ware systems have been discussed in more detail. The contribution of this thesis over previous work is the application of these methods for education, discussed in sections 4.4 and 4.6 and an additional method to build HMM-based workflow models, which is presented in section 4.5.3.

4.1. Applications of Workflow Modeling

In this section we will discuss some other applications and related work where workflow models are of use. The applications described in this section will not be addressed further in this

thesis. The use of workflow models in the context of the US simulator is described later in sections 4.4 and 4.6.

4.1.1. Skills Evaluation

One area related to workflow models, where much research has been done, is the evaluation of skills in medical simulators. As discussed in chapter 1, a standardized outcome of medical education is being considered as important part of future education methods. One important aspect to achieve this is standardized and objective performance measures and metrics for skills. One objective way to evaluate skills is using a statistical model describing how an expert performs a certain action or workflow. Using a distance or similarity measure between the workflow model of an expert and a recording of the performance of a trainee a performance measure can be established, by comparing how close the performance of a student is to the performance on an expert.

One example where such a method is used is a simulator for laparoscopic surgery [Rosen et al., 2000], which measures the forces and torques that are applied at the tool/hand interface of a laparoscopic grasper. In later work a sensor to measure grasping forces [Richards et al., 2000] and a tracking system, which measures position and orientation of the tools [Rosen et al., 2002], were added. Data was gathered from novice and experienced surgeons to compare their performances. The first version of the system did not have a very strong workflow model. The K-means algorithm was used to find the cluster centers of 14 different states and using statistical analysis a difference between novices and experts could be shown. In further work HMMs [Leong et al., 2006] were trained to obtain a stronger statistical model. The HMM models transitions between states and time intervals spent in each tool/tissue interaction. Difference in surgical skills is assessed by generating an HMM for each individual and measuring the statistical distance between HMMs. Besides the laparoscopic cholecystectomy, their system has been evaluated in further studies using contact force sensors on a pelvis simulator [Mackel et al., 2007] and during an animal study [Rosen et al., 2006].

For the same purpose of objective skill assessment the Imperial College Surgical Assessment Device (ICSAD) was developed [Datta et al., 2001]. Instrument positions are tracked using an electromagnetic tracking systems. The tracking data is processed to extract the number of movements made by each hand, distance traveled and time taken to complete the procedure. Using statistical analysis it has been shown that surgical skill is correlated to the number of hand movements and time taken to accomplish the task. This approach was extended using a 6-DOF tracker and HMMs [Leong et al., 2006].

HMMs have also been used by [Murphy et al., 2003] to assess skills of using laparoscopic tools in a VR setup. For the task of throwing a ball onto a target, ten different gestures have been identified manually. For each gesture a five-state HMM has been trained using position, velocity and force recorded from a 3DOF haptic device. Furthermore the position of the ball, distance between ball and tool tip and the status of the gripper have been used for training. The HMMs are used to identify the number of motions someone needs to fulfill the task. Based on wasted motions and pauses the skills of different subjects are assessed. In more recent work [Lin et al., 2006], 72 data points from the da Vinci robot, a telerobotic surgery system, have been used. The data included position, velocities, rotation, servo times and status of console buttons. The data is preprocessed using feature normalization and linear discriminant analysis. The motion is then classified into different classes using a Bayes classifier and the difference between data from experts and intermediates is analyzed based on the usage of the different

motion classes. A workflow model using HMMs for the same kind of data has been used later [Varadarajan et al., 2009, Reiley and Hager, 2009] for assessment of skills.

Similar approaches have also been shown by [Speidel et al., 2009] for skills assessment in laparoscopic surgery and by [Jayender et al., 2010] for skills assessment in colonoscopy. A review of related literature has been performed by [Reiley et al., 2011]. These approaches differ from the methods used in this work as they want to provide an objective and quantitative evaluation of skill. This is useful to assess students but only of limited use to support them in building up a mental model. Providing only a metric but not informing the student what he did wrong and how to improve is only a very weak feedback. Instead, in this thesis workflow models are used to provide feedback in a way such that the student can learn from it and correct her mental model.

4.1.2. Context-aware Systems in the OR

Developing context-aware ORs is another topic which is not related to medical education but that can benefit from statistical workflow models. Over the last decades, technological developments have changed surgical procedures. The use of intra-operative imaging has increased and computer-assistance systems have been introduced. Better and cheaper imaging modalities, pre-operative and intra-operative image processing, electronic patient recording and advances in computer graphics have led to a huge amount of information and different visualization options in the OR. In the future, the number of pre- and intra-operative images and other information is likely to further increase. While this development can help surgeons to perform better surgeries it also leads to information overflow. Surgeons have to filter and process an ever increasing amount of information, increasing their mental workload. Furthermore, they have to learn how to operate computer-assisted surgery (CAS) systems. While CAS systems help to assist surgeons, often they also place an additional burden onto the surgeon, as they have to carefully operate the system.

One method that could help to overcome these problems is to use context-adaptive systems. The goal is to develop computer systems that know the workflow of a type of surgery and that can detect the current phase of a running surgery. Such a system could provide the information that is relevant for the current phase of the surgery and provide a user interface that is optimal for the current situation. The modeling of such workflows and the detection of phases can be done based on signals that are obtained automatically from the OR. Such signals can come from anesthesia devices, video cameras, tracking systems or any other device which is inside the OR.

One example is the research by [Miyawaki et al., 2005]. They aim at developing a scrub nurse robot system that assists the surgeon and must therefore have a model of the workflow. Two models at a different scale were built manually, describing the workflow during a thoracoscopic surgery. The motions of a human scrub nurse were tracked using optical markers. In [Ohnumaa et al., 2006] motion analysis was used to recognize when the surgeon exchanges the surgical instruments. In [Hsu and Payandeh, 2006, Payandeh and Hsu, 2007] a system is proposed that recognizes surgical tool gestures from the image of a laparoscopic camera in a simulated setup. This information is then used to provide a context-sensitive user interface.

For a context-aware augmented reality system for dental implant surgery, [Katić et al., 2010] proposed a method to recognize the current phase of a surgery using a description-logic based ontology. They use fuzzy logics to describe situations and adopt the AR visualization in a VST-HMD based on the recognized situation. For cataract surgery [Lalys et al., 2012] pre-

sented a framework for recognition of phases based on video images where a statistical temporal model of different image features is used. For detecting phases in laparoscopic surgery [Bouarfa et al., 2011b] used HMMs to recognize high-level surgical tasks from observable low-level tasks. And in [Padoy and Hager, 2011] HMMs were used to model and recognize actions in a robotic MIS system to allow sub-tasks to be performed by the computer.

The methods presented in this thesis have been developed both for education and to enable context-aware systems. In this thesis we will focus on the use for education. However below some more advantages of context-aware systems are discussed.

4.1.2.1. Monitoring

Monitoring the surgery is the complementary task to providing a context-sensitive user interface. While the surgeon can only concentrate on a limited amount of information the computer can process huge amounts of data. A computer is certainly not capable of fully interpreting what happens during a surgery. But a computer can compare all signals to a model of the average surgery and inform the surgeon of any unusual measurements. So e.g. the blood pressure or the heart frequency of the patient could be monitored and any unusual blood pressure values can be reported. More advanced systems could also monitor the surgery based on intra-operatively acquired images. Using such methods it could be possible for the computer to distinguish between bleedings that are a normal part of the surgery and unusual bleedings. Certainly such a system can only notify about such events and the interpretation of the information remains the task of the surgeon.

4.1.2.2. Prediction

Predicting the remaining duration of a surgery is an application that does not reveal its full potential at the first glance. In fact this would be quite useful in order to support the planning of hospital wide time schedules. Even the duration of surgeries of the same type varies largely. While the staff inside the OR can forecast how long the surgery will take, they are usually too busy to inform people outside the OR. Therefore the schedule for following surgeries and other surgeons is a subject of great uncertainty and is quite often delayed. In a study it was shown that information regarding scheduled surgeries is in general only of low reliability and accuracy [Plasters et al., 2003]. To address this problem it was proposed to facilitate communication and provide tools for status monitoring. Solving this problem would result in a higher planning reliability and therefore in considerable cost savings. Automatically forecasting the remaining duration of the surgery again involves automatic analysis and a model of the surgery. If it is known in which phase a surgery is, the remaining time can be forecast. Such a system could even be coupled to a hospital-wide workflow management system, enabling a more sophisticated and accurate planning of all processes in the hospital.

Another area, where forecasting of the remaining duration of a surgery is of high importance, is anesthesia. As it is not known in advance when the surgery exactly starts, a higher dose of drugs has to be administered so that they are effective even for the case of a delay. This issue gets even more important since so called fast-track surgeries [Song et al., 1998] are gaining increased attention. Fast-track is a concept that tries to considerably reduce rehabilitation times e.g. by using minimally-invasive surgeries and minimizing doses of drugs for anesthesia.

4.1.2.3. Automatic Report Generation

Doctors spend a huge amount of their time on writing reports of surgeries. As highly educated doctors have to do this routine work, this is also an area where automation could lead to considerable time savings. A computer can certainly not write a report on its own. But a system capable of automatically analyzing a surgery can help to reduce this time by supporting the report generation. It would be possible to generate a template of a surgery, where all actions taken during the surgery are documented. This documentation could already include the time each action started and ended and unusual occurrences could be indicated. Instead of writing the whole documentation, doctors would only have to complete and correct the template. When generating the documentation in electronic form, video and audio data as well as all further signals obtained from the OR could be linked to the documentation. This would be very useful when it becomes necessary to further analyze a surgery afterwards e.g. when problems occur, for subsequent surgeries, training purposes or assurance cases.

4.1.3. Analysis of Workflows

The last section discussed the use of workflow models during medical procedures. A related field is analysis and optimization of surgeries using workflow models. An optimal integration of new technologies into the OR and into the workflow is crucial for enabling the surgeon to make use of the technology. Knowing the workflow is indispensable for designing new ORs, the spatial layout of an OR, new types of interventions, surgical tools or navigation systems. Other goals are redrafting, evaluation, statistical analysis and rapid prototyping of computer aided surgeries. For optimizing CAS procedures it is important to build workflow models that can be interpreted by humans. To model and describe the course of interventions [Burgert et al., 2006] built a top level ontology containing items, actions, concepts and relationships. During a surgery, actions are recorded and a mapping between the ontology and the workflow is used to give a meaning to the single items. They visualize the workflow to support preoperative planning, represent surgical work steps during a running intervention [Neumuth et al., 2006]. To do this a graphical representation is used, which helps a human to understand non-quantitative information. Time-related and logic-oriented visualizations are used to visualize position of tools, actions of the surgeons and other information. In [Neumuth et al., 2011] similarity measures for surgical process models have been discussed. Such metrics could be used to compare different ways to carry out a surgery e.g. when using different methods to perform a procedure or when analyzing the use of new technical assistance systems compared to traditional methods.

To analyze the place and role of images during a surgery [Jannin et al., 2001] modeled a surgical procedure and the use of image entities during the procedure using the Unified Modeling Language (UML). One goal of this work is to adapt the visualization to the workflow and to display only image entities relevant for the current working step. In later work, ontologies [Jannin and Morandi, 2007] are used to represent computer-assisted neurosurgeries. This data is used to predict aspects of the surgical procedure based on patient-related parameters.

In [Siddoway et al., 2006] workflow analysis was performed on interventional radiology. The purpose of this work was to identify inefficiencies in the workflow and to allow simulation, design and prototyping of new technologies. To perform the analysis, the workflow was recorded manually and visualized afterwards. In [Megali et al., 2006] first steps were taken to automatically model and get an understanding of surgical performance. An HMM was used to obtain information about differences in hand movement between experts and novices.
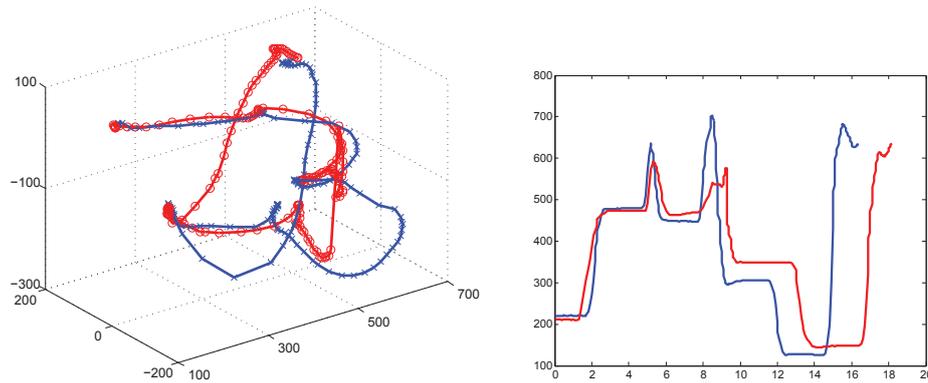
(a) The position data of two trajectories showing the same procedure performed twice.

(b) The same two trajectories, where the position on one axis is visualized over time.

Figure 4.1.: Data from the ultrasound simulator.

Being able to automatically build even more complex models of the surgical workflow would provide an objective and inexpensive way to analyze the workflow. A lot of information can be obtained from such a model. So it could be figured out how and for which purpose instruments are used. The average duration of different steps could be obtained. But also the variance of the duration could be useful to indicate whether an action is performed very systematic or if it is unpredictable. The most critical steps, where problems and irregularities occur more frequently could be identified. Using information about the eye-gaze [Nicolaou et al., 2004b] it could even be figured out where the attention of the surgeon is focused. Having access to all this information would of great value for understanding and improving the workflow.

4.2. Data

The methods that are described here have been developed using data from two different medical applications. One is data from the US simulator described in chapter 3. The other is data from a minimally invasive surgery. For MIS the aim was to automatically create human understandable statistical models of the workflow, which can be used for teaching. However, during research on these statistical models also the application of creating context-sensitive CAS systems has been considered. For context-sensitive systems the statistical model is used to recognize the current phase of a running surgery. In this thesis the use of statistical workflow models will only be discussed in the context of medical education and training. The application of these methods for context-sensitive CAS is discussed in detail in [Padoy et al., 2007, Blum et al., 2008, Padoy et al., 2008, Blum et al., 2010b, Blum et al., 2010a, Padoy et al., 2010] and will be discussed briefly towards the end of this chapter.

The data from the US simulator consists of position and orientation information of the relative pose of the US probe with regards to the phantom. This data is computed from the tracking data obtained from the optical tracking system and time stamps are attached to each data point. Two trajectories showing the movement of the US probe for two instances of the same procedure are shown in figure 4.1.

The second dataset has been acquired from laparoscopic cholecystectomies. This surgery is a standard procedure where the gallbladder of the patient is removed during a minimal-invasive procedure. With over 400.000 cholecystectomies per year in the United States
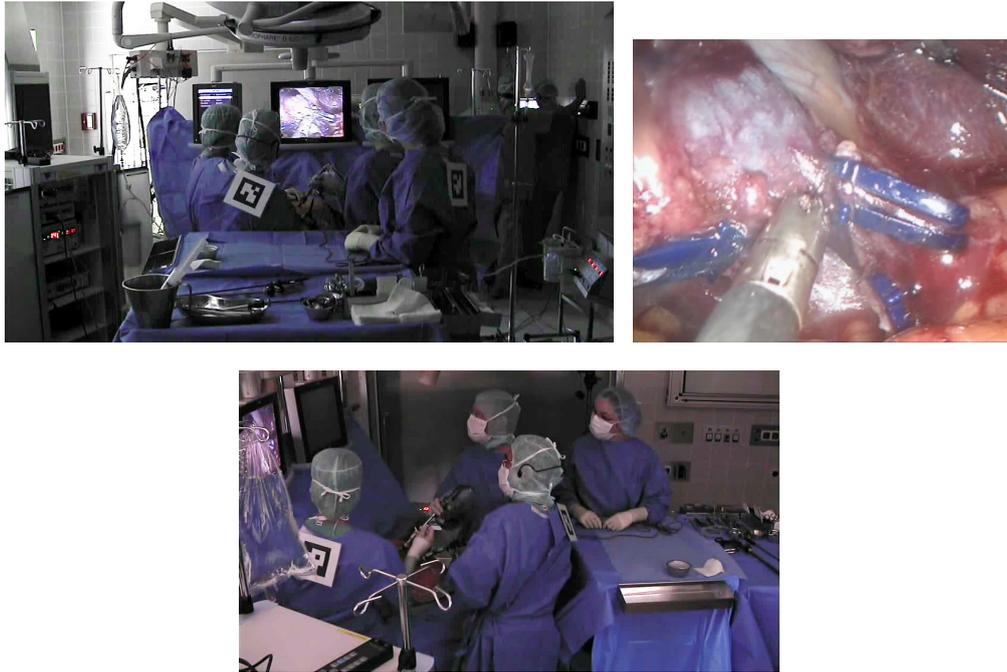
Figure 4.2.: Images from three synchronized video streams.

[Kozak et al., 2006] this is a quite common procedure. Today 95% are performed laparoscopically and so it is one of the most widespread minimal-invasive procedures. 10 laparoscopic cholecystectomies were recorded. All surgeries were performed at the surgical department at the hospital of the Technical University Munich (Klinikum Rechts der Isar, Technische Universiät München). They have been performed by four different surgeons with varying skill levels but from the same medical school.

The whole procedure was recorded using at least two standard camcorders that were mounted on tripods. One of the video cameras was used to record the laparoscopic image. One was used to get an overview of the whole surgery room. In some surgeries one or two additional cameras aimed at the patient from the right and the left side. To synchronize the video streams, an electronic stop watch was recorded at the beginning of the surgery with each camera. An example showing images from three synchronized video streams can be seen in figure 4.2. The videos were analyzed and a multi-dimensional signal representation was generated manually for each surgery. The signals are boolean and represent the use of the seventeen laparoscopic instruments during the surgery. These instruments include the laparoscopic camera, four trocars, which are used to insert other instruments, and several tools that are inserted into the patient. But also high frequency cutting and coagulation, which are both performed using one of the laparoscopic tools, are taken as additional signals. Per second one instrument vector was created. To generate this representation, a software was used, which allows to replay the synchronized videos and manually annotate which instruments are in use. The signals used during one exemplary surgery are shown in figure 4.3. Fourteen phases that occurred in every surgery have been identified and labeled manually in each surgery. A list of these phases is given in table 4.2.

While this data has been labeled manually, it would also be possible to automatically acquire such signals. Some signals such as switching on and off the light or the use of high frequency cutting could be recorded easily from the instruments. First experiments on attaching
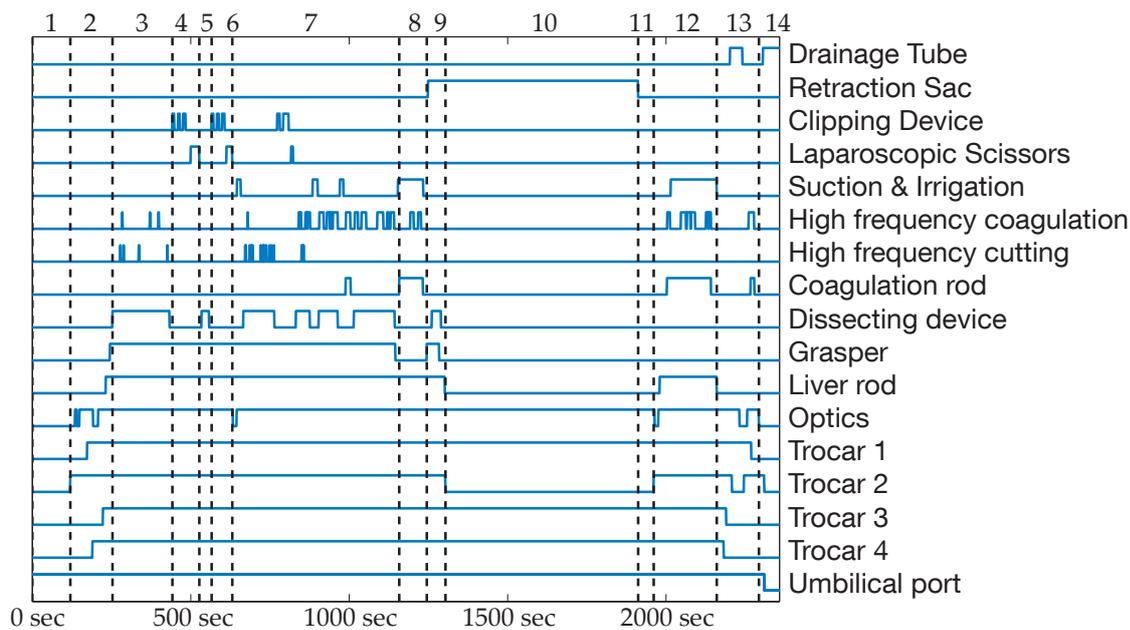
Figure 4.3.: Instrument use during one surgery represented as signal vectors.

barcodes to the instrument and detecting them via a camera that is attached to the trocar have shown promising results. But also other technologies such as RFID chips could be used to detect which instrument is used, as done by [Neumuth and Meißner, 2012] to detect and localize instruments in a surgery room. First work on tracking surgical motion only based on vision [Darolti et al., 2007] and optical range cameras [Yamauchi, 2007] has been conducted and could be used to obtain additional signals. Detection of surgical instruments using color markers has been shown in [Bouarfa et al., 2011a].

Both, the pose of the US probe with respect to the patient phantom, and the instrument use during the laparoscopic cholecystectomies are represented as time series $A = (a_1, \ldots, a_m)$. In the case of the ultrasound tracking data, a_i is a 7-dimensional vector of real-valued numbers where three values are used to represent the position of the probe and four values to represent the orientation using quaternions. In case of the laparoscopic surgery a_i is a 14-dimensional vector representing the use of the different instruments. While for a single instance of a surgery the values of a_i can only be 0 or 1, depending on whether the instrument is present or not, real values are used as later average representations will be computed which can take values between 0 and 1. In the remainder of this chapter both, ultrasound procedures and the laparoscopic surgery will be referred to as procedure. A single example of a procedure will be called instance.

4.3. Dynamic Time Warping

In this section we will describe how DTW can be used synchronize two instanced of a procedure and how to construct a statistical average of a procedure.

DTW is a time invariant similarity measure between two time series. While computing the similarity between two time series also a synchronization between both of them is performed. The method is well known for speech recognition [Sakoe and Chiba, 1978] and has also been

Number	Name	Number	Name
1	CO2 inflation	8	Liver bed coagulation 1
2	Trocar insertion	9	Packaging of gallbladder
3	Dissection phase 1	10	External gb. retraction
4	Clipping / cutting 1	11	External cleaning
5	Dissection phase 2	12	Liver bed coagulation 2
6	Clipping / cutting 2	13	Trocar retraction
7	Gallbladder detaching	14	Abdominal suturing

Table 4.1.: The fourteen phases that have been identified.

used for detection of movements in stroke rehabilitation [Tormene et al., 2009], gesture recognition [Corradini, 2001] and data analysis in bioinformatics [Yuan et al., 2011] and numerous other applications. DTW can be computed efficiently using dynamic programming and can be adapted to many problems just by defining a similarity measure between data points.

Using the DTW algorithm a distance measure between two time series $A = (a_1, \ldots, a_m)$ and $B = (b_1, \ldots, b_n)$ can be computed, while taking into account time variability. DTW can be used with any type of time series as long as a distance measure $d(a_m, b_n)$ between two points of the time series is defined. So it can be used on multivariate time series such as pose trajectories of US probes or signal vectors describing instrument usage during a surgery. While computing the distance, a warping path $w = ((g(1), h(1)), \ldots, (g(K), h(K)))$ is computed where the functions g and h define the mapping between the elements of the time series. K denotes the length of the warping path, which depends on the functions g and h. The warping path must associate each point in A with at least one point in B and vice versa. The warping path w is computed in a way to minimize the DTW distance

$$DTW(A, B) = \sum_{k=1}^{K} d(g(k), h(k)).$$ (4.1)

The DTW distance and the warping path can be computed recursively by

$$DTW(A, B) = d(a_m, b_n) + \\ min(DTW(A_{m-1}, B_{n-1}), DTW(A_{m-1}, B_n), DTW(A_m, B_{n-1})),$$ (4.2)

but it is usually computed using dynamic programming, which takes $O(mn)$, where O describes the limiting behavior following the Landau notation.

An example of this can be seen in figure 4.4. The first four phases of two surgeries are shown, where the dotted lines indicate the phase borders. The warping path shows a synchronization between both surgeries where the dotted lines are again the phase borders. A diagonal warping path shows that both surgeries were carried out at approximately the same speed. There are several vertical parts as one surgery has taken longer and multiple time steps are warped onto one point in the other surgery.

The DTW synchronization can be used for three purposes. First, it can be used to synchronize two instances of the same procedure. Second, synchronizing several instances of a procedure to a common timeline, and averaging them, an average model of the procedure can be obtained. Third, it can be used to synchronize one performance of a procedure to an average model of the procedure. Below, the computation of the average is discusses and in the next section the application of DTW to training and education is discussed.
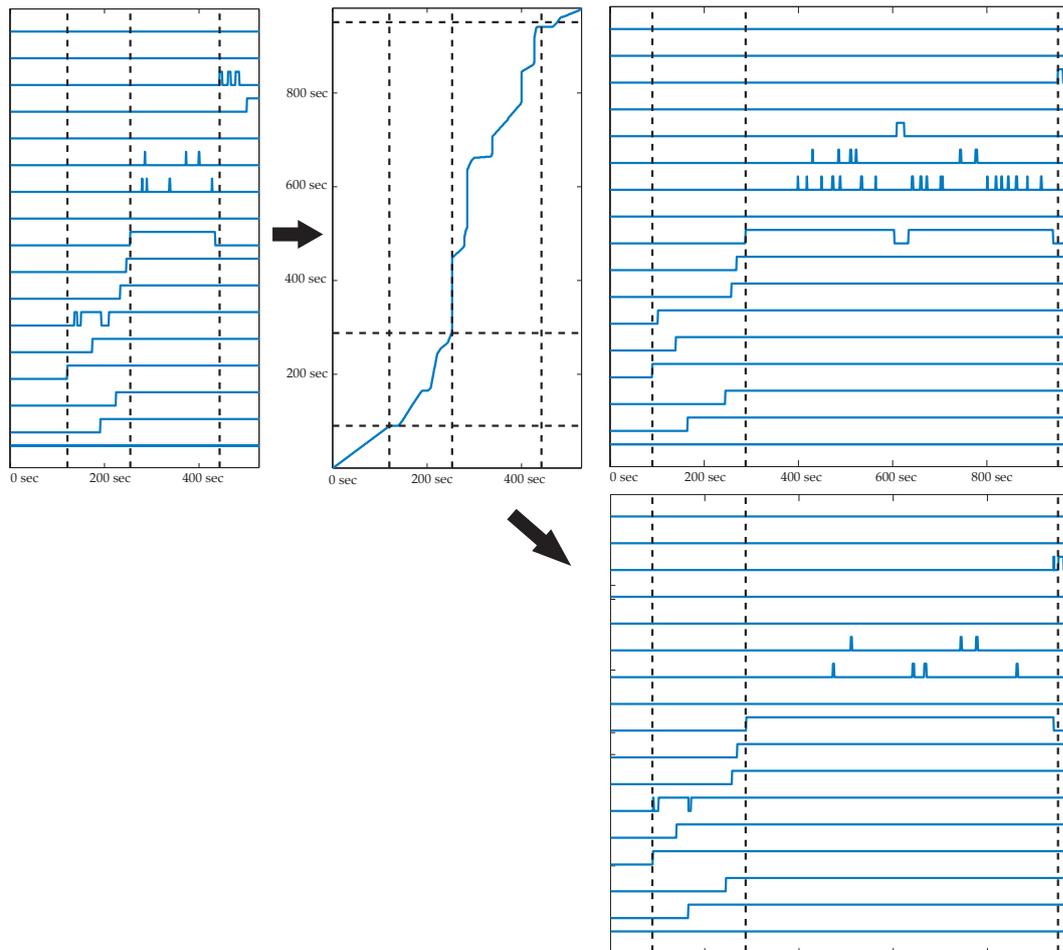
Figure 4.4.: The warping path (middle) synchronizes both surgeries onto each other. The x-axis of the warping path represents the timeline of the left surgery and the y-axis the timeline of the right surgery. The lower image shows the left surgery warped onto the right one using the warping path. As can be seen common actions are warped to the same point in time.

Multiple time series representing the same procedure are synchronized using the dynamic time warping algorithm to generate an average procedure on an average timeline, preserving the average length of the procedures, the phases and the actions. The set of instances that are used to compute the average is called training set. Similarly to [Wang and Gasser, 1997], we use three steps for the computation of the average procedure:

1. Compute an initial reference.

2. Compute the first average procedure.

3. Iterate the average procedure computation using the previous average instance as reference.

There are various ways to choose the initial reference. It could simply be an arbitrary instance. However, when some actions have not been performed in this specific instance, but in one or several of the other instances from the training set, it is likely that these actions are

also not represented correctly in the average. Experiments have shown that the following approach avoids this problem. First, the times series from the training set are averaged pairwise using DTW. The resulting average procedures are then again iteratively merged pairwise and the last one is taken as the initial reference.

The first average procedure is computed by computing the warping path from all procedures to the initial average. The warping paths are normalized and the average of all warping paths is calculated. Doing this, a mapping of an average timing onto the reference is obtained. Using the inversion of the average warping path all surgeries can be mapped onto a common time line and an average surgery can be computed. This averaging procedure is iterated until the average gets stable. An example of such an average is shown in figure 4.5.

Computing the DTW distance between two time series has a quadratic complexity. For the time series that are used in this thesis, DTW can be computed within a couple of seconds. However, the computation of an average of multiple time series involves many DTW computations. Therefore FastDTW, an approximation of the DTW with linear time and space complexity [Salvador and Chan, 2004] has been used. FastDTW performs a multi-scale computation of the DTW matrix. The algorithm starts with computing the DTW matrix at a very coarse level. Iteratively the matrix is computed at the next finer level where only parts of the matrix are computed that are close to the warping path at the higher level.

4.4. Dynamic Time Warping for Medical Training and Education

The first way to use DTW for education is the computation of an average model. Understanding the workflow of a procedure is difficult. A representation as seen in figure 4.5 can help to understand the order and average length of actions, how often an action occurs and the probability of actions to happen. One drawback of DTW is that it can only handle time series with a fixed order. The use of HMMs, which can handle more variability, to create another type of average representation is discussed later.

The second way to utilize DTW is to use it for an after action review (AAR). To goal of an AAR is to provide feedback to a trainee after she performed a procedure. AAR is a method that has been used before for training in particular for military training [Morrison and Meliza, 1999]. One way to provide an AAR is to compare the performance of a trainee and an expert. To achieve this in the US simulator a synchronized replay of the trainee and an expert is shown. This replay allows the student to study differences between both performances. Usually the expert and the trainee perform a procedure with different speed. When simply starting both replays at the same time it would be difficult to see translational and rotational differences. To synchronize the performance of an expert and a trainee DTW is used on the two time series. In figure 4.6, two trajectories of US probes and their synchronization are shown.

Performing an AAR using the HMD-based setup of the US simulator works as follows. Both, when the expert and the trainee perform the US procedure, no HMD is used and only the simulated US image but no additional information is shown. So the simulator acts like a real US machine. Only for the AAR additional information is displayed. The AAR is performed using the HMD. An image of this is shown in figure 4.7. The pose of the US slice of the expert and the trainee are shown by colored frames. During the AAR, the US probe is still tracked and can be used by the student to take different views of the region of interest. So the student can compare the views he has been taking before to the views of the expert. Virtual screens
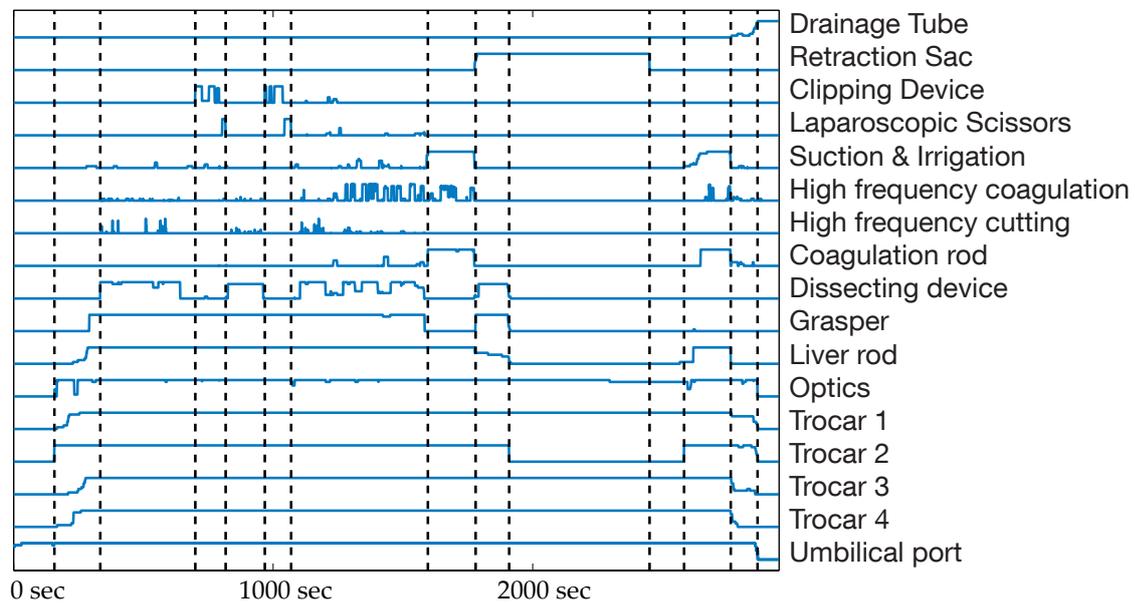
Figure 4.5.: Average laparoscopic cholecystectomy computed by DTW. The values represent the probabilities of an action to occure.
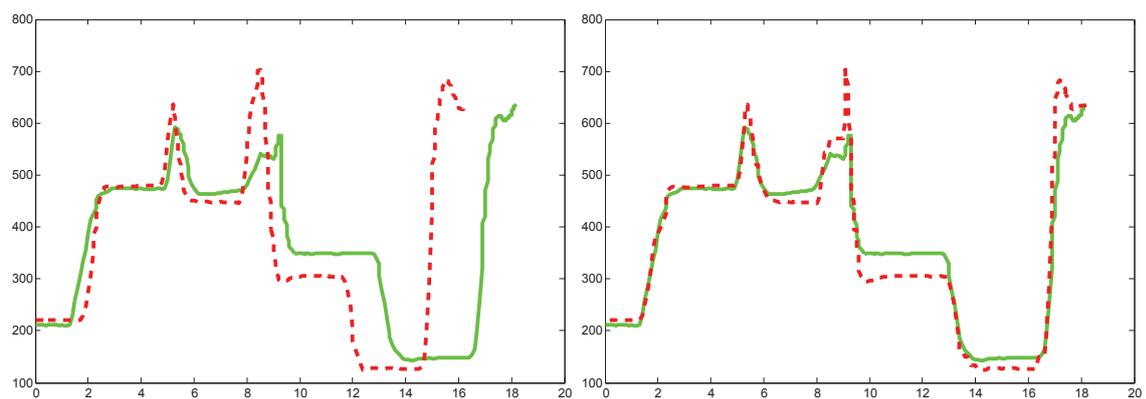
Figure 4.6.: On the left the position of the probe in one dimension for two instances of a US procedure is shown. On the right the same two instances are shown, but they have been synchronized using DTW.
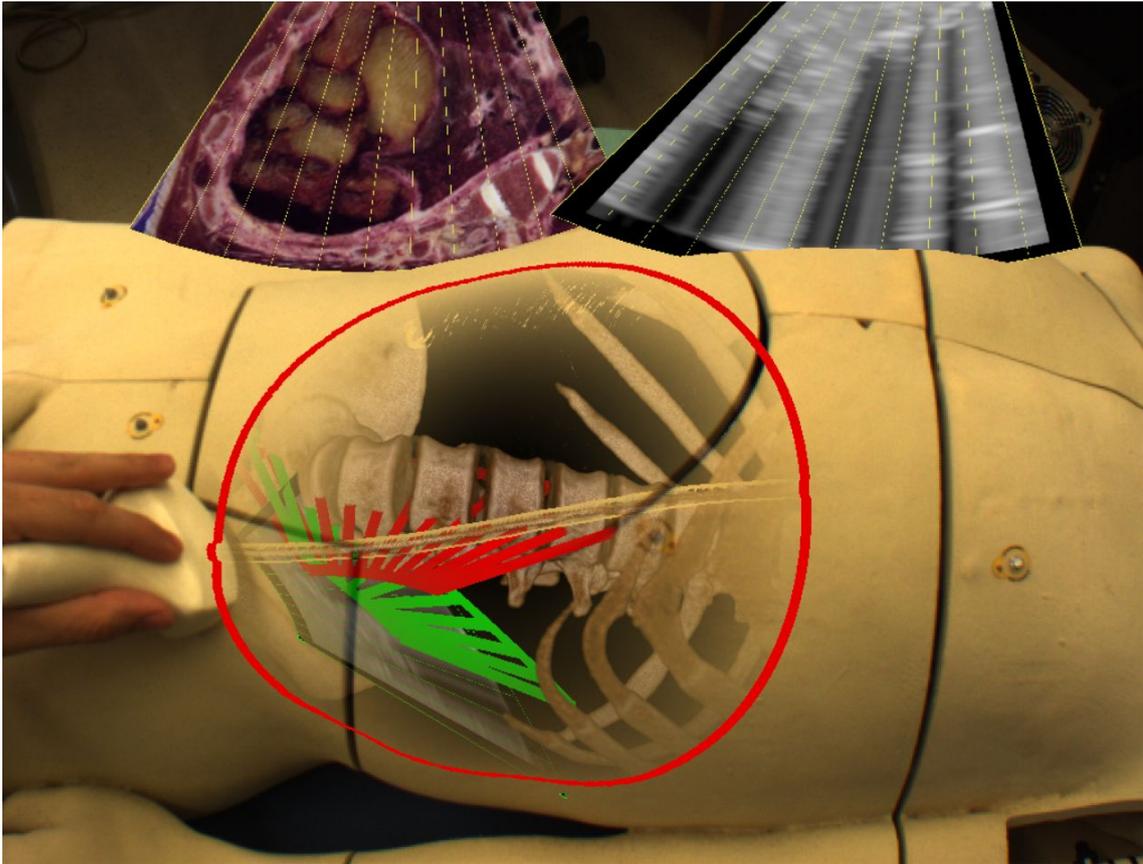
Figure 4.7.: Synchronized replay of expert and trainee. The green frame indicates the pose of the expert's US probe and the red frame the pose of the trainee's probe. The student can still move the US probe and sees the corresponding simulated US and photographic images on the two virtual screens.

can be placed in the scene. These virtual screen are used to show the simulated US, CT or photographic images.

As also discussed in [Quarles et al., 2008], AR has several important advantages over using a simple video AAR, which is the usual way to perform AAR today. AR allows a visualization of the performance in-situ from the perspective of the student instead of only using a monitor. The viewpoint can be changed, which makes it easier to study complicated situations in more detail. Using traditional video AAR, the replay of the student and the expert can only be shown on two monitors, dislocated from each other, while AR allows showing them colocated in one space. And using the DTW synchronization also a temporal colocation is obtained.

Alternatively the AAR can also be done without the HMD, which allows analyzing the performance on any PC without requiring AR hardware. The tracking data can be loaded to display the CT volume and the US slice in a virtual scene. All methods that are used for the AR version, such as in-situ visualization of the US slice in the CT volume and DTW synchronization are also available in the offline version. If the gaze-tracker was used while recording the performance of the expert, as discussed in section 3.2.7, the gaze-information can also be shown to help a trainee to understand the mental model of the expert. An example of the offline version is shown in figure 4.8. The performance of an expert and a trainee are shown synchronized and the CT and US images are displayed.
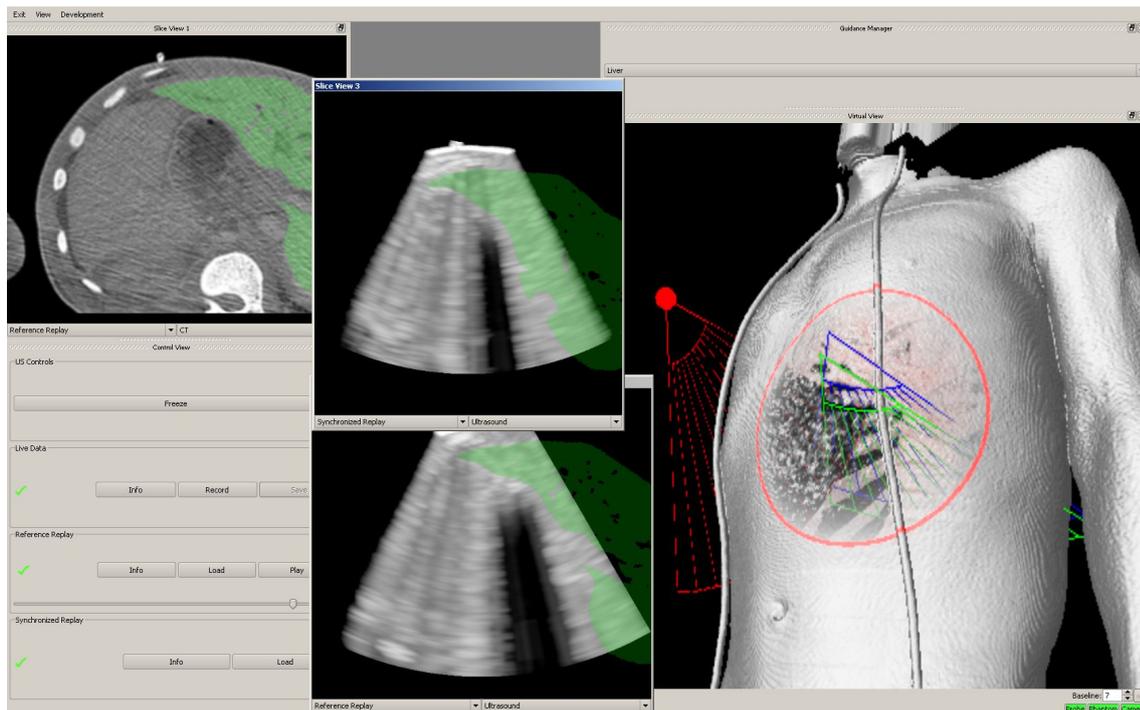
Figure 4.8.: User interface of the US simulator. In the *virtual view* the pose of the probe of a trainee (blue) and an expert (green) are shown, synchronized using DTW. The currently tracked probe is shown in red. Simulated US images of both performances and a CT of one performance are shown with an additional segmentation of the liver.

Instead of synchronizing two performances it is also possible to synchronize the performance of a student or and expert to an average model. Synchronizing the performance of a student to the average can be used similar to the synchronization to an expert so that to student can compare his own performance to the average. A synchronization between the performance of an expert and the average can be used to watch the replay of one instance, while being able to compare it to the average. An example of this, using HMMs instead of a DTW average will be discussed later.

Another way to use DTW for education is by synchronizing multiple videos. This allows to synchronously watch and compare the same scenes from multiple surgeries. For laparoscopic cholecystectomy this has been done by [Sielhorst et al., 2006] using DTW. A method to synchronize laparoscopic videos using visual features and only requiring manually annotation of instrument usage for a subset of the surgeries has been shown in [Blum et al., 2010a].

4.5. Hidden Markov Models

HMMs and similar statistical models are widely used for speech recognition [Gales and Young, 2008], gesture recognition [Elmezain et al., 2008], combinations of both [Nefian et al., 2002], bioinformatics [Alexandersson et al., 2003] and various other applications. It should be mentioned that HMMs share a lot of similarities with DTW. Where DTW computes the distance between an observed time series and a model, the HMM computes the probability of the model having generated the observation. One big advantage of HMMs

compared to DTW is that HMMs can model loops or alternative paths in the workflow.

For education and training, HMMs have a big advantage over other methods to model time series. They are intuitive to use and interpret for humans. Using HMMs, statistical information can be obtained from the data. An HMM can be represented as a graph, which is a very natural way to represent a workflow. Using such a visualization helps humans to understand the topology and embodied statistical information of an HMM. However, for big HMMs it is still challenging to construct HMMs with a topology that makes sense for humans. How to construct such HMMs will be discussed later in this section.

4.5.1. Introduction to HMMs

This introduction follows the notation of the classical HMM tutorial by [Rabiner, 1989]. An HMM consists of the following parts:

- A fixed number of N hidden states $S = \{S_1, S_2, \ldots, S_N\}$ the system can be in. The state at time t is given by q_t.

- A discrete alphabet made of M observation symbols $V = \{v_1, v_2, \ldots, v_M\}$ that can be emitted by each state.

- The state transition probability distribution $A = \{a_{ij}\}$ where

$$a_{ij} = P\left(q_{t+1} = S_j \mid q_t = S_i\right), \qquad 1 \leq i, j \leq N, \tag{4.3}$$

represents the probability of making a transition to state j if the HMM currently is in state i.

- The observation symbol probability distribution $B = \{b_j(k)\}$, where

$$b_j(k) = P\left(v_k \mid q_t = S_j\right), \qquad 1 \leq j \leq N, \quad 1 \leq k \leq M, \tag{4.4}$$

represents the probability of observing the symbol v_k when being in state j.

- The initial state distribution $\pi = \{\pi_i\}$, where

$$\pi_i = P\left(q_1 = S_i\right), \qquad 1 \leq i \leq N. \tag{4.5}$$

Given all parameters $\lambda = (A, B, \pi, N, M)$ the HMM can generate an observation sequence

$$\mathbb{O} = O_1 O_2 \ldots O_T, \tag{4.6}$$

where the observations O_t are taken from V.

We will follow the typical HMM introduction and use the weather as hidden states so that $S = \{Sun, Rain, Snow\}$. We assume that a friend lives at a distant place so that we cannot directly observe the weather at his place. We talk to him over the telephone once a day and he tells us which means of transportation he used to go to work. He chooses between three different options and his choice depends only on the weather. As he tells us which means of transportation he uses, these are our observation symbols $V = \{Bike, Car, Subway\}$. From long term weather statistics we know the transition probabilities between different states a_{ij}. The states and transition probabilities are visualized in figure 4.9. We know the probability that
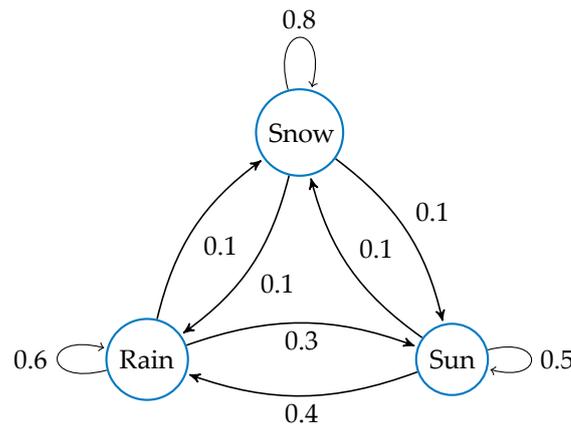
Figure 4.9.: Exemplary HMM showing the hidden states.

our friend chooses a certain means of transportation given the weather. So e.g. the observation symbol probability distribution for the state $Rain$ is fully given by $P\left(Bike \mid Rain\right) = 0.1$, $P\left(Car \mid Rain\right) = 0.6$ and $P\left(Subway \mid Rain\right) = 0.3$. How to automatically build such a model, based on a set of instances, including estimation of the right number of states, transition probabilities and observation symbol probability distributions is discussed later. Using an HMM, there are three typical problems that can be solved for this example. The first problem is to compute the probability of a certain sequence of observations given the model. The second problem is to compute the most probable sequence of states given a sequence of observations. The third problem is how to compute the parameters of the model, given one or several observation sequences. All three problems will be explained in more detail below. For education and training the third problem is the most relevant, as we are interested in obtaining a model of the average workflow, given a set of instances of the workflow.

4.5.1.1. Problem 1

The first problem is, given an observation sequence $\mathbb{O}$ and a model λ, how to compute the probability $P\left(\mathbb{O} \mid \lambda\right)$ that the model has generated this observation? This problem has to be solved when HMMs are used for classification. Having multiple HMMs it can be estimated for which one of them the probability is highest that it has generated the observed data. The problem can be solved efficiently using the forward-backward procedure. The forward variable $\alpha_t\left(i\right)$ gives the probability of being in state S_i at time t and having produced the output sequence $O_1 O_2 \ldots O_t$. It is defined as

$$\alpha_t\left(i\right) = P\left(O_1 O_2 \ldots O_t, q_t = S_i \mid \lambda\right), \tag{4.7}$$

and the probability of having generated the whole sequence is defined as

$$P\left(\mathbb{O} \mid \lambda\right) = \sum_{i=1}^{N} \alpha_T\left(i\right). \tag{4.8}$$

This can be computed efficiently using a dynamic programming approach, similar to the one used for DTW. After initializing

$$\alpha_1\left(i\right) = \pi_i b_i\left(O_1\right), \qquad 1 \leq i \leq N, \tag{4.9}$$
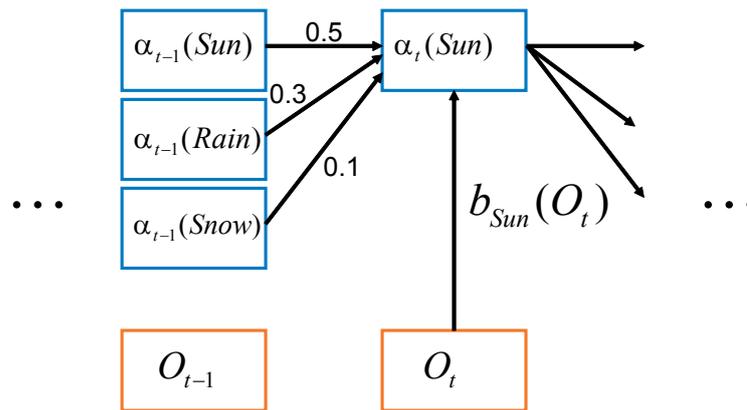
Figure 4.10.: Computation of the forward variable using dynamic programming.

$\alpha_2 (j) \dots \alpha_T (j)$ can be computed by induction using

$$\alpha_{t+1} (j) = b_j (O_{t+1}) \left[\sum_{i=1}^{N} \alpha_t (i) \, a_{ij} \right], \qquad 1 \leq t \leq T - 1, \quad 1 \leq j \leq N. \qquad (4.10)$$

The forward-algorithm is visualized for our example in figure 4.10. In each step t all $\alpha_t (i)$ are computed. Each of them depends on two parts. The probability of having been in any of the other states at $t - 1$ and doing a transition to state i. And on the probability of observing the current observation symbol. So $\alpha_t (i)$ is the probability of being in state i after having observed $O_1 O_2 \dots O_t$. The backward probability $\beta_t (i)$ is defined similarly as the probability of observing the rest of the observation sequence when being in state i,

$$\beta_t (i) = P (O_{t+1} \dots O_T \mid q_t = S_i, \lambda) . \qquad (4.11)$$

4.5.1.2. Problem 2

The second problem is, given an observation sequence $\mathbb{O}$ and a model λ how to compute the most likely sequence of states $Q = q_1 q_2 \dots q_T$ that has produced the observation sequence. This problem can be solved by modifying the forward-backward algorithm. For problem 1 the probability of entering state j at time t was computed by summing up the probabilities of having been in any other state at $t - 1$ and making a transition to state j. Now, the probability of the most likely path that enters state j at time t is computed. This is done by looking at the most likely path that entered state i at $t - 1$ and compute the probability that it will enter j at time t. Out of all states i the most probable of these paths is taken. So equation 4.10 is replaced by

$$\alpha_{t+1} (j) = b_j (O_{t+1}) \, max_{1 \leq i \leq N} \left[\alpha_t (i) \, a_{ij} \right], \qquad 1 \leq t \leq T - 1, \quad 1 \leq j \leq N. \quad (4.12)$$

Furthermore for each state the predecessor is stored. This is used later for backtracking. Starting at the last state, and iteratively going to the predecessor of the current state, the most probable state sequence is obtained. The computation using dynamic programming is very similar to the forward algorithm shown in figure 4.10. The solution of this problem is known as Viterbi algorithm and the path as Viterbi path. For some problems the Viterbi path is used as an approximation to the solution of problem one, which will also be done later in this work.

4.5.1.3. Problem 3

The last problem is, given a set of instances $\Omega = \{\mathbb{O}_1, \mathbb{O}_2, \ldots, \mathbb{O}_L\}$, where L is the number of instances, how to adjust the model parameters A, B and π to maximize $P(\Omega \mid \lambda)$. There are two different solutions to this problem. When training data is labeled so that for each observation symbol the state that produced it is known, parameters can be estimated directly by counting the number of transitions and observations. Usually data is not labeled as this would require extensive manual work. When unlabeled data is used, usually the expectation maximization (EM) algorithm is used. Below, only the basic idea of this algorithm is described. For details, the interested reader is referred to [Rabiner, 1989].

Given an observation sequence $\mathbb{O}$ it is assumed that it has been generated by the model λ. Using the forward and backward variables for each state the probability of having generated each observation is computed. This is used to update the transition probability distribution by

$$\bar{b}_j(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j}, \qquad (4.13)$$

where $\bar{b}_j(k)$ is the updated observation symbol probability distribution for state j. Transition and initial state probabilities are updated similarly.

One drawback of the EM is, that the HMM topology must be known. The number of states is fixed and transition probabilities that are zero will not be changed. For modeling of medical workflows this is a problem, as the number of states is not known beforehand. Solutions to estimate the topology of an HMM from the data are discussed later in chapter 4.5.2.

As the three problems described above will be used later, to estimate a HMM from a set of workflow instances, their computational complexity will be discussed briefly. In each time step t and state n the probability of being in n at t must be evaluated. This takes $O(N)$ each as the probability of being in any state at time $t-1$ and making a transition to state n have to be summed up. So the computational complexity of the first two problems is $O(N^2 T)$. The EM algorithm is in the order of $O(N^2 T L)$.

In our case, the transition probability for most pairs of states is 0. Instead of filling up the matrix that is shown in figure 4.10 it is reasonable to implement the HMM based on a graph structure. Here the states are represented as nodes and the transition probabilities as edges. Incoming and outgoing edges can be seen as parent and child relations between states/nodes. In each node all parent and child nodes are stored. For the child nodes also the transition probability is stored. The graph is very sparse and the number of ingoing edges is one or two for most states and did never exceed ten. Assuming the maximal number of parents per state to be a constant, computational complexity reduces to $O(NT)$ for the first two problems and to $O(NTL)$ for the EM algorithm.

4.5.2. Simple HMM Topologies

HMMs can have different topologies. A very simply topology is a left-to-right HMM, where every state has exactly one predecessor and one successor. The HMM states can only be visited in a fixed order and it is not possible to return to a state once it is left. For the laparoscopic surgery an example of such a left-to-right topology is illustrated in 4.11. For each of the 14 phases one state has been created and the observation symbol probability distribution is estimated from all samples of that state. For training and education this model is not very useful, as it only provides information on the underlying workflow on a very coarse level. From the model it can e.g. be seen that the clipping device is used more often in phase 3 than in phase

4, but the model does not reveal that usually it is used three or four times in state 3. While for the laparoscopic surgery it is obvious that using 14 HMM states is a reasonable choice to represent the 14 phases, for modeling the movement of an US probe it is not clear how many states the HMM should have. Therefore, more advanced methods are required the generate the HMM toplogy.

For most applications, HMMs with a higher number of states are used, where each state can have multiple predecessors and successors and the model can contain loops. An important problem is, how to construct such a HMM, as it is not obvious how many states such a HMM should have and how they should be connected. The standard way of constructing such a HMM is as follows. A fully connected HMM λ_i is initialized with random values and the EM algorithm is used to find a local maximum of $P\left(\Omega \mid \lambda\right)$. If additional information is available this can be used while training the model. For the data from the laparoscopic surgery, the phases are known. Therefore first sub-HMMs for each phase can be training, which are concatenated afterwards. An example of such a HMM is illustrated in figure 4.12. For data from the US simulator it is not obvious how to partition the time sequences into phases to train corresponding sub-HMMs.

While a model that has been constructed using a random initialization can be useful for task as recognizing the current state of a workflow, this kind of model is not suited for training and education. In most cases the HMM will not be easy to interpret by humans as different activities might be represented by one state and the HMM might contain loops where in the real workflow no loops exist.

Another model that can be derived directly from the training data would have one state for each distinct observation. The observation symbol probability distribution can be set directly to 1 for the corresponding observation. Figure 4.13 shows such a model representing one phase of the laparoscopic surgery. As each state can only produce one observation symbol, this can also be seen as a Markov chain. To see the difference to a left-to-right HMM with one state per phase we look at the example in figure 4.13 where a model of phase 4 is shown. In this model it is explicitly represented that the scissors must be used at the end of this phase. Looking at the left-to-right HMM in figure 4.11 it can be seen that in this model, this is not represented explicitly. Therefore it is much more difficult to interpret this model.

For data with a limited set of different observation symbols, such as the laparoscopic surgery, such a model can be constructed directly from the data. When using such a model to represent the movement of an US probe, a quantization method must be applied to the pose data to reduce it to a limited number of observation symbols. The appearance of the model will largely depend on the quantization method. Furthermore, the model is overfit to the training data. As overfitting and generalization of models will be very important in the next section, it is discussed shortly here. Overfitting occurs when a model has too many parameters compared to the amount of training data. While the model perfectly explains the training data it does not generalize well over new data. This is visualized for the problem of curve fitting in figure 4.14. In this example ten data points have been generated by adding Gaussian noise to a sinus curve. Using least-squares curve fitting, curves with different order have been generated. As can be seen, the curve of order one is to general to describe the data points. While the curve of order nine exactly explains the observed points, it will not generalize well over the rest of the curve. So a trade-off between a complex model that exactly describes the data and a more general model must be made.

Another motivation for using a different model shall be illustrated at the example of figure 4.13. For humans, one of the most obvious characteristics of phase 4 of the laparoscopic surgery

Legend labels (for the example phase waveforms shown above the table):

- Clipping Device
- Laparoscopic Scissors
- Suction & irrigation
- High frequency coagulation
- High frequency cutting
- Coagulation rod
- Dissecting Device
- Grasper

Instrument / Phase	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Drainage Tube	1	0	0	0	0	0	0	0	0	0	0	0	0.2098	0
Retraction Sac	0	0	0	0	0	0	0	0	0.8143	1	0.0008	0	0	0.9864
Clipping Device	0	0	0	0.3183	0	0.3670	0.0155	0.03380	0.0097	0	0	0	0	0
Lap. Scissors	0	0	0	0.1750	0.0092	0.2336	0.0410	0	0	0	0	0	0	0
Suction & Irrig.	0	0.0370	0.0233	0.0558	0.0635	0.0200	0.0971	0.9088	0.0271	0	0	0.8017	0.0797	0
HF Coagulation	0	0	0.0264	0.0221	0.0210	0.0019	0.2351	0.3509	0.0074	0	0	0.1806	0.0950	0.0293
HF Cutting	0	0	0.0252	0.0118	0	0	0.0374	0	0	0	0	0	0	0
Coagulation Rod	0	0	0	0.0221	0.0354	0.0463	0	0.8074	0.7584	0	0.1372	0.5903	0	0
Dissecting De.	0	0	0.9348	0.0096	0.7047	0.0305	0.5971	0.0944	0	0	0.0017	0.1075	0	0
Grasper	0	0.1076	1	1	1	1	0.9658	0.0080	0.9622	0	0.0017	0.0047	0	0
Liver Rod	0	0.1964	1	1	1	1	1	0.9828	0.4848	0.0017	0.0126	0.8711	0	0
Optics	0	0.7628	0.9941	0.9750	0.9817	1	0.9843	1	1	0.9557	0.9897	0.9509	0.9449	0
Torcar 1	0	0.6503	1	1	1	1	1	1	1	1	1	1	0.7210	0.0052
Torcar 2	0	0.2708	1	1	1	1	1	1	1	0.0003	0.0016	1	0.8968	0.0335
Trocar 3	0	0	1	1	1	1	1	1	1	1	1	1	0.4197	0
Trocar 4	0	0.5201	1	1	1	1	1	1	1	1	1	1	0.2145	0
Umbilical Port	0.9753	1	1	1	1	1	1	1	1	1	1	1	0.9672	0.0366

Transition probabilities of the 14-state left-to-right HMM:

State	Self-loop	Transition to next
1	0.9940	0.0060
2	0.9945	0.0055
3	0.9977	0.0023
4	0.9905	0.0095
5	0.9935	0.0065
6	0.9906	0.0094
7	0.9984	0.0016
8	0.9947	0.0053
9	0.9919	0.0081
10	0.9971	0.0029
11	0.9922	0.0078
12	0.9948	0.0052
13	0.9884	0.0116
14	1.0000	—

Figure 4.11.: For each of the fourteen phases the probability of the instruments being used is shown. Below, the transition probabilities of an 14-state left-to-right HMM are visualized. For comparison, an example of two of the phases taken from one instance of the surgery are shown above.
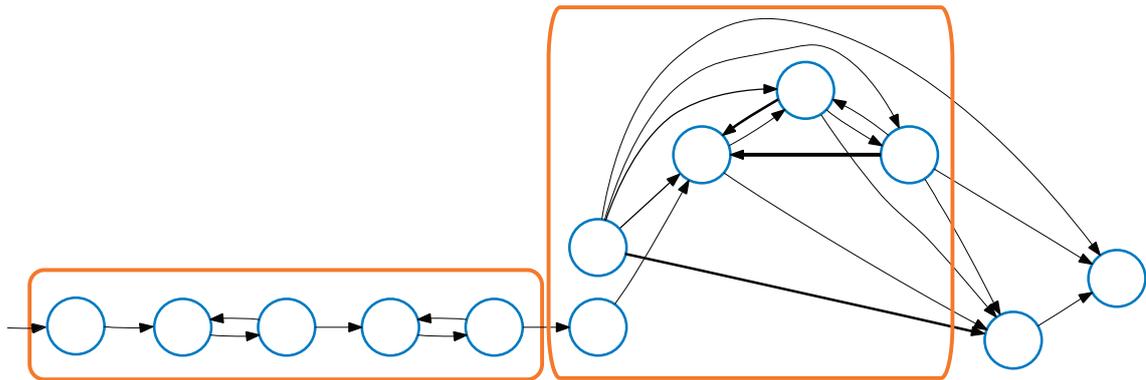
Figure 4.12.: Phases two and three of a model of the MIS, constructed using EM on five-state fully connected HMMs. Phase two is very sequential, which is reflected in the transitions. Phase three is of a less sequential nature.
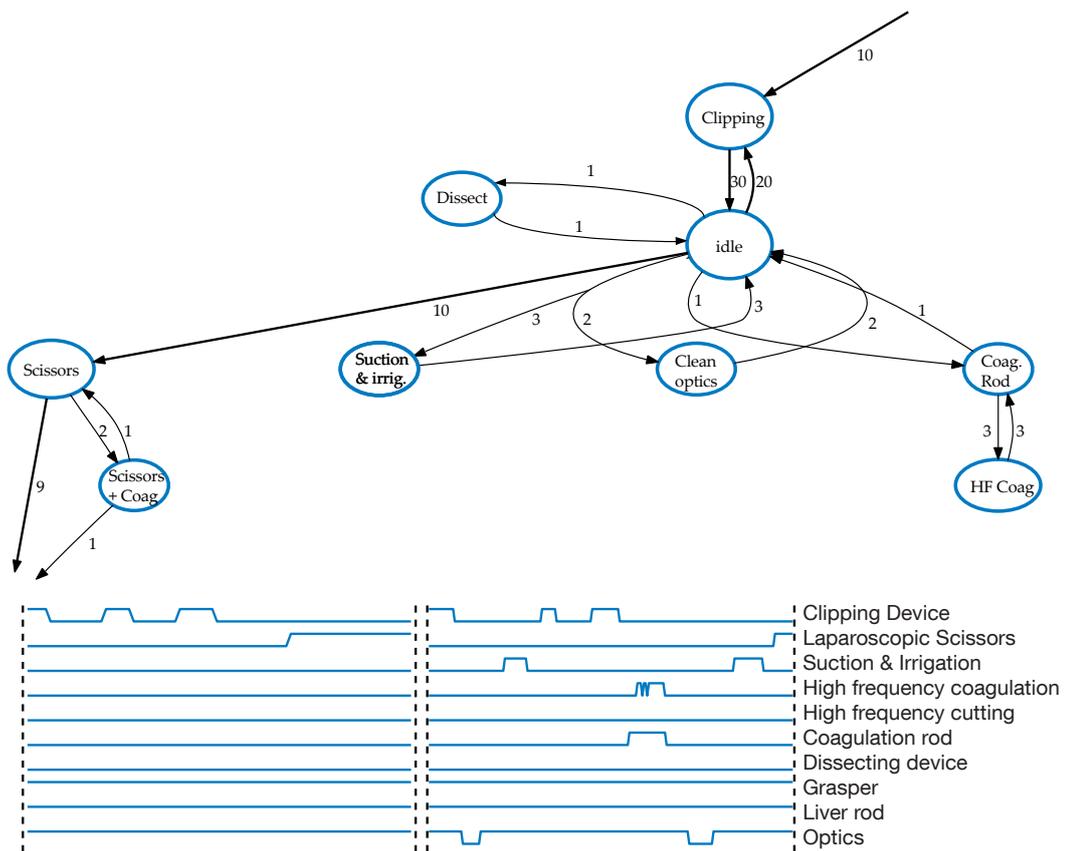
Figure 4.13.: HMM for phase 4 built from ten surgeries. For better conciseness the states are labeled with the instruments that are used in this state and are not used in the others. Instead of showing the transition probabilities, the number of transitions to other states are shown. Below two examples of this phase are shown. The left image shows a typical example of the phase, the right one is an exception. Both of them are represented in the HMM.
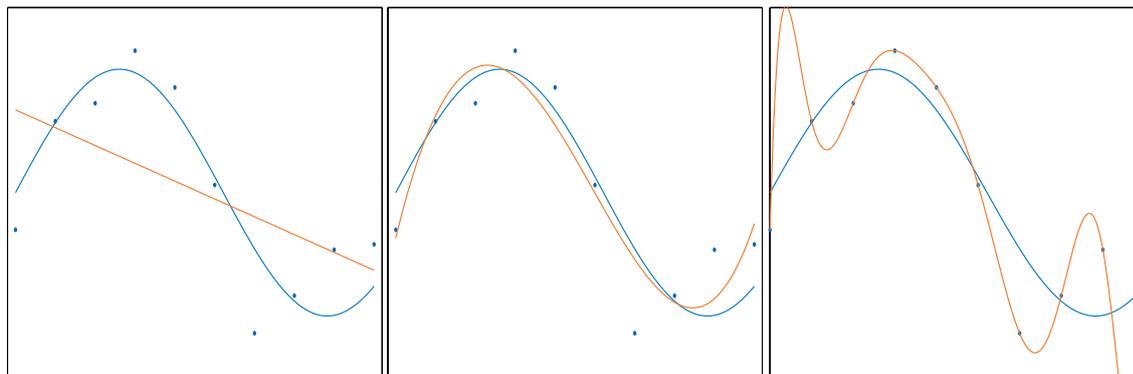
Figure 4.14.: The blue points have been generated by adding noise to the blue curve. The orange curves have been fitted to the data points. The left shows a curve of order one, which is underfitted. The middle curve is of order 4 and generalized well. The right curve is of order 9 and can perfectly fit all data points but is obviously overfitted.

is that the clipping device is used exactly three times. In eight of the ten training surgeries it was used three times while in one surgery it was used only used once and in one example four times. In the HMM this is modeled by the transition probabilities. But as first-order HMMs can only remember the last state, an HMM does not know whether the *clipping* state has already been visited several times. So a sequence of the actions *clipping, idle, scissors* has a higher probability than *clipping, idle, clipping, idle, clipping, idle, scissor* which happens more frequently. Here, action stands for several subsequent observations of the same kind. When generating a workflow model for training, we want that the model represents how often the instrument is used in a more explicit way. Such a workflow model cannot be derived trivially from the training data. How this problem can be solved is the topic of the next sections.

4.5.3. HMM Model Merging

Only few works have dealt with automatically deriving a model from the training data. Two approaches are model splitting and model merging. In this section the use of model merging to represent medical workflows will be discussed and afterwards the use of model splitting. Model merging is used to generate models representing laparoscopic cholecystectomy and model splitting for US workflows. The methods for model merging described in this section are mostly based on work presented in [Stolcke and Omohundro, 1994b] and [Stolcke and Omohundro, 1994a].

4.5.3.1. Model

The starting point is a model that can be obtained trivially from the workflow instances. A set of workflow instances $\Omega = \{\mathbb{O}_1, \mathbb{O}_2, \ldots \mathbb{O}_L\}$ is given. The initial model λ_0 is made of one state for each single observation of each instance. The transition probability for the next state is set to 1 and all others to 0. Using a sequence of *clipping* and *idle* observations the initial model is illustrated at the top of figure 4.15. For the first observation of each workflow instance the initial state probability is set to $\frac{1}{L}$. In other words, the model is built by adding one path per instance of the workflow where each path consists of one state per observation. To simplify the illustration in figure 4.15, only one observation of each action is used here. Using
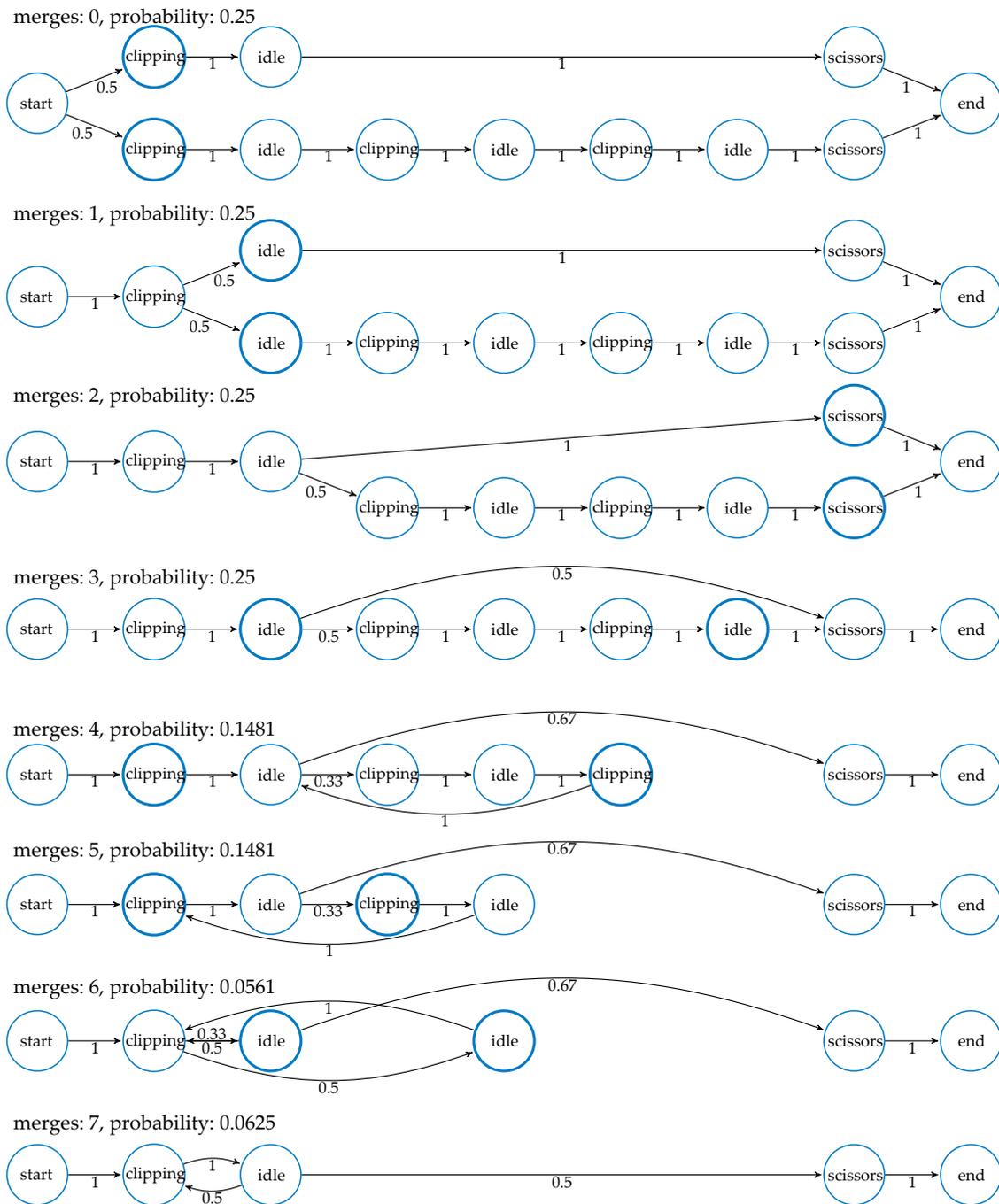
Figure 4.15.: The initial model has been generated from the sequences *{clipping, idle, scissors}* and *{clipping, idle, clipping, idle, clipping, idle, scissors}* . In each step two states are merged The states that are marked indicate that they are merged in the next step.

the real data there would be e.g. 20 subsequent states of *clipping* if the clipping action had a duration of 20 seconds and the sampling rate is 1Hz. This initial model is the model with the highest likelihood given the workflow instances, as it exactly explains each workflow instance. Comparing it to the curve fitting example in figure 4.14, this model corresponds to the curve of order 9 that exactly describes the data. Except for the special case that two workflow instances are identical, for each $\mathbb{O}_i \in \Omega$ the probability $P\left(\mathbb{O}_i \mid \lambda_0\right)$ is exactly $\frac{1}{L}$.

Obviously this model is extremely over-fitted. If such a model is created from a set of work-flow instance, the model will be huge and of no practical use. To generate a more compact workflow model we will iteratively merge two states. In each step we will create the model λ_{i+1} by merging two states from λ_i.

The initial model λ_0 does exactly explain the set of workflow instances. Merging states of the HMM, the probability $P\left(\Omega \mid \lambda_{i>0}\right)$ will generally decrease, as the model will not explain the set of workflow instances exactly. This can be seen in figure 4.15. This conforms to the example in figure 4.14 in the last section. A model that is more compact and generalizes better will be worse in explaining the data. One solution to solve the over-fitting problem in the example of curve fitting is to start with a high order curve and then iteratively reduce the order of the curve as long as the curve still explains the data points sufficiently. The model merging approach for HMMs follows the same idea.

4.5.3.2. Merging

Given a model λ_i, λ_{i+1} is generated by merging two states. This involves the following steps:

- Remove both states that are merged and replace them with a new state.

- All transitions to one of the deleted states are redirected to the new state. The outgoing transition probabilities of the new state are computed by merging the transition probabilities of the old states and normalizing them to 1.

- Also the observation symbol probability distribution of the new state is computed by merging the observation probabilities of the old states and normalizing them to 1.

The two states that are merged in each step are determined by a best-first heuristic. So in each step out of all possible merging candidates the two states are merged that give the highest probability $P\left(\Omega \mid \lambda_{i+1}\right)$. The merging is stopped based on a stopping criterion that will be discussed later.

4.5.3.3. Complexity and Implementation

The main problem when merging an HMM is the computational complexity. For an initial HMM with N states, $O\left(N\right)$ merging steps have to be done. In each step we have to evaluate each possible pair of states. The number of possible pairs is in $O\left(N^2\right)$. After merging two states, the transition and observation probabilities of the HMM have to be refined using EM, which takes $O\left(N^2 TL\right)$. Afterwards $P\left(\Omega \mid \lambda_i\right)$ has to be computed, which takes $O\left(N^2 TL\right)$ and is therefore of the same complexity as the EM step. So the overall complexity is $O\left(N^5 TL\right)$, which is far from being computable in reasonable time. So, several modifications and approximations are used to speed up the computation.

Again, most transition probabilities are 0 so that a graph implementation of the HMM saves time. One problem is the size of the initial model λ_0. The length of a surgery is typically around
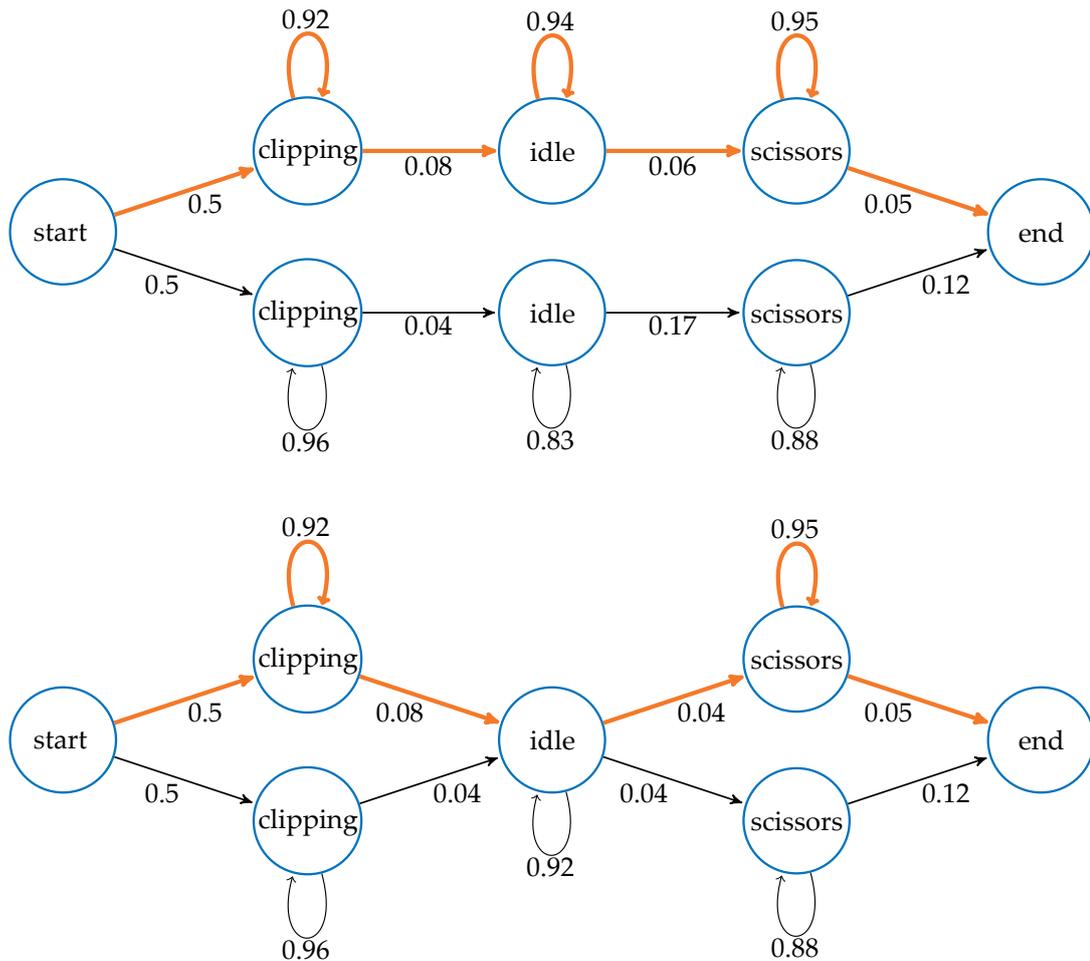
Figure 4.16.: This model has been created by generating only one state for multiple subsequent observation symbols. The two observation sequences that have been used to create this HMM are: *{26∗clipping, 12∗idle, 7∗scissors}* and *{12∗clipping, 17∗idle, 18∗scissors}* . The orange path is a Viterbi path. By merging two states (lower image), it is assumed that the Viterbi path is not changed.

50 minutes. When sampling at 1Hz we would have to create about 3000 states per training surgery. The size of the initial model λ_0 is reduced by not generating one state per second. Instead subsequent observations of the same symbol are grouped. The probability of staying in the state is then given by

$$a_{ii} = \frac{\text{number of repetitions of the same observation symbol}}{\text{number of repetitions of the same observation symbol} + 1}, \tag{4.14}$$

and the probability of making a transition to the next state is the complementary. For the tracked ultrasound probe this modification cannot be applied straightforward, as the number of observation symbols is much higher and a quantization would be required before.

The next modification is more complicated and involves several steps. The purpose is to simplify the computation of $P\left(\mathbb{O} \mid \lambda_{i+1}\right)$ given $P\left(\mathbb{O} \mid \lambda_i\right)$. After merging two states the first step would be to update the parameters of the HMM using EM. Assuming that a small change in the topology will not change the parameters much, this step is omitted. Next, instead of solving the first problem for HMMs and compute $P\left(\mathbb{O} \mid \lambda_{i+1}\right)$ this value is approximated

by solving the second problem and computing the most probable sequence of hidden states $Q = q_1 q_2 \ldots q_T$ given $\mathbb{O}$, known as Viterbi path. This is an approximation that is used quite often for different problems. When solving the first problem the probability of all possible sequences of hidden states is summed up. The approximation is justified by the assumption that the probability of all paths is governed by the most probable path and so all others can be neglected. An example of this is given in figure 4.16. This approximation alone does not save any time as computing the Viterbi path is just a slight modification of the first problem. Instead it is combined with another approximation called optimistic path updating. It is assumed that the only part of the Viterbi path that is affected by the merging operation are the two merged states. This can be seen in figure 4.16. So instead of having to re-estimate the Viterbi path, only the merged states are taken in account when approximating $P\left(\mathbb{O} \mid \lambda_{i+1}\right)$.

Using this approximation the search for the next merge candidates is done as follows. First for each workflow instance $\mathbb{O}_l = O_1 O_2 \ldots O_T$ the Viterbi path $Q_l = q_1 q_2 \ldots q_T$ is estimated. The probability of the Viterbi path is given by

$$P\left(\mathbb{O}_l, Q_l \mid \lambda_i\right) = \pi_{q_1} b_{q_1}\left(O_1\right) \prod_{j=1}^{T} a_{q_j q_{j+1}} b_{j+1}\left(O_{j+1}\right). \tag{4.15}$$

By decomposing the probability of the Viterbi paths the contribution of each state and each transition to this probability can be estimated. When merging two states the new probability $P\left(\mathbb{O}_l, Q_l \mid \lambda_{i+1}\right)$ can be computed efficiently by replacing all observation probabilities and transition probabilities in the Viterbi path that are affected by the merge.

Using these approximations the complexity of the model merging algorithm is $O\left(N^2 T L\right)$. Still, $O\left(N\right)$ merging steps are performed. In each step $P\left(\mathbb{O} \mid \lambda_i\right)$ is computed which takes $O\left(NTL\right)$ using a graph implementation. In each step additionally $O\left(N^2\right)$ possible pairs of states have to be evaluated. The initial model consists of at most one state per observation and so $N < TL$. Therefore $O\left(N^2\right) < O\left(NTL\right)$. The evaluation of each pair can be assumed to be fixed as the number of observation symbols and transitions per state is limited for the data from the laparoscopic surgery.

4.5.3.4. Model Generation

In this section, the method to generate a model of a surgery using model merging is described. The model is first build for each phase independently and afterwards it is concatenated. Doing this takes the following steps:

- Build a model of each phase using model merging.

- Refine the observation and transition probabilities of each phase using EM.

- Concatenate the model of all phases.

- Update observation and transition of the concatenated model using EM.

So, first for each phase a model is created based on all surgeries. This is done as described in the last section. Next, states of the model are merged iteratively. Only merges that do not create loops in the HMM are allowed. This is done to maintain an explicit representation of repeated events in the topology. E.g. in phase four, the clipping device is usually used three times. In the HMM shown in figure 4.13 this is not represented in the topology but only in the transition probabilities, which is a very weak modeling. An HMM without loops that was

derived from the same data is shown in figure 4.17. To avoid loops only states may be merged that cannot reach each other or that can only reach each other directly by one transition. This is checked using breadth first search in both directions, where direct transitions between source and target state are ignored. Doing this, the complexity of the algorithm increases to $O(N^3TL)$.

The point to stop merging is found by using a threshold value on $P(\mathbb{O} \mid \lambda)$. In figure 4.14 it can be seen that a curve of order 9 explains the training data exactly. A curve of order 4 still describes the data very well, although the model is far less complex. When using curves of lower order, the probability of having generated the training points reduces very fast. A similar behavior can be seen when looking at the probability during merging of an HMM in figure 4.18. Here the probability of having generated the training data is plotted against the number of merges. At the beginning, the probability reduces very slowly. Later it drops very fast. Merging is stopped when the derivative of this curve exceeds a threshold. Looking at some examples, the merging procedure was inspected manually by looking at each merging step. It could be seen that the slope in the curve coincides with merges that are subjectively unreasonable.

The next step is to refine the observation and transition probabilities. As explained in the last chapter the merging process involves approximations. So the observation and transition probabilities will not have the maximum likelihood, given the observed data. To address this problem the EM algorithm is run several times until the HMM parameters are stable. The initial model λ_0 was a global maximum for the training data. While the merging algorithm uses some approximations, the resulting HMM is still a very good initialization for EM and so it usually takes only two or three iterations until convergence.

Now, the phases are concatenated. Already for the last steps two additional states have been added to the model. A *start* state, where

$$a_{start\,i} = \pi_i, \tag{4.16}$$

for all i. So it is just another representation of the initial state distribution π. Furthermore an *end* state is introduced, which accounts for the probability that an observation sequence ends in a state. So the transition probability from each state i to the *end* state is defined by

$$a_{i\,end} = \frac{\text{expected number of observation sequences ending in state } i}{\text{expected number of times in state } i}. \tag{4.17}$$

When concatenating the HMMs of two phases, the transition probabilities for one state i from the first phase and one state j from the second phase are recomputed by

$$a_{ij} = a_{i\,end}a_{start\,j}. \tag{4.18}$$

4.5.4. Successive State Splitting

When using data from the US simulator, model merging has the drawback that the generation of the initial model is not straightforward. For the vectors representing instrument usage, building the initial model is done by creating a new HMM state every time the usage of one instrument changes. For the pose data from the US probe, it is not as easy. The tracking system tracks with sub-millimeter resolution, resulting in a huge number of different poses. In practice at every time step, the pose of the tracked probe will be different. As mentioned before one solution would be to do a quantization on the tracking data. Instead of doing this another method to directly infer a HMM topology from the training data is used, namely successive
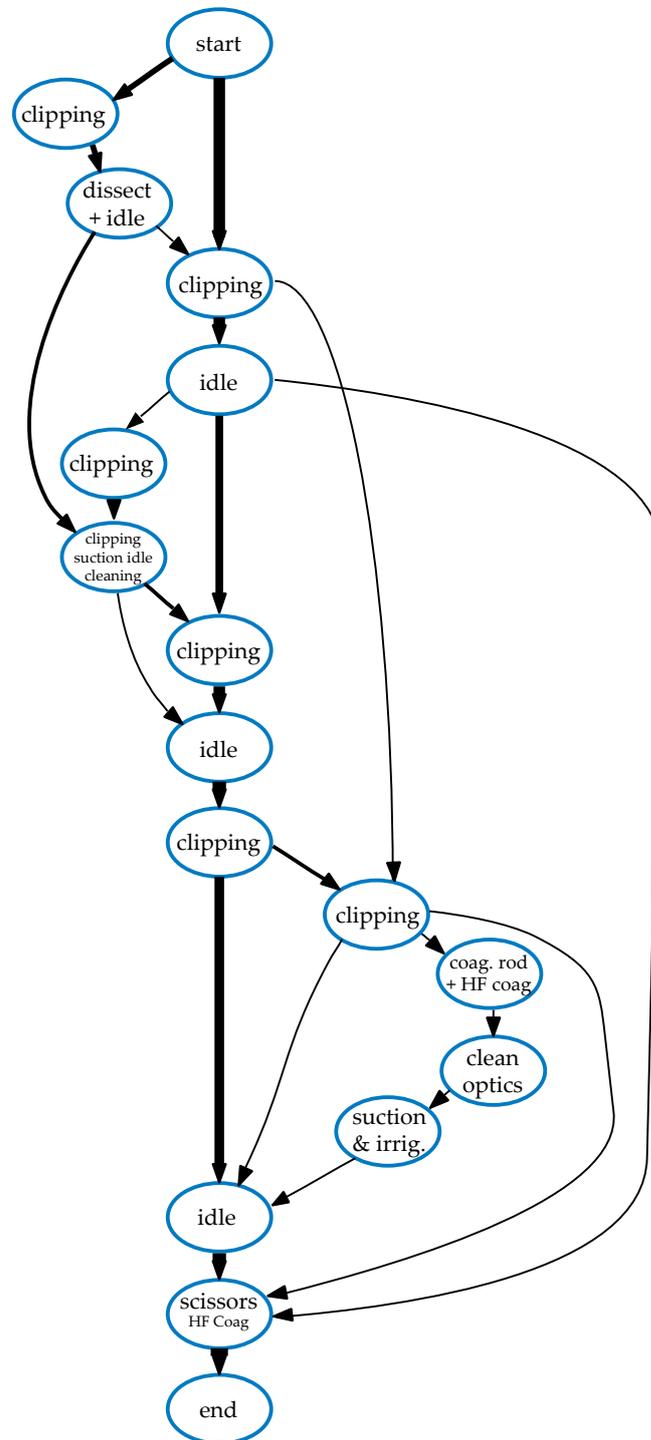
Figure 4.17.: An HMM of phase four obtained using model merging where loops are avoided. The thickness of the transitions represent their probability.
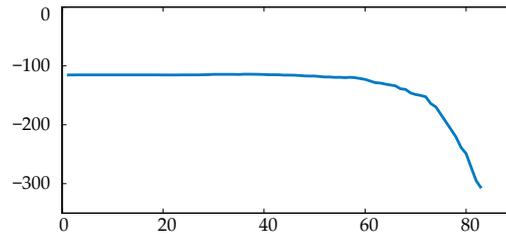
Figure 4.18.: The x-axis shows the number of merges. The y-axis shows $log_{10}P\left(\Omega \mid \lambda_x\right)$ for an exemplary phase and surgery.

state splitting [Ostendorf and Singer, 1997] (SSS). Simultaneous and independent from this work [Varadarajan et al., 2009] has been working on using SSS for automatic skill assessment in robotic MIS.

This method is similar to model merging, but works the other way round. The initial model λ_0 consists only of one state where $a_{00} = 1$ and $b_j(k)$ is estimated from all observations of Ω. The basic operation of SSS is splitting of a state. To compute the split, it is assumed that the observation symbols that are emitted by one state are generated by a two-mixture Gaussian. Two new states, each based on one mixture component, replace the old state. There are two different types of a split, a contextual and a temporal split. Both are illustrated in figure 4.19. In this example again the laparoscopic surgery is used to illustrate the method. A temporal split is performed if a state represents two or more actions that are carried out one after the other. A contextual split is done if one state represents actions where either one or the other action is performed. Given a model λ_i, the model λ_{i+1} is constructed by computing every possible contextual and temporal split and choosing the one that maximizes the probability $P\left(\Omega \mid \lambda_{i+1}\right)$. One drawback compared to model merging is that the topology that can be created is limited, as no model containing loops can be created. However, for representing linear workflows this poses no limitation and to build human-understandable models it can be desired to have models without loops.

A temporal split is done by performing the following steps:

- Remove the state that is split and replace it by two new states.

- All observations that have been assigned to the deleted state by the Viterbi path (see section 4.5.1.2) are split into two sets. New observation symbol probability distributions for both new states are computed from the two sets.

- For a temporal split, all transitions to the deleted state are redirected to the first of the new states. A transition from the first to the second new state is added. All outgoing transitions from the deleted state are assigned as outgoing transitions for the second new state. For a contextual split all transitions to the deleted state are redirected to both new states and all outgoing transitions from the deleted states are assigned as outgoing transitions to the new states.

- Transition and observation symbol probability distributions are refined using EM.

As for model merging, the state that is split is determined by a best-first heuristic.
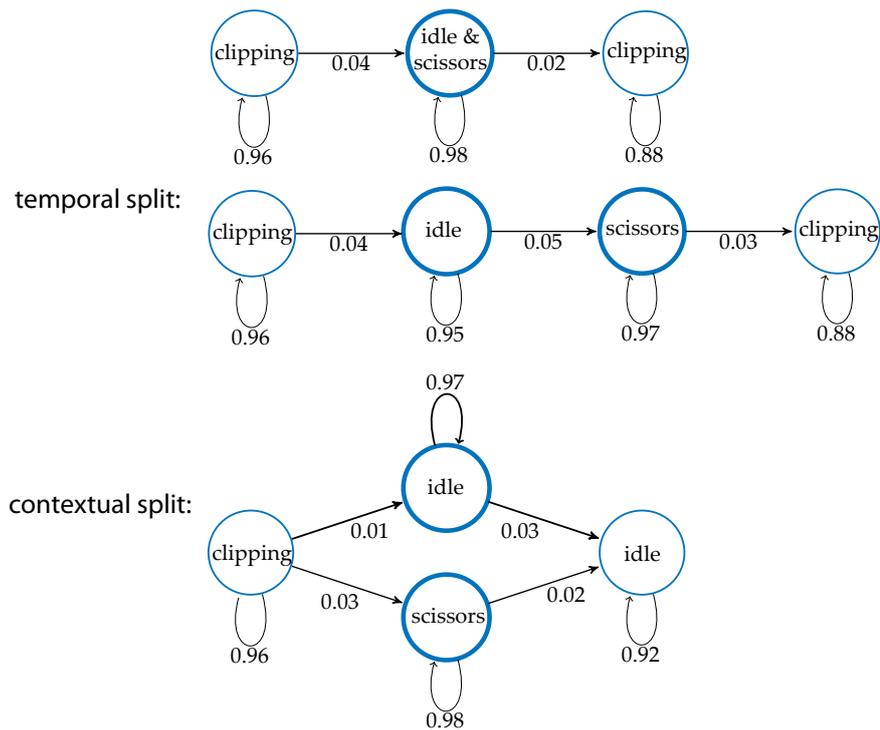
Figure 4.19.: Example of one splitting step. At the top the initial model is shown. In the middle a temporal and at the bottom a contextual split is illustrated.

4.5.4.1. Complexity and Implementation

In practice, splitting is faster than merging, as splitting starts with a very small model and therefore a smaller number of possible splits per iteration has to be considered, while in model merging a large number of possible merges must be considered. As we want to construct a HMM that generalizes well, we can assume that the number of states is bound by the length of the longest instance and therefore $O(L)$ splitting steps have to be performed. The number of possible splits is in $O(N)$. After splitting a state, the transition and observation probabilities of the HMM have to be refined using EM which takes $O(N^2TL)$. Afterwards $P\left(\Omega \mid \lambda_{i+1}\right)$ has to be computed which takes $O(N^2TL)$ and is therefore of the same complexity as the EM step. So the overall complexity is $O(N^3TL^2)$, which cannot be computed within a reasonable time.

As for the model merging, the complexity is reduced by approximating $P(\mathbb{O} \mid \lambda_{i+1})$ with an optimistic path updating of the Viterbi path. Using this approximation the splitting is computed in $O(NTL^2)$. Still, $O(L)$ splitting states are performed. In each step $P(\mathbb{O} \mid \lambda_{i+1})$ is computed, which takes $O(NTL)$ using a graph implementation. In each step additionally $O(N)$ possible states have to be evaluated.

For model merging, $P\left(\Omega \mid \lambda\right)$ is decreasing while the ability to generalize is growing. For SSS $P\left(\Omega \mid \lambda\right)$ is increasing while the ability to generalize is decreasing. The initial model generalizes very well, but does not explain the single instances of the procedure very well. Here again a stopping criteria is used and splitting is stopped when the increase in $P\left(\Omega \mid \lambda\right)$ drops below a threshold.

4.6. HMMs for Medical Education and Training

4.6.1. Visualization of Workflow Models

A graphical user interface (GUI) was developed to allow visual analysis of surgical workflow. Such a visual analysis can be used by students to study average workflows and to compare a single instance of a workflow to the average. The GUI can visualize the workflow and statistical properties of the HMM that was generated using the model merging approach. An example of this GUI is shown in figure 4. As HMMs use a graph-like structure, where states and transitions can be represented with nodes and edges, the natural way to visualize a HMM is drawing it as a graph. For generating the spatial layout of the nodes and edges GraphViz[1] [Ellson et al., 2004], an open source graph drawing tool, is used. The edges are weighted according to the transition probability they represent. Additional constraints are added to keep edges with a high weight short and straight and GraphViz builds the layout of the graph. GraphViz provides the position of the nodes and edges. For drawing the graph and a user interface Qt is used. Doing this, the typical workflow is visualized along a straight line, while uncommon actions are visualized to the left or right of the typical sequence of actions. The GUI allows to access the most important statistical properties of the HMM. The transition probabilities and average duration of a state are shown when moving the mouse over the corresponding node or edge. The average time spent in one state across all training surgeries is visualized by the size of the node.

A general problem in information visualization is to display the right amount of information. A graph, containing only few nodes for the most common actions, will be best suited to understand and analyze the most important aspects of the workflow. But it is not appropriate for detailed analysis or for examining uncommon events. Displaying every bit of information by using a graph with a huge number of states, allows a more detailed analysis of the workflow. But such a visualization will also include a lot of unimportant information and will be difficult to interpret. To deal with this problem, we reduce the number of visible elements by clustering. When doing clustering, the graph is simplified by merging several nodes into one. By allowing splitting nodes again, it is possible to analyze parts of the graph in more detail. Most mentions of clustering in graph visualization are purely structure-based [Herman et al., 2000], using only the graph structure to perform clustering. We use a content-based approach by utilizing the information obtained during the model merging process. Nodes can be expanded or merged using the merging information. This method is content-based as it does not rely on aspects such as neighborhood information in the graph, but on the merging process, which is driven by the data. A very compact visualization as seen in figure 4.20 on the left side can be expanded to a more detailed view as seen on the right side. While it would be possible to allow splitting states until the initial HMM λ_0 is reached, the GUI limits the possible splits. Only splits are allowed that significantly raise the probability $P(\Omega \mid \lambda)$. A split that does not raise this probability, does also not contribute to better explain the workflow as it does not add a significant amount of information. It must be noted that the order merges during the model generation does not restrict the order in which the nodes can be expanded in the GUI. So it is possible to expand one part of the graph, while viewing another part at a low level of detail. When splitting a state or merging two states, the number of states and therefore also the layout of the HMM will change. While the GUI is running the layout is recomputed after every merge or split using GraphViz and redrawn using Qt.

The GUI allows to simultaneously replay a video of the surgery and highlight the correspond-

---

[1]www.graphviz.org

Node	Avg. duration (s)	Probability of reaching(%	Probability of an instrument being used
First clipping	11.44	100.00	Clipping device = 100.00
First idle	12.71	77.78	No instrument = 100.00
Second clipping	9.00	77.78	Clipping device = 100.00
Second ilde	13.25	77.78	No instrument = 100.00
Third clipping	14.13	88.89	Clipping device = 100.00
Third idle	7.13	88.89	No instrument = 100.00
Scissors	21.33	100.00	Scissors = 100.00
Exception 1	29.00	22.22	Clipping device = 44.83
			No instrument = 55.17
Exception 2	20.00	11.11	Clipping device = 65.00
			No instrument = 35.00
Exception 3	27.00	11.11	Suction and irrigation = 75.00
			No instrument = 23.68
			Dissecting device = 76.32
Exception 4	38.00	11.11	HF cutting = 5.26
			No instrument = 81.58

Table 4.2.: Statistical properties embodied in the 11-state HMM shown on the right of figure 4.20

Node	Avg. duration (s)	Probability of reaching(%	Probability of an instrument being used
Clipping and idle	78.67	100.00	Clipping device = 47.38
			Suction and irrgation = 2.89
			Dissecting device = 4.13
			HF cutting = 0.28
			No instrument = 45.32
Scissors	21.33	100.00	No instrument = 100.00

Table 4.3.: Statistical properties embodied in the 2-state HMM shown on the left of figure 4.20
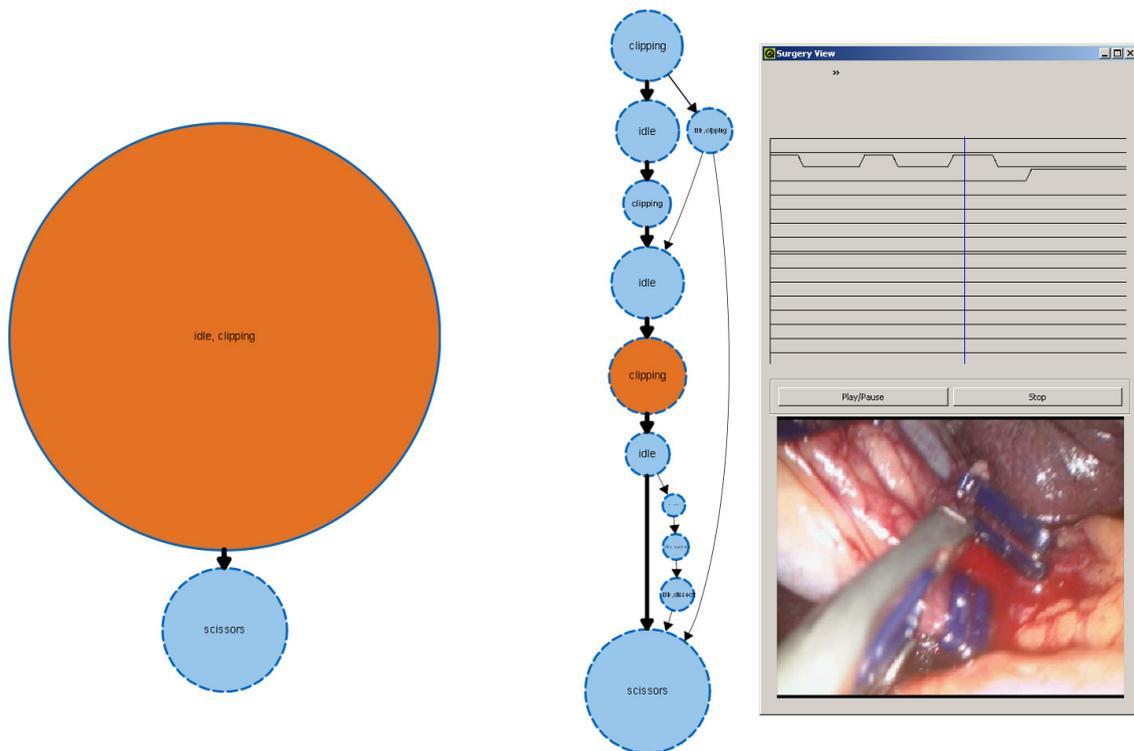
Figure 4.20.: GUI allowing to inspect a HMM representing one phase of a surgery. On the left a visualization of a compact HMM is shown. In the middle a more detailed HMM representing the same phase is visualized. On the upper right the instrument usage during one instance and on the lower right a video of the same instance is shown. The blue line in the upper right represents the current time step in this instance and the orange node represents the corresponding HMM state.

ing state of the HMM. This can be seen in figure 4.20. On top of the video, the instrument vector of this surgery is displayed. To synchronize the video of a surgery with the HMM, the Viterbi algorithm (see section 4.5.1.2) is used to estimate the most likely sequence of states in the model given the observations of one surgery. While the video is running, the level of detail of the HMM can still be changed. In addition to visualizing the HMM as a graph, the statistical parameters can be analyzed directly. Table 4.2 shows some of the parameters of the right HMM in figure 4.20. In this table, the average time spent in the states, the probability of reaching a state and the instrument use in the states are shown. Again we can make use of the merging, and display the statistical properties of a more compact HMM. The parameters of the compact visualization shown on the left side of figure 4.20 are given in table 4.3.

Using such a visualization can help a student to understand medical workflows, by analyzing the course of the workflow, studying the statistics and comparing the workflows of different experts. It is also possible to let a student perform a procedure and synchronize it to the model of an expert, so that the student can study differences.

4.6.2. Use of HMMs for the US Simulator

While for the laparoscopic cholecystectomies it has shown that model merging leads to better results, model splitting delivers better results for the data obtained from the US simulator.

While the laparoscopic cholecystectomies are represented by vectors consisting of binary signals, which represent instrument usage, for the US simulator 6 DOF tracking data is available. The US simulator tracks the pose of the phantom $_{Tracker}T^{Phantom}$, the US probe $_{Tracker}T^{Probe}$ and the camera $_{Tracker}T^{Camera}$ or HMD $_{Tracker}T^{HMD}$. For analysis of the workflow only the relative pose of the probe with respect to the phantom is relevant, which can be computed as $_{Phantom}T^{Probe} = (_{Tracker}T^{Phantom})^{-1} *_{Tracker} T^{Probe}$.

Instead of representing the pose by a 4 by 4 matrix in homogeneous coordinates, as done in the CAMPAR framework, for the workflow analysis it is represented by a 7-dimensional vector consisting of three values for the position and four values for a quaternion representation of the orientation. For the probability distribution function a Gaussian model is used. The motivation to use Gaussians is that the use of US consists of two different actions. Either the probe is placed at one location to take an image. Here only little movement occurs. Or the probe is moved from one location to another. Using a Gaussian model, placing the probe at one location can be represented by an HMM state with a Gaussian that only has little variance, while movement can be represented by one HMM state having high variance. Successive state splitting is very likely to automatically produce a HMM topology having these characteristics. We assume that the model λ_i has one state representing both, a movement and placing the probe at one location. Splitting this state into two states, where one represents the movement and one placing the probe at a location, will result in λ_{i+1} which has a much higher probability of representing the data $P(\mathbb{O} \mid \lambda_{i+1})$ than the previous model $P(\mathbb{O} \mid \lambda_i)$. Further splitting one of the two states will also result in a higher $P(\mathbb{O} \mid \lambda_i + 1)$, however the change will only be very low. Therefore such splits are likely to not occur before the stopping criterion is met.

Figure 4.21 shows the visualization of such a HMM in the US simulator. On the right the single steps of the workflow are shown. States with high variance, representing movement of the probe have been filtered out for this visualization. For every state, the mean position and duration are shown. On the left, a volume rendering of the CT volume is shown. The blue spheres represent the average position of the HMM states. While the student inspects the workflow she can still use the tracked probe. The red sphere represents the pose of the probe that is currently tracked and the white sphere represents a state of the HMM which has been selected. For the selected state the mean orientation is visualized by showing the geometry of the US plane in green.

Compared to the GUI that was presented in the last section, the advantage of this system is that it is integrated into the US simulator and the student can interact with the simulator while using the statistical model. The user can move the US probe and compare the pose to the pose from the HMM. All features of the simulator that haven been described in chapter 3, such as displaying co-registered CT slices, can be used while analyzing the statistical model. Also methods such as a replay of a procedure performed by an expert and synchronization of two procedures can be used.

4.7. Use of Workflow Models for Surgical Workflow Analysis

As mentioned before, the methods to generate DTW and HMM workflow models have not only been developed for medical training but also to enable systems that can automatically predict the current state of a running surgery. The ability to detect the current state of a running surgery has been evaluated in a series of articles. As this thesis is mainly about the use for
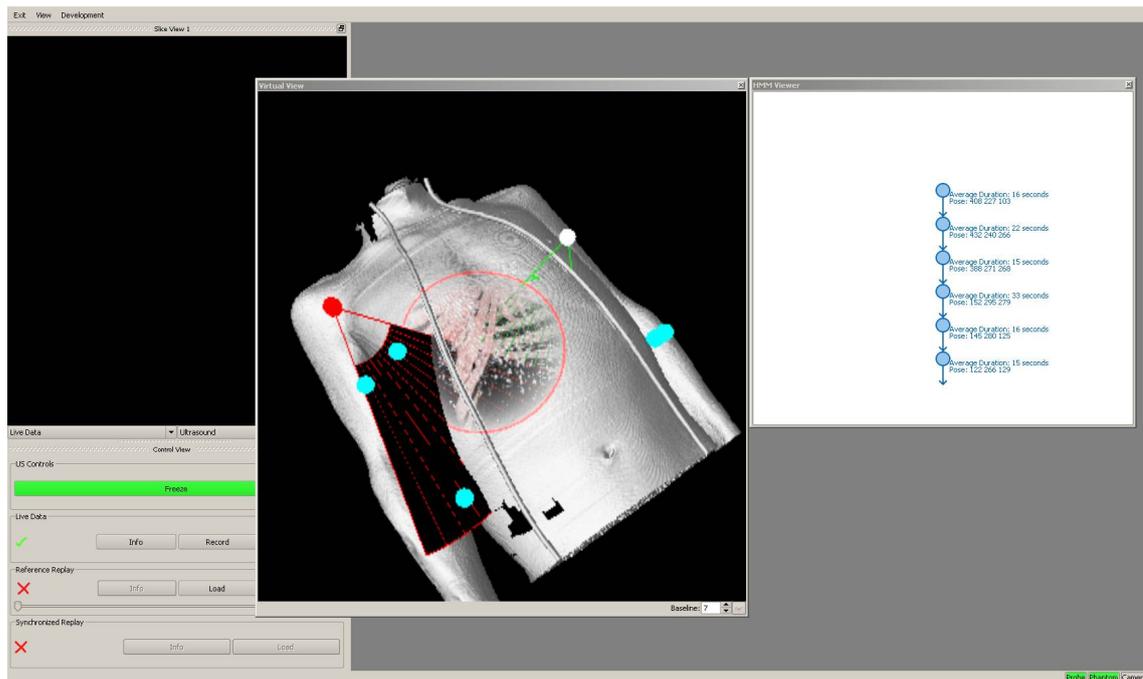
Figure 4.21.: Visualization of a HMM representing the workflow in the ultrasound simulator. On the right the HMM workflow model is shown. The spheres in the volume rendering in the middle correspond to the circles on the right, representing the HMM states. The red sphere and frame represent the currently tracked probe an the white sphere and green rectangle show the average pose of one HMM state that has been selected.

training and education only a brief summary of these results is provided here.

All evaluations are based on a full cross-validation where all but one surgeries have been used to construct the model. This model was used to predict the current phase for the remaining surgery. A list of all phases is provided in table 4.2. The error was always measured by predicting the current phase once per second and counting the prediction errors. The standard DTW algorithm as explained in section 4.3 and an extended version, which uses boosting for an adapted similarity measure have been evaluated in [Padoy et al., 2007]. The standard DTW showed an error of 0.8% and the adaptive DTW had an error of 0.3%. Both methods can only be used after a surgery has finished and not during a running surgery. Left-to-right HMMs using only a subset of the signals representing instrument usage have been evaluated in [Padoy et al., 2008]. This method had an error of 11.5% for recognition during a running surgery. The additional use of visual features from the laparoscopic video reduced the error rate to 7.6%. A comparison of 14 state left-to-right HMMs and models obtained by model merging showed error rates of 14.0% for the left-to-right HMM, respectively 6.7% for model merging [Blum et al., 2008]. In another article different methods to construct DTW and HMM based models for offline and online recognition have been compared [Padoy et al., 2010]. For offline recognition the error rates where 2.7% for DTW, 4.0% for left-to-right HMMs, 9.9% for randomly initialized HMMs and 6.1% for model merging. The errors for online recognition were 8.7% for left-to-right HMMs, 11.3% for randomly initialized HMMs and 11.9% for model merging.

While the methods for recognition of phases have not yet been integrated into the US simu-

lator, there are also possible uses for medical education. If the simulator can detect the current phase of a procedure that is performed by a student, it could trigger actions. So the simulator could give an advise for the next step or simulate a change in aspects such as blood pressure. For a system that also simulates communication with the patient, this could be used to trigger questions or comments by the patient.

5. mirracle, an Augmented Reality Magic Mirror

While the last two chapters focused on training of ultrasound and the use of workflows for education, in this chapter a third area of medical education and training will be addressed, namely teaching human anatomy. Knowledge about human anatomy is an important issue for everyone working in the field of medicine. But it is also an important part of the general education and it is relevant for many other professions related e.g. to healthcare or sports. The human anatomy is very complex and it does not only involve knowledge about the single organs, 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 big 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. In this chapter a novel way to intuitively teach human anatomy using an augmented reality magic mirror system that displays anatomical structures overlaid onto the body of the user is presented. First, the system will be explained in section 5.1. Then the user interface, a novel metaphor for gesture-based interaction and a first user study on the UI are described in section 5.2. The use of the magic mirror for anatomy education and to visualize US workflows is discussed in section 5.3.

Augmented reality systems for visualization of anatomy have been shown before. [Davis et al., 2002] presented a system that augments a 3D model of the anatomical airways onto a patient phantom using a HMD. Another system that used a HMD to visualize human anatomy onto a phantom has been shown by [Juan et al., 2008]. Their system allows students to open the abdomen of the phantom and it visualizes different organs on the phantom. Furthermore, AR has been used in different medical simulation systems e.g. in systems for simulating an anesthesia machine [Quarles et al., 2008], sedation [Hwang et al., 2009] or forceps delivery [Lapeer et al., 2004].

Previous systems on AR visualization of anatomy used expensive systems involving HMDs. The system presented in this chapter is an inexpensive and easy to use AR system, which makes use of the magic mirror concept to present information about human anatomy. It presents anatomical data augmented onto the user and it shows additional 2D and 3D information.

The system presented here builds upon previous work on medical AR visualization using a HMD at the Chair for Computer Aided Medical Procedures & Augmented Reality [Sielhorst, 2008, Kutter, 2010, Bichlmeier, 2010]. The main contribution of this thesis was to extend the concept of medical AR for training and education by using a novel and inexpensive alternative to HMDs and the development of new concepts for gesture-based interaction with the system.

5.1. System

The first version of the mirracle system has mainly been developed for education of anatomy in classrooms, museums or exhibitions. It focuses on a small number of important organs of
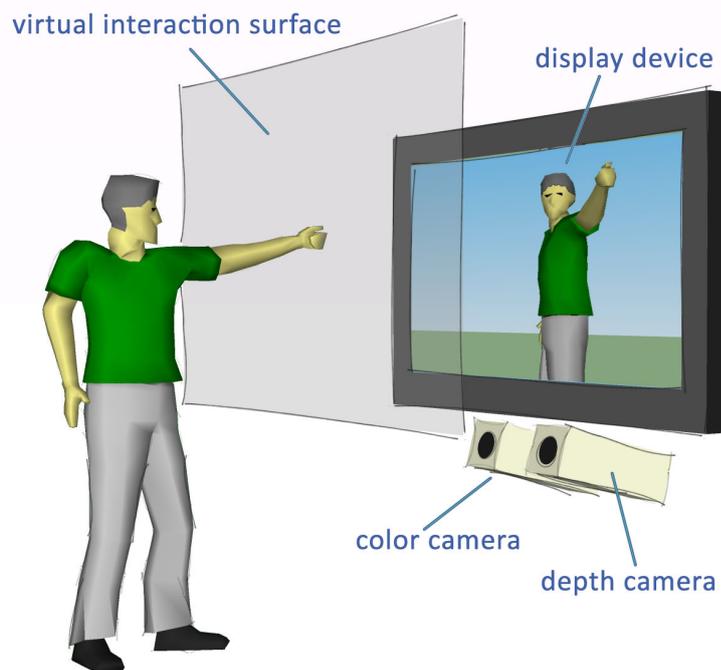
Figure 5.1.: The system consists of a display device, a color camera and a depth camera. There is a virtual interaction plane between the user and the display. This plane is used for the interaction between user and system.

the abdomen, namely liver, lungs, pancreas, stomach, small intestine and bones. A version for anatomy education for medical students and for visualizing US workflows will be discussed later.

5.1.1. Hardware Setup

An illustration of the mirracle hardware setup can be seen in figure 5.1. 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. The third component is a depth camera which is placed next to the color camera and which has a similar field of view and viewing direction as the color camera. The current system uses the Microsoft Kinect (Microsoft Corporation, Redmond, United States), which is sold as an add-on for the Xbox 360 video game console to enable games using gestures and body movement as input. It consists of a color and a depth camera that are assembled into one housing. The depth camera is an infrared camera that uses structured light, which is emitted by an additional infrared projector to estimate depth values for each pixel.

5.1.2. Software Framework

In figure 5.2 the software framework is illustrated. The rectangles with italic text represent the three different visualization modes of the system. They will be discussed in more detail later. The system uses Qt for window management and basic user interface elements. For the AR in-
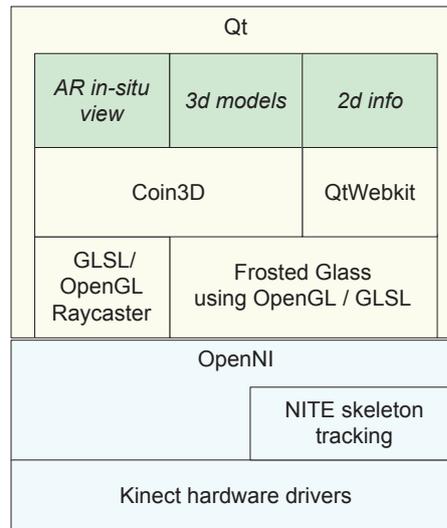
Figure 5.2.: Illustration of the framework that is used for the magic mirror. The lowest two layers represent external software that is used to access the hardware and to perform the skeleton tracking. The middle two layers represent external software integrated into our framework. The boxes with italic text represent the different data presentation modes. The boxes below represent software components that are used in the different modes.

situ visualization of medical volumes the medical raycaster is used that was also used for the AR visualization in the US simulator [Kutter et al., 2008], which is based on OpenGL and the OpenGL Shading Language (GLSL). 2D text and image information is displayed using QtWebkit, a port of the WebKit web-browser engine [1] to Qt. Polygonal 3D models of anatomy are drawn using Coin3D [2], an OpenGL implementation of scene graphs. The frosted glass visualization, which is described later, is implemented in OpenGL and GLSL. To access the Kinect the system uses OpenNI [3], which is an open source software framework that allows retrieving color and depth images from the Kinect. The depth camera 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. This functionality is implemented in OpenNI. 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. The pose and orientation of skeleton joints are projected into the color image. For skeleton tracking the magic mirror uses NITE [1] a software by PrimeSense (PrimeSense, Tel Aviv, Israel) that performs gesture recognition and skeleton tracking based on depth images. NITE can be used with the Kinect through the OpenNI framework.

5.1.3. AR In-situ Visualization of Human Anatomy

This section describes the visualization in the *AR in-situ view* mode. For an intuitive visualization of organs the concept of a magic mirror is used. A camera that is mounted next to a display is taking images of the user standing in front of the display. The images are flipped horizontally and shown on the screen such that the user has the impression of stand-

---

[1] `www.webkit.org`
[2] `www.coin3d.org`
[3] `www.openni.org`
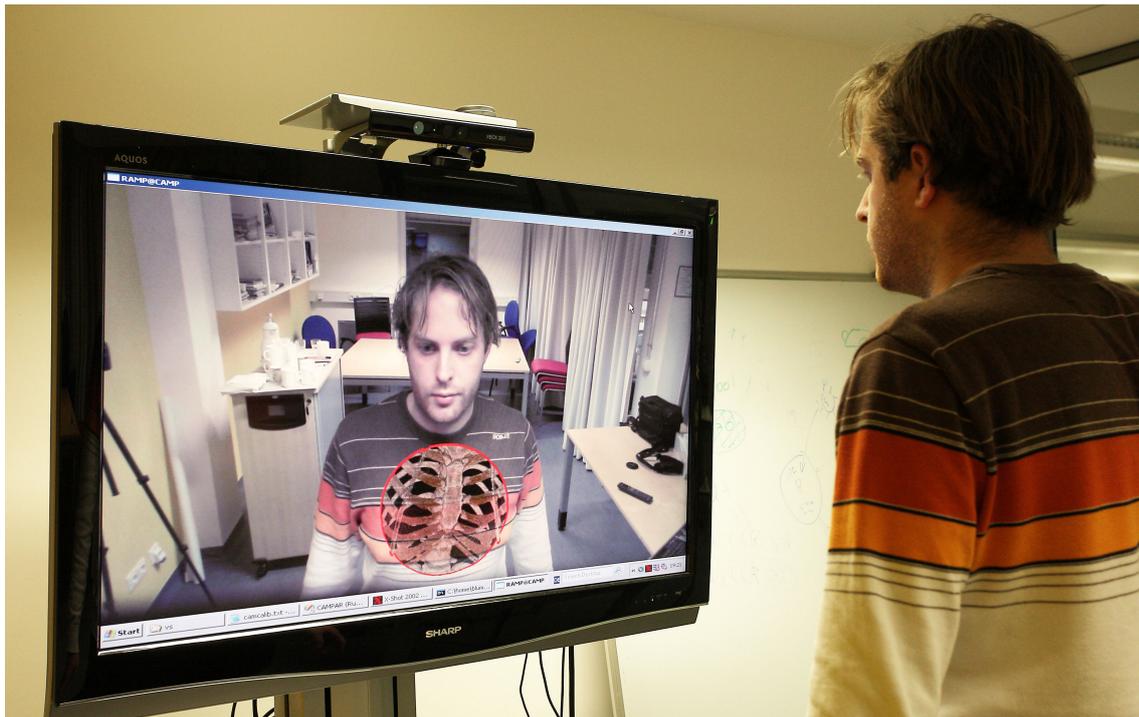
Figure 5.3.: The user is tracked and a CT dataset is augmented onto the user.

ing in front of a mirror. Virtual objects can be added to the image of the real scene. The magic mirror concept has been used previously to augment virtual shoes [Eisert et al., 2008, Mottura et al., 2007], shirts [Ehara and Saito, 2006, Hilsmann and Eisert, 2009] or knight's armors [Fiala, 2007] onto the user. For tracking the user, previous systems have been using markers [Fiala, 2007, Ehara and Saito, 2006, Mottura et al., 2007] or required to wear a shirt with a rectangular highly textured region [Hilsmann and Eisert, 2009]. For shoes, a vision-based approach has been presented by [Eisert et al., 2008].

While previous systems have augmented objects onto the user, this system extends the magic mirror concept for education of anatomy. It creates the illusion that the user can look inside his own body. An image of this can be seen in figure 5.3. To achieve this visualization, the magic mirror augments a volume visualization of a CT dataset onto the user. To allow a correct augmentation of the CT, the pose of the user has to be tracked. This is done based on the depth image using the NITE skeleton tracking.

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 CT volume is scaled to the size of the user and augmented onto the user. One drawback of the current system is that the dataset is not deformed. So if the user bends, this is not reflected in the visualization of the CT and also movements of the limbs are not visualized correctly. While later, in section 6.2.2.3 possible solutions to address this issue are discussed, for the current system, which focuses on the abdominal area, this is a minor problem.

For the augmentation the same focus and context visualization [Bichlmeier et al., 2007b] as in the US simulator (see section 3.3.2) is used, such that the virtual objects are only shown through a circular focus window. This can be seen in figure 5.3. This leads to a better perception of depth, compared to a simple augmentation of the whole CT. While the system could use a

Figure 5.4.: The magic mirror shows a simulated X-ray (DRR). For comparison, the image printed on the sheet of paper shows a real X-ray of the same person.

CT scan of the user, for medical education it is not possible to acquire a CT scan of the user if it is not required for medical reasons. Therefore we augment the CT of another person onto the user. For this purpose again the VKH dataset [Park et al., 2005] is used, which was also used to provide photographic images for the US simulator. More details on the dataset have been provided in section 3.2.2.

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. A visualization of the bones from the CT using direct volume rendering (DVR) is shown in figure 5.3. For other applications, such as patient-doctor communication it would also be possible to augment a dataset of the user. An example of this is shown in figure 5.4, where a CT of the user is augmented onto the user. In this example instead of a DVR a digitally reconstructed radiograph (DRR) is used, which simulates the appearance of an X-ray image.

Visualizing structures other than bones from the CT is more challenging. In a first attempt the segmentation that is available for the VKH dataset 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 dataset [4] 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 5.5.

---

[4] `www.anatomium.com`
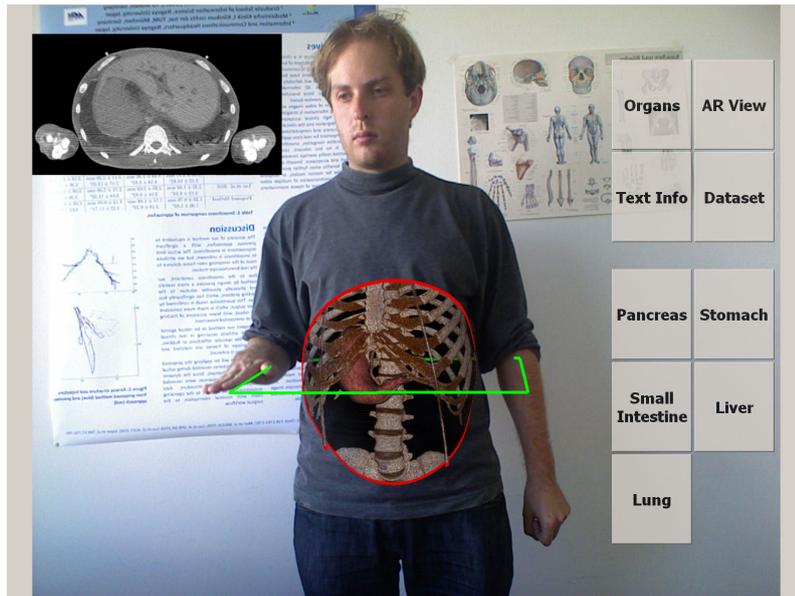
Figure 5.5.: Simultaneous visualization of bone from a CT volume and a polygonal model of the stomach.

5.2. User Interaction

In this section different aspects related to user interaction will be discussed. First, the interaction for selection of slices will be presented, which helps users to relate anatomical images to their own body. Afterwards a new paradigm for gesture-based interaction will be discussed and a study on different options for the user interface will be described.

5.2.1. Gesture-based Interaction for Slices

In the magic mirror system gesture-based touch-free interaction is used to enable the user to interact with additional 2D and 3D information about organs. This section describes the user interaction to display additional anatomical slices in the *AR in-situ view* mode.

When using traditional user interfaces often 3D information about the real world is visualized on a 2D display and a 2D input device is used to interact with this 3D information. This forces the user to perform multiple mental projections between different spaces. One central idea of AR is to have the real information, the virtual information and the user interface in the same 3D space. Therefore, gesture-based interaction has been explored for various AR applications. One example is marker-based finger tracking to recognize interaction with virtual objects in HMD-based AR [Buchmann et al., 2004]. For non-AR applications video-based detection of pointing gestures has been used [Cheng and Takatsuka, 2009]. [Soutschek et al., 2008] used a depth camera to detect different gestures of the fingers and hand movement in a user interface for medical imaging applications.

Medical volumes are usually visualized by showing slices that are aligned with the axes of the volume. Figure 5.6 illustrated how a user can select a sagittal plane in the magic mirror system. A sagittal plane is a plane that passes from the front of the user to her back. A volume can be seen as a stack of sagittal slices starting from the left side and going to the right. When the system is in sagittal slice mode the user selects a sagittal slice by moving her hand from left to right. To switch to a transverse slice mode the user has to move the hand up or down,
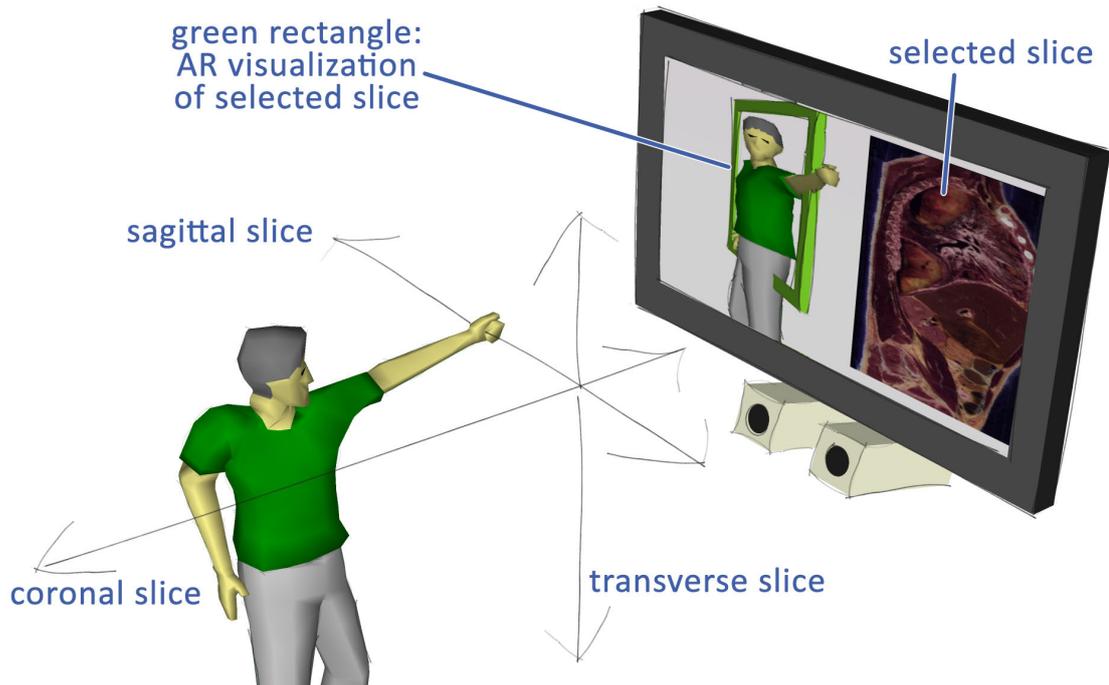
Figure 5.6.: This illustration shows the interaction for selecting slices. By hand movement sagittal, coronal and transverse slices can be selected. A green rectangle augmented on the user shows the current slice.

respectively front or back for the coronal slice mode. To avoid that the system switches between transverse, sagittal and coronal slice mode due to minor movements of the hand, it only switches to another mode when the movement of the hand along an axis exceeds a threshold. The current slice is shown on the left or right side of the monitor while a green rectangle is augmented onto the user. This rectangle visualizes the slice that is currently selected. When using the VKH dataset we can show either slices from the CT or the photographic volume.

5.2.2. Frosted Glass Interaction Metaphor

In addition to the AR in-situ visualization of anatomy we want to display additional text, images and 3D models to provide more information about anatomical structures. To do this, the system switches to another mode, where no magic mirror visualization is used, but the whole screen is used to display additional information. This section describes the interaction metaphor of using frosted glass, which is used in the modes *2D info* and *3D models*.

Introducing new interaction paradigms is difficult as the user has to learn how to use them. Many user interfaces require some time to learn how to use them. This is acceptable for some application. But as the magic mirror should be used with children and visitors of exhibitions we want to avoid using an interface that requires time to learn. Over the last years, multi-touch surfaces have become very popular. As they are used in many mobile phones people have become familiar with this kind of interaction. Gestures, such as zooming by framing a target area by two fingers and moving these fingers inwards or outwards are known to most people. At the first glance, it seems like interaction using depth cameras enables using the same kind of gestures in a touch-free setup. However, when examining the interaction in more detail there are important differences.

Select area to zoom	→	Start zooming	→	Perform zoom	→	End zooming

| Move fingers above surface | → | Touch surface | → | Move fingers on surface | → | Release surface |

Figure 5.7.: Interaction of zooming on a surface that is touched. The upper layer models the higher level interaction and the lower layer the corresponding physical actions.

In figure 5.7, the gesture of zooming in with a multi-touch surface is modeled on two layers. The upper layer represents the higher lever interaction and the lower layer represents the physical action that is required to perform the interaction. The first step is selecting the area to zoom. This step does not yet involve touching the surface but the user does only decide which area to zoom and moves the fingers towards the points framing this area. Directing the fingers to these points is easy as the image and the fingers are in the same 3D space. The second step is confirming the selection and starting to zoom. This is done by touching the surface with both fingers. As a physical surface is touched the user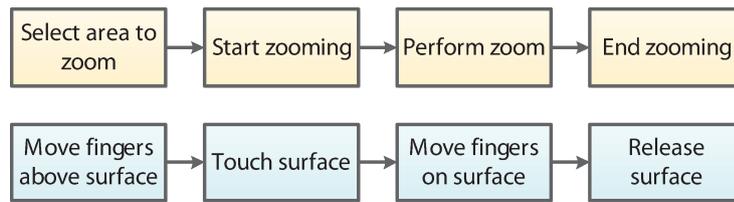 receives haptic feedback about this action. The zooming itself is performed by moving both fingers and the zoom interaction is ended by taking the fingers away from the display, which again results in a haptic feedback for the user.

A similar interaction metaphor can be used in touch-free interfaces by introducing a virtual interaction plane as seen in figure 5.1. A common method is to use a virtual representation of the hand on the screen e.g. by an icon of a hand which follows the movement of the real hand. Such an interface introduces several problems compared to touch-based interaction. While in touch-based interaction points on the display device are directly touched, in the touch-free interface the hands and the image are at different locations in space. Therefore selecting the two points that frame the zoom area is less intuitive. Furthermore the haptic feedback when starting and ending the zoom gesture is missing. This is not only a problem for zooming but also for other interactions such as pressing a button.

We propose a novel interaction method to address these problems. We are using the metaphor of frosted glass. When looking at frosted glass, objects that are far behind the glass are not seen. Objects that are closer to the glass are seen blurred and with low contrast and objects that touch the glass are seen well. When using the frosted glass metaphor for a user interface, information such as text information about an organ is drawn onto the frosted glass. The hand of the user is drawn as if it would be behind the frosted glass. Based on the distance between the virtual interaction plane and the hand, opacity and blur are modified. The depth camera is used to estimate the distance of each image pixel to the virtual plane. Pixels from the color image that are close to the virtual interaction plane, or have passed it, are drawn as if they would touch the frosted glass. Other pixels are drawn with blur and opacity depending on their distance to the virtual interaction plane. An example of this visualization is shown in figure 5.8.

Using the frosted glass metaphor we can resemble the interactions of a multi-touch surface. When the hands are distant from the virtual plane, only the information on the frosted glass is visible. To select the zoom region the hands are put some centimeters distant from the virtual plane and they are drawn blurred and semi-transparent. So, the information on the screen can still be seen while the user can see her own hands relative to information on the screen. To confirm the selection and start zooming, the user moves the hands closer to the virtual plane,
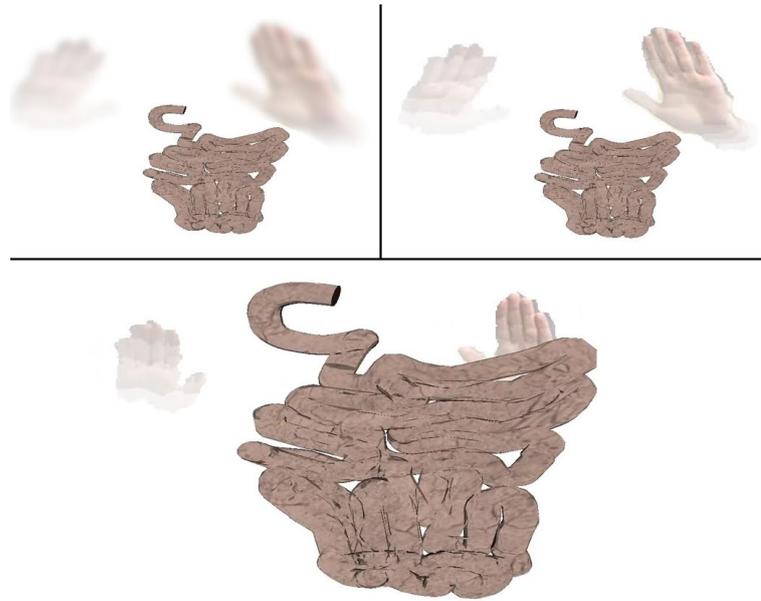
Figure 5.8.: These images show a zoom interaction in our system. In the upper left image the hands of the user did not touch the virtual interaction plane, yet. In the upper right the zoom gesture has started, in the lower image the 3D model of the small intestine has been zoomed in by a gesture.

whereby the hands become completely visible without blur. By using this interaction metaphor the hands of the user and the display are brought into the same space, such that actions such as selecting a point are more intuitive. Using the frosted glass metaphor, a concept that is known to everyone from the real world is used to encode the distance between the hands and the virtual interaction plane. The haptic feedback of touching the surface is replaced by a visual feedback.

This visualization is implemented in OpenGL / GLSL. The color and the depth image are loaded into the graphics memory and a shader program is used to blur the image. Details on the method used for generating the blur are described in appendix A.1.2.3 where the same method is used to generate artificial out-of-focus blur for a study on the effect of missing blur in AR and VR systems. The amount of blur depends on the distance of each pixel to the virtual plane. In addition to the blur, the opacity of the pixels is changed depending on the distance to the virtual plane.

One problem with this metaphor is that the user moves his hand towards the frosted glass, while in the visualization it looks like the hand of the user would come from the other side of the glass. However, as in the AR mode the metaphor of a magic mirror is used, it has shown that users immediately understand this interaction concept. A photo of a zoom gesture performed using the frosted glass metaphor is shown in figure 5.9.

5.2.3. Presentation of Additional Information

While the AR visualization is suited well for showing the position of different organs and 2D slices from the CT or photographic volume, it is not suited well to display more detailed information. Therefore the user can switch to the *2D info* or the *3D models* mode, which both use the frosted glass metaphor described before. In this mode either a 3D view of organs or
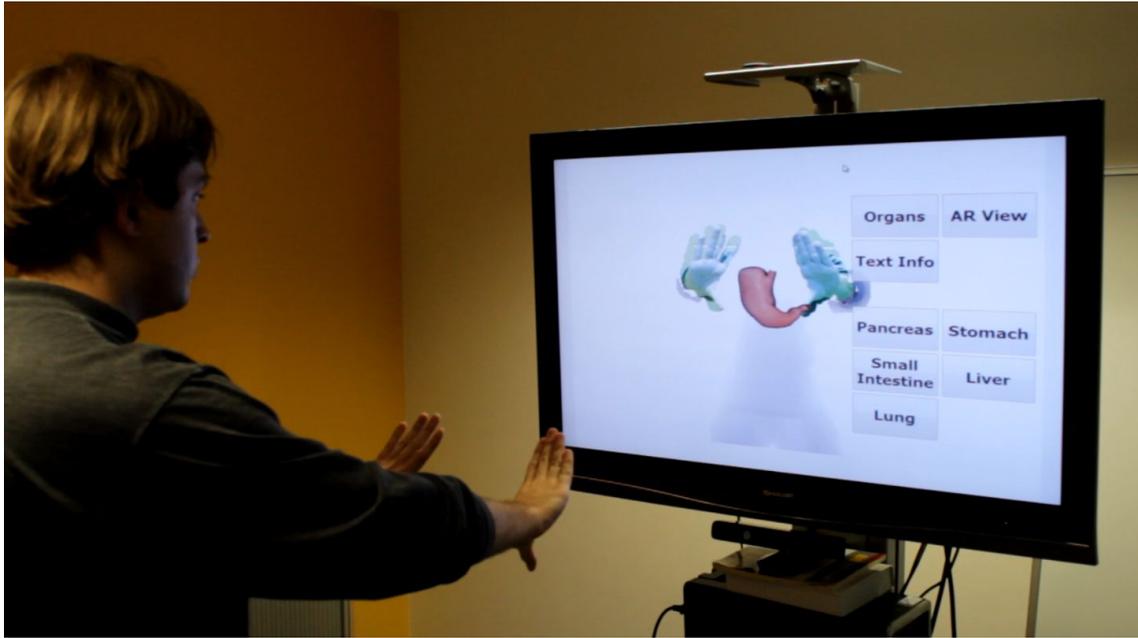
Figure 5.9.: Photo of the system while zooming is performed using the frosted glass interaction metaphor.

2D information about the organs can be displayed.

In the AR mode the organs are augmented onto the user, and therefore they are only shown rather small. In the *3D models* mode a fullscreen version of the polygonal organ models are shown such that details can be studied. As an example, the small intestine taken from the Anatomium dataset can be seen in figure 5.8. Using the frosted glass interaction metaphor, the user can rotate and zoom the organ models.

The system can also show additional text information and images about the organs in the *2D info* mode. The software framework of the magic mirror allows to overlay different Qt widgets onto the frosted glass visualization. For the 2D information QtWebkit is used, which can render html websites. Currently we show the corresponding articles from Wikipedia, but this could be replaced with any other information in html format. Examples of this are presented in the figures 5.10 to 5.12. As most text processing programs can export to html, it is easy to author new content, which can be shown in the magic mirror system and using programming languages such as JavaScript it would even be possible to integrate interactive content or connect the system to a database.

5.2.4. Selection

The frosted glass metaphor is a general metaphor for gesture-based interaction. Using this metaphor, different user interfaces, using e.g. different ways to push a button or perform a zoom gesture can be implemented. This section describes three different methods that were developed and tested to select items. The first one uses an element that is familiar to most users: buttons. These buttons can be clicked by performing click gestures. The second interface is similar to a slider, which gets activated as the users hand moves over it. Selection is done by moving the hand along the slider, and finally the user confirms the selection by moving her hand outside the slider's region. The last interface can be regarded as a mixture of the previous

two interfaces. The user pushes icons out of the screen. If this sliding button is pushed by a certain amount, the button is considered as pressed. The three different methods are explained in more detail below.

The first idea was to rely on user interfaces that most people know from their everyday work with PCs. Buttons are used that are approximately as big as the user's hand on the screen, and can be clicked via performing a click-like gesture. The click gesture consists of two stages. First the user's hand passes the interaction plane. At this stage, the button over which the hand is placed gets highlighted. Then the user has to perform a movement towards the screen, followed by a movement backwards. If this movement is completed, the gesture is considered as successfully completed. There are two sets of buttons. The first set allows to switch between the AR visualization and visualization of 3D models and web pages. The second set of buttons allows changing the organs. An exemplary view of this user interface with active *2D info view* is shown in figure 5.10.
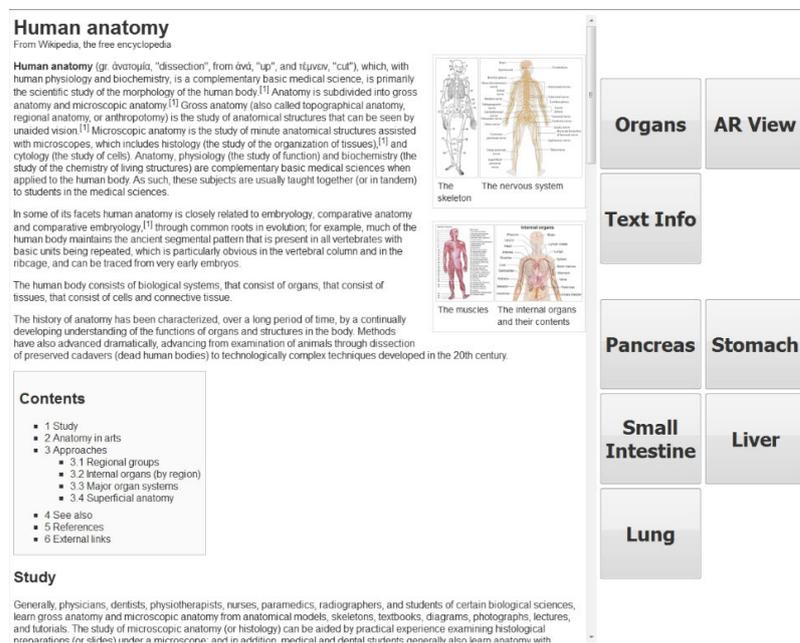
Figure 5.10.: The button user interface with active *2D info* view.

The second graphical user interface is using the slider metaphor. Sliders are displayed on the left and right side of the application, each slider containing multiple entries. For example the left slider contains three entries for the different visualization modes. Figure 5.11 shows the slider user interface during interaction. Interaction with a slider consists of three steps. As the user's hand is moved over the slider it is activated and highlighted by painting its background darker. Within the area of the slider the user can move his hand up and down to select the desired entry. As the hand moves outside the slider region again, the currently selected entry is confirmed.

The third interface is called sliding buttons and can be considered as a mixture of the previous two. Instead of having a slider covering the complete left or right side of the screen each element is considered as an individual sliding button. These buttons differ from the buttons in the first interface. Instead of pushing into the direction of the screen, the user has the push them to the side. While the user pushes a button, the button moves towards the screen border. This creates the illusion of pushing the button outside the screen. If a button has been pushed
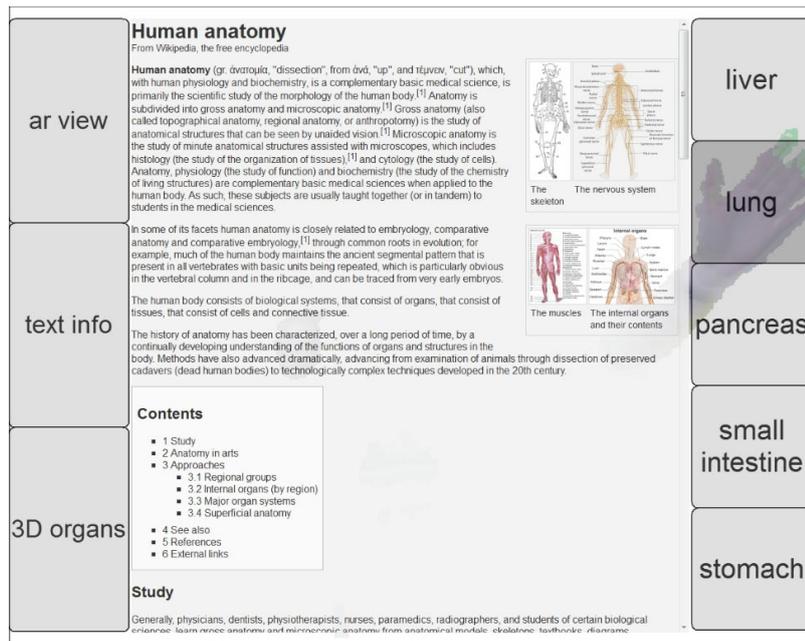
Figure 5.11.: The slider user interface with active *2D info* view.

half to the side, the selection is considered as confirmed. Figure 5.12 shows a user while selecting the *small intestine* button. The user can always abort the selection by moving her hand outside of the button area, or by moving over another element.

5.2.5. Gestures

In addition to selecting from different options, as described in the last section, gestures are used for zooming and rotating. The gestures are similar to the gestures most users know from touch screen interfaces. These gestures were implemented on top of Qt's built-in gestures and their gesture recognizers.

5.2.5.1. Pinch Gesture

This gesture is used for zooming web pages and 3D organ models, as well as zooming and moving the focus window in the *AR view*. The user starts by moving both hands past the interaction plane. Then he moves the hands away from each other for zooming in, or towards each other for zooming out. During this movement, the distance between both hands is tracked. This distance is used as scale factor. Further, the center of the hands is tracked. To finish the gesture, the user simply pulls back his hands outside the interaction plane. Figure 5.13 shows an illustration of these three steps.

If the currently active view mode is the *AR view*, the scale factor determines the size of the focus window, and the center point is used to move the focus window. In the *3D models* view, the scale factor is directly transferred to the scale of the currently visible organ. During the *2D info* view, the scale factor cannot be transferred directly to the zoom factor. While current implementations of html renderer on mobile phones are optimized for gesture-based interaction and allow seamless zooming, desktop html render engines change the text size and scale images step-wise and rearrange the layout. Therefore we pick a screenshot as the
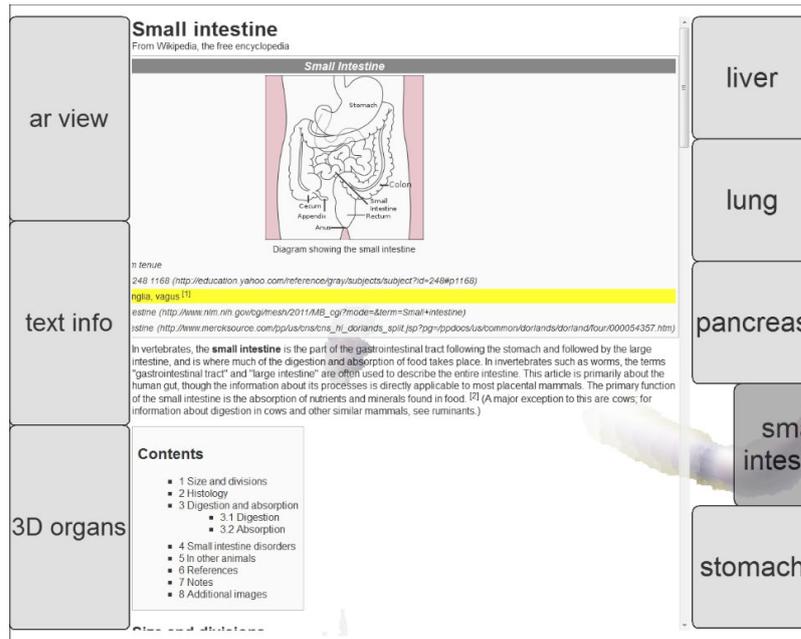
Figure 5.12.: The sliding button user interface with active *2D info* view.

| Hands passed the interaction plane $\Rightarrow$ start positions are recorded. | Hands moves $\Rightarrow$ the scale and the center are recomputed. | Hands left the interaction plane $\Rightarrow$ the interaction ends. |
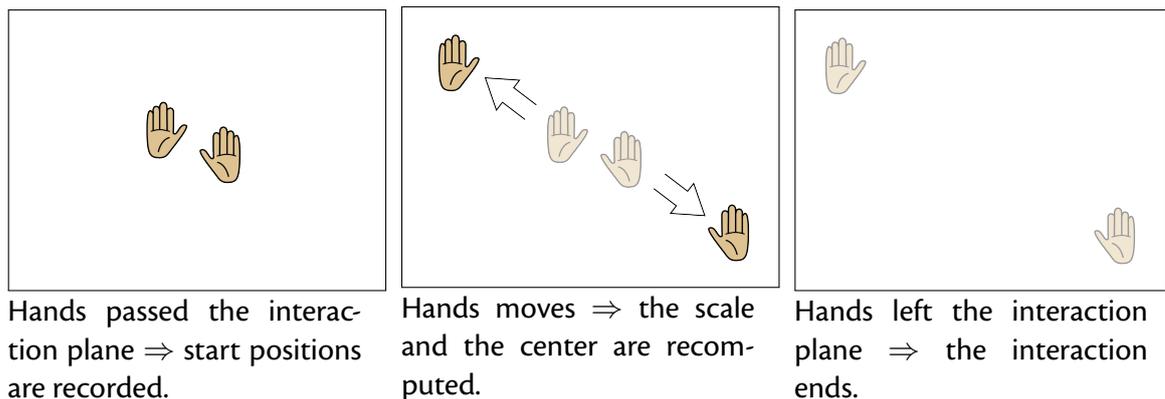
Figure 5.13.: Illustration of the three steps that form the pinch gesture.

gesture is started, and replace the web page with its screenshot while the gesture is active. The screenshot is scaled according the gestures scale factor. When the gesture is finished, the current scale factor is applied to the web page, and the screenshot is replaced by the scaled web page.

5.2.5.2. Swipe Gesture

The swipe gesture is used in the *3D models* and *2D info* views. It can be operated with either the left or right hand. The user starts again by moving his hand past the interaction plane. From that point on, the movement of the hand is tracked, and the distance relative to the start location is recorded as the swipe length. Furthermore, the angle between start and current location is saved as swipe angle. An example illustration of a horizontal swipe is shown in figure 5.14. In the *3D models* view, the swipe length in horizontal direction is translated into a rotation angle around the x-axis. The swipe length in vertical direction is translated into a rotation angle around the y-axis. In the *2D info* view, the swipe length in horizontal direction is translated into

Hand passed the interaction plane ⇒ start position is recorded.	Hand moves ⇒ distance to start position and angle are computed.	Hand left the interaction plane ⇒ the interaction ends.
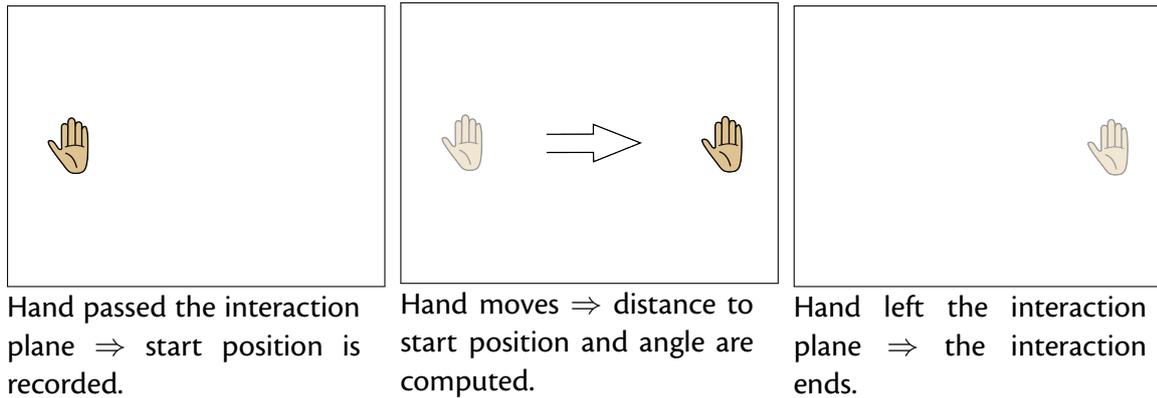
Figure 5.14.: Illustration of the three steps of a horizontal swipe gesture.

	AR view	3D models	2D info
pinch	size and position of focus window	organ scale	web page zoom
swipe	(not used)	organ rotation	web page scroll
click	only used in the button interface for clicking the buttons		

Table 5.1.: Summary of the associated effects of each gesture depending to the current view mode.

a horizontal scroll, and the vertical swipe length is translated into a vertical scroll.

5.2.6. User Study

To compare different methods for the user interface we conducted a preliminary user study. The goals of this user study where:

- Identify the problems of each user interface.

- Find out which user interface feels most comfortable to the user.

- Find out if interaction via gestures is satisfying for the user.

5.2.6.1. Study Setup

To evaluate the implemented options we decided to give each user three different tasks, that represent interaction in each of the three view modes.

- Task 1: In the *AR view*, the user is instructed to inspect her hip bone. For completing this task, she has to move the focus window, or scale it to a very large size.

- Task 2: This assignment is about an organ, e.g. the user has to find out how many disease symptoms are listed for the liver. This information can be found somewhere on the web page of the specific organ. The user has to switch to the *2D info* view, open the page for the requested organ, and navigate through the page to find the requested information.

- Task 3: For the last assignment, the user is instructed to inspect a 3D organ model from all sides. Therefore the user has to switch to the *3D model* view, select the requested organ, and rotate it.

	task 1	task 2	task 3	comfort
Slider / frosted glass	2.09 (0.94)	2.09 (1.58)	1.45 (0.52)	2.00 (0.77)
Slider / hand icons	1.45 (0.52)	1.27 (0.47)	1.55 (0.69)	1.55 (0.52)
Buttons	1.73 (0.90)	4.73 (1.62)	4.73 (1.79)	4.18 (1.17)
Sliding buttons	2.36 (1.43)	2.64 (1.21)	1.91 (0.83)	2.55 (0.93)

Table 5.2.: Average (and standard deviation) values of the user test.

The users are instructed to spend not more than 2 minutes on each task. After each question the user is asked if he was able to complete the task, and to rate the tasks difficulty on a scale between 1 (very easy) and 5 (very difficult). All three methods for selection were used with the frosted glass metaphor. Additionally the selection using a slider was used without the frosted glass visualization, but using hand icons that become visible it the user`s hands passes the interaction plane. The users had to perform these tasks in randomized order. The specific details of tasks two and three were varied for each different user interface. After completing all tasks in all user interfaces, the user is asked if he has a medical education, or experience with AR applications, and if he owns a smartphone. Then, the user is asked to rate how comfortable she felt with each user interface on a scale between 1 (very good) and 5 (very bad).

A total of eleven persons participated in the study. Three of them told us, that they used an augmented reality application before. The same number stated, that they have a medical education. Six persons own a smartphone, and are therefore familiar with using gestures. During the test an instructor told the users how the particular user interface works, e.g. that the sliding buttons have to be pushed outside the screen. The instructor also explained the virtual interaction plane and the gestures. Furthermore, the instructor read the tasks to the users, and wrote down their answers, so the users were not disturbed in their interaction with the application.

5.2.6.2. Results

The slider UI received best grades in comfort and for the specific tasks, both with the frosted glass visualization and the hand icons. The hand icons got the best grades, because most people told us that the feedback is very clear and precise. The user interface with the lowest ratings is the one with the buttons. Seven people were unable to complete the given tasks, because of problems with the click gesture. The results are provided in table 5.2. In the case of the button UI, where some users were unable to complete the given tasks, this was rated with the value 6. Otherwise 1 means very good, or very comfortable, and 5 means very bad or very uncomfortable.

The main problem with the button user interface was that the users first had to pass the interaction surface and then had to perform an additional click gesture. This did not work for many participants. Our first implementation of the click gesture was also not consistent with the frosted glass metaphor, as the user first had to touch the virtual interaction plane and then perform an additional click. After the study, the user interface for clicking buttons was changed such that touching the interaction surface is considered as clicking and no second movement is required. Another problem that was identified during the study is that the interaction plane was at a fixed distance from the depth camera. Users often had problems figuring out where they have to stand. After the study we modified the system such that the interaction plane is at a fixed distance from the user and moves when the user moves. This turned out to be more intuitive. One problem with the frosted-glass metaphor was that some participants did

not recognize when they touched the virtual interaction surface. After the study we added additional color feedback. So in addition to reducing the blur when the user's hand comes closer to the virtual surface, the parts that are touching the virtual surface are drawn greenish. In addition we added audio feedback, such that an audio file is played as soon as the user touches the virtual interaction surface.

5.2.7. Calibration of the User

The NITE skeleton tracking algorithm, which is used in the magic mirror system, 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 can detect a user before the user is calibrated. It cannot estimate the poses of the skeleton joints, but provides a pixel map, where all pixels that belong to a person are labeled. 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 and the systems prints the message *Detected unknown life form*. Now the user has to take a calibration pose where both arms have to be put into the air. The system augments a silhouette of a person taking this pose onto the magic mirror and asks the user to take the same position. During the calibration procedure, which last a couple of seconds, a red laser scanning the user from top to bottom is augmented onto the mirror. After calibration finished the systems prints a message telling that the life form has been recognized as human and the *AR view* will be turned on.

The system uses version 1.4.1.2 of the NITE skeleton tracking libraty. Newer versions, starting from version 1.5.0.2 do not require the calibration pose, however their tracking has shown to be less stable. While for the older version the distance between joints does not change, for the version not requiring the calibration, the joint distances change, which leads to a jittering augmentation of the CT volume. An alternative to NITE is the Kinect for Windows SDK by Microsoft, which does not need a calibration, but does only provide a good tracking as long as the user is facing directly towards the depth camera. As the user of the magic mirror should be able to move freely, the use of this tracking algorithm has shown to be problematic.

5.3. Use of mirracle for Medical Education

5.3.1. Education of Anatomy

The first version of the mirracle system was shown beginning of 2011 and did include the *AR in-situ view* of bones but not the gesture-based interaction or the other views. It attracted huge media attention resulting in multiple online articles, magazine articles and over 200.000 views of a video on YouTube [5]. This resulted in several contacts with medical doctors who are interested in using such a system for teaching of physiotherapy, anatomy, radiology, sports medicine, children's medicine and for general education in science centers. Discussions with these contacts showed that there is a wide range of potential applications of such a system.

The system does not require expensive hardware and can be set up with different display devices. Therefore we have created a demo setup consisting only of a Kinect and a laptop that we can use to easily set up a demo of the system. Starting from February 2011 the system has

---

[5]`http://campar.in.tum.de/Chair/ProjectKinectMagicMirror`

been shown to the public at several occasions. It was shown during the open day of a hospital, in a school and during conferences. In later demos of the system the modes *2D info* and *3D models* have also been shown. Over the course of several months the system has iteratively been improved, made more stable and tested during different public events. In August 2012 the system has been used for two weeks during the IdeenPark, a big public event where technological innovations have been shown to the public. During the IdeenPark the system was used by more than 2.000 people.

The feedback of the users, in particular from children has been very positive. While the AR in-situ visualization attracts the attention of people, it turned out that most users would spend much more time using the volume slicing feature to understand where different organs are. 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 chosen well, calibration problems are reduced largely.

To further develop the system for education of medical students a collaboration with the Academic Medical Center (AMC), the hospital of the University of Amsterdam, has been established. A copy of the system was set up at the AMC such the MDs can directly use it. Using remote desktop connection the system is regularly updated to the latest version of the software such that the medical partners can provide feedback on changes to the system. A concept for using the magic mirror for anatomy education of medical students has been developed in cooperation with the medical partners and a first prototype has been implemented. The flowchart of the concept is shown in figure 5.15.

For the system a list of question are developed by the medical partners. Up to now a first set of questions for the prototype has been generated. Each question asks for the location of an anatomical structure. In each question either the sagittal, coronal or transverse slice passing through an anatomical structure must be found. In the first step the student has to position the focus window at the correct location by moving his hand there. 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. After a position that is close to the structure has been selected, the focus window is fixed and a mode is activated where a more fine grained selection is possible. Here the movement of the hand is scaled down and the user can select a slice in the area of the focus window. Again the user has to confirm her selection, and has to reselect a slice if the selected one was too far from the correct one. After a set of questions has been answered, a score is presented, which is based on the number of tries a student required to find the correct slice.

While gathering questions about the anatomy is has shown that the VKH, while having the advantage that photographic images are available, also has a significant drawback. The dataset has many pathologies, which limits the use for education of anatomy. As an alternative we started evaluating the Visible Human Dataset (VHD) [Spitzer et al., 1996]. In addition we consider using normal CT volumes, where no photographic slices are available.

5.3.2. Visualization of Ultrasound Workflows

In addition to showing anatomy, the magic mirror can also be used to visualize medical workflows. A screenshot of this is shown in figure 5.16. In addition to the augmentation of bones, a replay of an US procedure and an US workflow are augmented. A workflow model has been built as described in section 4.7. This workflow model is visualized onto the user. In addition, a
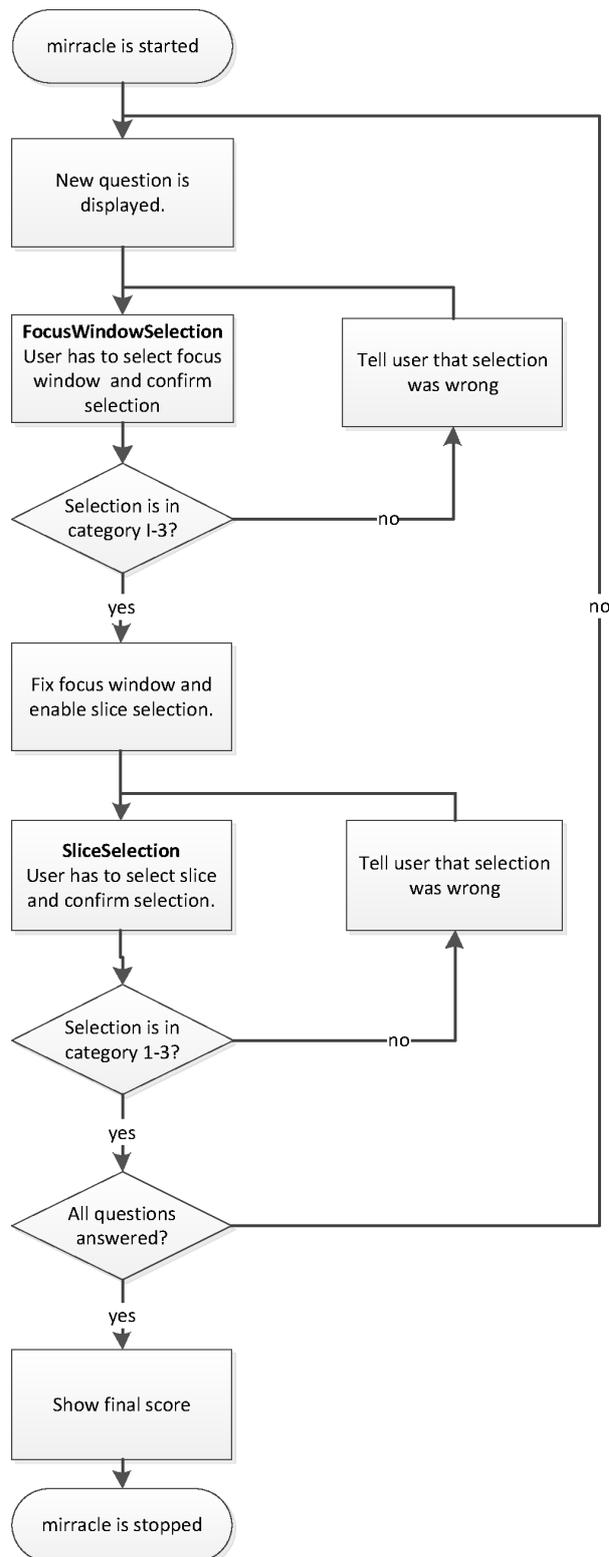
Figure 5.15.: Flowchart of the system for education of anatomy
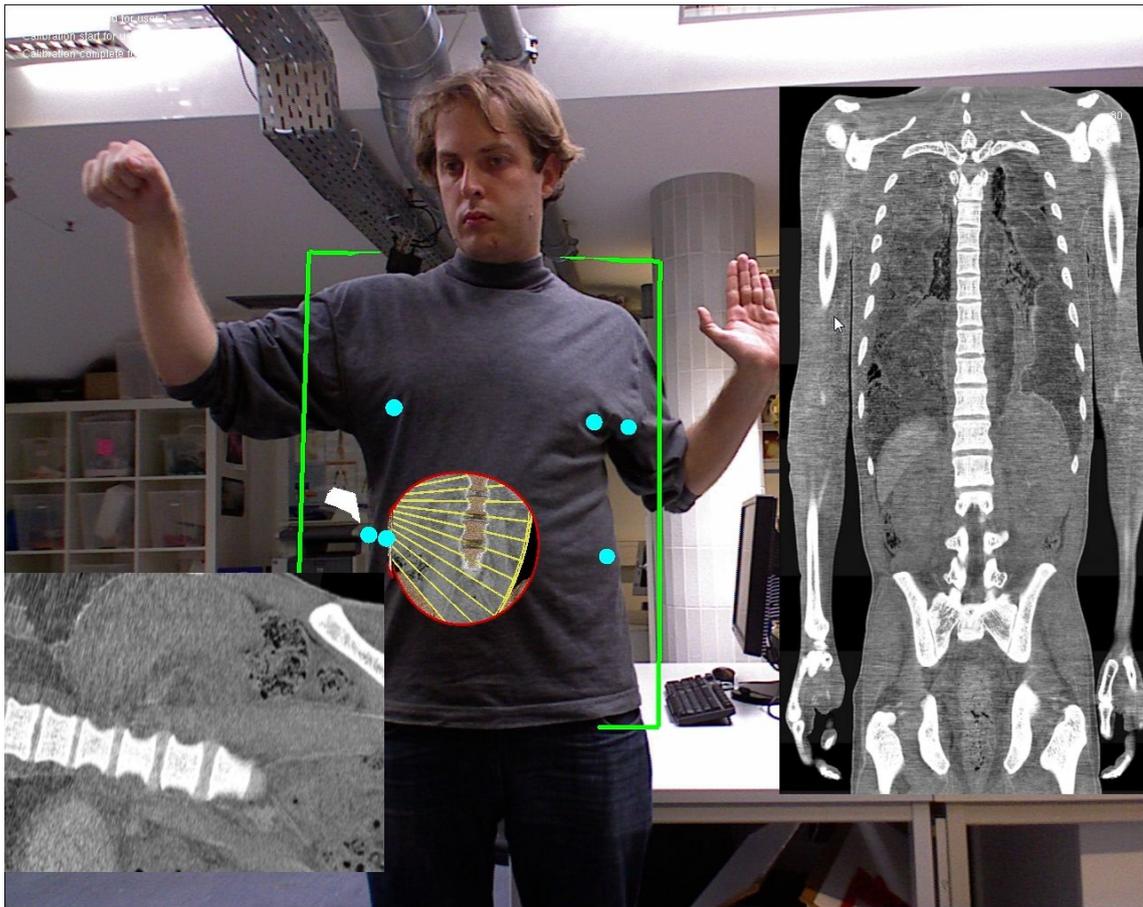
Figure 5.16.: Screenshot of the magic mirror showing an US workflow. The light blue balls represent the workflow steps as in figure 4.21. On the left side a white US probe is augmented. On the lower left the US image plane is shown. In this image the corresponding CT image is shown. And on the right a CT slice is shown, which can be selected by hand movement.

replay of one instance of the procedure is shown. By moving the left hand up or down the user can replay or rewind the replay. Using the right hand, the user can select slices from the CT or photographic dataset as described in section 5.2.1. Such a system could be used by medical students to learn and better understand medical workflows. Using the magic mirror it is possible to relate information to the own body. In addition to using it for medical education such a system could also be used to educate patients about certain medical procedures and for doctor-patient communication.

6. Conclusion

6.1. Summary

In this thesis several aspects of human-computer interaction for medical education and training have been discussed. In chapter 1 an overview of historic and recent developments in medical education with special attention to computer-based education was given. Computer-based simulation and e-learning have become state-of-the-art in some branches of medicine and in the last decades a lot of research has been done in the area. Computer-based education and training has the potential of playing an important role in future medical education, in particular for individualized and competency-based education. In chapter 2 the first exhaustive literature review on computer-based training of ultrasound has been presented. Methods to build US simulators and concepts of using such systems for training and education have been discussed. The main difference to traditional training is that computer-based simulators can offer novel ways to interact with the user and help students to build up a mental model. They enable advanced visualization and new ways to provide feedback. An implementation of an AR ultrasound simulator has been presented in chapter 3. The system uses a generative method for simulation of US images and physical phantoms of patient and probe. Two setups have been developed, one using a webcam and one using a HMD for the AR visualization. To detect where the user is looking, a gaze-tracking device was integrated. Different teaching concepts have been proposed. The AR visualization allows using arbitrarily shaped objects to learn the appearance of basic objects in US. The VKH dataset is used to display co-registered slices of simulated US, CT and photographic images. And the data from the gaze tracking device has been used to visualize the gaze of a recorded US procedure. In chapter 4, methods for modeling and analyzing medical workflows have been discussed. Several applications such as evaluation of skills, context-sensitive user interfaces, monitoring of surgeries and analysis of workflows have been discussed. The use of DTW and HMM for modeling of medical workflows has been described. DTW allows generating average representations from multiple examples of a medical workflow, by synchronizing them. A synchronization of two recorded US procedures can be used to provide an after action review. HMMs can be used to generate more complex workflow models that can contain different paths or loops. One problem is to construct HMMs that can be interpreted by humans. The use of successive state splitting and model merging to generate such models has been discussed. Two applications of such workflow models were introduced: Visualization of surgical workflows and visualization of US workflows in an US simulator. And in chapter 5 a magic mirror system for education of anatomy has been presented. In the context of this system, a novel metaphor for gesture-based interaction has been proposed. One big advantage of the magic mirror is that it uses commercial off-the-shelf hardware and therefore it allows bringing many of the concepts that have been discussed throughout this thesis into real use. The system has been shown to the public many times and a copy of it has been set up at a partner hospital, which allows MDs to use it.

In appendix A two aspects related to user interaction in HMD-based AR are discussed. In the context of stereo displays for AR, a study on the effect of missing out of focus blur is presented.

This study has shown that missing out of focus blur in virtual and augmented images can cause visual discomfort. A first prototype of a system that estimates the depth from two cameras and adds out of focus blur is shown. Furthermore, first steps towards using brain-computer interface devices for AR are been shown and a pilot study with medical experts showed promising results.

6.2. Discussion & Future Work

6.2.1. Augmented Reality Ultrasound Simulator

6.2.1.1. Ultrasound Image Simulation

One shortcoming of the current system is the quality of the simulated US images. While the images are realistic enough to teach the general appearance and basics of US, it is not realistic enough to educate students on pathologies. Looking at the commercial systems it can be seen that only systems that teach echocardiography use generative image simulation methods and all others use interpolative methods (see section 2.2.1 for a comparison of the different methods). Systems using generative methods are usually based on CT, MR or computer models and use this data also for volume visualization. Systems using interpolative methods do usually not have additional images that can be used for a volume visualization. In echocardiography the appearance of a very complex anatomical structure has to be learned. Here the advantage of using computer graphics is very important and the advantage of volume visualization outweighs the drawback of less realistic US images. However, for teaching of pathologies, very realistic images are required, which currently cannot be generated in real-time using generative methods. As we want to use advanced AR visualization we rely on a model that can be used for 3D visualization.

There are two ways to address this problem. One would be a combination of interpolative methods and an additional 3D model or volume. 3D US volumes acquired from patients could be registered with a CT or a virtual physiological human (VPH) model to combine the advantages of both. The other way to address the problem would be to improve the quality of generative methods.

One problem of generative methods is that a lot of simplifications and approximations have to be applied to the physical model that is used in the image generation process to allow real-time performance. Current methods only cast rays through the volume, whereas in real US, waves are propagated through the tissue and complex interactions such as multiple reflections occur. While wave-based simulation using Field II [Jensen, 1996], which is still considered as the gold standard software, takes several days or even weeks, more recent work has shown that realistic simulation can today be done within minutes [Karamalis et al., 2010] or seconds [Hergum et al., 2009, Gao et al., 2009]. While this is not sufficient for a simulator, we can expect further performance increases due to parallelization of the computation.

6.2.1.2. Retrospective Think Aloud

Most material for basic education is designed by teachers, e.g. by writing a book or lecture slides. This however does not necessarily help to transfer the mental model of the expert to the students. Methods to reveal the mental model of experts have been studied in large details, and a standard method is the think aloud or retrospective think aloud method
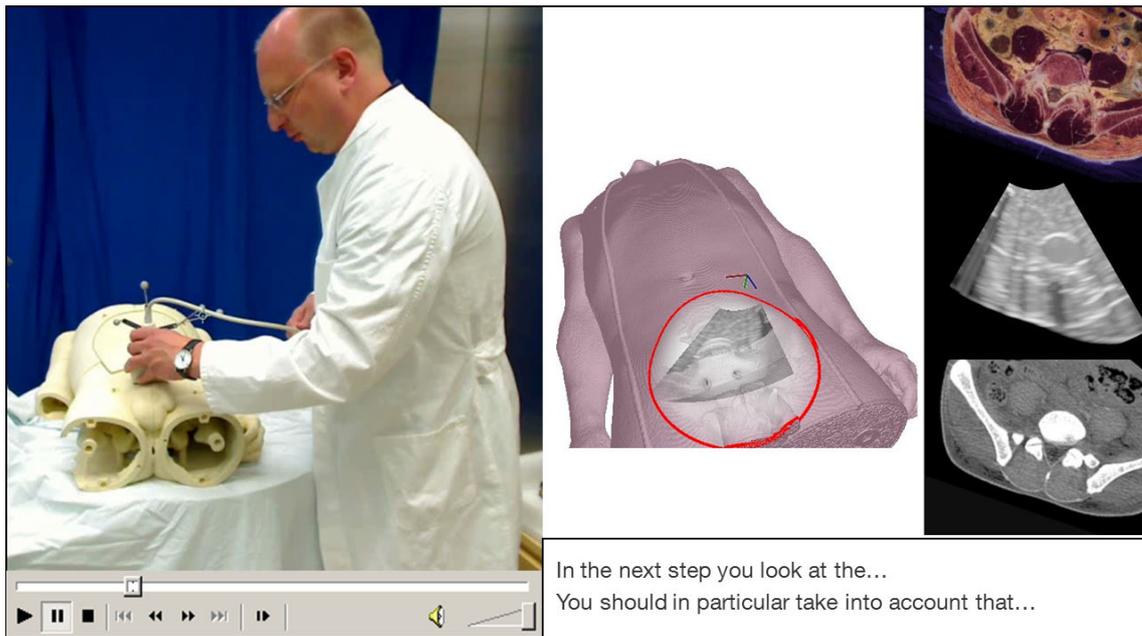
Figure 6.1.: Mock-up of a software to teach the use of US by replaying a retrospective think aloud session.

[van Someren et al., 1994, Nielsen et al., 2002]. In the think aloud method, an expert is asked to verbally express her thoughts while performing a task. For many applications the think aloud method is not used as the expert might be too concentrated on his work to explain his thoughts. For such tasks an alternative is the retrospective think aloud method. When using this method, the task is recorded, e.g. by video, and afterwards the expert formulates what he has been thinking. An US simulator as presented in this thesis could be beneficial both for recording and replaying retrospective think aloud sessions. As discussed in section 3.2.8.4 and shown in figure 3.12 it is possible to record and replay the gaze of the user. Providing a visualization of the gaze during a retrospective think aloud session could help the expert to precisely describe each step of an US procedure. Compared to just showing a replay of a video, a replay of the gaze can help the expert to remember his actions and thoughts more precisely. The same visualization could also be used during a replay for a student. In figure 6.1 a mock-up of a software to teach US procedures by replays of retrospective think aloud is shown. The idea is that an expert is using the simulator to perform a procedure. While doing the procedure a video and the tracking information are recorded. Then he records his thoughts in a retrospective think aloud session. The student does not use the simulator, but only a software that replays the video, displays the thoughts of the expert either as text or as audio recording and shows a 3D scene of the phantom and the US image plane. Additionally co-located slices from the CT and if available the photographic dataset can be shown. For students, the advantage over videotaping real US procedures is that by using the 3D scene, co-located images of other modalities and the gaze-tracking data, a lot of additional information is available.

6.2.1.3. Gaze Analysis

The US simulator presented in chapter 3 allows tracking the 2D gaze position on the monitor and therefore to compute the 3D position inside the CT volume. While the current version

of the system only detects and visualizes fixations, this could be used for more sophisticated analysis of how MDs use and interpret US images. In previous work, information from gaze-tracking devices have been used to predict text relevance by analyzing gaze-patterns while reading a text [Pfeiffer et al., 2005, Lepola, 2005] and HMMs have been used to classify different tasks while reading a text [Simola et al., 2008]. Using such methods on the 3D gaze position it would be possible to recognize which parts of the anatomy are relevant for a diagnosis. This could be helpful e.g. to develop automatic diagnosis systems. Previous work in the medical domain only focused on analyzing search patterns on 2D images, but did not take into account the 3D gaze location inside the image volume. Studies to analyze eye-movement have been done for diagnostic pathology [Krupinski et al., 2006], mammogram interpretation [Nodine et al., 2002, Kundel et al., 2007], laparoscopic surgery [Nicolaou et al., 2004a] and colonoscopy [Almansa et al., 2011]. The advantage of using the US simulator is that the CT that was used for the simulation is available. So an analysis of gaze-patterns could also take into account this additional information.

Another research direction where gaze-tracking data is used is skill assessment, which is a very relevant topic for a simulator system. Gaze tracking has been used in a laparoscopic surgery simulator to analyze differences between experts and novices using the simulator [Law et al., 2004]. They have shown that novices spend more time on looking at tools in order to estimate their position while experts focus only on the target. A similar study, also using a laparoscopic simulator [Wilson et al., 2010], showed very similar results. Using such an analysis it would be possible to develop automatic skill assessment methods, which could be used in the US simulator to rate students or provide feedback. For functional endoscopic sinus surgery [Ahmidi et al., 2010] presented a method that uses the motion of tools and gaze-tracking data for skills evaluation.

6.2.2. Magic Mirror

The magic mirror is not only interesting for anatomy education but also for other medical applications. One example is communication between patients and MDs. This is very important, however MDs sometimes have problems in correctly communication issues to the patient. Recently, an AR system using a mobile projector to augment anatomy on the skin of the patient has been proposed by Ni et al. [Ni et al., 2011] for this purpose. Also the magic mirror could be used for patient communication and patient education allowing the doctor e.g. illustrating different steps of a surgery on the patient.

For the frosted glass metaphor we believe that it will also be important for applications other than in the medical area. Touch-free gesture-based interaction is one of the most important trends in user interfaces, especially as currently a huge amount of devices that enable such interfaces is finding their way into living rooms. The Microsoft Kinect has been sold 8 million times within the first 60 days after market introduction and is the fastest-selling consumer electronics device up today [1]. Therefore we believe that system such as the one described in this paper will find their way into daily use very fast and advanced methods to provide intuitive user interfaces will be of high importance.

We presented the system to many MDs. While the general feedback was very positive, there are also some aspects where the system has to be improved. Several MDs who used the system

---

[1]`http://community.guinnessworldrecords.com/_Kinect-Confirmed-As-Fastest-Selling-Consumer-Electronics-Device/blog/3376939/7691.html`, accessed April 23rd, 2012

stressed the importance of visualizing motion of the anatomy, in particular of the muscles.

6.2.2.1. Serious Games & Motor Rehabilitation

As already discussed in section 1.2.6, serious games are a topic of recent interest. The magic mirror could also be a platform for developing serious games. The advantage of the magic mirror is that the user can relate information to her own body. A related topic is motor rehabilitation where several AR and VR systems exist that make use of gaming elements to make rehabilitation exercises more interesting for the user. Examples of this are VR games for motor rehabilitation [Flynn et al., 2007, Jadhav et al., 2006, Deutsch et al., 2008, Schönauer et al., 2011], balance rehabilitation [Lange et al., 2010, Fitzgerald et al., 2010] and for visually impaired [Morelli et al., 2010]. For motor rehabilitation also different AR based systems have been developed [Gaggioli et al., 2005, Jang et al., 2005]. Another example of an AR system has been developed by [Regenbrecht et al., 2012] for patients who have problems moving one hand. It is a digital version of a mirror box, one method that is used in rehabilitation, which mirrors the healthy arm. Similar concepts could also be implemented using the magic mirror concept. To implement such a system, a 3D model of the healthy hand could be extracted from the video and depth images and this model could be used for augmentation of a virtual hand.

6.2.2.2. Treatment of Phobias

Another application where the magic mirror could be used it the treatment of phobias. Several VR systems for spider exposure therapy have been presented, where the patient is confronted with virtual spiders [Hoffman et al., 2003, Garcia-Palacios et al., 2002, Bouchard et al., 2006]. For people with a cockroach phobia an AR system has been proposed where the user wears a HMD and can see cockroaches crawling over the own hand [Botella et al., 2005, Juan et al., 2006]. A magic mirror system could be used similarly to place virtual spiders or cockroaches onto the mirror image of the user.

6.2.2.3. Real-time Animation of Anatomy

In the current version of the magic mirror a rigid CT volume is augmented onto the user. As soon as the user moves her extremities or bends, the augmentation does not correspond to the real user. Correct visualization of movement is required for several potential future applications of this system, such a teaching of the musculoskeletal system.

To allow movement of extremities, bending and to simulate breathing, different methods could be used. The first solution would be to deform the CT volume-based on the tracked pose of the user. The skeleton tracking algorithm provides the pose of several joints of the user. The pose of these joints could be used to deform the volume. Non real-time deformation of MR volumes has been done by [Rhee et al., 2007] and non real-time full animation of data from the Visible Human Project based on a skeleton has been done by [Gagvani and Silver, 2001]. A real-time deformation of the CT datasets has been proposed by [Walton and Jones, 2007]. This method uses user-defined points for the deformation and allows e.g. to cut the CT volume into two parts. However this method is not suitable to generate plausible deformations that correspond to the movement of the user.

To achieve plausible real-time deformation of the dataset, a GPU-based approach to deform the dataset based on tracked points could be used. The magic mirror system uses the

Visible Korean Human dataset, which includes CT, MR, photographic data and a full segmentation. This segmentation could be used to infer constraints, restricting e.g. the deformation of bone. One problem is that current methods to track the pose from depth images only deliver the pose of certain joints. The movement of the skeletal systems has a much higher degree of freedom than the number of joints that can be tracked. To allow a realistic deformation, constraints could be added such that only a defined set of movements is possible.

Another method to achieve real-time animation of organs would be to replace the volumetric dataset by a polygonal dataset and to animate this by rigging. For rigging, joints are defined in the polygonal dataset and assigned to joints from the user tracking. This has advantages and drawbacks compared to visualization from CT. An advantage of polygonal datasets is that they usually include much more details. The advantage of a deformation of CT volumes is that anatomical slices and radiological images can be extracted from the deformed CT.

6.2.2.4. Modeling and Visualization of Muscle Activity

Visualization of muscle activity was one issue that many MDs who saw our system asked for. Understanding how muscles work is very complex and a system that visualizes muscle activity could be used e.g. to provide a better understanding of muscles to patients that have to do rehabilitation exercises. Such a system would also be interesting for students of sports who have to know about the locomotor system and muscles.

Usually several muscles are involved in moving one bone. Therefore the degrees of freedom for muscle activity are even higher than for bone movement. While there are different methods to simulate muscle activity from motion data, the problem is ill-posed as there are more actuators than degrees of freedom and real-time simulation is not possible to date [Murai et al., 2010].

Modeling and simulation of muscle activity from a Visible Human dataset has been done before [Teran et al., 2005]. It must be noted that this method requires four minutes to simulate one frame. A method for real-time approximation of muscle activity has been shown by [Murai et al., 2010]. However, this method requires motion data and information from electromyographic sensors that are attached to the user. This is not desired in our system as attaching sensors to the user would make usage of the system very time-consuming. Achieving a real-time, realistic simulation of muscle activity using motion capture data does not seem to be a realistic option at the moment, especially as motion estimation from depth images does not provide enough information e.g. about whether the arm is twisted or not.

Instead of providing a full simulation of muscle activity for arbitrary poses, an alternative solution would be to define poses that are of interest in a certain, defined, learning context. The user is asked to perform a motion where the user is instructed to use e.g. a certain twist of the arm. Muscle activity for this motion could be pre-computed using software such as OpenSim [Delp et al., 2007] or modeled for a specific joint, as done by [Baillot et al., 2000] in an AR system to visualize knee joint motions.

6.2.2.5. Advanced Gesture-based 3D Interaction

The current implementation of the frosted glass interaction metaphor takes the 3D position of the hand as input, but reduces it to a 2D interaction on a surface. The advantage is that this allows to use interaction metaphors that most users know from multi-touch displays. On the other hand the user interaction is restricted to 2D. The main concept of the frosted glass interaction metaphor is to use a mixed reality visualization to bring the hands of the user

these objects are not directly visible to the user. image on the screen is created as if real occlusion between hand and objects would occur

image as seen by user on the screen: the hand is in front of the yellow ball, the part of the arm that passed the surface is ocludded by the red box, the rest of the user is behind the surface and therefor not visible

body parts behind surface are not drawn on display

hand

Figure 6.2.: Concept of mixed reality 3D interaction using a depth camera.

and the objects on the screen into the same space. Building on the same idea more advanced user interfaces could be realized. In figure 6.2 the extension of this concept to 3D interaction is illustrated. Instead of showing the parts of the arm that are close to a virtual interaction surface, the parts of the arm that passed a boundary could be shown on the screen. The parts that passed that boundary are not drawn as if they would be behind frosted glass, but as if the arm would be reaching into a virtual space. The video image of the arm is rendered into the virtual scene and occlusion between the real and the virtual objects can occur. This would allow the user e.g. to grasp objects and perform 3D interactions. As gestures are not only relevant for medical education systems, such a concept could also be used for many other applications.

Appendix

A. Perception and User Interfaces for HMD-based AR

While the use of HMDs allows to directly augment anatomy and other data onto the patient or a phantom there are several problems related to HMDs. One is perception of depth and simulator sickness. A more detailed discussion of issues that are related to depth perception in medical AR can be found in [Bichlmeier, 2010]. The other problem is developing appropriate user interfaces. Using mouse and keyboard is no option for most medical AR systems as the user is not sitting but standing or moving. In this appendix, two specific issues related to perception and user interfaces for medical AR are discussed. First in section A.1 a study on the effect of missing blur in stereo visualization is presented. Then in section A.2 first steps towards using brain-computer interfaces are described. Both issues are not only relevant when using AR for medical education but for all AR applications using a HMD.

A.1. The Effect of Missing Blur in Stereo Visualization

Providing a good perception is important for medical AR and VR simulators. Some problems with correct depth perception have already been mentioned before when explaining the use of focus and context visualization in section 3.3.2. In this section one specific problem related to depth perception when using VST-HMDs is examined in more detail, namely the effect of missing our-of-focus blur. First, we will discuss the problem of missing blur, then a study is done to investigate the problem and a first prototype system to address missing blur in VST-HMDs is presented.

It is crucial for stereo AR systems to offer good stereoscopic vision that does not lead to simulator sickness, minimizes visual discomfort and at the same time offers a realistic perception of depth. This is an issue not yet solved completely. In particular for VST-HMD as used for the US simulator it is difficult to achieve stereo vision without visual discomfort and realistic depth perception, as many parameters such as position, orientation and FOV of cameras and monitors have to be chosen correctly. Many effects related to stereo vision in HMDs have been studied in the past. So the vergence-accommodation conflict, which is considered to be a major source of visual discomfort, has been studied extensively [Peli, 1999]. [Min and Jense, 1994] studied the effect of angle and distance between the cameras and the FOV of cameras. [Livingston et al., 2009] investigated the influence of contrast and disparity on the accuracy of stereo perception. [Kooi and Toet, 2004] estimated threshold values for a wide range of causes for viewing discomfort in stereo images.

Beneath other effects such as the vergence-accommodation conflict, high disparities are another cause for visual discomfort in stereo vision [Peli, 1999, Lambooij et al., 2009]. As an effect of the separation s between both eyes or cameras, objects are seen at different positions in the left and the right eye as illustrated in figure A.1. On the left side, the case for normal
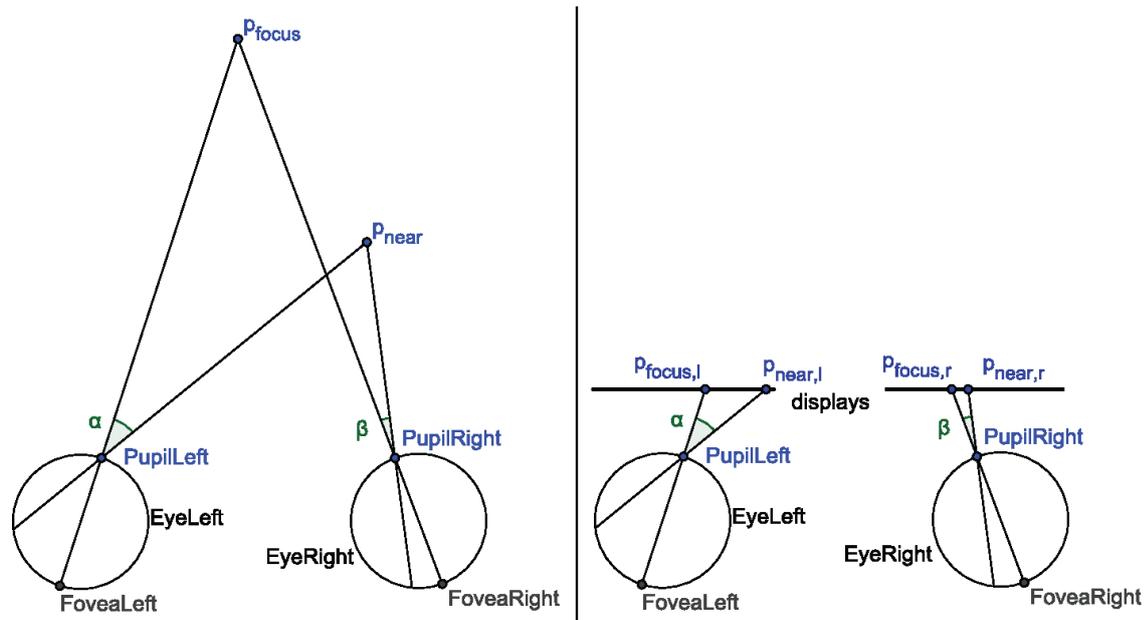
Figure A.1.: On the left the effect of disparity in normal vision is illustrated and on the right when using a HMD.

vision is illustrated and on the right side when using a HMD. Both eyes will always converge towards the object that is currently fixated such that this object is projected to the fovea, which is the part of the eye where sharp vision is possible. Objects that lie at a different depth than the fixated point will be projected to different locations in the left and the right eye. Let p_{focus} be the point that is focused and p_{near} be the point in the image that is closest to the eyes. The disparity can be measured as angle, $d_a = \alpha - \beta$. When using a, HMD let $p_{focus,l}$ be the position where the point p_{focus} is shown on the left display and the other points be defined the same way. Here the disparity can also be defined as distance on the display $d_i = (p_{focus,l} - p_{focus,r}) - (p_{near,l} - p_{near,r})$. The disparity for the far point is defined analogous. On the one hand this effect helps to perceive depth, on the other hand high disparities lead to visual discomfort and high disparities reduce the capability of fusing both images to one. The maximum angular disparity where images can be fused has been estimated as being only $0.1°$ in the fovea and $0.66°$ in the peripheral visual field [Lambooij et al., 2009].

The negative effect of high disparities on the ability to fuse images is not only a problem of stereo displays but also for the human vision. In the real world this effect is only perceived in extreme cases e.g. when focusing on a far object while putting a finger close to the eye a double image of the finger will be seen. The occurrence of double vision is called diplopia. It might surprise that disparity is considered as a problem when using stereo displays, as this effect does not really pose a problem for real vision.

The human eye always focuses on a certain distance. As the pupil has a certain diameter, objects that are at a different distance as the fixated point are blurred and therefore areas where disparities occur are blurred. This can again be seen when focusing on a far object while putting the finger close to the eye. Double vision of the finger is perceived but the finger will be blurred. As soon as one focuses on the finger it will be in focus, but the double image will disappear as the eyes converge towards the finger. The hypothesis of this study is that on stereo displays the effect of double vision is more problematic as in most AR or VR systems only very little or no out-of-focus blur is present. AR systems use mainly small cameras with small sensors

and small aperture widths which both leads to a large depth of field. For virtual objects, blur is usually missing completely as rendering is in most cases done with camera models that do not produce out-of-focus blur.

Several papers have investigated this or related effects. For stereo monitors, [Nagata, 1996] performed an experiment where a foreground object was show in front of a background consisting of random black and white boxes. One version where the background was blurred and one where the background was sharp have been presented to three test subjects. It was shown that blurring the background increases the limits of binocular disparity where the two images could be fused.

[Blohm et al., 1997] investigated on a stereo monitor the effect of adding different amounts of artificial blur to stereo images and video sequences. The test subjects were asked to look either on the foreground while the background was blurred or vice versa. Pairs of images with different amounts of blur were presented to the subject. The experiment showed that higher amounts of blur are preferred.

The effect of adding synthetic depth of field to motion sequences was also investigated by [Sun and Holliman, 2009]. A scene with a flying spaceship was shown, where the fixation point was always on the spaceship and the fore- and background were blurred. In this study, subjects preferred the non-blurred version. It was assumed that the depth of the scene, and therefore the disparities, were not large enough to create diplopia and therefore the effect of adding blur did not improve the perceived quality. Another reason could be that the test subjects were not explicitly asked to always focus at the spaceship. Therefore it is possible that subjects focused at parts of the image that were blurred.

In a survey on visual discomfort, [Lambooij et al., 2009] suggested to build a system that uses a gaze-tracker to estimate the focus point of the subject and add artificial out-of-focus blur based on this. However they did not implement such a system.

While in previous work the effect was only investigated for completely virtual images, missing out-of-focus blur is also a problem for AR. Similar to the solution proposed by [Lambooij et al., 2009] the problem could be addressed in AR. A gaze-tracker, integrated into a HMD, can be used to estimate the 2D gaze position of the user on the display. Using a soft- or hardware solution to obtain a depth image it is possible to estimate the 3D gaze position. Knowing the 3D focus point and using the depth map, artificial out-of-focus blur can be added to the images. In this section four contributions on using out-of-focus blur to reduce visual discomfort are presented. First, the first implementation and evaluation of a VR system is presented that uses an gaze-tracker to estimate the focus point, adds artificial blur to non fixated layers, and presents the image using a stereo monitor, as suggested by [Lambooij et al., 2009]. Second, the first evaluation on the effect of out-of-focus blur on a stereo HMD is shown. Third, a first prototype of an AR system is presented that recovers depth information from the stereo cameras of a VST-HMD and uses this information to add out-of-focus blur. Fourth, the first experiments not only using virtual images, but also photographed scenes are presented.

A.1.1. Related Work

Several other methods to address the problem of diplopia are known in computer graphics. A comparison of different methods can be found in [Wartell et al., 2002]. These methods either involve a static or dynamic change in the separation of the virtual cameras or some sort of image or view scaling to reduce the amount of disparity. For most AR applications these methods are not well suited, because they either change the camera separation which will lead to a

wrong perception of depth, or they change the FOV, which is known to have negative effects on stereo vision [Min and Jense, 1994]. Using another FOV as the one of the HMD would also lead to a magnification or demagnification of the image, which is usually not desired in AR.

For AR, artificial blur has been applied to virtual objects in order to better merge them with camera images. [Fischer et al., 2006] added artificial motion blur and synthetic noise to virtual objects such that they look more similar to real objects recorded by cameras. An advanced camera model that includes motion blur, distortions and effects from the image processing pipeline of a camera has been shown by [Klein and Murray, 2008]. However, they did not model out-of-focus blur. [Okumura et al., 2006] tracked a marker and estimated out-of-focus blur and motion blur from the appearance of the marker in the image. Virtual objects close to the marker are blurred based on the estimated blurring of the marker. Using this method a better integration of virtual objects into the camera image can be achieved, but it is not possible to change the focus distance or apply artificial blur to real objects in the camera image. A template matching-based tracking that can handle motion blur has been presented by [Park et al., 2009] and has been used to add artificial motion blur to virtual objects.

The effect of out-of-focus blur as depth cue has been studied in [Mather and Smith, 2000] by showing images that only consist of noise, where some parts are blurred. Their study showed that blur is only a minor depth cue. The use of blur as a way to reduce visual discomfort has not been considered in their work.

A.1.2. Methods

A.1.2.1. Objective and Hypothesis

The objective of this study is to investigate the effect of out-of-focus blur on visual discomfort when using stereo displays. We performed experiments both for a HMD and a stereo monitor. The main hypothesis is that applying blur to non-fixated layers in a scene will lead to lower visual discomfort. We also assume that the effect will be stronger in images with higher disparities.

A.1.2.2. Subjects

A total of 18 subjects participated in the test, where one subject only performed the experiment using the stereo monitor. Four were female and 14 were male and the average age is $25.44(\pm3.13)$. The participation of all subjects was voluntary and they have not been compensated. All subjects were made familiar with the two different visualization methods, but were not aware of the objective and hypothesis of the evaluation. One subject did not pass the stereo vision test that was performed at the beginning and the results from this subject were excluded. For four people the gaze-tracker did not work precise enough when using the monitor. For these people the focus distance was set manually as done when using the HMD.

A.1.2.3. Apparatus and Stimuli

We used two different setups, one with a stereo monitor and one with a VST-HMD. The first setup consists of a stereo monitor using polarized glasses. Additionally we used a gaze-tracker device, which in most cases worked without problems when wearing the polarized glasses. Instead of showing images from video cameras we have chosen to use a virtual scene which gives us full control over all parameters, such as separation and angle of the cameras, FOV and

position of all objects. The same 22" stereo monitor and the same gaze-tracking device as used in the US simulator have been used (see 3.2.6 and 3.2.7 for details on the devices). The subject was positioned at a distance of 80cm in front of the screen, resulting in a horizontal field of view of $32°$. Behind the subject a black fabric was placed to avoid reflection on the monitor, which is very glossy. The gaze-tracker was used to estimate the gaze point of the user on the stereo monitor. The virtual scene was rendered for both eyes using quad-buffering. Knowing the 2D gaze position on the monitor, the 3D gaze point in the virtual scene can be obtained by reading out one pixel in the depth buffer.

In previous work on reducing visual discomfort due to high disparities, different methods have been used. [Nagata, 1996] did not aim at generating out-of-focus blur but only investigated the general effect of sharp or blurred background. They used a low-pass filter to generate the blurred images. A filtering approach that is based on a geometric optical model has been used by [Blohm et al., 1997] to blur video sequences. While this method provides good results, it is too slow for real-time processing. An accumulation buffer method, where the image is rendered multiple times always moving the camera slightly was used by [Sun and Holliman, 2009]. We first tried this method but did not use it for the experiments, because performance for more complex scenes is low and this method only works on purely virtual scenes and not for AR scenes. Other methods for adding synthetic depth of field are discussed in [Demers, 2004]. We have chosen to use a reverse-mapped z-buffer method. The implementation we used is similar to the one described in [Riguer et al., 2003]. It is implemented on the GPU and uses two subsequent rendering passes. In the first rendering pass the scene is drawn without blur and at the same time for every pixel the blurriness is computed based on the distance to the gaze point in z-direction, where the z-axis points into the screen. This is done by reading the depth from the depth buffer and comparing it to the depth of the gaze point. The blurriness for every pixel is stored in an additional texture. In the second pass each pixel is blurred by sampling neighboring pixels using stochastic jittering. The distance of the sampled points is scaled by the blurriness of the pixel and every sampled point is weighted by the blurriness of that point. The sampling of the points is done in the world coordinate system and not in image coordinates. As an effect of this, a point that is closer to the viewer will be blurred with a higher radius, as it is the case in reality. The depth of each point is also considered to avoid that background objects blur into foreground objects.

Eleven different virtual scenes were shown to each subject. In every scene, the foreground object was at a distance of 80cm and the distance of the background image varied between 102cm and 294cm resulting in on-screen disparities between 3% and 10% of the screen width. Scenes with different contrast and different fore- and background images have been used. Some of the images are photographs some consist only of a texture. In addition to the virtual scenes we also used real, static scenes where two images with a baseline of ∼7cm have been taken using a photo camera. The photos always contain one foreground object and one or several background objects where the distance to all background objects is very similar. The photos have been segmented manually into fore- and background. Using the same methods as for the virtual scenes, the fore- and background can be blurred based on the gaze of the user. An example of a virtual and a real scene can be seen in figure A.2.

In the second setup the same OST-HMD that is used in the US simulator is used. It is an nVisor SX OST-HMD with a resolution of 1280x1024 and a horizontal field of view of $48°$. While gaze-trackers that can be integrated into HMDs are commercially available (e.g. by Applied Science Laboratories, Bedford, United States) we do not have a HMD equipped with an gaze-tracker. Therefore the focus distance in this setup is set manually. For visualization, this setup
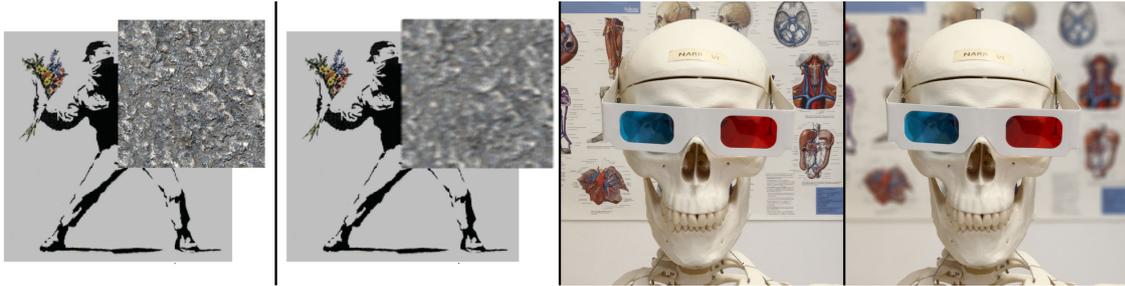
Figure A.2.: Images with and without simulated out-of-focus blur. On the left a virtual scene is shown, on the right a photographed one.

uses the same virtual scenes, respectively the same photographic images as in the other setup. Also the blurring is performed using the same technique.

A.1.2.4. Procedure

For the test method we followed the ITU-R BT.500-11 recommendation [BT, 2002] for subjective assessment of the quality of television pictures. Two previous studies on the effect of blur on visual discomfort followed the same recommendations [Sun and Holliman, 2009, Blohm et al., 1997]. For assessment of stereoscopic images, the simultaneous double stimulus for continuous evaluation method is recommended by the International Telecommunication Union (ITU) and was used in our experiment. First, all subjects had to undergo a stereo vision test to see if they can perceive stereo images. For this, a random-dot-stereogram was used. Next it was decided randomly to start either with the stereo monitor or the HMD. For the monitor a calibration of the gaze-tracking system was performed. Afterwards one scene was shown to illustrate the two different visualization methods. Next, eleven virtual scenes were shown to the subject. Every scene was presented in a version with blur and one without blur. Whether first the blurred or the non-blurred version was shown has been randomized for every image. In alternating manner, the subject was asked to look at the fore- and the background, for four seconds each. This was repeated twice, then a mid-gray level screen was shown for five seconds and the same scene was shown again using the other visualization mode. After having seen both modes the subject could either choose to rate their quality or to see both sequences again. To rate the images the subject was asked to put one mark for each version on a vertical continuous scale, which is divided into five equal lengths, corresponding to the ITU-R five-point quality scale. The five parts on the scale correspond to quality ratings ranging from bad to excellent, but the subjects were free to put there mark at any place of the scale. Afterwards the mark of the subject was measured and the results are converted to a scale between 0 and 100. The first three of the eleven scenes were only used to stabilize the result and were not used for the evaluation. The eight scenes that were used for the statistics are always the same scenes. After the virtual scenes were shown, the three photographed scenes were presented to the subject. The order in which the images were shown was randomized for every subject. After finishing the experiment on the HMD and the monitor the subjects had to fill out a short questionnaire on how often they use different types stereo displays and whether they wear glasses.

	no blur	blur	p-value
HMD virt.	$63.56(\pm12.15)$	$71.55(\pm9.15)$	**p=0.008**
HMD real	$71.15(\pm16.77)$	$73.47 \pm (17.06)$	p=0.359
Monitor virt.	$62.04(\pm11.35)$	$70.21(\pm10.42)$	**p=0.026**
Monitor real	$70.35(\pm19.82)$	$79.02 \pm (12.97)$	p=0.079

Table A.1.: Means, standard deviations and p-values of the quality ratings. P-values that can be considered as significant are drawn bold.

A.1.3. Results and Discussion

For every subject we averaged the scores for all eight virtual, respectively all three real scenes. The results of all subjects were screened for outliers based on the ITU recommendation but no one had to be eliminated. The quality ratings have been tested for normal distribution using the Kolmogorov-Smirnov test and they can be considered as having normal distribution. A paired one-tailed t-test was performed on the average ratings. The results are shown in table A.1 and the distribution is shown in figure A.3. For both, the monitor and the HMD, the difference is significant at a level of $p < 0.05$. However, the difference between the blurred and non-blurred version is lower than we expected. We realized that, after having finished with the first few subjects. We started to do interviews with some of the subjects after they finished the experiment. In the interviews some subjects reported that they did not perceive double vision at all, even for the images that have high disparities. With some subjects we performed a threshold test, where we showed a virtual scene consisting of a fore- and background object. At the beginning the camera separation was set to 0. We step-wise increased the camera separation and asked the subject to report any inconvenience or double vision to us. Some subjects reported double vision very soon while others did not perceive double vision even with on-screen disparities above 30% of the screen width. We noticed that people perceiving double vision very soon, did prefer the blurred version. For the subjects that did not perceive double vision, the results were not that clear. However, as we only performed the threshold test for few subjects, we could not obtain any significant results for that. Some subjects reported that they did not experience any discomfort and they rated the image quality based on other criteria. Some people liked that they are not distracted from the foreground when the background is blurred. Others liked it more to see the whole image sharply.

We assumed that the blurred version would be preferred in particular in scenes with high disparities. The average rating for every single image is shown in figure A.4, where the disparity increases from left to right. As can be seen in figure A.4 this assumption could not be confirmed. We believe that there are two reasons why this assumption did not hold true. First, as discussed above, many people did not perceive double vision in the first place. Second, additional effects such as the contrast of the images seem to affect the perceived quality very much. So the image with 6% disparity is the image with the highest contrast. This image got the lowest quality rating of all images while most subjects did clearly prefer the blurred version. There have been two previous studies where it was shown that subjects preferred artificial depth of field [Blohm et al., 1997] respectively blurring the background [Nagata, 1996]. In both studies, images with very high contrast were used. In another study, where no significant difference between using artificial depth of field or not could be shown [Sun and Holliman, 2009], low contrast images have been used. These results suggest that the use of artificial depth of field
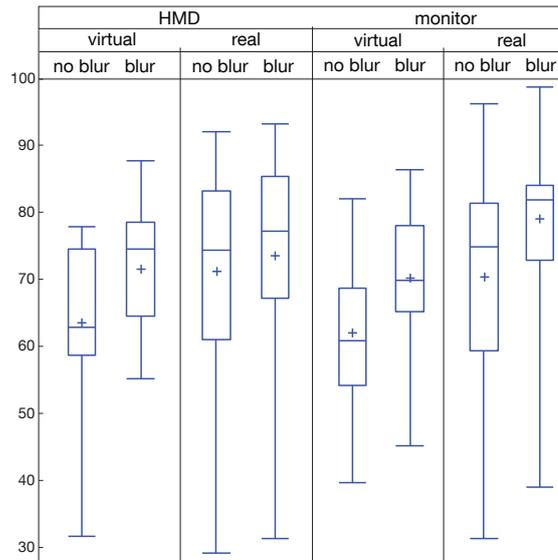
Figure A.3.: Boxplot of the mean quality ratings per subject, showing median, upper and lower quartiles and the mean values (cross).
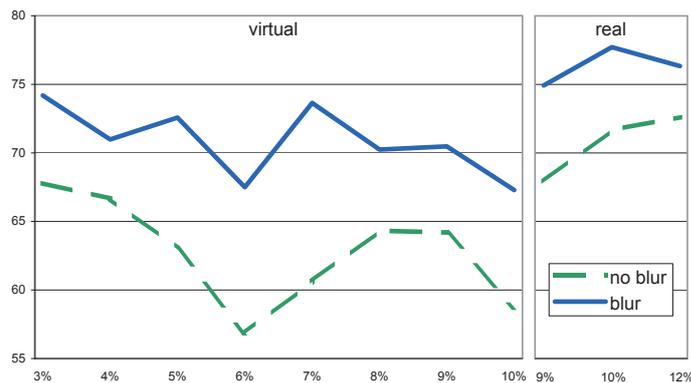
Figure A.4.: The x-axis shows the disparity in percent of the screen width. On the y-axis the quality rating is shown. The values are averages from both, the monitor and the HMD.

would be preferred in particular in high contrast images, which can be confirmed by our results.

There is no significant difference between using the stereo monitor and the HMD. The difference for the real images is not as big as for the virtual scenes as can be seen in table A.1. While there is a difference between the mean quality ratings, this difference cannot be considered as significant. The ratings are less consistent and have a higher standard deviation compared to the virtual scenes. While for the virtual scenes many subjects did only experience smaller differences between the two versions, for the real scenes more people had a strong preference towards the blurred or non-blurred version. Therefore we can assume that additional effects that are present in the real but not in the virtual images influence the results. But it must also be noted that none of the photographs had strong contrast.

We made some additional interesting observations. Three subjects said that they prefer a sharp foreground with a blurred background over a completely sharp scene, but they prefer the

completely sharp scene over a sharp background and a blurred foreground. Sun and Holliman [Sun and Holliman, 2009] reported the same comment by two subjects in a study on depth of field in animated stereo scenes. Some subjects mentioned that they did not like the transition when the focus distance changes and two people assumed that the effect of adding depth of field would be more beneficial in moving scenes.

A.1.4. Prototype of an AR System Using Adaptive Blur

There are several ways how such a system could be built for AR. To do this, three issues have to be addressed. First, the gaze of the user has to be tracked. This can done using gaze-tracking, which is commercially available for OST-HMDs. Second, the depth of the gaze point has to be known. In a stereo HMD this information can be obtained from a stereo depth reconstruction. Another solution would be to use a depth camera such as a time-of-flight (TOF) camera or a structured light camera such as the Kinect, which is used in the magic mirror system in chapter 5 in another context. Calibrating the depth camera and the optical camera to each other, depth values can be obtained for all pixels in the video image. The third problem is to blur the image. This can be done using a camera with small depth of field and a focus that can be controlled from a computer. Another solution could be to use coded aperture cameras [Levin et al., 2007] or plenoptic cameras [Ng et al., 2005] that can both recover the depth information and change the focus after the image has been obtained. However, all methods that use real out-of-focus blur generated by the lens and the sensor, have the problem that in AR usually small cameras with small lenses are used, which only produce a low amount of out-of-focus blur. Therefore the use of artificial blur is to be preferred, which requires a depth map of the scene. See [Demers, 2004] for an overview of different methods to simulate depth of field.

To avoid the use of additional hardware we have decided to estimate the depth by a stereo depth reconstruction using the two cameras of the HMD. In order to compute the depth of a pixel, the disparity of the pixel between the left and the right image has to be computed by a disparity matching. Prior to the disparity matching the two camera images of the HMD have to be rectified as described by [Loop and Zhang, 1999]. Our current disparity matching solution is implemented in OpenCL[1], a C++ library for parallel programming, in order to be executed on the GPU. For efficient execution we utilize shared memory as well as GL/CL sharing. The matching algorithm is based on local normalized cross-correlation (LNCC) with 9x9 windows. For correct matching, even within homogeneous regions, we use a Gaussian pyramid based approach as used by [Sizintsev and Wildes, 2010] with four to five levels. In the current version we only classify pixels into fore- or background. For this purpose we decided to calculate the disparity for all but the highest level. Both components on their own are able to achieve nearly real-time performance. Due to this, further implementations could be based on two GPUs for separate calculation to achieve real-time performance.

In order to enable the use of the gaze-tracker, in the current implementation the users sees the camera images only on a monitor and not on the HMD itself. The gaze is tracked and the parts of the image that are at a different depth as the focus point are blurred, using the same methods as described before. Adding virtual objects and also blurring them is straightforward and requires only setting the reconstructed depth map as depth buffer before drawing an object. Our current prototype is promising, but the quality of the depth reconstruction and the speed have to be further improved to build a reliable system.

---

[1] `www.khronos.org/opencl`

A.1.5. Conclusion

In this section an experiment on using artificial depth of field to reduce visual discomfort when using stereo displays was presented. Our main hypothesis was that out-of-focus blur would reduce visual discomfort due to double vision. Results showed a significantly better perceived quality when using out-of-focus blur in virtual scenes. However, for photographed scenes results were not significant. Taking into account that many subjects reported that they did not perceive double vision at all we could not confirm our initial hypothesis without doubt. Looking at the results and considereing the interviews we did with some of the subjects, we assume that there is a multitude of reasons why some subjects like or dislike the use of blur. In particular for the photographed scenes the results are not very consistent among the subjects.

So, when building an AR system the use of artificial out-of-focus blur should be considered, but it should not be assumed that all users prefer blur. As mentioned before, several subjects preferred only blurring the background when looking at the foreground, but not the other way round. It might be a good solution to build a system that only blurs the background when the user looks at nearby objects. We must further investigate the interplay between different aspects of the image. So the contrast seems to influence the perceived image quality strongly. Also the influence of further characteristics of the images such as color, homogeneity or repetitiveness should be investigated in more detail. Future studies should take into account differences between the subjects. We investigated only whether wearing glasses or working regularly with stereo displays affects the perception of visual discomfort due to disparities and we did not find any significant relation. Other parameters that could influence the perception are long vision, short vision or ocular dominance of one eye. While in our study only static scenes with fixed viewpoint have been shown, it has to be investigated whether the same assumptions hold true for dynamic scenes. Next steps for building a working system using a HMD could be an improvement of the depth reconstruction algorithm or the additional use of depth cameras, as the results obtained with the stereo images are currently not reliable enough.

In particular, when combining artificial depth of field with other methods to improve stereo vision in AR, as for example virtual convergence [State et al., 2001] or simulation of camera artifacts [Klein and Murray, 2008], we have to further investigate the different effects that lead to comfortable stereo vision. Insight on how to do this could also be gained from other areas where stereo vision is important. So, in stereo photography, which is already done for over hundred years, many rules of thumb exist. So e.g. the maximum on screen disparity for stereo photography is usually considered to be 3%. Another area that is currently very active is stereo movies, where extensive use of blur and other methods to reduce visual discomfort is made.

A.2. Towards Brain-Computer Interfaces for Medical Augmented Reality

After having examined one problem related to perception and displays for AR, in this section one specific user interface for HMD-based AR will be discussed. The ultimate vision of medical AR is enabling a Superman-like X-ray vision into the body of the patient or a phantom. The use of display devices such as HMDs to augment pre- or intra-operative images onto patients or phantoms comes already very close to this idea. Just like Superman, MDs do not want to use the X-ray vision constantly as often they have to see the skin of the patient, their hands or instruments outside the patient body. The same is true for medical education. To switch
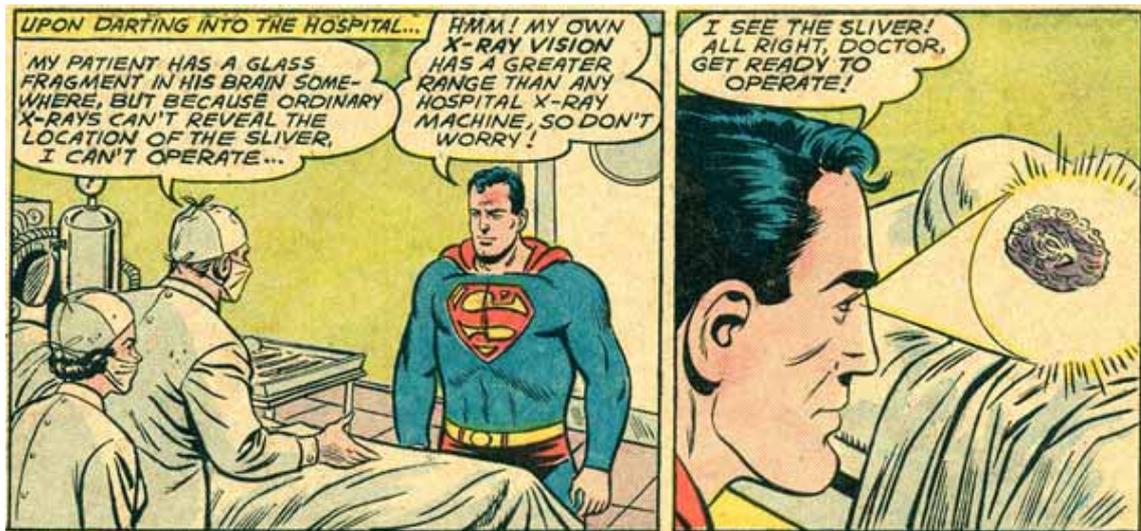
Figure A.5.: Illustration of superman not using a traditional user interfaces to turn on his X-ray vision [Finger and Platino, 1961].

between standard and X-ray vision or change parameters of the visualization, a user interface is required. Developing appropriate UIs for such tasks is a major problem for AR. While for medical education it might be acceptable to use traditional user interfaces such as mouse and keyboard, it limits the user as he always has to move to the keyboard for any user interaction. In a surgical setup, traditional user interfaces are even more problematic as the surgeon has to stay sterile and most surgeries require the surgeon to use both hands. As is commonly known, Superman does not need a traditional UI, but can control the X-ray vision with his mind (see figure A.5 for an illustration of this). In this section first steps towards transferring the concept of mind-controlled X-ray vision from comic books to medical AR by using gaze-tracking in combination with brain-computer interface (BCI) devices that measure bioelectric signals are discussed.

Augmenting structures inside a human body, while at the same time showing structures outside the patient, is a common problem in medical AR. Most systems use only simple transparency [Kersten-Oertel et al., 2010] which leads to bad depth perception and requires an UI to change the transparency. An alternative to transparency are virtual windows into the patient, which is a visualization technique that has already been used by one of the very first medical AR systems [Bajura et al., 1992]. A similar F+C visualization has been discussed before for the US simulator. While such a visualization improves depth perception, the virtual window hides real objects and this technique also requires an UI to change visualization parameters. For non-medical AR, different solutions to visualize occluded objects have been proposed using e.g. visual saliency [Sandor et al., 2010] to preserve information on the occluding object while showing the occluded object. While such methods can improve the perception, in particular for medical AR it is crucial that the user can control the visualization and switch to a mode where only the real view or only the augmented object of interest is shown. Using methods that automatically decide which features of the real and the augmented images should be presented are problematic in medical AR, as they can influence the perception and decisions of the MD.

BCI devices have been used in combination with AR by [Faller et al., 2010] and by [Takano et al., 2011] in steady-state visual evoked potential (SSVEP) BCI systems. SSVEP sys-
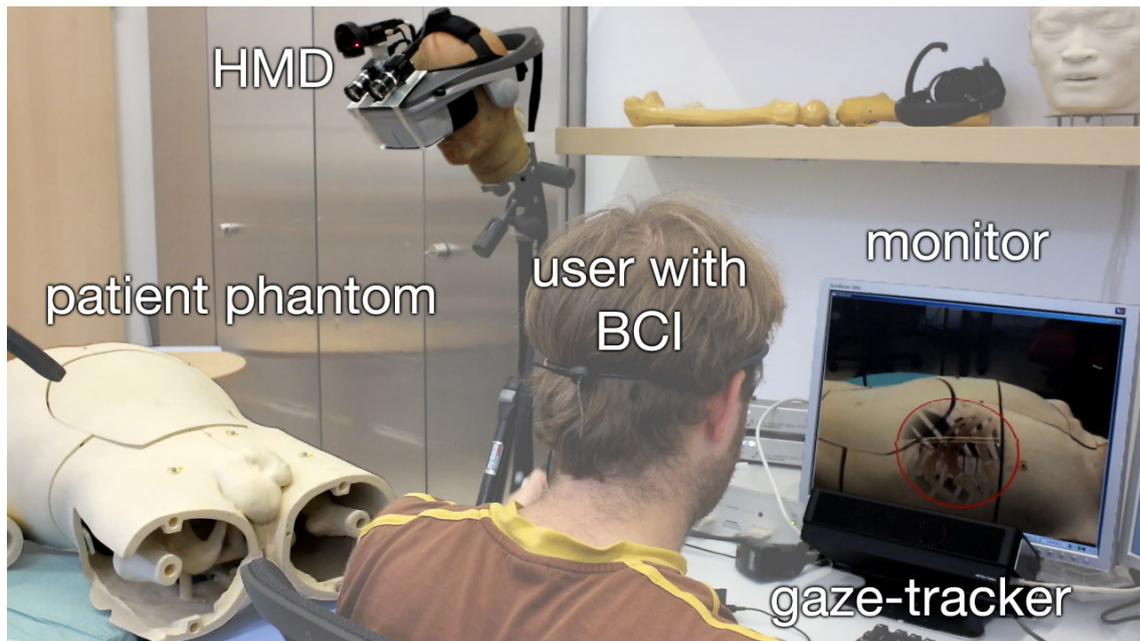
Figure A.6.: Setup using a HMD, a monitor-based visualization, a gaze-tracker and a BCI.

tems use active stimuli such as flickering lights or patterns and can detect if the user is looking at them. In both systems AR was used to augment active stimuli that allow triggering interactions with real objects.

A.2.1. Methods

For this study again the CAMPAR framework is used, which was also used for the US simulator. In addition to the gaze-tracker (see section 3.2.7) a BCI device was integrated. We used the Neural Impulse Actuator (NIA) (OCZ Technology, San Jose, United States). The NIA is a headband that can read bioelectric signals. It can be used to read alpha and beta brain waves, electromyographic (EMG) and electrooculographic signals. The device is marketed for computer games and is based on technology for hands-free computer access for people with disabilities. After initial tests it has shown that learning how to control the alpha and beta waves requires a longer learning phase. The goal of our pilot study was to get first feedback from MDs on the use of BCI. Although alpha and beta waves would allow controlling additional parameters [McFarland et al., 2008] we decided not to use them and to avoid the learning phase for the MDs. When using the electrooculographic signals, which are triggered by eye movement, the user has to look to the left or right. Because the gaze position is important, as will be detailed later, the use of eye movement showed to be difficult. For this study we focused on the use of EMG signals as controlling them can be learned within few minutes. EMG signal raise with muscle tension. It has shown that the easiest and most robust way for the user to increase the EMG signal is to raise the eyebrows. As an alternative BCI device we tried the NeuroSky mindset (NeuroSky, San Jose, United States), a similar device that can detect brainwaves, concentration and eye-blinks. We decided not to use the mindset in this study for different reasons. The mindset was too big to fit under the HMD. Furthermore it was difficult to differentiate between normal eye-blinks and eye-blinks that should trigger an action. The use of the concentration value has also shown to be difficult, as during a task such as medical education or

surgery it is difficult for the user to control the concentration.

The BCI and the gaze-tracker were used in two different setups. The first setup is the one that was also used for the HMD-based US simulator discussed in section 3.3, using an optical tracking system and a video see-through HMD to augment a volume rendering of a CT on a phantom using a focus and context visualization. This setup is shown in figure A.6. We implemented two different methods to use the BCI. In the first method, the focus window and the augmented anatomy is only shown as long as the EMG signal is above a threshold. When using the HMD, the focus window is fixed to the center of the view and can thus be controlled by moving the head. We also presented the images from the HMD on a screen to the participants of the study. In this setup the gaze-tracker is used to control the position of the focus window. In the second method, a peak in the EMG signal is used to switch between three different visualization modes. In the first mode, no augmentation is shown. By an EMG peak the second mode is activated where the position of the focus window is controlled by the gaze. In the next mode the focus window is fixed to the current gaze-position. The next EMG peak will switch off the augmentations again.

The second system we integrated the BCI is CamC [Navab et al., 2010], a monitor-based medical AR system where an optical camera is mounted next to a mobile X-ray device. Using a mirror construction the optical camera has the same view on the patient as the X-ray device. The X-ray and the video image are overlaid by a one-time calibration. CamC is also implemented using the CAMPAR framework and has been used in over 40 real surgeries. The standard visualization method of CamC combines the X-ray and the optical image using simple alpha-blending as can be seen in figure A.7. A slider on a touch-screen controls the alpha value. Using the touch-screen during a surgery is difficult as the surgeon has to stay sterile and in most OR setups the screen is not within reach of the surgeon. Usually it can only be operated indirect via communication with a nurse. The BCI was integrated similar as in the HMD-setup. Using peaks in the EMG signal the user can switch between three modes, where in the first mode only the X-ray is shown, in the second mode the video is augmented in a circular area around the current gaze-position, and in the third mode, the area where the video is shown is fixed.

The system was shown to 9 medical professionals, 3 of them female, and with an average age of 32.7. 6 of them were last year medical students. We explained the technologies to them and familiarized them with the BCI and the gaze-tracker. All participants used the HMD-based setup once wearing the HMD and once using the monitor and the gaze-tracker. The use of the BCI with CamC was illustrated to them. All participants were also made familiar with the possibility to capture additional brain waves and use them to control additional parameters. Afterwards they answered several questions on a scale between 1 (I totally disagree) and 5 (I totally agree).

A.2.2. Results

For the majority of the participants the BCI was intuitive to use (4.6 ± 0.7). Surprisingly they rated the possible use of BCI for non-AR applications higher (4.0 ± 0.7) than for HMD-based (3.6 ± 1.0) and monitor-based (3.8 ± 0.7) AR. The additional use of gaze-tracking was seen as very valuable (4.7 ± 0.7) and most participants preferred using a peak in the EMG signal to trigger an interaction over using constant muscle activity (4.2 ± 1.0). Most participants could imagine using a BCI for a short time during a surgery (4.6 ± 0.5) while there was no clear tendency on whether the use during a whole surgery is acceptable (2.7 ± 1.2). Most
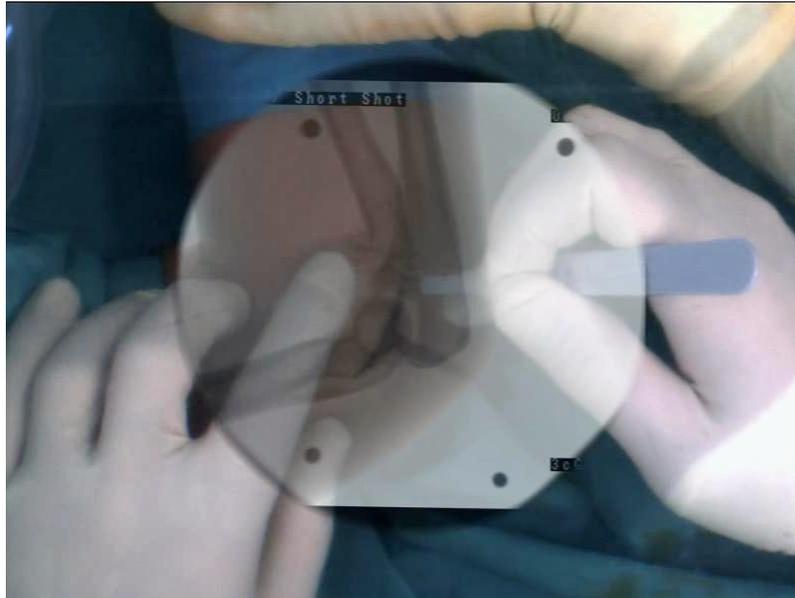
Figure A.7.: Example of an image produced by the CamC system when not using the BCI. A camera and a X-ray image are overlaid using transparency.

participants stated that the recognition of the BCI events has to be improved (4.1 ± 1.1) and the majority would be willing to spend time to learn how to use brain waves to control additional parameters (4.2 ± 0.7).

A.2.3. Discussion

In this section an integration of a BCI device and a gaze-tracking into a medical AR system and a pilot study to assess the potential of the technology was presented. One drawback of the study is that no brain waves but only EMG signals have been used to avoid a learning phase for the MDs. However an important result from the questionnaire was that most participants would be willing to spend time to learn how to interact using brain waves. While the questionnaire was mainly targeting the use of BCI during surgery, the same concepts can be transfered to AR systems for medical education and training straightforward.

   We encountered some problems with using the BCI device in combination with the HMD. Sometimes the HMD put too much pressure on the sensor which led to problems in obtaining a signal from the BCI device. However BCI devices are interesting in particular for HMD-based AR as they could be integrated into a HMD. For two users the BCI device did not work immediately, but we had to change the activation thresholds as their EMG signals were constantly higher compared to the other users. Apart from this, both the BCI and the gaze-tracker worked for all participants.

   While this was only a pilot study to assess the possible use of BCI in medical AR the results are promising. We only used very simple hardware and a very short training period. Nevertheless the majority of participants was in favor of this new kind of UI and found this Superman-like mind-controlled X-ray vision very intuitive. Using more advanced devices and trained users, it would be possible to control additional parameters. While BCIs could also be beneficial for other AR applications, they are interesting in particular for medical AR, where the use of the hands for human-computer interaction is often not possible. For medical education

such an user interface could be beneficial in particular for training concepts such as showing co-registered images from different modalities (see section 2.3.2). Students could easily switch between displaying images from two different modalities while carrying out a procedure.

B. Authored and Co-Authored Publications

1. T. Blum, A. Rieger, N. Navab, H. Friess, M. E. Martignoni. A Review of Computer-Based Simulators for Ultrasound Training. In *Simulation in Healthcare, Volume 8, Issue 2, April 2013, pp. 98-108*

2. T. Blum, R. Stauder, E. Euler, N. Navab. Superman-like X-ray Vision: Towards Brain-Computer Interfaces for Medical Augmented Reality. In *The 11th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, USA, November 2012*

3. A. Bigdelou, R. Stauder, T. Benz, A. Okur, T. Blum, R. Ghotbi, N. Navab. HCI Design in the OR: A Gesturing Case-study. In *MICCAI 2012 Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI), Nice, France, October 2012*

4. N. Navab, T. Blum, A. Okur, T. Wendler. First Deployments of Augmented Reality in Operating Rooms. In *IEEE Computer, Volume 45, Issue 7, July 2012, pp. 48-55*

5. N. Padoy, T. Blum, A. Ahmadi, H. Feußner, M.O. Berger, N. Navab. Statistical Modeling and Recognition of Surgical Workflow. In *Medical Image Analysis (MIA), Volume 16, Issue 3, April 2012, pp. 632-641*

6. T. Blum, V. Kleeberger, C. Bichlmeier, N. Navab. mirracle: An Augmented Reality Magic Mirror System for Anatomy Education. In *IEEE Virtual Reality 2012 (VR), Orange County, USA, March 2012*

7. R.R. Shamir, L. Joskowicz, I. Tamir, M. Horn, T. Blum, N. Navab, J. Mehrkens, Y. Shoshan. A Method for Reduced Risk Trajectory Planning in Image-Guided Keyhole Neurosurgery. In *Conference of the Society for Medical Innovation and Technology (SMIT 2011), TelAviv, Israel, September 2011*

8. R.R. Shamir, M. Horn, T. Blum, J. Mehrkens, Y. Shoshan, L. Joskowicz, N. Navab. Trajectory Planning with Augmented Reality for Improved Risk Assesment in Image-Guided Keyhole Neurosurgery. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2011), Chicago, Illinois, USA, March 2011*

9. T. Blum, M. Wieczorek, A. Aichert, R. Tibrewal, N. Navab. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *The 9th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2010), Seoul, Korea, October 2010*

10. C. Bichlmeier, E. Euler, T. Blum, N. Navab. Evaluation of the Virtual Mirror as a Navigational Aid for Augmented Reality Driven Minimally Invasive Procedures. In *The 9th IEEE and ACM International Symposium on Mixed and Augmented Reality, Seoul (ISMAR 2010), Korea, October 2010*

11. T. Blum, H. Feußner, N. Navab. Modeling and Segmentation of Surgical Workflow from Laparoscopic Video. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2010), Beijing, China, September 2010*

12. T. Blum, N. Navab, H. Feußner. Methods for Automatic Statistical Modeling of Surgical Workflow. In *7th International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior 2010), Eindhoven, The Netherlands, August 2010*

13. L. Wang, J. Landes, S. Weidert, T. Blum, A. von der Heide, E. Euler, N. Navab. First Animal Cadaver Study for Interlocking of Intramedullary Nails under Camera Augmented Mobile C-arm: A Surgical Workflow Based Preclinical Evaluation. In *The 1st International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), Switzerland, June 23 2010*

14. T. Blum, S.M. Heining, O. Kutter, N. Navab. Advanced Training Methods using an Augmented Reality Ultrasound Simulator. In *8th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2009), Orlando, USA, October 2009, pp. 177-178*

15. T. Blum, M. Martignoni, H. Friess, N. Navab. An Ultrasound Simulator for Training using CT-based Simulation of Ultrasound Images. In *5th Russian-Bacarian Conference on Bio-Medical Engineering, Munich, Germany, July 2009*

16. N. Fritz, T. Meyer, T. Blum, H.U. Lemke, M. Ilzkovitz, Y. Nevatia, M. Nolden, I. Wegner, M. Weinlich, R. Breitkreutz, W. Wein, M. Lazerges, O. Angerer, N. Navab. CAMDASS: An Augmented Reality Medical Guidance System for Spaceflights. In *Proceedings of Computer Assisted Radiology and Surgery (CARS 2009), Berlin, Germany, June 2009*

17. T. Blum, N. Padoy, H. Feußner, N. Navab. Workflow Mining for Visualization and Analysis of Surgeries. In *International Journal of Computer Assisted Radiology and Surgery, Volume 3, Number 5, November 2008, pp. 379-386*

18. T. Blum, N. Padoy, H. Feußner, N. Navab. Modeling and Online Recognition of Surgical Phases using Hidden Markov Models. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2008), New York, USA, September 2008, pp. 627-635*

19. N. Padoy, T. Blum, H. Feußner, M.O. Berger, N. Navab. On-line Recognition of Surgical Activity for Monitoring in the Operating Room. In *Proceedings of the 20th Conference on Innovative Applications of Artificial Intelligence (IAAI 08), Chicago, USA, July 2008, pp. 1718-1724*

20. T. Blum, N. Padoy, H. Feußner, N. Navab. Workflow Mining for Visualization and Analysis of Surgeries. In *Proceedings of Computer Assisted Radiology and Surgery (CARS 2008), Barcelona, Spain, June 2008, pp. 134-135*

21. N. Padoy, T. Blum, I. Essa, H. Feußner, M.O. Berger, N. Navab. A Boosted Segmentation Method for Surgical Workflow Analysis. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007), Brisbane, Australia, October 2007, pp. 102-109*

22. N. Padoy, T. Blum, H. Feußner, N. Navab. Towards Automatic Surgical Workflow Analysis: Method and Application. In 6. *Jahrestagung der Deutschen Gesellschaft für Computer-und Roboter-Assistierte Chirurgie (CURAC 2007), Karlsruhe, Germany, October 2007*

23. T. Blum, T. Sielhorst, N. Navab. Advanced Augmented Reality Feedback for Teaching 3D Tool Manipulation. In *New Technology Frontiers in Minimally Invasive Therapies, 2007, pp. 223-236*

24. T. Blum, O. Häberle, M. Appel and H. Krcmar. Estimating the Financial Consequences of Using Augmented Reality in the Construction of Power Plants. In *Dritter Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR, Koblenz, Germany, September 2006*

25. O. Häberle, T. Blum and H. Krcmar. Evaluating an Innovative Technology in the Presence of Uncertainty. In *12th Americas Conference on Information Systems (AMCIS 2006), Acapulco, México, August 2006, pp. 541-549*

26. T. Sielhorst, T. Blum and N. Navab. Synchronizing 3D Movements for Quantitative Comparison and Simultaneous Visualization of Actions. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2005), pp. 38-47, Vienna, Austria, October 2005, pp. 38-47*

27. T. Sielhorst, T. Blum and N. Navab. Quantitative comparison of movements for medical training. In *4. Jahrestagung der Deutschen Gesellschaft für Computer-und Roboter-Assistierte Chirurgie (CURAC 2005), Berlin, Germany, September 2005*

C. Abstracts of Major Publications not Discussed in the Dissertation

Trajectory Planning with Augmented Reality for Improved Risk Assesment in Image-Guided Keyhole Neurosurgery

R.R. Shamir, M. Horn, T. Blum, J. Mehrkens, Y. Shoshan, L. Joskowicz and N. Navab

We present a new preoperative planning method for reducing the risk associated with insertion of straight tools in image-guided keyhole neurosurgery. The method quantifies the risks of multiple candidate trajectories and presents them on a physical model of a head using Augmented Reality (AR) to assist the neurosurgeon in selecting the safest path. The surgeon can then define and/or revise the trajectory in the physical space with AR visualization of risk structures (e.g. blood vessels and ventricles), tool placement uncertainty, and quantitative risk measurements. Then, the neurosurgeon can revise the selected path on the 2D MRI image slices to incorporate all relevant information. Finally, a simulation of the surgery can be performed on the physical head model for a more detailed examination of the possible risks. Our preliminary results on clinical data show that in complex situations the method can improve risk assessment.

Modeling and Segmentation of Surgical Workflow from Laparoscopic Video

T. Blum, H. Feußner and N. Navab

Modeling and analyzing surgeries based on signals that are obtained automatically from the operating room (OR) is a field of recent interest. It can be valuable for analyzing and understanding surgical workflow, for skills evaluation and developing context-aware ORs. In minimally invasive surgery laparoscopic video is easy to record but it is challenging to extract meaningful information from it. We propose a method that uses additional information about tool usage to perform a dimensionality reduction on image features. Using Canonical Correlation Analysis (CCA) a projection of a high-dimensional image feature space to a low dimensional space is obtained such that semantic information is extracted from the video. To model a surgery based on the signals in the reduced feature space two different statistical models are compared. The capability of segmenting a new surgery into phases only based on the video is evaluated. Dynamic Time Warping which strongly depends on the temporal order in combination with CCA shows the best results.

First Animal Cadaver Study for Interlocking of Intramedullary Nails under Camera Augmented Mobile C-arm A Surgical Workflow Based Preclinical Evaluation

L. Wang, J. Landes, S. Weidert, T. Blum, A. von der Heide, E. Euler and N. Navab

The Camera Augmented Mobile C-arm (CamC) system that augments a regular mobile C-arm by a video camera provides an overlay image of X-ray and video. This technology is expected to reduce radiation exposure during surgery without introducing major changes to the standard surgical workflow. Whereas many experiments were conducted to evaluate the technical characteristics of the CamC system, its clinical performance has not been investigated in detail. In this work, a workflow based method is proposed and applied to evaluate the clinical impact of the CamC system by comparing its performance with a conventional system, i.e. standard mobile C-arm. Interlocking of intramedullary nails on animal cadaver is chosen as a simulated clinical model for the evaluation study. Analyzing single workflow steps not only reveals individual strengths and weaknesses related to each step, but also allows surgeons and developers to be involved intuitively to evaluate and have an insight into the clinical impact of the system. The results from a total of 20 pair cases, i.e. 40 procedures, performed by 5 surgeons show that it takes significantly less radiation exposure whereas operation time for the whole interlocking procedure and quality of the drilling result are similar, using the CamC system compared to using the standard mobile C-arm. Moreover, the workflow based evaluation reveals in which surgical steps the CamC system has its main impact.

CAMDASS: An Augmented Reality Medical Guidance System for Spaceflights

N. Fritz, T. Meyer, T. Blum, H.U. Lemke, M. Ilzkovitz, Y. Nevatia, M. Nolden, I. Wegner, M. Weinlich, R. Breitkreutz, W. Wein, M. Lazerges, O. Angerer and N. Navab

In any long term space mission, medical care has to be considered because a crew health problem can significantly impact the mission. Availability of a physician onboard cannot be guaranteed at any time, so additional means to provide crew members with medical assistance have to be investigated. One way to achieve this objective is to use an Augmented Reality (AR) system for guiding non-expert crew members through medical procedures. AR is considered as one of the key future technologies for medical assistance, knowledge maintenance and training. Especially for space these activities are highly critical with respect to long duration missions, exploration and experimentation. One of the goals of the CAMDASS project (Computer Aided Medical Diagnostic and Surgery System) is to develop and implement such an AR system in order to guide astronauts in emergencies and scientific medical procedures. Specifically, this system shall offer feedback to the user (who may or may not be an expert) regarding the quality of the performed procedure in order to provide some training means, to optimise the scientific output and to react correctly in emergency cases. During the first phase of the project, the developments will focus on demonstrating feasibility and pertinence through an example medical application that will allow performing assisted ultrasound (US) examination.

D. Abbreviations

AAR   After Action Review
AR   Augmented Reality
AV   Augmented Virtuality
BCI   Brain Computer Interface
CAS   Computer-Assisted Surgery
CRL   Crown-Rump Length
CT   Computed Tomography
DOF   Degrees of Freedom
DTW   Dynamic Time Warping
DVR   Direct Volume Rendering
EM   Electromagnetic
EMG   Electromyographic
F+C   Focus and Context
FAST   Focused Abdominal Sonography for Trauma
FEM   Finite Element Method
FOV   Field of View
GPGPU   General Purpose GPU
GPU   Graphics Processing Unit
GUI   Graphical User Interface
HCI   Human-Computer Interaction
HMD   Head-Mounted Display
HMM   Hidden Markov Model
ICP   Iterative Closest Point
IVUS   Intravascular Ultrasound
LED   Light-Emitting Diode
LNCC   Local Normalized Cross-Correlation
MD   Medical Doctor
MIS   Minimally Invasive Surgery
MR   Magnetic Resonance
NT   Nuchal Translucency Thickness
OR   Operating Room
OST   Optical See-Through
TOF   Time-of-Flight
TTE   Transthoracic Echocardiography
TOE   Transesophageal Echocardiography
SSS   Successive State Splitting
SSVEP   Steady-State Visual Evoked Potential
US   Ultrasound
VR   Virtual Reality
VHD   Visible Human Dataset
VKH   Visible Korean Human
VPH   Virtual Physiological Human
VST   Video See-Through

E. Bibliography

[Abkai et al., 2007] C. Abkai, N. Becherer, J. Hesser, and R. Männer (2007). Real-time simulator for intravascular ultrasound (IVUS). In *SPIE Medical Imaging*.

[Abolmaesumi et al., 2004] P. Abolmaesumi, K. Hashtrudi-Zaad, D. Thompson, and A. Tahmasebi (2004). A haptic-based system for medical image examination. In *Engineering in Medicine and Biology Society (IEMBS)*, volume 1.

[Ahmadi et al., 2006] A. Ahmadi, T. Sielhorst, R. Stauder, M. Horn, H. Feußner, and N. Navab (2006). Recovery of surgical workflow without explicit models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 420--428.

[Ahmidi et al., 2010] N. Ahmidi, G. Hager, L. Ishii, G. Fichtinger, G. Gallia, and M. Ishii (2010). Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2010*, pages 295--302.

[Aiger and Cohen-Or, 1998] D. Aiger and D. Cohen-Or (1998). Real-time ultrasound imaging simulation. *Real-Time Imaging*, 4(4):274.

[Aiger and Cohen-Or, 2000] D. Aiger and D. Cohen-Or (2000). Mosaicing ultrasonic volumes for visual simulation. *IEEE Computer Graphics and Applications*, 20(2):53--61.

[Alankus et al., 2010] G. Alankus, A. Lazar, M. May, and C. Kelleher (2010). Towards customizable games for stroke rehabilitation. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 2113--2122. ACM.

[Alexandersson et al., 2003] M. Alexandersson, S. Cawley, and L. Pachter (2003). SLAM: Cross-species gene finding and alignment with a generalized pair hidden Markov model. *Genome research*, 13(3):496--502.

[Almansa et al., 2011] C. Almansa, M. Shahid, M. Heckman, S. Preissler, and M. Wallace (2011). Association between visual gaze patterns and adenoma detection rate during colonoscopy: a preliminary investigation. *The American journal of gastroenterology*, 106(6):1070--1074.

[Alterovitz et al., 2003a] R. Alterovitz, K. Goldberg, J. Pouliot, R. Taschereau, and I. Hsu (2003a). Needle insertion and radioactive seed implantation in human tissues: Simulation and sensitivity analysis. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 2.

[Alterovitz et al., 2003b] R. Alterovitz, J. Pouliot, R. Taschereau, I. Hsu, and K. Goldberg (2003b). Simulating needle insertion and radioactive seed implantation for prostate brachytherapy. In *Medicine Meets Virtual Reality (MMVR)*, pages 19--25. IOS Press; 1999.

[Andermann et al., 2007] P. Andermann, S. Schlögl, U. Mäder, M. Luster, M. Lassmann, and C. Reiners (2007).   Intra- and interobserver variability of thyroid volume measurements in healthy adults by 2D versus 3D ultrasound. *Nuklearmedizin*, 46(1):1--7.

[ap Cenydd et al., 2009] L. ap Cenydd, F. Vidal, N. John, D. Gould, E. Joekes, and P. Littler (2009). Cost Effective Ultrasound Imaging Training Mentor for use in Developing Countries. In *Medicine Meets Virtual Reality (MMVR)*, page 49. Ios Pr Inc.

[Arkhurst, 2005] W. Arkhurst (2005). *Ein interaktiver Atlas für die Sonographie und Anatomie des Säuglingsgehirns*. PhD thesis, University Hamburg.

[Arkhurst et al., 2001] W. Arkhurst, A. Pommert, E. Richter, H. Frederking, S. Kim, R. Schubert, and K. Höhne (2001). A virtual reality training system for pediatric sonography. In *International Congress Series*, volume 1230, pages 483--487. Elsevier.

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

[Baillot et al., 2000] Y. Baillot, J. Rolland, K. Lin, and D. Wright (2000). Automatic modeling of knee-joint motion for the virtual reality dynamic anatomy (VRDA) tool. *Presence: Teleoperators & Virtual Environments*, 9(3):223--235.

[Bajura et al., 1992] M. Bajura, H. Fuchs, and R. Ohbuchi (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In *ACM SIGGRAPH Computer Graphics*, pages 203--210. ACM.

[Basdogan et al., 2004] C. Basdogan, S. De, J. Kim, M. Muniyandi, H. Kim, and M. Srinivasan (2004). Haptics in minimally invasive surgical simulation and training. *Computer Graphics and Applications, IEEE*, 24(2):56--64.

[Bates, 2008] A. Bates (2008). Indecent and demoralising representations: Public anatomy museums in mid-victorian england. *Medical history*, 52(1):1.

[Beaulieu, 2007] Y. Beaulieu (2007). Bedside echocardiography in the assessment of the critically ill. *Critical Care Medicine*, 35(5):S235.

[Beckh et al., 2002] S. Beckh, P. Bölcskei, and K. Lessnau (2002). Real-Time Chest Ultrasonography. *Chest*, 122(5):1759.

[Berger et al., 2007] M. Berger, S. Störmann, and M. Fischer (2007). Eine studentische wiki-bibliothek für unterrichtsbegleitende materialien: Konzeption, implementierung und evaluation für das medizinische curriculum münchen (mecum) a student-driven wiki-library for educational materials: Concept, implementation, and evaluation for the medical curriculum munich (mecum)). *GMS Z Med Ausbild*, 24:4.

[Berlage, 1997] T. Berlage (1997). Augmented reality for diagnosis based on ultrasound images. In *Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*, page 262. Springer-Verlag.

[Berlage, 2008] T. Berlage (2008). Improving Human Skills in Echocardiography through Scene-Based Enabling Systems. Technical report, German National Research Center for Information Technology (GMD).

[Berlage et al., 1996] T. Berlage, T. Fox, G. Grunst, and K. Quast (1996). Supporting Ultrasound Diagnosis Using An Animated 3D Model of the Heart. In *Proceedings of IEEE Multimedia Systems*, pages 34--39. Citeseer.

[Betker et al., 2007] A. Betker, A. Desai, C. Nett, N. Kapadia, and T. Szturm (2007). Game-based exercises for dynamic short-sitting balance rehabilitation of people with chronic spinal cord and traumatic brain injuries. *Physical therapy*, 87(10):1389--1398.

[Bichlmeier, 2010] C. Bichlmeier (2010). *Immersive, Interactive and Contextual In-Situ Visualization for Medical Applications*. PhD thesis, Technische Universität München (TUM), Chair for Computer Aided Medical Procedures.

[Bichlmeier et al., 2007a] C. Bichlmeier, S. Heining, M. Rustaee, and N. Navab (2007a). Virtually extended surgical drilling device: Virtual mirror for navigated spine surgery. In *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention-Volume Part I*, pages 434--441. Springer-Verlag.

[Bichlmeier et al., 2008] C. Bichlmeier, B. Ockert, O. Kutter, M. Rustaee, S. M. Heining, and N. Navab (2008). The visible korean human phantom: Realistic test & development environments for medical augmented reality. In *International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery (AMI-ARCS 2008)*, New York, USA. MICCAI Society.

[Bichlmeier et al., 2007b] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab (2007b). Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 129--138.

[Birkfellner et al., 2002] W. Birkfellner, M. Figl, K. Huber, F. Watzinger, F. Wanschitz, J. Hummel, R. Hanel, W. Greimel, P. Homolka, R. Ewers, et al. (2002). A head-mounted operating binocular for augmented reality visualization in medicine-design and initial evaluation. *Medical Imaging, IEEE Transactions on*, 21(8):991--997.

[Birkfellner et al., 2008] W. Birkfellner, J. Hummel, E. Wilson, and K. Cleary (2008). Tracking devices. In T. Peters and K. Cleary, editors, *Image-Guided Interventions: Technology and Applications*, chapter 2, pages 23--44. Springer.

[Blohm et al., 1997] W. Blohm, I. Beldie, K. Schenke, K. Fazel, and S. Pastoor (1997). Stereoscopic image representation with synthetic depth of field. *Journal of the Society for Information Display*, 5:307.

[Blum, 2005] T. Blum (2005). Synchronisierung von 3d bewegungen für einen quantitativen vergleich und zeitgleiche visualisierung von abläufen. Softwareentwicklungsprojekt (SEP). Technische Universität München.

[Blum, 2007] T. Blum (2007). Surgical workflow analysis: Representation and application to monitoring. Master's thesis, Technische Universität München. Diploma Thesis.

[Blum et al., 2010a] T. Blum, H. Feussner, and N. Navab (2010a). Modeling and segmentation of surgical workflow from laparoscopic video. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Beijing, China.

[Blum et al., 2010b] T. Blum, N. Navab, and H. Feußner (2010b). Methods for automatic statistical modeling of surgical workflow. In *Proceedings of 7th International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior 2010)*, Eindhoven, The Netherlands.

[Blum et al., 2008] T. Blum, N. Padoy, H. Feussner, and N. Navab (2008). Modeling and online recognition of surgical phases using hidden markov models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 627--635, New York, USA.

[Bockholt et al., 2003] U. Bockholt, A. Bisler, M. Becker, W. Müller-Wittig, and G. Voss (2003). Augmented reality for enhancement of endoscopic interventions. In *Proceedings of the IEEE Virtual Reality 2003*, page 97. IEEE Computer Society.

[Bommersheim et al., 2005] S. Bommersheim, U. Tiede, E. Burmester, M. Riemer, and H. Handels (2005). Simulation von Ultraschallbildern für ein virtuelles Trainingssystem für endoskopische Longitudinal-Ultraschalluntersuchungen. In *Bildverarbeitung für die Medizin*, pages 450--454. Springer.

[Bose et al., 2009] R. Bose, R. Matyal, P. Panzica, S. Karthik, B. Subramaniam, J. Pawlowski, J. Mitchell, and F. Mahmood (2009). Transesophageal Echocardiography Simulator: A New Learning Tool. *Journal of Cardiothoracic and Vascular Anesthesia*, 23(4):544--548.

[Bose et al., 2010] R. Bose, R. Matyal, H. Warraich, J. Summers, B. Subramaniam, J. Mitchell, P. Panzica, S. Shahul, and F. Mahmood (2010). Utility of a Transesophageal Echocardiographic Simulator as a Teaching Tool. *J Cardiothorac Vasc Anesth*.

[Botella et al., 2005] C. Botella, M. Juan, R. Baños, M. Alcañiz, V. Guillen, and B. Rey (2005). Mixing realities? an application of augmented reality for the treatment of cockroach phobia. *Cyberpsychology & Behavior*, 8(2):162--171.

[Bouarfa et al., 2011a] L. Bouarfa, O. Akman, A. Schneider, P. Jonker, and J. Dankelman (2011a). In-vivo real-time tracking of surgical instruments in endoscopic video. *Minimally Invasive Therapy & Allied Technologies*.

[Bouarfa et al., 2011b] L. Bouarfa, P. Jonker, and J. Dankelman (2011b). Discovery of high-level tasks in the operating room. *Journal of biomedical informatics*, 44(3):455--462.

[Bouchard et al., 2006] S. Bouchard, S. Côté, J. St-Jacques, G. Robillard, and P. Renaud (2006). Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3d games. *Technology and Health Care*, 14(1):19--27.

[Boulos et al., 2006] M. Boulos, I. Maramba, and S. Wheeler (2006). Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. *BMC medical education*, 6(1):41.

[Bradley, 2006] P. Bradley (2006). The history of simulation in medical education and possible future directions. *Medical Education*, 40(3):254--262.

[Brauer et al., 2005] V. Brauer, P. Eder, K. Miehle, T. Wiesner, H. Hasenclever, and R. Paschke (2005). Interobserver variation for ultrasound determination of thyroid nodule volumes. *Thyroid*, 15(10):1169--1175.

[Brown et al., 2008] C. Brown, B. McNicholl, and R. Wright (2008). Ultrasound simulator for venous access. *Emergency Medicine Journal*, 25(2):122.

[BT, 2002] I. BT (2002). 500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT. 500-11, ITU Telecom. *Sector of ITU.*

[Buchmann et al., 2004] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn (2004). FingARtips: gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 212--221. ACM.

[Bude and Adler, 1995] R. Bude and R. Adler (1995). An easily made, low-cost, tissue-like ultrasound phantom material. *Journal of Clinical Ultrasound*, 23(4):271--273.

[Bürger et al., 2008] B. Bürger, C. Abkai, and J. Hesser (2008). Simulation of dynamic ultrasound based on CT models for medical education. In *Medicine Meets Virtual Reality (MMVR)*, page 56. Ios Pr Inc.

[Burgert et al., 2006] O. Burgert, T. Neumuth, F. Lempp, R. Mudunuri, J. Meixensberger, G. Strauß, A. Dietz, and P. J. amd H. U. Lemke (2006). Linking top-level ontologies and surgical workflows. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, pages 437--438.

[Cardiac, 2004] Q. Cardiac (2004). American Society of Echocardiography recommendations for use of echocardiography in clinical trials. *Journal of the American Society of Echocardiography*, 17(10):1086--1119.

[Casebeer et al., 2002] L. Casebeer, N. Bennett, R. Kristofco, A. Carillo, and R. Centor (2002). Physician internet medical information seeking and on-line continuing education use patterns. *Journal of Continuing Education in the Health Professions*, 22(1):33--42.

[Chan et al., 2010] W. Chan, D. Ni, W. Pang, J. Qin, Y. Chui, S. Yu, and P. Heng (2010). Learning Ultrasound-Guided Needle Insertion Skills through an Edutainment Game. *Transactions on Edutainment IV*, pages 200--214.

[Cheitlin et al., 2003] M. Cheitlin, W. Armstrong, G. Aurigemma, G. Beller, F. Bierman, J. Davis, P. Douglas, D. Faxon, L. Gillam, T. Kimball, et al. (2003). ACC/AHA/ASE 2003 guideline update for the clinical application of echocardiography: summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/ASE Committee to Update the 1997 Guidelines for the Clinical Application of Echocardiography). *Journal of the American College of Cardiology*, 42(5):954.

[Cheng and Takatsuka, 2009] K. Cheng and M. Takatsuka (2009). Initial evaluation of a bare-hand interaction technique for large displays using a webcam. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*, pages 291--296. ACM.

[Choules, 2007] A. Choules (2007). The use of elearning in medical education: a review of the current situation. *Postgraduate Medical Journal*, 83(978):212.

[Coletta et al., 2006] C. Coletta, E. De Marchis, M. Lenoli, S. Rosato, M. Renzi, A. Sestili, P. Romano, T. Infusino, R. Ricci, and V. Ceci (2006). Reliability of cardiac dimensions and valvular regurgitation assessment by sonographers using hand-carried ultrasound devices. *European Journal of Echocardiography*, 7(4):275.

[Cooke et al., 2010] M. Cooke, D. Irby, B. O'Brien, and L. Shulman (2010). *Educating physicians: a call for reform of medical school and residency*, volume 16. Jossey-Bass.

[Cooper and Taqueti, 2004] J. Cooper and V. Taqueti (2004). A brief history of the development of mannequin simulators for clinical education and training. *Quality & safety in health care*, 13(Suppl 1):i11.

[Corradini, 2001] A. Corradini (2001). Dynamic Time Warping for Off-Line Recognition of a Small Gesture Vocabulary. *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS)*, pages 82--89.

[Costantino et al., 2005] T. Costantino, E. Bruno, N. Handly, and A. Dean (2005). Accuracy of emergency medicine ultrasound in the evaluation of abdominal aortic aneurysm. *Journal of Emergency Medicine*, 29(4):455--460.

[Costantino et al., 2003] T. Costantino, W. Satz, S. Stahmer, and A. Dean (2003). Predictors of success in emergency medicine ultrasound education. *Academic Emergency Medicine*, 10(2):180--183.

[Counselman et al., 2003] F. Counselman, A. Sanders, C. Slovis, D. Danzl, L. Binder, and D. Perina (2003). The status of bedside ultrasonography training in emergency medicine residency programs. *Academic Emergency Medicine*, 10(1):37--42.

[Cox et al., 2006] M. Cox, D. Irby, M. Cooke, D. Irby, W. Sullivan, and K. Ludmerer (2006). American medical education 100 years after the flexner report. *New England Journal of Medicine*, 355(13):1339--1344.

[Crosbie et al., 2008] J. Crosbie, S. Lennon, M. McGoldrick, M. McNeill, J. Burke, and S. McDonough (2008). Virtual reality in the rehabilitation of the upper limb after hemiplegic stroke: a randomised pilot study. *Proc. 7th ICDVRAT with ArtAbilitation, Maia, Portugal*, pages 229--235.

[Darolti et al., 2007] C. Darolti, R. Bouchard, S. Farke, A. Condurache, and U. Hofmann (2007). Measurement of surgical motion with a marker free computer vision based system. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, pages 213--215.

[Datta et al., 2001] V. Datta, S. Mackay, M. Mandalia, and A. Darzi (2001). The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model1. *Journal of the American College of Surgeons*, 193(5):479--485.

[d'Aulignac et al., 2005] D. d'Aulignac, C. Laugier, J. Troccaz, and S. Vieira (2005). Towards a realistic echographic simulator. *Medical Image Analysis*, 10(1):71--81.

[Davis et al., 2002] L. Davis, F. Hamza-Lup, J. Daly, Y. Ha, S. Frolich, C. Meyer, G. Martin, J. Norfleet, K. Lin, C. Imielinska, et al. (2002). Application of augmented reality to visualizing

anatomical airways. In *Proceedings of SPIE, the International Society for Optical Engineer-ingProceedings of SPIE, the International Society for Optical Engineering*, volume 4711, pages 400--405. Citeseer.

[de Lange et al., 2010] P. de Lange, Y. Takata, H. Kim, H. Liao, E. Kobayashi, M. Ono, S. Kyo, S. Takamoto, S. Ishii, T. Asano, et al. (2010). Real-time epicardial excitation time map overlay. *Medical Imaging and Augmented Reality*, pages 521--530.

[Delp et al., 2007] S. Delp, F. Anderson, A. Arnold, P. Loan, A. Habib, C. John, E. Guendelman, and D. Thelen (2007). Opensim: open-source software to create and analyze dynamic simulations of movement. *Biomedical Engineering, IEEE Transactions on*, 54(11):1940--1950.

[Demers, 2004] J. Demers (2004). Depth of field: A survey of techniques. *GPU Gems*, 1:375--390.

[Deutsch et al., 2008] J. Deutsch, M. Borbely, J. Filler, K. Huhn, and P. Guarrera-Bowlby (2008). Use of a low-cost, commercially available gaming console (wii) for rehabilitation of an adolescent with cerebral palsy. *Physical Therapy*, 88(10):1196--1207.

[Duchowski, 2007] A. Duchowski (2007). *Eye tracking methodology: Theory and practice*. Springer-Verlag New York Inc.

[Edwards et al., 2000] P. Edwards, A. King, C. Maurer Jr, D. De Cunha, D. Hawkes, D. Hill, R. Gaston, M. Fenlon, A. Jusczyzck, A. Strong, et al. (2000). Design and evaluation of a system for microscope-assisted guided interventions (magi). *Medical Imaging, IEEE Transactions on*, 19(11):1082--1093.

[Eggers et al., 2005] G. Eggers, G. Sudra, S. Ghanai, T. Salb, R. Dillmann, R. Marmulla, and S. Hassfeld (2005). Augmented-reality-guided biopsy of a tumor near the skull base: the surgeon's experience. In *Proceedings of SPIE*, volume 5744, page 338.

[Ehara and Saito, 2006] J. Ehara and H. Saito (2006). Texture overlay for virtual clothing based on PCA of silhouettes. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 139--142. IEEE Computer Society.

[Ehler and Carney, 2001] D. Ehler and D. Carney (2001). Guidelines for cardiac sonographer education: recommendations of the American Society of Echocardiography Sonographer Training and Education Committee. *Journal of the American Society of Echocardiography*, 14(1):77--84.

[Ehricke, 1998] H. Ehricke (1998). SONOSim3D: a multimedia system for sonography simulation and education with an extensible case database. *European Journal of Ultrasound*, 7(3):225--230.

[Eisert et al., 2008] P. Eisert, P. Fechteler, and J. Rurainsky (2008). 3-D Tracking of shoes for Virtual Mirror applications. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1--6. IEEE.

[Ellson et al., 2004] J. Ellson, E. Gansner, E. Koutsofios, S. North, and G. Woodhull (2004). Graphviz and dynagraph -- static and dynamic graph drawing tools. In M. Junger and P. Mutzel, editors, *Graph Drawing Software*, pages 127--148. Springer-Verlag.

[Elmezain et al., 2008] M. Elmezain, A. Al-Hamadi, and B. Michaelis (2008). Real-time capable system for hand gesture recognition using hidden markov models in stereo color image sequences. *Journal of WSCG*, 16(1):65--72.

[Faller et al., 2010] J. Faller, G. Müller-Putz, D. Schmalstieg, and G. Pfurtscheller (2010). An application framework for controlling an avatar in a desktop-based virtual environment via a software ssvep brain-computer interface. *Presence: Teleoperators and Virtual Environments*, 19(1):25--34.

[Feinstein and Cannon, 2002] A. Feinstein and H. Cannon (2002). Constructs of simulation evaluation. *Simulation & Gaming*, 33(4):425.

[Feuerstein et al., 2008] M. Feuerstein, T. Mussack, S. Heining, and N. Navab (2008). Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *Medical Imaging, IEEE Transactions on*, 27(3):355--369.

[Fiala, 2007] M. Fiala (2007). Magic Mirror System with Hand-held and Wearable Augmentations. In *Virtual Reality Conference, 2007. VR'07. IEEE*, pages 251--254. IEEE.

[Finger and Platino, 1961] B. Finger and A. Platino (1961). Superman's thoughest day, action comics no 282.

[Fischer et al., 2005] G. Fischer, A. Deguet, D. Schlattman, R. Taylor, L. Fayad, S. Zinreich, and G. Fichtinger (2005). Mri image overlay: applications to arthrography needle insertion. *Studies in health technology and informatics*, 119:150.

[Fischer et al., 2006] J. Fischer, D. Bartz, and W. Straßer (2006). Enhanced visual realism by incorporating camera image effects. In *IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006*, pages 205--208.

[Fitzgerald et al., 2010] D. Fitzgerald, N. Trakarnratanakul, B. Smyth, and B. Caulfield (2010). Effects of a wobble board-based therapeutic exergaming system for balance training on dynamic postural stability and intrinsic motivation levels. *journal of orthopaedic & sports physical therapy*, 40(1):11--19.

[Flexner, 1910] A. Flexner (1910). *Medical education in the United States and Canada: a report to the Carnegie Foundation for the Advancement of Teaching*. Carnegie Foundation for the Advancement of Teaching.

[Flynn et al., 2007] S. Flynn, P. Palma, and A. Bender (2007). Feasibility of using the sony playstation 2 gaming platform for an individual poststroke: a case report. *Journal of neurologic physical therapy*, 31(4):180.

[Forest et al., 2007] C. Forest, O. Comas, C. Vaysiere, L. Soler, and J. Marescaux (2007). Ultrasound and needle insertion simulators built on real patient-based data. In *Medicine Meets Virtual Reality (MMVR)*, page 136. Ios Pr Inc.

[Gaba and DeAnda, 1988] D. Gaba and A. DeAnda (1988). A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology*, 69(3):387.

[Gaggioli et al., 2005] A. Gaggioli, F. Morganti, A. Meneghini, M. Alcaniz, J. Lozano, J. Montesa, J. Sáez, R. Walker, I. Lorusso, and G. Riva (2005). The virtual reality mirror: mental practice with augmented reality for post-stroke rehabilitation. *Annual Review of CyberTherapy and Telemedicine*, 4:199--207.

[Gagvani and Silver, 2001] N. Gagvani and D. Silver (2001). Animating volumetric models. *Graphical Models*, 63(6):443--458.

[Gales and Young, 2008] M. Gales and S. Young (2008). The application of hidden markov models in speech recognition. *Foundations and Trends in Signal Processing*, 1(3):195--304.

[Gamberini et al., 2007] L. Gamberini, L. Breda, and A. Grassi (2007). Videodope: applying persuasive technology to improve awareness of drugs abuse effects. *Virtual Reality*, pages 633--641.

[Gao et al., 2009] H. Gao, H. Choi, P. Claus, S. Boonen, S. Jaecques, G. Van Lenthe, G. Van Der Perre, W. Lauriks, and J. D'hooge (2009). A fast convolution-based methodology to simulate 2-dd/3-d cardiac ultrasound images. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 56(2):404--409.

[Garcia-Palacios et al., 2002] A. Garcia-Palacios, H. Hoffman, A. Carlin, T. Furness, C. Botella, et al. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy*, 40(9):983--993.

[Glossop et al., 2003] N. Glossop, C. Wedlake, J. Moore, T. Peters, and Z. Wang (2003). Laser projection augmented reality system for computer assisted surgery. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 239--246.

[Goksel and Salcudean, 2009] O. Goksel and S. Salcudean (2009). B-mode ultrasound image simulation in deformable 3-D medium. *IEEE transactions on medical imaging*, 28(11):1657.

[Goksel and Salcudean, 2010] O. Goksel and S. Salcudean (2010). Haptic Simulator for Prostate Brachytherapy with Simulated Ultrasound. In *International Symposium on Biomedical Simulation*, pages 150--159. Springer.

[Gold and Rangarajan, 1996] S. Gold and A. Rangarajan (1996). A graduated assignment algorithm for graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(4):377--388.

[Greenberg et al., 1999] L. Greenberg, D. Ochsenschlager, R. O'Donnell, J. Mastruserio, and G. Cohen (1999). Communicating bad news: a pediatric department's evaluation of a simulated intervention. *Pediatrics*, 103(6):1210.

[Guirlinger, 2007] S. Guirlinger (2007). *Simulation to Augment Standardized Patients in Obetetric Ultrasound Training*. PhD thesis, Old Dominion University.

[Hacker et al., 2002] S. Hacker, U. Tiede, E. Burmester, T. Leineweber, and K. Heinz (2002). Ein virtuelles Trainingssystem für endoskopische Longitudinal-Ultraschalluntersuchungen. In *Bildverarbeitung für die Medizin*, page 149. Springer.

[Hajnal et al., 2001] J. Hajnal, D. Hawkes, and D. Hill (2001). *Medical image registration*. CRC.

[Harness and Wisher, 2001] J. Harness and D. Wisher (2001). *Ultrasound in surgical practice: basic principles and clinical applications.* Wiley-Liss.

[Heer et al., 2004] I. Heer, K. Middendorf, S. Muller-Egloff, M. Dugas, and A. Strauss (2004). Ultrasound training: the virtual patient. *Ultrasound in Obstetrics and Gynecology*, 24(4):440--444.

[Heer et al., 2001] I. Heer, A. Strauss, S. Müller-Egloff, and U. Hasbargen (2001). Telemedicine in ultrasound: new solutions. *Ultrasound in Medicine & Biology*, 27(9):1239--1243.

[Hergum et al., 2009] T. Hergum, S. Langeland, E. Remme, and H. Torp (2009). Fast ultrasound imaging simulation in k-space. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 56(6):1159--1167.

[Herman et al., 2000] I. Herman, G. Melançon, and M. Marshall (2000). Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1):24--43.

[Hertzberg et al., 2000] B. Hertzberg, M. Kliewer, J. Bowie, B. Carroll, D. DeLong, L. Gray, and R. Nelson (2000). Physician training requirements in sonography: how many cases are needed for competence? *American Journal of Roentgenology*, 174(5):1221.

[Hilsmann and Eisert, 2009] A. Hilsmann and P. Eisert (2009). Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications. In *Computer Vision/Computer Graphics Collaboration Techniques: 4th International Conference, MIRAGE 2009, Rocquencourt, France, May 4-6, 2009, Proceedings*, page 94. Springer-Verlag New York Inc.

[Hirji, 2006] S. Hirji (2006). *Real-time and interactive virtual Doppler ultrasound.* PhD thesis, The University of Western Ontario.

[Hoffman et al., 2003] H. Hoffman, A. Garcia-Palacios, A. Carlin, T. Furness Iii, and C. Botella-Arbona (2003). Interfaces that heal: Coupling real and virtual objects to treat spider phobia. *International Journal of Human-Computer Interaction*, 16(2):283--300.

[Holden et al., 1999] M. Holden, E. Todorov, J. Callahan, and E. Bizzi (1999). Virtual environment training improves motor performance in two patients with stroke: case report. *Journal of Neurologic Physical Therapy*, 23(2):57.

[Hostettler et al., 2005] A. Hostettler, C. Forest, A. Forgione, L. Soler, and J. Marescaux (2005). Real-time ultrasonography simulator based on 3D CT-scan images. In *Medicine Meets Virtual Reality (MMVR)*, volume 111, pages 191--193. IOS Press.

[Hsu and Payandeh, 2006] J. Hsu and S. Payandeh (2006). Toward tool gesture and motion recognition on a novel minimally invasive surgery robotic system. In *Proceedings 2006 IEEE International Conference on Robotics and Automation (ICRA)*, pages 631--636.

[Huang et al., 2007] G. Huang, R. Reynolds, and C. Candler (2007). Virtual patient simulation at us and canadian medical schools. *Academic Medicine*, 82(5):446.

[Hwang et al., 2009] Y. Hwang, S. Lampotang, N. Gravenstein, I. Luria, and B. Lok (2009). Integrating conversational virtual humans and mannequin patient simulators to present mixed reality clinical training experiences. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 197--198. IEEE Computer Society.

[Imani et al., 2002] F. Imani, K. Reinig, V. Spitzer, Y. Chen, and C. Lee (2002). Simulation of Medical Ultrasound Images using Visible Human Dataset. In *Visible Human Conference*.

[Irby et al., 2010] D. Irby, M. Cooke, and B. O'Brien (2010). Calls for reform of medical education by the carnegie foundation for the advancement of teaching: 1910 and 2010. *Academic Medicine*, 85(2):220.

[Issenberg et al., 1999] S. Issenberg, W. McGaghie, I. Hart, J. Mayer, J. Felner, E. Petrusa, R. Waugh, D. Brown, R. Safford, I. Gessner, et al. (1999). Simulation technology for health care professional skills training and assessment. *JAMA: the journal of the American Medical Association*, 282(9):861.

[Issenberg et al., 2005] S. Issenberg, W. Mcgaghie, E. Petrusa, D. Gordon, and R. Scalese (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher*, 27(1):10--28.

[Jack et al., 2001] D. Jack, R. Boian, A. Merians, M. Tremaine, G. Burdea, S. Adamovich, M. Recce, and H. Poizner (2001). Virtual reality-enhanced stroke rehabilitation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 9(3):308--318.

[Jadhav et al., 2006] C. Jadhav, P. Nair, and V. Krovi (2006). Individualized interactive home-based haptic telerehabilitation. *Multimedia, IEEE*, 13(3):32--39.

[Jang et al., 2005] S. Jang, S. You, M. Hallett, Y. Cho, C. Park, S. Cho, H. Lee, and T. Kim (2005). Cortical reorganization and associated functional motor recovery after virtual reality in patients with chronic stroke: an experimenter-blind preliminary study. *Archives of physical medicine and rehabilitation*, 86(11):2218--2223.

[Jannin and Morandi, 2007] P. Jannin and X. Morandi (2007). Surgical models for computer-assisted neurosurgery. *NeuroImage*.

[Jannin et al., 2001] P. Jannin, M. Raimbault, X. Morandi, and B. Gibaud (2001). Modeling surgical procedures for multimodal image-guided neurosurgery. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MIC-CAI)*, pages 565--572. Springer-Verlag London, UK.

[Jayender et al., 2010] J. Jayender, R. Estépar, K. Obstein, V. Patil, C. Thompson, and K. Vosburgh (2010). Hidden markov model for quantifying clinician expertise in flexible instrument manipulation. *Medical Imaging and Augmented Reality*, pages 363--372.

[Jensen, 1996] J. Jensen (1996). Field: A program for simulating ultrasound systems. In *10th Nordicbaltic Conference on Biomedical Imaging*. Citeseer.

[Juan et al., 2008] C. Juan, F. Beatrice, and J. Cano (2008). An augmented reality system for learning the interior of the human body. In *Eighth IEEE International Conference on Advanced Learning Technologies*, pages 186--188. IEEE.

[Juan et al., 2006] M. Juan, D. Joele, R. Baños, C. Botella, M. Alcañiz, and C. van der Mast (2006). A markerless augmented reality system for the treatment of phobia to small animals. In *Presence Conference, Cleveland, USA*.

[Kahol et al., 2009] K. Kahol, M. Vankipuram, and M. Smith (2009). Cognitive simulators for medical education and training. *Journal of Biomedical Informatics*, 42(4):593--604.

[Kamin et al., 2003] C. Kamin, P. O'Sullivan, R. Deterding, and M. Younger (2003). A comparison of critical thinking in groups of third-year medical students in text, video, and virtual pbl case modalities. *Academic Medicine*, 78(2):204.

[Karamalis et al., 2010] A. Karamalis, W. Wein, and N. Navab (2010). Fast ultrasound image simulation using the westervelt equation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 243--250. Springer.

[Kassebaum and Eaglen, 1999] D. Kassebaum and R. Eaglen (1999). Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Academic Medicine*, 74(7):842.

[Katić et al., 2010] D. Katić, G. Sudra, S. Speidel, G. Castrillon-Oberndorfer, G. Eggers, and R. Dillmann (2010). Knowledge-based situation interpretation for context-aware augmented reality in dental implant surgery. *Medical Imaging and Augmented Reality*, pages 531--540.

[Kelly et al., 2001] S. Kelly, K. Harris, E. Berry, J. Hutton, P. Roderick, J. Cullingworth, L. Gathercole, and M. Smith (2001). A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma. *Gut*, 49(4):534.

[Kempny and Piórkowski, 2010] A. Kempny and A. Piórkowski (2010). CT2TEE--a novel, internet-based simulator of transoesophageal echocardiography in congenital heart disease. *Kardiologia Polska*, 68:374--379.

[Kendall et al., 2007] J. Kendall, S. Hoffenberg, and R. Smith (2007). History of emergency and critical care ultrasound: The evolution of a new imaging paradigm. *Critical Care Medicine*, 35(5):S126.

[Kersten-Oertel et al., 2010] M. Kersten-Oertel, P. Jannin, and D. Collins (2010). Dvv: towards a taxonomy for mixed reality visualization in image guided surgery. *Medical Imaging and Augmented Reality*, pages 334--343.

[Khoshniat et al., 2005] M. Khoshniat, M. Thorne, T. Poepping, S. Hirji, D. Holdsworth, and D. Steinman (2005). Real-time numerical simulation of Doppler ultrasound in the presence of nonaxial flow. *Ultrasound in Medicine & Biology*, 31(4):519--528.

[Kizakevich et al., 2006] P. Kizakevich, R. Furberg, R. Hubal, and G. Frank (2006). Virtual reality simulation for multicasualty triage training. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2006. NTSA.

[Klein and Murray, 2008] G. Klein and D. Murray (2008). Compositing for small cameras. In *Proc. Seventh IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'08)*, Cambridge.

[Kliewer et al., 1999] M. Kliewer, D. Sheafor, E. Paulson, R. Helsper, B. Hertzberg, and R. Nelson (1999). Percutaneous liver biopsy: a cost-benefit analysis comparing sonographic and CT guidance. *American Journal of Roentgenology*, 173(5):1199.

[Knudson and Sisley, 2000] M. Knudson and A. Sisley (2000). Training residents using simulation technology: experience with ultrasound for trauma. *The Journal of Trauma*, 48(4):659.

[Köhn et al., 2004] S. Köhn, R. Van Lengen, G. Reis, M. Bertram, H. Hagen, G. Kaiserslautern, S. Koehn, and R. van Lengen (2004). VES: Virtual Echocardiography System. In *Visualization, Imaging, And Image Processing: Fourth IASTED International Conference Proceedings*. Citeseer.

[Kooi and Toet, 2004] F. Kooi and A. Toet (2004). Visual comfort of binocular and 3D displays. *Displays*, 25(2-3):99--108.

[Kozak et al., 2006] L. Kozak, C. DeFrances, and M. Hall (2006). National hospital discharge survey: 2004 annual summary with detailed diagnosis and procedure data. *Vital Health Stat 13*, 162:1--209.

[Krüger et al., 2006] J. Krüger, J. Schneider, and R. Westermann (2006). Clearview: An interactive context preserving hotspot visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, pages 941--948.

[Krupinski et al., 2006] E. Krupinski, A. Tillack, L. Richter, J. Henderson, A. Bhattacharyya, K. Scott, A. Graham, M. Descour, J. Davis, and R. Weinstein (2006). Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human pathology*, 37(12):1543--1556.

[Kundel et al., 2007] H. Kundel, C. Nodine, E. Conant, and S. Weinstein (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study1. *Radiology*, 242(2):396--402.

[Kutter, 2010] O. Kutter (2010). *Visual Computing for Computer Assisted Interventions*. PhD thesis, Technische Universität München (TUM), Chair for Computer Aided Medical Procedures.

[Kutter et al., 2008] O. Kutter, A. Aichert, C. Bichlmeier, J. Traub, S. M. Heining, B. Ockert, E. Euler, and N. Navab (2008). Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In *International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery (AMI-ARCS 2008)*, New York, USA. MICCAI Society.

[Kutter et al., 2009] O. Kutter, R. Shams, and N. Navab (2009). Visualization and GPU-accelerated simulation of medical ultrasound from CT images. *Computer methods and programs in biomedicine*, 94(3):250--266.

[Lalys et al., 2012] F. Lalys, L. Riffaud, D. Bouget, and P. Jannin (2012). A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. *Biomedical Engineering, IEEE Transactions on*, 59(4).

[Lambooij et al., 2009] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53:030201.

[Lane et al., 2001] J. Lane, S. Slavin, and A. Ziv (2001). Simulation in medical education: A review. *Simulation & Gaming*, 32(3):297.

[Lange et al., 2010] B. Lange, S. Flynn, C. Chang, A. Ahmed, Y. Geng, K. Utsav, M. Xu, D. Seok, S. Cheng, and A. Rizzo (2010). Development of an interactive rehabilitation game using the nintendo® wii fit balance board for people with neurological injury. *Proceedings of ICDVRAT, Chile, August.*

[Lapeer et al., 2004] R. Lapeer, M. Chen, and J. Villagrana (2004). An augmented reality based simulation of obstetric forceps delivery. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 274--275. IEEE Computer Society.

[Law et al., 2004] B. Law, M. Atkins, A. Kirkpatrick, and A. Lomax (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 41--48. ACM.

[Lee et al., 1999] D. Lee, J. Cottrell, R. Sanders, C. Meyers, E. Wulfsberg, and C. Sun (1999). OEIS complex (omphalocele-exstrophy-imperforate anus-spinal defects) in monozygotic twins. *American Journal of Medical Genetics Part A*, 84(1):29--33.

[Leong et al., 2006] J. Leong, M. Nicolaou, L. Atallah, G. Mylonas, A. Darzi, and G.-Z. Yang (2006). Hmm assessment of quality of movement trajectory in laparoscopic surgery. In *Medical Image Computing and Computer Aided Intervention (MICCAI)*, pages 752--759.

[Leong et al., 2003] S. Leong, C. Baldwin, and A. Adelman (2003). Integrating web-based computer cases into a required clerkship: development and evaluation. *Academic Medicine*, 78(3):295.

[Lepola, 2005] T. Lepola (2005). Inferring relevance from eye movements using generic neural microcircuits. In *Workshop at NIPS 2005, in Whistler, BC, Canada, on December 10, 2005.*, page 31.

[Leung, 2002] W. Leung (2002). Competency based medical training: review. *BMJ: British Medical Journal*, 325(7366):693.

[Levi et al., 1995] S. Levi, J. Schaap, P. De Hava, R. Coulon, and P. Defoort (1995). End-result of routine ultrasound screening for congenital anomalies: The Belgian Multicentric Study 1984-92. *Ultrasound in Obstetrics and Gynecology*, 5(6):366--371.

[Levin et al., 2007] A. Levin, R. Fergus, F. Durand, and W. Freeman (2007). Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3):70.

[Liao et al., 2004] H. Liao, N. Hata, S. Nakajima, M. Iwahara, I. Sakuma, and T. Dohi (2004). Surgical navigation by autostereoscopic image overlay of integral videography. *Information Technology in Biomedicine, IEEE Transactions on*, 8(2):114--121.

[Lieberman, 2001] D. Lieberman (2001). Management of chronic pediatric diseases with interactive health games: Theory and research findings. *The Journal of Ambulatory Care Management*, 24(1):26.

[Lin et al., 2006] H. Lin, I. Shafran, D. Yuh, and G. Hager (2006). Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery*, 11(5):220--230.

[Litchfield et al., 2008] D. Litchfield, L. Ball, T. Donovan, D. Manning, and T. Crawford (2008). Learning from others: effects of viewing another person's eye movements while searching for chest nodules. In *Proceedings of SPIE*, volume 6917, page 691715.

[Liu et al., 2010] Y. Liu, N. Glass, and R. Power (2010). New Teaching Model for Practicing Ultrasound-Guided Regional Anesthesia Techniques: No Perishable Food Products! *Anesthesia & Analgesia*, 110(4):1233.

[Livingston et al., 2009] M. Livingston, Z. Ai, and J. Decker (2009). A user study towards understanding stereo perception in head-worn augmented reality displays. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 53--56. IEEE Computer Society.

[Loop and Zhang, 1999] C. Loop and Z. Zhang (1999). Computing rectifying homographies for stereo vision. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:1125.

[Lorensen et al., 1993] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli, and L. Gleason (1993). Enhancing reality in the operating room. In *Visualization, 1993. Visualization'93, Proceedings., IEEE Conference on*, pages 410--415. IEEE.

[Ludmerer, 2003] K. Ludmerer (2003). The internal challenges to medical education. *Transactions of the American Clinical and Climatological Association*, 114:241.

[Luo et al., 2005] X. Luo, T. Kline, H. Fischer, K. Stubblefield, R. Kenyon, and D. Kamper (2005). Integration of augmented reality and assistive devices for post-stroke hand opening rehabilitation. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 6855--6858. IEEE.

[Lys et al., 1989] F. Lys, P. De Wals, I. Borlee-Grimee, A. Billiet, M. Vincotte-Mols, and S. Levi (1989). Evaluation of routine ultrasound examination for the prenatal diagnosis of malformation. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 30(2):101--109.

[Mackel et al., 2007] T. Mackel, J. Rosen, and C. Pugh (2007). Markov model assessment of subjects' clinical skill using the e-pelvis physical simulator. *Biomedical Engineering, IEEE Transactions on*, 54(12):2133--2141.

[Magee and Kessel, 2005] D. Magee and D. Kessel (2005). A Computer Based Simulator for Ultrasound Guided Needle Insertion Procedures. In *International Conference on Visual Information Engineering*, pages 301--308.

[Magee et al., 2007] D. Magee, Y. Zhu, R. Ratnalingam, P. Gardner, and D. Kessel (2007). An augmented reality simulator for ultrasound guided needle placement training. *Medical and Biological Engineering and Computing*, 45(10):957--967.

[Mahmood et al., 2008] F. Mahmood, A. Christie, and R. Matyal (2008). Transesophageal echocardiography and noncardiac surgery. *Seminars in Cardiothoracic and Vascular Anesthesia*, 12(4):265.

[Markov-Vetter et al., 2009] D. Markov-Vetter, J. Muehl, D. Schmalstieg, E. Sorantin, and M. Riccabona (2009). 3D Augmented Reality Simulator for Neonatale Cranial Sonography. In *Computer Assisted Radiology and Surgery (CARS)*.

[Marquardt et al., 2007] R. Marquardt, A. Scharf, and B. Stoiber (2007). Sonotrainer-Kurse jetzt auch zur Mammasonographie. *Frauenarzt*, 48(2):151--152.

[Mather and Smith, 2000] G. Mather and D. Smith (2000). Depth cue integration: stereopsis and image blur. *Vision Research*, 40(25):3501--3506.

[Maul et al., 2004] H. Maul, A. Scharf, P. Baier, M. Wüstemann, H. Gunter, G. Gebauer, and C. Sohn (2004). Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound in Obstetrics and Gynecology*, 24(5):581--585.

[Maul et al., 2006] H. Maul, A. Scharf, and C. Sohn (2006). Was kann der Sonotrainer-Ultraschallsimulator? *Der Gynäkologe*, 39(11):870--877.

[McFarland et al., 2008] D. McFarland, D. Krusienski, W. Sarnacki, and J. Wolpaw (2008). Emulation of computer mouse control with a noninvasive brain--computer interface. *Journal of neural engineering*, 5:101.

[Megali et al., 2006] G. Megali, S. Sinigaglia, O. Tonet, F. Cavallo, and P. Dario (2006). Understanding expertise in surgical gesture by means of Hidden Markov Models. *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 625--630.

[Memel et al., 1996] D. Memel, G. Dodd 3rd, and C. Esola (1996). Efficacy of sonography as a guidance technique for biopsy of abdominal, pelvic, and retroperitoneal lymph nodes. *American Journal of Roentgenology*, 167(4):957.

[Mercier et al., 2005] L. Mercier, T. Langø, F. Lindseth, and L. Collins (2005). A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Medicine & Biology*, 31(2):143--165.

[Milgram and Kishino, 1994] P. Milgram and F. Kishino (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77:1321--1321.

[Min and Jense, 1994] P. Min and H. Jense (1994). Interactive stereoscopy optimization for head-mounted displays. *Proc. SPIE Stereoscopic displays and VR*.

[Miyawaki et al., 2005] F. Miyawaki, K. Masamune, S. Suzuki, K. Yoshimitsu, and J. Vain (2005). Scrub nurse robot system-intraoperative motion analysis of a scrub nurse and timed-automata-based model for surgery. *IEEE Transactions on Industrial Electronics*, 52(5):1227--1235.

[Moberg and Whitcomb, 1999] T. Moberg and M. Whitcomb (1999). Educational technology to facilitate medical students' learning: background paper 2 of the medical school objectives project. *Academic Medicine*, 74(10):1146.

[Monsky et al., 2002] W. Monsky, D. Levine, T. Mehta, R. Kane, A. Ziv, B. Kennedy, and H. Nisenbaum (2002). Using a sonographic simulator to assess residents before overnight call. *American Journal of Roentgenology*, 178(1):35.

[Moore et al., 2004] C. Moore, S. Gregg, and M. Lambert (2004). Performance, training, quality assurance, and reimbursement of emergency physician-performed ultrasonography at academic medical centers. *Journal of Ultrasound in Medicine*, 23(4):459.

[Morelli et al., 2010] T. Morelli, J. Foley, and E. Folmer (2010). Vi-bowling: a tactile spatial exergame for individuals with visual impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, pages 179--186. ACM.

[Morrison and Meliza, 1999] J. Morrison and L. Meliza (1999). *Foundations of the after action review process*, volume 6. United States Army Research Institute for the Behavioral and Social Sciences.

[Mottura et al., 2007] S. Mottura, L. Greci, E. Travaini, G. Viganò, and M. Sacco (2007). MagicMirror & FootGlove: a new system for the customized shoe try-on. *The Future of Product Development*, pages 441--450.

[Murai et al., 2010] A. Murai, K. Kurosaki, K. Yamane, and Y. Nakamura (2010). Musculoskeletal-see-through mirror: Computational modeling and algorithm for whole-body muscle activity visualization in real time. *Progress in biophysics and molecular biology*, 103(2):310--317.

[Murphy et al., 2003] T. Murphy, C. Vignes, D. Yuh, and A. Okamura (2003). Automatic motion recognition and skill evaluation for dynamic tasks. In *Eurohaptics 2003*, pages 363--373.

[Nagata, 1996] S. Nagata (1996). The binocular fusion of human vision on stereoscopic displays: field of view and environment effects. *Ergonomics*, 39(11):1273--1284.

[Nakamoto et al., 2002] M. Nakamoto, Y. Sato, M. Miyamoto, Y. Nakamjima, K. Konishi, M. Shimada, M. Hashizume, and S. Tamura (2002). 3d ultrasound system using a magneto-optic hybrid tracker for augmented reality visualization in laparoscopic liver surgery. *Medical Image Computing and Computer-Assisted Intervention⊠MICCAI 2002*, pages 148--155.

[Nalanagula et al., 2006] D. Nalanagula, J. Greenstein, and A. Gramopadhye (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International journal of industrial ergonomics*, 36(4):289--300.

[Navab et al., 2010] N. Navab, S. Heining, and J. Traub (2010). Camera augmented mobile c-arm (camc): Calibration, accuracy study, and clinical applications. *Medical Imaging, IEEE Transactions on*, 29(7):1412--1423.

[Nefian et al., 2002] A. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy (2002). A coupled HMM for audio-visual speech recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2.

[Neri et al., 2007] L. Neri, E. Storti, and D. Lichtenstein (2007). Toward an ultrasound curriculum for critical care medicine. *Critical Care Medicine*, 35(5):S290.

[Neumuth et al., 2011] T. Neumuth, F. Loebe, and P. Jannin (2011). Similarity metrics for surgical process models. *Artificial Intelligence in Medicine*.

[Neumuth and Meißner, 2012] T. Neumuth and C. Meißner (2012). Online recognition of sur-
gical instruments by information fusion. *International Journal of Computer Assisted Radiol-
ogy and Surgery*, pages 1--8.

[Neumuth et al., 2006] T. Neumuth, S. Schumann, G. Strauß, P. Jannin, J. Meixensberger, A. Di-
etz, H. U. Lemke, and O. Burgert (2006). Visualization options for surgical workflows. In
International Journal of Computer Assisted Radiology and Surgery, pages 438--440.

[Newey et al., 2003] V. Newey, D. Nassiri, A. Bhide, and B. Thilaganathan (2003). Nuchal
translucency thickness measurement. *Ultrasound in Obstetrics and Gynecology*, 21(6):596--
601.

[Ng et al., 2005] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan (2005).
Light field photography with a hand-held plenoptic camera. *Computer Science Technical
Report CSTR*, 2.

[Ni et al., 2009] D. Ni, W. Chan, J. Qin, Y. Chui, I. Qu, S. Ho, and P. Heng (2009). A Virtual
Reality Simulator for Ultrasound Guided Organ Biopsy Training. *Computer Graphics and
Applications*, 99.

[Ni et al., 2008] D. Ni, W. Chan, J. Qin, Y. Qu, Y. Chui, S. Ho, and P. Heng (2008). An Ultrasound-
Guided Organ Biopsy Simulation with 6DOF Haptic Feedback. In *Medical Image Computing
and Computer-Assisted Intervention (MICCAI)*, page 559. Springer-Verlag.

[Ni et al., 2011] T. Ni, A. K. Karlson, and D. Wigdor (2011). Anatonme: facilitating doctor-
patient communication using a projection-based handheld device. In *Proceedings of the
2011 annual conference on Human factors in computing systems*, CHI '11, pages 3333--3342,
New York, NY, USA. ACM.

[Nicolaou et al., 2004a] M. Nicolaou, A. James, A. Darzi, and G. Yang (2004a). A study of sac-
cade transition for attention segregation and task strategy in laparoscopic surgery. *Medical
Image Computing and Computer-Assisted Intervention--MICCAI 2004*, pages 97--104.

[Nicolaou et al., 2004b] M. Nicolaou, A. James, A. Darzi, and G.-Z. Yang (2004b). A Study of
Saccade Transition for Attention Segregation and Task Strategy in Laparoscopic Surgery. In
*7th International Conference on Medical Image Computing and Computer-Assisted Interven-
tion (MICCAI)*, pages 97--104.

[Nielsen et al., 2002] J. Nielsen, T. Clemmensen, and C. Yssing (2002). Getting access to what
goes on in people's heads?: reflections on the think-aloud technique. In *Proceedings of the
second Nordic conference on Human-computer interaction*, pages 101--110. ACM.

[Nodine et al., 2002] C. Nodine, C. Mello-Thoms, H. Kundel, and S. Weinstein (2002). Time
course of perception and decision making during mammographic interpretation. *American
Journal of Roentgenology*, 179(4):917--923.

[Notelaers et al., 2010] S. Notelaers, T. De Weyer, K. Robert, C. Raymaekers, and K. Coninx
(2010). Design aspects for rehabilitation games for ms patients. In *Design and Engineer-
ing of Game-like Virtual and Multimodal Environments (DEnG-VE)*.

[Ohnumaa et al., 2006] K. Ohnumaa, K. Masamuneb, K. Yoshimitsua, T. Sadahiroa, J. Vainc, Y. Fukuia, and F. Miyawakia (2006). Timed-automata-based model for laparoscopic surgery and intraoperative motion recognition of a surgeon as the interface connecting the surgical scenario and the real operating room. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, pages 442--445.

[Okumura et al., 2006] B. Okumura, M. Kanbara, and N. Yokoya (2006). Augmented reality based on estimation of defocusing and motion blurring from captured images. In *IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006*, pages 219--225.

[Ones and Creager, 2003] M. Ones and M. Creager (2003). ACC/AHA Clinical Competence Statement on Echocardiography. *Journal of the American College of Cardiology*, 41(4).

[Ostendorf and Singer, 1997] M. Ostendorf and H. Singer (1997). HMM topology design using maximum likelihood successive state splitting. *Computer Speech & Language*, 11(1):17--41.

[Owen, 2012] H. Owen (2012). Early use of simulation in medical education. *Simulation in Healthcare*.

[Padoy et al., 2010] N. Padoy, T. Blum, A. Ahmadi, H. Feussner, M. Berger, and N. Navab (2010). Statistical modeling and recognition of surgical workflow. *Medical Image Analysis*.

[Padoy et al., 2007] N. Padoy, T. Blum, I. Essa, H. Feußner, M.-O. Berger, and N. Navab (2007). A boosted segmentation method for surgical workflow analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Brisbane, Australia. to appear.

[Padoy et al., 2008] N. Padoy, T. Blum, H. Feussner, M.-O. Berger, and N. Navab (2008). On-line recognition of surgical activity for monitoring in the operating room. In *Proceedings of the 20th Conference on Innovative Applications of Artificial Intelligence*, pages 1718--1724.

[Padoy and Hager, 2011] N. Padoy and G. Hager (2011). Human-machine collaborative surgery using learned models. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 5285--5292. IEEE.

[Park et al., 2005] J. Park, M. Chung, S. Hwang, Y. Lee, D. Har, and H. Park (2005). Visible korean human: Improved serially sectioned images of the entire body. *IEEE transactions on medical imaging*, 24(3):352--360.

[Park et al., 2009] Y. Park, V. Lepetit, and W. Woo (2009). ESM-Blur: Handling & rendering blur in 3D tracking and augmentation. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 163--166. IEEE Computer Society.

[Pathak and Lees, 2009] S. Pathak and C. Lees (2009). Ultrasound structural fetal anomaly screening: an update. *Archives of disease in childhood. Fetal and neonatal edition*, 94(5).

[Payandeh and Hsu, 2007] S. Payandeh and J. Hsu (2007). Toward surgical tool motion and gesture recognition. In *International Journal of Computer Assisted Radiology and Surgery*, page 492.

[Peli, 1999] E. Peli (1999). Optometric and perceptual issues with head-mounted displays. *Visual instrumentation: Optical design and engineering principles*, pages 205--276.

[Pfeiffer et al., 2005] M. Pfeiffer, A. AA, and A. Juffinger (2005). Predicting text relevance from sequential reading behavior. In *Workshop at NIPS 2005, in Whistler, BC, Canada, on December 10, 2005.*, page 25.

[Philip, 1986] J. Philip (1986). Gas man®: An example of goal oriented computer-assisted teaching which results in learning. *International Journal of Clinical Monitoring and Computing*, 3(3):165--173.

[Pieper et al., 1997] S. Pieper, M. Weidenbach, and T. Berlage (1997). Registration of 3D ultrasound images to surface models of the heart. In *Proceedings Interfaces to Real & Virtual Worlds*, pages 28--30. Citeseer.

[Piron et al., 2005] L. Piron, P. Tonin, F. Piccione, V. Iaia, E. Trivello, and M. Dam (2005). Virtual environment training therapy for arm motor rehabilitation. *Presence: Teleoperators & Virtual Environments*, 14(6):732--740.

[Plasters et al., 2003] C. Plasters, F. Seagull, and Y. Xiao (2003). Coordination Challenges in Operating-Room Management: An In-Depth Field Study. *AMIA Annual Symposium Proceedings*, pages 524--8.

[Quarles et al., 2008] J. Quarles, S. Lampotang, I. Fischler, P. Fishwick, and B. Lok (2008). Collocated AAR: Augmenting After Action Review with Mixed Reality. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 107--116.

[Queisser-Luft et al., 2002] A. Queisser-Luft, G. Stolz, A. Wiesel, K. Schlaefer, and J. Spranger (2002). Malformations in newborn: results based on 30940 infants and fetuses from the Mainz congenital birth defect monitoring system (1990--1998). *Archives of gynecology and obstetrics*, 266(3):163--167.

[Rabiner, 1989] L. Rabiner (1989). A tutorial on hidden Markov models and selected applications inspeech recognition. *Proceedings of the IEEE*, 77(2):257--286.

[Regenbrecht et al., 2012] H. Regenbrecht, S. Hoermann, G. McGregor, B. Dixon, E. Franz, C. Ott, L. Hale, T. Schubert, and J. Hoermann (2012). Visual manipulations for motor rehabilitation. *Computers & Graphics*.

[Reichl et al., 2008] T. Reichl, J. Passenger, O. Acosta, and O. Salvado (2008). Ultrasound goes GPU: real-time simulation using CUDA. In *SPIE Medical Imaging*.

[Reiley and Hager, 2009] C. Reiley and G. Hager (2009). Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2009*, pages 435--442.

[Reiley et al., 2011] C. Reiley, H. Lin, D. Yuh, and G. Hager (2011). Review of methods for objective surgical skill evaluation. *Surgical endoscopy*, 25(2):356--366.

[Reis et al., 2003] G. Reis, M. Bertram, R. van Lengen, and H. Hagen (2003). Adaptive Volume Construction from Ultrasound Images of a Human Heart. In *Eurographics/IEEE TCVG Visualization Symposium Proceedings*, pages 321--330. Citeseer.

[Reis et al., 2006] G. Reis, B. Lappé, S. Köhn, C. Weber, M. Bertram, and H. Hagen (2006). To-
    wards a Virtual Echocardiographic Tutoring System. *Visualization in Medicine and Life Sci-
    ences*, pages 99--119.

[Rhee et al., 2007] T. Rhee, J. Lewis, U. Neumann, and K. Nayak (2007). Soft-tissue deformation
    for in vivo volume animation. In *Computer Graphics and Applications, 2007. PG'07. 15th
    Pacific Conference on*, pages 435--438. IEEE.

[Richards et al., 2000] C. Richards, J. Rosen, B. Hannaford, C. Pellegrini, and M. Sinanan (2000).
    Skills evaluation in minimally invasive surgery using force/torque signatures. *Surgical en-
    doscopy*, 14(9):791--798.

[Riguer et al., 2003] G. Riguer, N. Tatarchuk, and J. Isidoro (2003). Real-time depth of field
    simulation. *ShaderX2: Shader Programming Tips and Tricks with DirectX*, 9:529--556.

[Rosen et al., 2002] J. Rosen, J. Brown, M. Barreca, L. Chang, B. Hannaford, and M. Sinanan
    (2002). The Blue DRAGON--a system for monitoring the kinematics and the dynamics of
    endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment.
    Studies in Health Technology and Informatics, 85:412--8.

[Rosen et al., 2006] J. Rosen, J. Brown, L. Chang, M. Sinanan, and B. Hannaford (2006). Gener-
    alized Approach for Modeling Minimally Invasive Surgery as a Stochastic Process Using a
    Discrete Markov Model. *IEEE Transactions on Biomedical Engineering*, 53(3):399--413.

[Rosen et al., 2000] J. Rosen, M. Solazzo, B. Hannaford, and M. Sinanan (2000). Objective eval-
    uation of laparoscopic surgical skills using hidden markov models based on haptic informa-
    tion and tool/tissue interactions. *American College of Surgeons Annual Meeting-Washington
    State Chapter*, 1.

[Rosen, 2008] K. Rosen (2008). The history of medical simulation. *Journal of critical care*,
    23(2):157--166.

[Rosenthal et al., 2001] M. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E. Pisano,
    M. Jiroutek, K. Muller, and H. Fuchs (2001). Augmented reality guidance for needle biopsies:
    A randomized, controlled trial in phantoms. In *Medical Image Computing and Computer-
    Assisted Intervention--MICCAI 2001*, pages 240--248. Springer.

[Ruiz et al., 2006] J. Ruiz, M. Mintzer, and R. Leipzig (2006). The impact of e-learning in medical
    education. *Academic medicine*, 81(3):207.

[Sadasivan et al., 2005] S. Sadasivan, J. Greenstein, A. Gramopadhye, and A. Duchowski (2005).
    Use of eye movements as feedforward training for a synthetic aircraft inspection task. In
    Proceedings of the SIGCHI conference on Human factors in computing systems, pages 141--
    149. ACM.

[Sakoe and Chiba, 1978] H. Sakoe and S. Chiba (1978). Dynamic programming algorithm op-
    timization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal
    Processing*, 26(1):43--49.

[Salb et al., 2002] T. Salb, O. Burgert, T. Gockel, J. Brief, S. Hassfeld, J. Muehling, and R. Dillmann (2002). Risk reduction in craniofacial surgery using computer-based modeling and intraoperative immersion. *Medicine meets virtual reality 02/10: digital upgrades, applying Moore's law to health*, 85:441.

[Salen et al., 2001] P. Salen, R. O'Connor, B. Passarello, D. Pancu, S. Melanson, S. Arcona, and M. Heller (2001). Fast education: a comparison of teaching models for trauma sonography1. *Journal of Emergency Medicine*, 20(4):421--425.

[Salisbury et al., 2004] K. Salisbury, F. Conti, and F. Barbagli (2004). Haptic rendering: Introductory concepts. *Computer Graphics and Applications, IEEE*, 24(2):24--32.

[Salvador and Chan, 2004] S. Salvador and P. Chan (2004). Fastdtw: Toward accurate dynamic time warping in linear time and space. In *KDD Workshop on Mining Temporal and Sequential Data*, pages 70--80.

[Sandars and Schroter, 2007] J. Sandars and S. Schroter (2007). Web 2.0 technologies for undergraduate and postgraduate medical education: an online survey. *Postgraduate medical journal*, 83(986):759.

[Sandor et al., 2010] C. Sandor, A. Cunningham, A. Dey, and V. Mattila (2010). An augmented reality x-ray system based on visual saliency. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 27--36. IEEE.

[Sauer et al., 2001a] F. Sauer, A. Khamene, B. Bascle, and G. Rubino (2001a). A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for imri-guided nuerosurgery. In *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2001*, pages 707--716. Springer.

[Sauer et al., 2001b] F. Sauer, A. Khamene, B. Bascle, L. Schinunang, F. Wenzel, and S. Vogt (2001b). Augmented reality visualization of ultrasound images: system description, calibration, and features. In *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*, pages 30--39. IEEE.

[Sauer et al., 2008] F. Sauer, S. Vogt, and A. Khamene (2008). Augmented reality. In T. Peters and K. Cleary, editors, *Image-Guided Interventions: Technology and Applications*, chapter 4, pages 81--120. Springer.

[Schlögl et al., 2001] S. Schlögl, E. Werner, M. Lassmann, J. Terekhova, S. Muffert, S. Seybold, and C. Reiners (2001). The use of three-dimensional ultrasound for thyroid volumetry. *Thyroid*, 11(6):569--574.

[Schmid-Wendtner and Burgdorf, 2005] M. Schmid-Wendtner and W. Burgdorf (2005). Ultrasound scanning in dermatology. *Archives of dermatology*, 141(2):217.

[Schönauer et al., 2011] C. Schönauer, T. Pintaric, and H. Kaufmann (2011). Full body interaction for serious games in motor rehabilitation. In *Proceedings of the 2nd Augmented Human International Conference*, page 4. ACM.

[Sclaveranoa et al., 2009] S. Sclaveranoa, G. Chevreaua, L. Vadcardc, P. Mozerb, and J. Troccaza (2009). BiopSym: a simulator for enhanced learning of ultrasound-guided prostate biopsy. In *Medicine Meets Virtual Reality (MMVR)*, page 301. Ios Pr Inc.

[Seward et al., 2002] J. Seward, P. Douglas, R. Erbel, R. Kerber, I. Kronzon, H. Rakowski, L. Sahn, E. Sisk, A. Tajik, and S. Wann (2002). Hand-carried cardiac ultrasound (HCU) device: recommendations regarding new technology. A report from the Echocardiography Task Force on New Technology of the Nomenclature and Standards Committee of the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, 15(4):369--373.

[Shams et al., 2008] R. Shams, R. Hartley, and N. Navab (2008). Real-time simulation of medical ultrasound from CT images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 734--741. Springer.

[Siddoway et al., 2006] D. Siddoway, M. Ingeholm, O. Burgert, T. Neumuth, V. Watson, and K. Cleary (2006). Workflow in interventional radiology: nerve blocks and facet blocks. *Proceedings of the SPIE Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display*, 6145:66--73.

[Sielhorst, 2008] T. Sielhorst (2008). *New Methods for Medical Augmented Reality*. PhD thesis, Technische Universität München (TUM), Chair for Computer Aided Medical Procedures.

[Sielhorst et al., 2008] T. Sielhorst, M. Feuerstein, and N. Navab (2008). Advanced medical displays: A literature review of augmented reality. *Display Technology, Journal of*, 4(4):451--467.

[Sielhorst et al., 2004] T. Sielhorst, T. Obst, R. Burgkart, R. Riener, and N. Navab (2004). An augmented reality delivery simulator for medical training. In *International Workshop on Augmented Environments for Medical Imaging-MICCAI Satellite Workshop*, volume 141. Citeseer.

[Sielhorst et al., 2006] T. Sielhorst, R. Stauder, M. Horn, T. Mussack, A. Schneider, H. Feussner, and N. Navab (2006). Simultaneous replay of automatically synchronized videos of surgeries for feedback and visual assessment. *International Journal of Computer Assisted Radiology and Surgery,Supplement 1*, 2:433--434.

[Simola et al., 2008] J. Simola, J. Salojärvi, and I. Kojo (2008). Using hidden markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4):237--251.

[Sizintsev and Wildes, 2010] M. Sizintsev and R. P. Wildes (2010). Coarse-to-fine stereo vision with accurate 3d boundaries. *Image Vision Comput.*, 28(3):352--366.

[Smith Jr et al., 1998] J. Smith Jr, M. Bergmann, R. Gildersleeve, and R. Allen (1998). A simple model for learning stereotactic skills in ultrasound-guided amniocentesis. *Obstetrics & Gynecology*, 92(2):303.

[Solberg et al., 2007] O. Solberg, F. Lindseth, H. Torp, R. Blake, and T. Nagelhus Hernes (2007). Freehand 3D ultrasound reconstruction algorithms--a review. *Ultrasound in Medicine & Biology*, 33(7):991--1009.

[Soler and Marescaux, 2008] L. Soler and J. Marescaux (2008). Patient-specific surgical simulation. *World Journal of Surgery*, 32(2):208--212.

[Song et al., 1998] D. Song, G. Joshi, and P. White (1998). Fast-track eligibility after ambulatory anesthesia: a comparison of desflurane, sevoflurane, and propofol. *Anesthesia & Analgesia*, 86(2):267.

[Soutschek et al., 2008] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber (2008). 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1--6. IEEE.

[Speidel et al., 2009] S. Speidel, T. Zentek, G. Sudra, T. Gehrig, B. Müller-Stich, C. Gutt, and R. Dillmann (2009). Recognition of surgical skills using hidden markov models. In *Proceedings of SPIE*, volume 7261, page 726125.

[Spitzer et al., 1996] V. Spitzer, M. Ackerman, A. Scherzinger, and D. Whitlock (1996). The visible human male: a technical report. *Journal of the American Medical Informatics Association*, 3(2):118.

[Staboulidou et al., 2010] I. Staboulidou, M. Wüstemann, B. Vaske, M. Elsässer, P. Hillemanns, and A. Scharf (2010). Quality assured ultrasound simulator training for the detection of fetal malformations. *Acta obstetricia et gynecologica Scandinavica*, 89(3):350--354.

[Stallkamp and Wapler, 1998] J. Stallkamp and M. Wapler (1998). UltraTrainer--a training system for medical ultrasound examination. In *Medicine Meets Virtual Reality (MMVR)*, volume 50, page 298.

[Standen et al., 2010] P. Standen, D. Brown, S. Battersby, M. Walker, L. Connell, A. Richardson, F. Platts, K. Threapleton, and A. Burton (2010). A study to evaluate a low cost virtual reality system for home based rehabilitation of the upper limb following stroke. *International Journal on Disability and Human Development*, 10(4).

[State et al., 2001] A. State, J. Ackerman, G. Hirota, J. Lee, and H. Fuchs (2001). Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space. *Augmented Reality, International Symposium on*, 0:137.

[Stengel et al., 2001] D. Stengel, K. Bauwens, J. Sehouli, F. Porzsolt, G. Rademacher, S. Mutze, and A. Ekkernkamp (2001). Systematic review and meta-analysis of emergency ultrasonography for blunt abdominal trauma. *British Journal of Surgery*, 88(7):901--912.

[Stetten et al., 2003] G. Stetten, A. Cois, W. Chang, D. Shelton, R. Tamburo, J. Castellucci, and O. Von Ramm (2003). C-mode real time tomographic reflection for a matrix array ultrasound sonic flashlight. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 336--343.

[Stolcke and Omohundro, 1994a] A. Stolcke and S. Omohundro (1994a). Best-first model merging for hidden markov model induction. Technical Report Technical Report TR-94-003, ICSI, Berkeley, CA.

[Stolcke and Omohundro, 1994b] A. Stolcke and S. Omohundro (1994b). Inducing probabilistic grammars by Bayesian model merging. *Grammatical Inference and Applications: Proceedings of the Second International Colloquium on Grammatical Inference*, pages 106--118.

[Sun and McKenzie, 2008] B. Sun and F. McKenzie (2008). Medical student evaluation using virtual pathology echocardiography (VPE) for augmented standardized patients. In *Medicine Meets Virtual Reality (MMVR)*, volume 132, page 508.

[Sun and Holliman, 2009] G. Sun and N. Holliman (2009). Evaluating methods for controlling depth perception in stereoscopic cinematography. In *Proceedings of SPIE*, volume 7237, page 72370I.

[Sunguroff and Greenberg, 1978] A. Sunguroff and D. Greenberg (1978). Computer generated images for medical applications. *ACM SIGGRAPH Computer Graphics*, 12(3):196--202.

[Sutherland, 1968] I. Sutherland (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757--764. ACM.

[Szpala et al., 2005] S. Szpala, M. Wierzbicki, G. Guiraudon, and T. Peters (2005). Real-time fusion of endoscopic views with dynamic 3-d cardiac images: a phantom study. *Medical Imaging, IEEE Transactions on*, 24(9):1207--1215.

[Tafra, 2001] L. Tafra (2001). The learning curve and sentinel node biopsy. *The American Journal of Surgery*, 182(4):347--350.

[Tahmasebi et al., 2007] A. Tahmasebi, P. Abolmaesumi, and K. Hashtrudi-Zaad (2007). A Haptic-based Ultrasound Examination/Training System. In *IEEE International Conference on Robotics and Automation (ICRA)*.

[Tahmasebi et al., 2008] A. Tahmasebi, P. Abolmaesumi, D. Thompson, and K. Hashtrudi-Zaad (2008). A Framework for the Design of a Novel Haptic-based Medical Diagnostic Simulator. *IEEE Transactions on Information Technology in Biomedicine*, 12(5):658--666.

[Takano et al., 2011] K. Takano, N. Hata, and K. Kansaku (2011). Towards intelligent environments: an augmented reality--brain--machine interface operated with a see-through head-mount display. *Frontiers in neuroscience*, 5.

[Teran et al., 2005] J. Teran, E. Sifakis, S. Blemker, V. Ng-Thow-Hing, C. Lau, and R. Fedkiw (2005). Creating and simulating skeletal muscle from the visible human data set. *Visualization and Computer Graphics, IEEE Transactions on*, 11(3):317--328.

[Terkamp et al., 2003] C. Terkamp, G. Kirchner, J. Wedemeyer, A. Dettmer, J. Kielstein, H. Reindell, J. Bleck, M. Manns, and M. Gebel (2003). Simulation of abdomen sonography. Evaluation of a new ultrasound simulator. *Ultraschall in der Medizin*, 24(4):239.

[Tjomsland and Baskett, 2002] N. Tjomsland and P. Baskett (2002). Asmund s. laerdal. *Resuscitation*, 53(2):115--120.

[Tong et al., 1998] S. Tong, H. Cardinal, R. McLoughlin, D. Downey, and A. Fenster (1998). Intra-and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. *Ultrasound in Medicine & Biology*, 24(5):673--681.

[Tormene et al., 2009] P. Tormene, T. Giorgino, S. Quaglini, and M. Stefanelli (2009). Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial Intelligence in Medicine*, 45(1):11--34.

[Traub et al., 2006] J. Traub, P. Stefan, S.-M. M. Heining, T. Sielhorst, C. Riquarts, E. Euler, and N. Navab (2006). Towards a hybrid navigation interface: Comparison of a slice based navigation system with in-situ visualization. In *Proceedings of MIAR 2006*, LNCS, pages 364--372, Shanghai, China. Springer.

[Triola et al., 2006] M. Triola, H. Feldman, A. Kalet, S. Zabar, E. Kachur, C. Gillespie, M. Anderson, C. Griesser, and M. Lipkin (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of general internal medicine*, 21(5):424--429.

[Troccaz et al., 2000] J. Troccaz, D. Henry, N. Laieb, G. Champleboux, J. Bosson, and O. Pichot (2000). Simulators for medical training: application to vascular ultrasound imaging. *The Journal of Visualization and Computer Animation*, 11(1):51--65.

[Trochim, 2002] S. Trochim (2002). *Situiertes Lernen in Augmented-Reality-basierten Trainingssystemen am Beispiel der Echokardiographie.* PhD thesis, Universitätsbibliothek Bielefeld.

[Tsai and Lenz, 1988] R. Tsai and R. Lenz (1988). Real time versatile robotics hand/eye calibration using 3d machine vision. In *Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on*, pages 554--561. IEEE.

[Tuceryan et al., 2002] M. Tuceryan, Y. Genc, and N. Navab (2002). Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators & Virtual Environments*, 11(3):259--276.

[Tuceryan et al., 1995] M. Tuceryan, D. Greer, R. Whitaker, D. Breen, C. Crampton, E. Rose, and K. Ahlers (1995). Calibration Requirements and Procedures for a monitor-based augmented reality system. *IEEE Transactions on Visualization and Computer Graphics*, 1(3):255--273.

[Tworetzky et al., 1999] W. Tworetzky, D. McElhinney, M. Brook, V. Mohan Reddy, F. Hanley, and N. Silverman (1999). Echocardiographic diagnosis alone for the complete repair of major congenital heart defects. *Journal of the American College of Cardiology*, 33(1):228--233.

[van Someren et al., 1994] M. van Someren et al. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes.* Academic Press, Inc.

[Varadarajan et al., 2009] B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager (2009). Data-derived models for segmentation with application to surgical assessment and training. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2009*, pages 426--434.

[Vidal et al., 2005] F. Vidal, N. Chalmers, D. Gould, A. Healey, and N. John (2005). Developing a needle guidance virtual environment with patient-specific data and force feedback. In *International Congress Series*, volume 1281, pages 418--423. Elsevier.

[Vidal et al., 2008] F. Vidal, A. Healey, D. Gould, and N. John (2008). Simulation of ultrasound guided needle puncture using patient specific data with 3D textures and volume haptics. *Computer Animation and Virtual Worlds*, 19(2):111--127.

[Vidani et al., 2010] A. Vidani, L. Chittaro, and E. Carchietti (2010). Assessing nurses' acceptance of a serious game for emergency medical services. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2010 Second International Conference on*, pages 101--108. IEEE.

[Voit et al., 2001] C. Voit, T. Mayer, M. Kron, A. Schoengen, W. Sterry, L. Weber, and T. Proebstle (2001). Efficacy of ultrasound B-scan compared with physical examination in follow-up of melanoma patients. *Cancer*, 91(12):2409--2416.

[Vozenilek et al., 2004] J. Vozenilek, J. Huff, M. Reznek, and J. Gordon (2004). See one, do one, teach one: advanced technology in medical education. *Academic Emergency Medicine*, 11(11):1149--1154.

[Wachinger and Navab, 2009] C. Wachinger and N. Navab (2009). Alignment of Viewing-Angle Dependent Ultrasound Images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

[Wachinger et al., 2008] C. Wachinger, W. Wein, and N. Navab (2008). Registration Strategies and Similarity Measures for Three-dimensional Ultrasound Mosaicing. *Academic Radiology*, 15:1404--1415.

[Wacker et al., 2005] F. Wacker, S. Vogt, A. Khamene, F. Sauer, M. Wendt, J. Duerk, J. Lewin, and K. Wolf (2005). Mr image-guided needle biopsies with a combination of augmented reality and mri: A pilot study in phantoms and animals. In *International Congress Series*, volume 1281, pages 424--428. Elsevier.

[Wallace, 1997] P. Wallace (1997). Following the threads of an innovation: the history of standardized patients in medical education. *Caduceus*, 13(2):5.

[Walton and Jones, 2007] S. Walton and M. Jones (2007). Interacting with volume data: Deformations using forward projection. In *Medical Information Visualisation-BioMedical Visualisation, 2007. MediVis 2007. International Conference on*, pages 48--53. IEEE.

[Wang and Gasser, 1997] K. Wang and T. Gasser (1997). Alignment of Curves by Dynamic Time Warping. *The Annals of Statistics*, 25(3):1251--1276.

[Warnes et al., 2008] C. Warnes, R. Williams, T. Bashore, J. Child, H. Connolly, J. Dearani, P. del Nido, J. Fasules, T. Graham Jr, Z. Hijazi, et al. (2008). ACC/AHA 2008 Guidelines for the Management of Adults With Congenital Heart Disease. *Circulation*, 118(23):e714.

[Wartell et al., 2002] Z. Wartell, L. Hodges, and W. Ribarsky (2002). A geometric comparison of algorithms for fusion control in stereoscopic HTDs. *IEEE Transactions on Visualization and Computer Graphics*, pages 129--143.

[Weidenbach et al., 2007] M. Weidenbach, H. Drachsler, F. Wild, S. Kreutter, V. Razek, G. Grunst, J. Ender, T. Berlage, and J. Janousek (2007). EchoComTEE--a simulator for transoesophageal echocardiography. *Anaesthesia*, 62(4):347--353.

[Weidenbach et al., 2009] M. Weidenbach, V. Razek, F. Wild, S. Khambadkone, T. Berlage, J. Janoušek, and J. Marek (2009). Simulation of congenital heart defects: a novel way of training in echocardiography. *Heart*, 95(8):636.

[Weidenbach et al., 2004] M. Weidenbach, S. Trochim, S. Kreutter, C. Richter, T. Berlage, and G. Grunst (2004). Intelligent training system integrated in an echocardiography simulator. *Computers in biology and medicine*, 34(5):407--425.

[Weidenbach et al., 2000] M. Weidenbach, C. Wick, S. Pieper, K. Quast, T. Fox, G. Grunst, and D. Redel (2000). Augmented reality simulator for training in two-dimensional echocardiography. *Computers and biomedical research*, 33(1):11--22.

[Weidenbach et al., 2005] M. Weidenbach, F. Wild, K. Scheer, G. Muth, S. Kreutter, G. Grunst, T. Berlage, and P. Schneider (2005). Computer-based training in two-dimensional echocardiography using an echocardiography simulator. *Journal of the American Society of Echocardiography*, 18(4):362--366.

[Wein, 2007] W. Wein (2007). *Multimodal Integration of Medical Ultrasound for Treatment Planning and Interventions*. PhD thesis, Technische Universität München (TUM), Chair for Computer Aided Medical Procedures.

[Wein et al., 2008] W. Wein, S. Brunke, A. Khamene, M. Callstrom, and N. Navab (2008). Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Image Analysis*, 12:577--585.

[Wein et al., 2007] W. Wein, A. Khamene, D. Clevert, O. Kutter, and N. Navab (2007). Simulation and fully automatic multimodal registration of medical ultrasound. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer.

[Wendler et al., 2006] T. Wendler, J. Traub, S. Ziegler, and N. Navab (2006). Navigated three dimensional beta probe for optimal cancer resection. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2006*, pages 561--569.

[Whitworth et al., 2010] M. Whitworth, L. Bricker, J. Neilson, and T. Dowswell (2010). Ultrasound for fetal assessment in early pregnancy. *Cochrane database of systematic reviews*, 4.

[Wick et al., 1999] C. Wick, S. Pieper, M. Weidenbach, D. Redel, T. Berlage, and F. Publica (1999). First experiences with cardiological teleconsultation using 3D ultrasound data sets. In *Proceedings of Computer Aided Radiology and Surgery (CARS)*, volume 99.

[Wilson et al., 2010] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters (2010). Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surgical endoscopy*, 24(10):2458--2464.

[Wu, 2001] J. Wu (2001). Tofu as a tissue-mimicking material. *Ultrasound in Medicine & Biology*, 27(9):1297--1300.

[Wüstemann et al., 2008] M. Wüstemann, G. Jehle, R. Schwerdtfeger, and K. Mühlhaus (2008). Virtueller Herzultraschall erstmals möglich: Fetale echokardiographie am ultraschallsimulator. *Frauenarzt*, 49(1):46--49.

[Wüstemann et al., 2002] M. Wüstemann, A. Scharf, H. Maul, P. Baier, and C. Sohn (2002). Der Ultraschallsimulator: Eine effektive Trainingsmethode zur Steigerung der Untersucherkompetenz bei der Bestimmung der fetalen Nackentransparenz. *Geburtshilfe und Frauenheilkunde*, 62(12):1183--1187.

[Yamauchi, 2007] Y. Yamauchi (2007). A motion data recording of surgical staff by optical range camera. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, page 496.

[Yuan et al., 2011] Y. Yuan, Y. Chen, S. Ni, A. Xu, L. Tang, M. Vingron, M. Somel, and P. Khaitovich (2011). Development and application of a modified dynamic time warping algorithm (dtw-s) to analyses of primate brain expression time series. *BMC bioinformatics*, 12(1):347.

[Zhu and Salcudean, 2010] M. Zhu and S. E. Salcudean (2010). Real time ultrasound needle image simulation using multi-dimensional interpolation. In T. Jiang, N. Navab, J. P. Pluim, and M. A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 429--436. Springer.

[Zhu et al., 2006] Y. Zhu, D. Magee, R. Ratnalingam, and D. Kessel (2006). A virtual ultrasound imaging system for the simulation of ultrasound-guided needle insertion procedures. In *Proceedings of Medical Image Understanding and Analysis (MIUA)*, volume 1, pages 61--65. Citeseer.

[Zhu et al., 2007] Y. Zhu, D. Magee, R. Ratnalingam, and D. Kessel (2007). A training system for ultrasound-guided needle insertion procedures. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 566--574. Springer.

[Ziv et al., 2003] A. Ziv, P. Wolpe, S. Small, and S. Glick (2003). Simulation-based medical education: an ethical imperative. *Academic Medicine*, 78(8):783.