Computer Aided Medical Procedures
Prof. Dr. Nassir Navab

Dissertation

Multi-Modal Visualization Paradigms for RGBD Augmented X-ray Imaging

Séverine Ginette Geneviève Habert

Fakultät für Informatik
Technische Universität München

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin

Multi-Modal Visualization Paradigms for RGBD Augmented X-ray Imaging

Séverine Ginette Geneviève Habert

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.

Vorsitzende(r) : Prof. Bernd Brügge, Ph.D.

Prüfer der Dissertation : 1. Prof. Dr. Nassir Navab

2. Prof. Pascal Fallavollita

Die Dissertation wurde am 12.10.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 11.02.2018 angenommen.

Séverine Ginette Geneviève Habert
Multi-Modal Visualization Paradigms for RGBD Augmented X-ray Imaging
Dissertation, Version 1.0

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 Garching bei München

Cover Image Credit :
The Operating Room of the Future, public domain

Abstract

Since the discovery in 1895 by Wilhelm Röntgen of X-ray radiation, X-ray imaging has become a valuable and indispensable imaging tool in the Operating Room (OR). It allows obtaining an immediate and actualized feedback on the internal anatomy during the surgery. Since the introduction of Image-Guided Surgery, the research community has investigated how to utilize this precious information at its best to provide shorter, less invasive, less radiative, more precise surgery than the sole use of the 2D X-ray image could propose without additional sensors and computational tools. Such improvement goals are beneficial for the three main actors of hospital; for the patient, by impacting positively the recovery time and surgery aftermath; for the surgeon and surgical crew, by reducing exposure to radiation and therefore improving working conditions; for the hospital, by increasing cost-efficiency regarding patient stay and OR turnover. For best integration of new X-ray imaging based technology in the OR, minimum disruption to the surgical workflow is also an aim to achieve.

In this thesis, we propose setups composed of an X-ray imaging device, namely a C-arm, on which is attached a consumer-device RGBD camera. The first setup, composed of a mirror construction placed in the housing of the X-ray source, allows for an exact overlay of X-ray image over video image complemented with depth information, which contextualizes the X-ray image in the surgical environment. However, the mirror construction requires heavy engineering on the device due to its location inside the C-arm housing, which contradict with the minimal disruption aim. Therefore, we propose a second mirrorless setup that consists of two RGBD cameras attached to the C-arm, but outside of the C-arm housing, that provides the same output in terms of visualization as the first setup, thanks to the RGBD data, allowing an effortless integration of the setup in the OR. The second part of the thesis presents clinical applications using the proposed setup, focused on providing intelligible visualizations to the surgeon regarding the surgical context. We propose 2 applications whose goals are to extend the contextualization of the X-ray image from its 2D context (video) to either 3D using 3D reconstruction from RGBD data or to 2D multi-layers, providing the surgeon the maximal information about the surgical environment in a single visualization. Then, we develop a sensibilization tool about the radiation exposure targeting the surgeon, realized by estimating and displaying the X-ray scattering information through an augmented reality heat map visualization. We also propose an assistive tool that provides an intelligible metric of back surface deformation during scoliosis minimally invasive surgery, that complements the metrics provided by X-ray imaging for spinal deformations, aiming at both better surgical and aesthetics outcomes. Finally, we propose a mixed-reality approach for C-arm based surgeries that combines patient-based 3D printed anatomy and simulated X-ray imaging with a real C-arm to complement traditional training and new technology assessment. This thesis has explored novel C-arm augmentations and novel visualization paradigms with the main goal to provide the surgeon with a more intelligible and safer OR environment that will benefit the surgical crew, the patient as well as the hospital.

Zusammenfassung

Seit der Entdeckung von Röntgenstrahlen durch Wilhelm Röntgen 1895, wurde das Röntgen zu einer wertvollen und unabdingbaren bildgebenden Werkzeug im Operationssaal (OP). Es ermöglicht eine sofortige und aktuelle Darstellung der internen Anatomy während eines Eingriffs. Seit der Einführung von bildgestützten Operationen hat die wissenschaftliche Gemeinschaft untersucht wie diese wertvolle Information bestmöglichst verwendet werden kann um eine kürzere, weniger invasive, weniger radioaktive und präzisere Operation zu ermöglichen in dem über das alleinige 2D Röntgenbild ohne zusätzliche Sensoren oder computergestützte Werkzeuge hinausgegangen wird. Diese Verbesserungen sind vorteilhaft für die drei Hauptaktoren eine Krankenhauses: für den Patient, in dem die Wiedergenesung und Nachwirkungen der Operation positiv beeinflusst werden; für den Chirurgen und sein Team, in dem die Strahlenbelastung reduziert und Arbeitsbedingungen verbessert werden; für das Krankenhaus, in dem die Kosteneffizient durch einen kürzeren Patientaufhalt und eine höhere OP Auslastung verbessert wird. Für eine bestmögliche Integration neuer Röntgenbasierter Technology im OP, ist auch ein Ziel die Störung des chirurgischen Arbeitsflusses minimal zu halten.

In dieser Doktorarbeit stellen wir verschiedene Konfigurationen eines Röntgengerätes vor, in diesem Fall ein C-Bogen, an dem eine herkömmliche RGBD Kamera befestigt ist. Die erste Konfiguration, bei der ein Spiegel im Gehäuse der Röntgenquelle montiert wird, erlaubt eine exakte Überlagerung von Röntgenbildern über Videobildern erweitert durch Tiefeninformationen, was eine Kontextualisierung des Röngtenbildes in die chirurgische Umgebung ermöglicht. Allerdings benötigt die Spiegelaufhängung durch die Positionierung im Gehäuse einen größeren technischen Eingriff, was einen Widerspruch zum Ziel der minimalen Störung steht. Deshalb stellen wir eine zweite, spiegellose Konfiguration vor, die aus zwei RGBD Kameras besteht, die außerhalb des C-Bogen Gehäuses befestigt werden, die dank der RGBD Daten die selbe Visualisierung wie die erste Konfiguration ermöglichen aber mühelos in den OP integriert werden kann. Der zweite Teil der Doktorarbeit stellt klinische Anwendungsfälle der vorgestellter Konfiguration vor, mit Fokus auf eine intelligente Visualisierung des chirurgischen Kontexts für den Chirurgen. Wir schlagen zwei Anwendungen vor deren Ziel es ist, den Kontext des Röntgenbildes von 2D (Video) entweder nach 3D, durch 3D Rekonstruktion der RGBD Daten oder nach mehrschichtigem 2D zu erweitern. Dies liefert dem Chirurgen die maximalen Informationen über die chirurgische Umgebung in einer einzigen Visualisierung. Danach entwickeln wir ein Werkzeug zur Sensibilisierung der Strahlenbelastung des Chirurgen, realisiert durch eine Abschätzung und Darstellung der zerstreuten Röntgenstrahlung in einer Virtuellen Realität Visualisierung in Form einer "Heatmap". Desweiteren führen wir ein technisches Hilfsmittel vor, das eine verständliche Metrik für die Rückgratverkrümmung während einer minimal-invasiven Skoliose Operation liefert. Diese Metrik ergänzt die Metriken welche durch die Röntgenbilder der Rückgratverkrümmung geliefert werden und zielt auf einen besseren chirurgischen und ästhetischen Ausgang der Operation. Schlussendlich demonstrieren wir

einen gemischte Realitätsansatz für C-Bogen basierte Operationen, welcher Patienten basierte, 3D gedruckte Anatomy und simulierte Röngtenstrahlung mit einem echtem C-Bogen kombiniert um übliche Trainingsmethoden und Bewertungen von neuen Technologien zu ergänzen. Diese Doktorarbeit hat neuartige C-Bogen Erweiterungen und neue Visualisierungstechniken untersucht, mit dem Hauptziel dem Chirurgen eine verständlichere und sichere OP Umgebung zu bieten von dem das chirurgische Personal, der Patient und auch das Krankenhaus profitieren.

Acknowledgments

I want to express my deep gratitude to Prof. Navab for welcoming me into his laboratory in 2013. I have not only discovered a terrific team to research with but also a great PhD supervisor to listen, discuss and work with. I would like also to thank the two NARVIS senior researchers: Pascal Fallavollita and Ulrich Eck, that immensely helped me during the PhD as well as Federico Tombari for providing me good advices in Computer Vision that considerably contributed to my project.

I want to acknowledge the different fundings that helped me during my PhD: the Bayerische Forschungsstiftung during the first three years (obtained thanks to Olivier Pauly), the TUM Diversity Office to provide me the Laura Bassi Award for my fourth year, and finally the Deutsche Forschungsgemeinschaft for covering my last months.

I am grateful to the huge CAMP team; the Narvis crew: Meng, Felix, Patrick, Philipp, Xiang, Hemal; the teaching colleagues: Keisuke, Christian SzB, Christian R., Hessam; and of course, all my other CAMP colleagues, too many to be all cited, that made my stay in the lab nice, fun and interesting.

Finally, I want to express my affection to my parents and my brother for supporting me in everything I do, even though it means to be far from home. I would like to thank Markus's family for welcoming me in Munich and making me feel at home, even far from it. Finally, I owe my deepest gratitude to Markus for his constant support and listening.

Understand well as I may, my comprehension can only be an infinitesimal fraction of all I want to understand.

— Ada Lovelace

Contents

Part I

Introduction

1 Motivation

> *sur·ger·y,*
> *from Latin chirurgiae; Greek cheir, hand, + ergon, work*
> *1. The branch of medicine concerned with the treatment of disease, injury, and deformity by physical operation or manipulation.*
> *2. The performance or procedures of an operation.*
>
> **— Farlex Partner Medical Dictionary**

The word Surgery derives from the Greek χειρουργική cheirourgikē (composed of χείρ, "hand", and ἔργον, "work"), via the Latin word: chirurgiae, meaning "hand work". The origins of the word Surgery reflects what is the core of surgery until nowadays, a manual work performed by the surgeon on a patient body. As the Greek origin of the word hints, the art of surgery is ancient, as old as the humans who first learned to create and handle tools. The brutality and risks of opening a living person's body have for a long time stayed considerable. Until the industrial revolution in the XIXth century, surgeons were incapable of overcoming the three main impediments that plagued the surgical profession from its ancient start: bleeding, pain, and infection.

1.1 An Historical Perspective on Modern Surgery

In the Middle Ages, surgery practice, due to its manual nature, was not recognized as a noble medical art, overlooked by physicians that would prefer more intellectual practice such as consulting or observation. This left the field free to barber-surgeons, with little medical knowledge but a sharp razor, that would take care of treating the visible affections: trauma and orthopedic (going until amputations as shown in Figure 1.1), skin disease, etc. Mortality of surgery would then be quite high due to loss of blood and infection.

The surgery practice wins one's spurs as scientific innovations revolutionize the field and solve the main culprits of surgery at the time, i.e. bleeding, pain and infection. The surgical revolutions are namely the anesthetics, the antiseptics, and the imaging.

1.1.1 Anesthesia

Modern pain control via anesthesia was developed in the middle of the XIXth century. Before that, surgery would remain a traumatic and painful procedure that requires the surgeon to

Fig. 1.1. A feldsher (barber-surgeon in Germanic countries) performing an amputation. Engraving from 1540, in US public domain

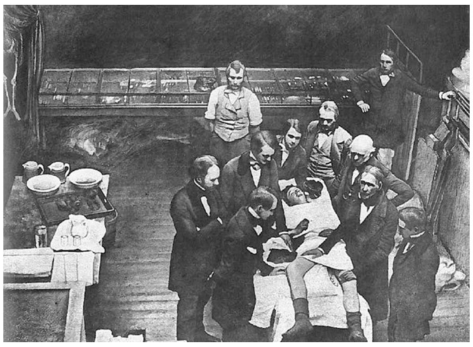


Fig. 1.2. Operation Being Performed with the Use of Ether Anesthesia in Spring 1847, the first public demonstration of surgical anesthesia occurred in the same room on October 16, 1846, reproduced with permission from [47], ©Massachusetts Medical Society.

perform the procedure as fast as possible to minimize patient suffering, limiting the surgical applications to few such as amputations or external growth removals. Beginning in the 1840s, the discovery of effective and practical anesthetic chemicals such as ether and chloroform would allow the practice of general anesthesia as shown in Figure 1.2 that will not only relieve patient suffering but also allow more intricate operations inside the human body. In addition, the discovery of muscle relaxants such as curare allowed for safer applications.

1.1.2 Antiseptics

Infections are the invasion of the human body by microorganisms such as virus or bacteria which can lead to lesions on the human body. Based on the contemporary works of Pasteur on those microorganisms, Joseph Lister discovered in the middle of the XIXth century that spraying carbolic acid, now known as phenol, on his instruments would reduce significantly the incidence of gangrene [90]. Following this work, he would also realize that infection during surgery could be better avoided by preventing bacteria from getting into surgical site and wounds in the first place. This led to the development of sterile surgery with the use of clean gloves and hand/surgical instruments washing in 5% carbolic solution before and after operations. He also introduced the steam sterilizer to sterilize equipment.

1.1.3 Imaging and Minimally-Invasive Surgery

The discovery in 1895 by Wilhelm Röntgen of X-ray radiation has started the revolution of the XXth century in the surgical field which is the development of medical imaging. Since then, numerous imaging technologies have been conceived: Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, Endoscopy, Elastography, Tactile Imaging, Thermography, Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT). The ability to observe structure and function within the human body has considerably improve the diagnosis before surgery as well as the navigation during surgery to reach the right surgical target. In particular, Endoscopy has helped the development of

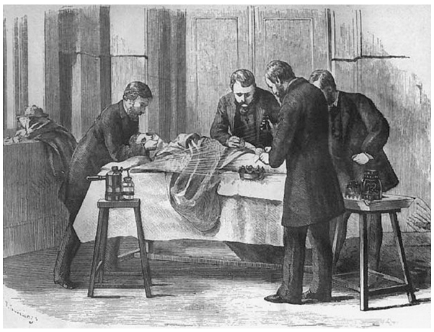


Fig. 1.3. Introduction of Carbolic Acid for Antisepsis, reproduced with permission from [47], ©Massachusetts Medical Society.

Fig. 1.4. Physicians perform laparoscopic stomach surgery, public domain in the USA

minimally-invasive surgeries in combination with the imaging modalities in the late 1980's (see Figure 1.4). This has allowed reducing the surgical incision size, limiting the bleeding and the healing time.

1.2 Ongoing Innovation

1.2.1 Innovation in Surgery

We have shown, in the previous section, few revolutions in surgical procedures that drastically changed the practice. They were so-called innovation, defined as "(the use of) a new idea or method", of the surgical practice. Riskin et al. [132] classify the innovations in new surgical technology by their clinical impact, describing two types of innovation: expanding and refining. They define an expanding period of innovation as "a time when technology develops rapidly and patient care is significantly altered"[132]. Then a refining period of innovation is defined as "a time when existing technologies are improved upon, but patient care is changed little by these improvements. A refining innovation generally either increases efficiency, lessens the labor or device costs for a procedure, or slightly improves outcome" [132]. Innovation falling into the expanding category would easily and fast integrate hospitals, as the use of surgical endoscopes for MIS introduced in the 1980's and that became rapidly a standard. However, the history of surgical innovation follows an ebb-and-flow pattern. The expanding technology leads to a rapid expansion of medical capabilities and procedures, then followed by a slower period of technical refinement and consolidation of approaches as shown in Figure 1.5 with the pink curve. This shows that even though searching for the expanding innovation is an aim, the refining period must not be ignored and is a necessary time to bring technologies and techniques to the maximum of their capacity.

1.2.2 Need for Innovation

Based on the developments that led to Modern Surgery and the tremendous progress that has been made since two centuries, surgical protocols and workflows have been established for most of the pathologies. The question is therefore how to innovate to change those existing procedures and reduce the still existing risks inherent to surgery, resulting in even better outcomes. The primary risk for the patient is of course that the surgery did not improve its condition or even make it worse - this being very surgery-specific. They are secondary risks inherent to surgeries that are shared by every patient entering the OR. In Table 1.1, we have described those secondary risks, as well as the strategy to fight against those risks. Those risks are not negligible, Klevens et al. [76] reports an incidence of 2% of surgical site infections for all surgical procedures in the USA, while Haller et al. [54] explains that one patient on ten will have face an anesthesia incident.

Surgical risk	Strategy against the risk
Infection due to open exposed surgical area	Reduce the size of the open surgical area
Damage of healthy structure	Provide navigation to only operate on surgical target
Anesthesia Effects	Reduce time of surgery and invasiveness
Muscle Atrophy due to long hospital stays	Reduce the recovery time
Radiation exposure (X-ray imaging, nuclear imaging)	Reduce use of radiative imaging to the minimum
Unaesthetic scar	Reduce the size of the open surgical area

Tab. 1.1. Risks inherent to surgery for the patient and strategy against those risks

Innovating can help to improve the efficiency of a surgery by fighting against the primary risk, but as well against the secondary risks with final benefits for the patient to have shorter, safer and hospitals stays with no or minimum side effect, reducing the risk of future complications. The reduction of surgical risk does not only benefit the patient but profits hospitals and surgical crew as well. In the perspective of cost-efficiency required by most of the hospitals in Western countries [60], shortening every hospital procedures from the surgery to the patient stay allows higher bed hospitals and OR use turnover, improving the hospital profitability. Safer surgeries also reduce the legal risk for the hospital. From the surgical crew perspective, working conditions can also be improved with reduced time in the OR, less exposure to radiation, even more lethal for them than to the patient, as they are exposed potentially every day to it as in the case of X-ray imaging.

1.2.3 Innovation in Modern-Day Hospitals

The OR has always been a place that welcomed the innovation. However, most of the Operating Rooms in the world are not equipped with state of the art technologies. The cost of new technology is often a stopper for its immediate and full integration in the OR. Indeed, there is a balance of cost/clinical impact to consider the worth of such technology. The term cost includes here, of course, the direct expenses such as technology price, but also the indirect ones such as the disruption of the clinical workflow. Stronger the disruption is, the more indirect cost the technology involves, penalizing it in the balance cost/clinical impact. The disruption can lead to new constraints: longer surgery, more complex technology requiring training and longer learning phase, increase in radiation exposure, degradation in working conditions if the case of cumbersome technology. Therefore, not only the price of technology but also any negative/positive changes to the clinical workflow must be considered in the balance cost/clinical impact.

If we consider the imaging in the OR for orthopedic and trauma surgery, CT is the imaging modality providing the most information and could, therefore, be of best use in the OR. However, the price of this device, the additional radiation (which forces the surgical staff to leave the OR room at every acquisition) makes it not the primary choice for surgical modality in comparison to the C-arm, which presents less information but is cheaper (of factor five at least), involves less radiative and less breaks in the workflow. C-arm devices can be found in every OR while even in developed countries such as Germany, the access to CT imaging is limited. As an example, the Klinikum Innenstadt in Munich only has one CT scanner for the full trauma and orthopedic facility, which is used for diagnostic CT as well as surgery. The CT will then only be used when its added value is significant compared to the C-arm, when the surgery is difficult or when 3D visualization is necessary.

We have described previously that innovation can be of different types: expanding or refining, depending on their clinical impact. Not all innovations are equal in the outcomes they bring to the surgery and it is the cost of this innovation versus the benefits (for the patient/surgical staff/hospital), i.e. the profitability that must be calculated. At equal cost, the expanding innovation will find better its way to the Operating Room. However, due to the slow pace of innovation, expanding will come more rarely than refining innovation. To still bring innovation in the OR at a fast pace, it is the “economical” and minimally workflow disruptive refining innovation that must also be targeted by the research community. In the case of C-arm, this means that even though it is in its refining period, innovations on this device must still be pursued, while the effort to research on the next expanding innovation such as for example Cone Beam Computer Tomography - CBCT (in a cheaper version than now/ or more efficient) that may replace C-arm.

1.2.4 Refining Technologies

Refining technologies such as imaging devices can take several paths, through device modification or device supplementation. The first solution generally does not lead to bringing innovation rapidly in the OR, as its introduction would involve buying a new device which is not economical at all for hospitals. Also, devices in the Operating Room have a lifespan which is sometimes over 20 years. For profitability reason, the devices must be used all along their lifespan even if rapidly outdated by new innovation. If the innovation can only be brought every time the device is replaced, then a long period of time happens where the clinical outcomes for the patient are not improved. The gap is then widening with the state of the art technology as the device ages as it can be seen in Figure 1.5 where we show the theoretical pace of innovation (in pink) versus the one as seen in the OR if the only innovation happens when devices are replaced/upgraded (in blue).

A refining innovation for C-arm devices was the introduction of flat-panel detectors (FPD) in 2006 by Ziehm, however, this increases the price of the device by two, FPD costs between

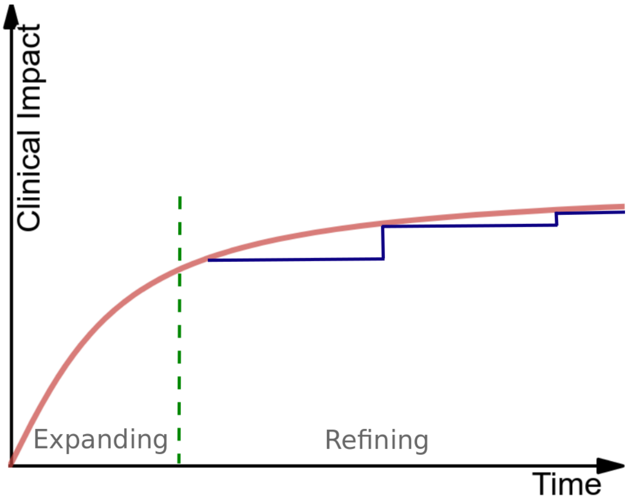


Fig. 1.5. Clinical Impact versus Time for Surgical Innovation, in pink, the pace of innovation, in blue, the pace of innovation as seen in the OR

280000$ and 300000$ and standard C-arm cost between 120000$ and 170000$ [36]. However, Dubinsky in their analysis of the FDP C-arm market [36] explains that it is only a question of time before the FDP totally replace the standard C-arm. To bring the innovation in the OR in a faster manner, supplementing the existing devices is, therefore, a possible solution. It can fill the gap in between the current state of the art (the pink curve in Figure 1.5) and the available device in the OR (the blue curve in Figure 1.5). Based on the technology already in the Operating Room, the introduction can be a lot faster.

1.3 Challenges of C-arm based Surgeries

We have already introduced the terms of C-arm in the previous section. A C-arm is an X-ray imaging device designed to be used during surgery. We refer the reader to the Background Chapter 3 for more details on C-arm. To resume, the X-ray image output by C-arm is a 2D accumulative projection of a complex 3D volume with numerous organs of different densities overlaid. This leads to several challenges when using C-arms such as a difficult mental mapping from 3D to 2D, exposure to radiation as well as the device cumbersomeness. All the challenges are illustrated in one image that I have captured, shown in Figure 1.6, taken in the Operating Room during an angioplasty procedure in Pasing Klinikum in 2014.

We will use this image as an illustration along as we describe the different challenges:

- Radiation Exposure (blue circle)
- C-arm collision (orange circle)
- Lack of Internal context (green circle)
- Lack of External context (red circle)
- Interpretation of X-ray image (green circle)

Fig. 1.6. Photography acquired during an angioplasty procedure

1.3.1 Safety in the OR

Radiation Exposure

The radiation emitted by C-arm can be lethal at long term (more details in Background Chapter, Section 3.2.3). Therefore, even though the surgical staff is protected with lead aprons and glasses, this does not protect the full body such as the head and the hands as shown in Figure 1.6. In that regards, the FDA points out in their regulations that the ALARA principle: "As Low As Reasonably Achievable" should always be followed. As one of their recommendation to help following that principle, the FDA advises that the "Imaging teams [...] should [...] develop protocols and technique charts (or use those available on the equipment) that optimize exposure for a given clinical task and patient group" [105]. Therefore, the challenge with radiation exposure in the Operation Room is to reduce the amount of radiation used for surgeries, and as well as to optimize as its best the one that is still required.

The surgical staff is the main target for the radiation exposure reduction work as they are exposed to it daily, whereas the patient is exposed only for the very few surgeries it will go under. However, the radiation exposure can also be of concern in certain category of persons such child where the FDA also recommends following the ALARA principle [123].

C-arm Collision

C-arm devices are cumbersome objects capable of ample motion at their 5 joints. In a cluttered environment such as the OR, as shown in Figure 1.6, the patient and its table, the surgeon and the surgical crew, surgical lights, table holding surgical tools are all present on the C-arm path and be crashed in case of C-arm motions. The C-arm motion is either left at the discretion of the surgical crew that must then take care that collisions will not happen when moving the C-arm or performed robotically. In both cases, for the safety of the persons present in the OR, the potential collisions of the C-arm with its environment must be detectable beforehand to be prevented.

1.3.2 Lack of Context

Even though already valuable for the doctor, the X-ray image is often shown on a screen with no relation with its context and environment. Figure 1.6 shows a common situation in the Operating Room, with the surgeon is looking up at the screen the X-ray image while the patient is down its sight. The only context is the C-arm position compared to the patient at the time of acquisition which is at the best robotically placed or in the worst case, manually positioned by the surgical staff at the order of the surgeon. In Figure 1.6, the surgeon dictates the nurse how to move the C-arm to acquire its desired views as the controls are on the left side of the patient table (red circle). The patient-to-C-arm is a difficult mental exercise for the surgeon that needs to relate the 3D environment to the final desired 2D image. The lack of external context can lead to wrong or imprecise C-arm positioning, requiring numerous C-arm repositions and therefore X-ray images until the desired position. In the scene depicted in Figure 1.6, several iterations of orders and executions were necessary to reach the surgeon desired view. Moreover, the content of the X-ray image itself also lacks context and its localization in the surgeon's environment and actions heavily rely on the experience of the surgeon to perform the mental correlation to the surgical area and actions. Lack of internal context also leads to multiple X-ray images acquisition in order to help the surgeon to relate its surgical actions and their consequences on the anatomy only visible through X-ray imaging. In the case of the angioplasty procedure depicted in Figure 1.6, a catheter is passed through the patient artery and only a new X-ray image can allow the surgeon to track the catheter path along the artery until its target. This requires an important amount of X-ray images as new X-ray images are shot at high frequency in fluoroscopy mode to follow the surgical actions.

1.3.3 X-ray Image Interpretation

X-ray images are a 2D accumulative projection of a complex 3D volume with numerous organs of different densities overlaid. Although being most of the widely used medical modality, X-ray images still require time, training, and experience from the surgeon to develop the perceptual and cognitive skills to search for the desired information and how to interpret that information [34]. The interpretation of the X-ray image from the 3D real data to its representation into a 2D image is a difficult mental mapping, localization context from another modality can help the interpretation in the 2D image plane, however the third dimension along the X-ray beam held most of the ambiguities due to the accumulative nature of the X-ray image, for example to perceive the right ordering of the structures present in the image. Lifting the ambiguities in the projective direction would help the interpretation of the X-ray image.

1.3.4 Current Context and Objectives

Those challenges are not solved yet in the manner surgery happens every day. Often, the surgeon only disposes of the C-arm as tool which means little to none help to retrieve the context or the interpretation of the X-ray image. This leads to numerous X-ray images being acquired in order to the surgeon to perceive and interpret the X-ray image, which is far from the target goal recommended by the FDA. As we have seen before, the innovation necessary to improve on those issues cannot wait until the purchase of the new state of the art device. Supplementing the C-arm device already present in the OR via “economical” and minimally workflow disruptive technology is a mean to integrate into a faster manner innovation in the OR such as the surgeons can improve on the challenges described earlier. There are different directions to take for performing so: by software, by hardware or by hybrid software-hardware supplements. In the scope of this thesis, we will only describe the hybrid methods for reasons discussed in the State of the Art Chapter 2.1, the works described in this thesis also fitting this category. At the end of the State of the Art Chapter, we will compare the objectives described in this chapter regarding “economical” and minimally workflow disruptive technology versus the different works/directions existing in the literature and we will draw which tracks fit better those objectives.

2 State of the Art

In the previous chapter, we have explored the challenges given by C-arm based surgeries. Since the introduction of C-arm in the Operating Room, the research community has investigated its full potential and has for this purpose explored in detail how to supplement it in order to optimize the best outcome for the patient, surgeon, surgical crew, and hospital.

2.1 Introduction

Supplementing a C-arm can take three main directions: via software, via hardware and via a hybrid combination of both. Hardware supplement, in the sense of a physical object brought to the OR to help on the surgery, includes for example the use of newer surgical tools or new modality imaging such as ultrasound. This type of supplement brings new possibilities and perspectives that would be impossible with a C-arm only. The direction of hardware-only is not going to be explored in this thesis as it rapidly reaches its limits when not fused to the C-arm data to tackle the challenges described earlier. Software-only supplement is a full field of study on its own that tries to maximize the acquired data in order to present and extract the most information as possible to the surgeon given the existing data. The workflow can be modified but no hardware is added to the Operating Room (except the computer to process the data). The limitation in that direction is that no new perspective can be bought from the data, it only optimizes on the seen data. We will not detail in that direction since we are interested in works that both combine hardware and software in order to complement C-arm, providing the best of both hardware and software worlds: bring a new perspective and optimize the acquired data. We will, therefore, describe in this chapter the works supplementing C-arm using hybrid combinations of software and hardware. We have classified the different works according to the aforementioned challenges that they address: safety in the OR, X-ray image contextualization, and interpretation. Along the section, we will explicit the hierarchy of the different works we present under the form of a hierarchical tree. In Figure 2.1, we show an overview of the classification of the different works. For some categories, we will divide furthermore the classification.

2.2 X-ray Image Contextualization

Earlier in the Motivation Chapter in Section 1.3.2, we have made the distinction between the internal and external context of the X-ray image. This distinction can also be found in the works supplementing C-arm. We will, therefore, first describe the works allowing to give the external context of the X-ray image, by tracking the C-arm or computing the relationship C-arm to patient. Then, we will describe the works that give internal context to the X-ray

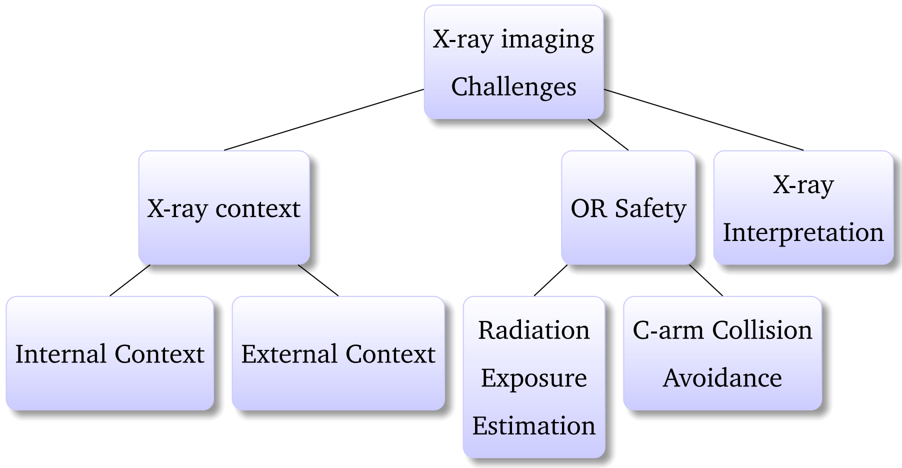


Fig. 2.1. Hierarchical tree of Challenges with X-ray images

image. All of them use another modality, which can be of pre-operative or intra-operative nature. Numerous works can provide at the same time external and internal context of the X-ray. For those cases, we classify the works according to the main use attended by the author.

We will enter many sub-categories of the state of the art works concerning X-ray context and for clarity of presentation, we show in Figure 2.2 the hierarchical tree on those works categories that we subjectively choose.

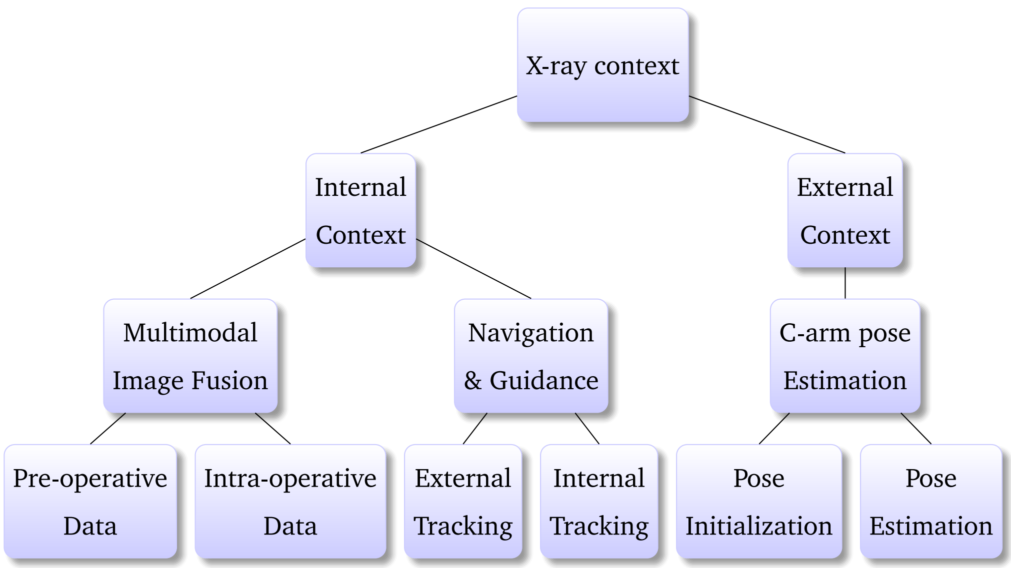


Fig. 2.2. Hierarchical tree of the works contextualizing X-ray images

2.2.1 External X-ray Context

The external context of the X-ray image is the knowledge of the position of the C-arm into its environment (to patient, to the table, to the surgeon, etc.). The works can target to recover this position by performing C-arm pose estimation or to assist the surgeon to define the next pose given some constraints on radiation exposure or structure visibility as we will see in the next section.

C-arm Pose Assistance

Positioning the C-arm to target the desired structure requires experience and planning from the surgeon. Often, the surgeon will use pre-operative data such as CT to mentally compute the C-arm poses required during the surgery. Fallavollita et al. [41] bridge the gap between the planning and the C-arm positioning by providing an image-based C-arm positioning mobile device interface. The surgeon using the pre-operative CT can observe its desired X-ray image, generated using Digitally Reconstructed Radiograph (DRR), to further use during surgery and the corresponding C-arm poses are computed with inverse kinematics. Figure 2.3 compares the views and poses chosen by the surgeon on the mobile device compared to the ones during surgery, showing that they are similar.

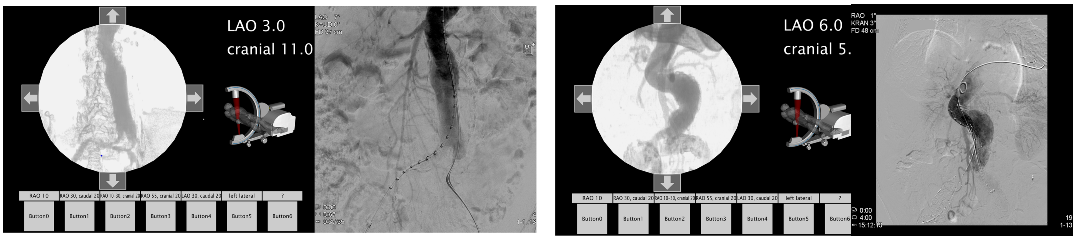


Fig. 2.3. Comparison of the view chosen by the surgeon with the similar C-arm view acquired during surgery, reproduced with permission from [41], ©Springer

Rodas et al. [134] aim to output the C-arm pose that minimizes the radiation exposure of the surgical staff while maintaining internal structure visibility in the image. Using the setup described later in this chapter in Section 2.4.1, they record the current OR situation such as surgical staff positioning and input this in a cost function parametrized on the C-arm joints, which aims at minimizing the radiation exposure given the OR situation. The output is the DRR image of the optimal pose, the latter pose can be used by the surgeon if the DRR is suitable.

C-arm Pose Estimation

All the works targeting to estimate the C-arm pose do it with respect to a referential, which can be the patient, fiducials reference or initial pose. Therefore, C-arm pose estimation encompasses multiple end-goals that we retrieve in the different works: 3D reconstruction, patient to C-arm positioning. The final application is one criterion of division among the works, however, we present those works according to another division criteria, which is if their technology is markerless or not.

Marker-based Tracking

Using fiducials in order to recover the 3D pose of the C-arm has been explored for decades. The first class of algorithm relies on fiducials placed in a 3D known configuration. When imaging this configuration, the 2D-3D correspondences between the fiducials on the 3D configuration and on the 2D images need to be computed. Then, depending on the nature of the configuration (planar or not), different computer vision algorithms can be used to recover the pose such as the standard camera calibration algorithm from Zhang et al. [186] or the Direct Linear Transform. The complicated task is to recover the 2D/3D correspondence. To perform automatic pose estimation, the following works have encoded their configuration to simplify the matching. Navab et al. [112] use fiducials placed on a cylinder shape following a codeword pattern around the patient head to recover the poses of several cerebral angiography images at different angulations and then, perform 3D reconstruction of the cerebral vessels. To also recover the 3D pose of radioactive seeds during brachytherapy, Jain et al. [64] also design an encoded fiducial object shown in Figure 2.4, with a unique representation at every viewpoint. Kainz et al. [69] place a plane of fiducials under the patient with X-ray fiducials placed in a unique configuration shown in Figure 2.5 in order to recover the pose of C-arm. Steger et al. [155] use projective invariant cross-ratio to recover the correspondences and apply the recovered pose during bronchoscopy by overlaying the 3D visualizations of airways (segmented from pre-operative CT) over X-ray image. Amini et al. [6] also use a unique planar pattern of fiducials in combination with IMU (a system that we will describe in the markerless tracking section) to track the C-arm in comparison to the patient, with the final goal to measure on the patient the distance between two patient anatomical landmarks.

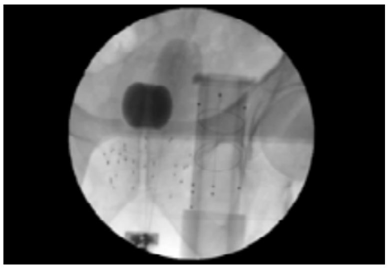


Fig. 2.4. Encoded fiducial object to recover 3D pose of radioactive seeds during brachytherapy from Jain et al. [64], also used by Fallavollita et al. [40], reproduced with permission from [40], ©Springer

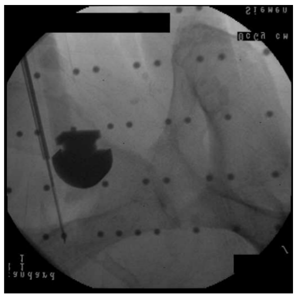


Fig. 2.5. Plane of fiducials under the patient with X-ray fiducials placed in a unique configuration, reproduced with permission from [69], ©Springer

Fallavollita et al. [40] use a different, image-based technique to perform the 3D/2D matching with the goal of radioactive seeds 3D reconstruction. It uses the 3D shape of the fiducial object designed by Jain et al. [64], shown in Figure 2.4, to create a DRR of this object from the estimated pose. This is compared to the real X-ray image containing the fiducial object using an image similarity metric. This process is repeated through optimization over the pose converging to the generated DRR fitting the X-ray image.

The second class of algorithms still relies on markers but also introduces another modality in combination with the X-ray image to track those. Wang et al. [168] use a video camera mounted on the C-arm, known as the CamC system which we will describe later in more details with its original publication by Navab et al. [111] in the Section 2.2.2. They use this video camera to track AR markers placed under the patient table and, therefore, compute the C-arm pose with respect to the patient table. In the end, the aim is to perform the stitching of X-ray images. Fuerst et al. [44] use the same video system to track the laser, often present on newer C-arm to help patient positioning, with SURF feature detection to perform CBCT volume stitching.

Markerless Tracking

Several works recover the C-arm pose with markerless techniques. A large part of those works uses sensors directly affixed on the C-arm such as accelerometers or Inertial Measurement Unit (IMU) composed of gyroscope, accelerometer, and magnetometer. The former only recovers angle while the latter can recover a full motion (rotation and translation). All those devices use mechanical motion compared to the Earth gravity to recover the Degrees of Freedom. Grzeda et al. [50] were the first work to use accelerometers to recover the C-arm rotations. Later, this work has been used for the case of brachytherapy [180]. They could show that their tracking was more accurate than the built-in sensor of the C-arm. Amiri et al. [7] use IMU to recover the C-arm pose, they report the accuracy of their system for several applications such as 2D–3D registration, 3D landmark localization, and for panoramic stitching. Moataz et al. [107] also use IMU sensors combined with a custom-made engine system to perform CT-like 3D reconstruction, the full device construction is shown in Figure 2.6.

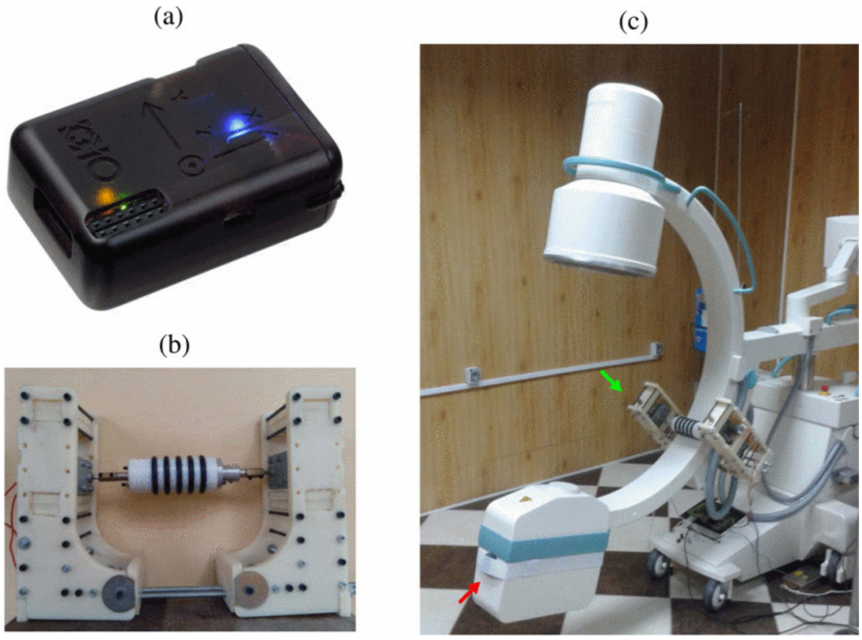


Fig. 2.6. (a) IMU sensor (b) engine system (c) C-arm with IMU sensor attached (red arrow) and engine system (green arrow) from [107] ©[2016] IEEE

Another class of works uses active infrared light for recovering the C-arm pose. Schaller [142] use Time of Flight camera attached on the C-arm to position the patient with respect to the C-arm. Using the ToF camera, they recreate the patient surface which they split into anatomical regions. For each region, they compute the region iso-center. For patient positioning, this region iso-center needs to be aligned with the C-arm iso-center. They compare their work to manual positioning, which will only roughly estimate the region iso-center.

2.2.2 Internal X-ray Context

The internal context of X-ray image consists of placing the anatomical structures present in the X-ray image into their respective environment. This is possible using information coming from an external source such as another modality fused with the X-ray image using techniques that will be described first in this section. The main goal is to provide the surgeon with a visualization that will reduce the mental exercise of structure localization with another modality more informative on the localization. A subset of works goes further than the multimodal image fusion and provides to the surgeon navigation and guidance to the target area. We will describe those works in a second time.

Multimodal Image Fusion

We first describe the works fusing X-ray data with another modality in order to provide context to the anatomical structure in the X-ray image. We will separate the works into classes according to the nature of the other modality (intra-operative or pre-operative), each class deals with different problematics. Inside each class, we will discuss how the different modalities are registered as well as their visualization.

Using Pre-Operative Modality

Often, pre-operative information is acquired in order for the surgeon to better plan the surgery. Modalities such as CT, CBCT, or MRI give valuable information about the 3D anatomy. However, most of those modalities are very complicated to acquire intra-operatively due to their high cost regarding the equipment, and also due to radiation (for the CT and CBCT). CBCT can be treated as CT in terms of the registration/visualization challenges as the modalities are similar, however, CBCT is often acquired pre-surgery (shortly before the surgery starts) thanks to CBCT-enabled C-arm which can change the registration workflow as we will observe in the next paragraph.

Works investigating the fusion of pre-operative data with the intra-operative X-ray image are a good comprise in order to use this precious anatomical information from pre-operative data intra-operatively. However, to be of interest, the pre-operative data needs to be placed in the coordinate system of the intra-operative modality. The registration can also allow updating the pre-operative data to current anatomical situation which can deform during surgery. First, we will look at the different registration techniques from the pre-operative data to the X-ray image, then we will look at the different visualization paradigms used to fuse them.

Registration

Markelj et al. [102] provide an extensive review of the works performing registration from pre-operative data to 2D intra-operative X-ray image registration. They report three types of modality, all of them 3D: CT (and by extension CBCT), MRI, and anatomical model (statistical or geometrical). Markelj et al. first report a classification criterion based on the use of fiducials. We will describe the works according to this criterion. They also use a classification criterion looking at the dimensionality relationship: projection technique (projecting the 3D modality to 2D), back-projection technique (bringing the 2D modality to 3D) or reconstruction technique (using multiple 2D images to bring the 2D modality to 3D).

The first mean of registration is to use fiducials placed in the images. This type of registration technique, qualified as extrinsic by Markelj et al. [102], requires stereotactic frames [68] or fiducials (placed on the bone [45], soft tissue [148] or skin [152]) to be present in both modalities at the same location. The three types of dimensionality relationship criteria have been used for extrinsic registration. Shirato et al. [148] project the detected 3D markers to 2D which allows to perform the registration in 2D. Back-projection was used by Tang et al. [157] where rays are created from the detected 2D markers in the image to the X-ray source. The registration is performed by minimizing the distance between the rays and the 3D markers. Finally, the multi-view reconstruction is used by Litzenberg et al. [92] where the 3D reconstruction of the 2D markers is computed from several X-ray images and compared to the 3D markers from the pre-operative modality. This technique is also used by Ma et al. [99] to register MRI with X-ray images using a catheter as fiducial. The catheter is considered here as a fiducial as it is inserted for that purpose only. The drawback of the extrinsic registration methods is that the fiducials must be present in both modalities, which is problematic for pre-operative data which can be acquired a long time before the surgery such as CT or MRI.

The second type of registration reported by Markelj et al. [102], qualified as intrinsic, use the information from the data to perform the registration. Markelj et al. report the main classes of intrinsic registration algorithms: feature-, intensity-, and gradient-based methods. Feature-based works extract 3D features from the pre-operative data to register them to corresponding 2D intra-operative features. The features can be points, curves or surfaces. Point feature based works often include anatomical landmarks identifiable in both modalities [13]. The main issue is, however, the segmentation and matching the correspondences between the two modalities. Correspondence search can be simplified by performing curve-to-curve feature registration or 3D surface-to-curve. Curves segmented in 2D can be back-projected to 3D using the previously mentioned ray technique and compared with a 3D surface [84] using distance minimization. Surfaces from 3D can be projected to 2D and the resulting curve compared to the curve segmented feature in the X-ray image [187]. The main issues of such techniques remain that it depends highly on the segmentation accuracy. Intensity-based works apply mainly to CT to X-ray registration. Projective intensity-based techniques are an intensively researched topic where the CT data is projected to 2D Digitally Reconstructed Radiographs (DRR) images, which are then compared to the 2D X-ray images using image similarity metrics. Numerous of those metrics have been investigated: Normal Cross-Correlation [72], Mutual Information [73], Sum of Squared Differences [43] for the most common of them. Few works performed reconstruction in intensity-based works, in this case, multiple X-ray images are used to fill a CT-like 3D reconstruction that is compared to the CT using a 3D similarity metric [159]. Intensity-based registration methods are possible in between X-ray and CT due to their interleaved nature and common origin. This method, therefore, does not work well for MRI to X-ray which both are of very different nature. Van der Bom et al. [19] develops a machine-learning method that generates a pseudo-CT data from MRI, this pseudo-CT can be used as a CT in the previously mentioned works in order to register the MRI with the X-ray image. Finally, works using gradient-based methods use the fact that edges (described by the gradient map and volume) are both present in the 2D and 3D modalities, including in MRI. In projective gradient-based works, a 3D gradient volume is projected and compared to the X-ray 2D gradient map [176]. In back-projected works, they use the fact that rays from 2D edges should be tangent to surface in 3D, therefore, should pass through the maximum in the 3D gradient volume [160].

Beyond the direct registration from the pre-operative modality to X-ray, several works use an intermediate modality such as a CBCT acquired shortly before surgery. The pre-operative modality such as MRI, CT or PET is registered to the CBCT which is then by design registered to the X-ray image as we will discuss in the next section. The commercial system XperGuide from Philips[1] allows performing such registration and several works can be reported for CT, MRI [89] and PET [1]. The registration between the pre-operative modality and the intermediate modality is performed manually using fiducials or anatomical landmarks.

Visualization

Once registered, the pre-operative data can be visualized in different manners on the intra-operative X-ray image: planned path, segmented organs, segmented slice. Often, 3D pre-operative data such as CT, PET or MRI is used to plan a surgical path to follow during surgery. The planned path, known in the reference system of the 3D operative modality, can then be projected and visualized on the intra-operative X-ray image during surgery. Abi et al. [1] use the combination of the pre-operative PET with the CBCT to compute the surgical path to be displayed during the surgery. This type of visualization is often found in navigation systems, which we will describe later. A visualization that provides more context consists on displaying segmented organs onto the intra-operative X-ray. The full organ can be overlaid such as in the commercial system EP Navigator[2] from Philips that we show in Figure 2.7.

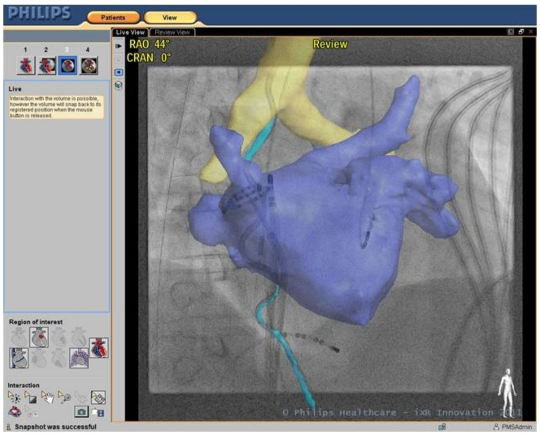


Fig. 2.7. Overlay by transparency of segmented heart model on the X-ray image in EP Navigator from Philips, by Philips Communications, distributed under CC BY-NC-ND 2.0 license, Link to Image

A processed version of the segmented organ can also be overlaid such as performed by Ma et al. [99] who segmented the heart model from MRI, that is then split into 16 standardized segments. The split model is then overlaid by transparency to the X-ray, every segment is color-coded for context. Using a shortly acquired cardiac MRI before surgery, Behar et al. [9] colorize the standardized segments according to the target location of the affections (heart scar and dyssynchrony), facilitating the guidance to those locations. The visualization can be observed on Figure 2.8.

The 3D nature of the pre-operative modalities causes that no work tries to overlay them entirely to the X-ray image, however, the XperGuide system from Philips provides the overlay of a CT or MRI slice over the intra-operative X-ray image. Van et al. [18] use this system to overlay an MRI slice and the intermediate modality CBCT slice over the intra-operative

1. http://www.philips.ca/healthcare/product/HCOPT06/xperguide-live-3d-image-needle-guidance
2. http://www.usa.philips.com/healthcare/product/HCNCVC419/ep-navigator

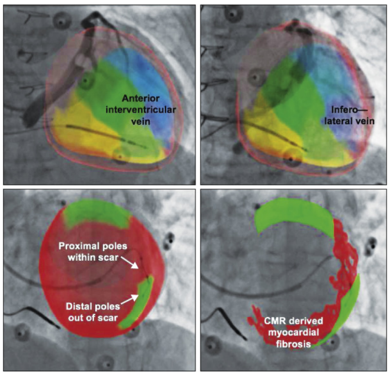


Fig. 2.8. (Top) Overlay by transparency of segmented heart model with standard colorized segment on the X-ray image, (Bottom) Colorization in green of the targets, from Behar et al. [9] distributed under CC BY 4.0 license, Link to Image

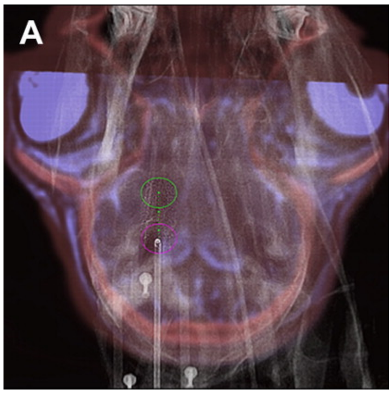


Fig. 2.9. Real-time x-ray images (gray scale) are overlaid on CBCT (red) and MRI (blue), the entry point (pink circles), target point (green circles) and planned path (green dots), from Van et al. [18] distributed under CC BY-NC 2.0 license, Link to Image

X-ray. In addition, the entry point, target point, and planned path are also segmented from the pre-operative data and added to the overlay as we show in Figure 2.9.

Using Intra-Operative Modality

While the pre-operative data was limited to 3D data, the diversity in the intra-operative data going from 2D ultrasound, nuclear imaging to video image, makes the challenges in terms of registration and visualization more extensive than with pre-operative data.

Registration

The first straightforward intra-operative registration is the by-design registration where the modalities are already directly in the same framework due to the construction of the C-arm or if the other modality only needs a one-time calibration not performed during surgery. The first case applies to CBCT devices which can produce both an intra-operative CT-like reconstruction and live X-ray images. Since they are created from the same device, their registration is known by design if the C-arm keeps the relative motion between the CT acquisition and the acquired X-ray image. Racadio et al. [129] and Leschka et al. [87] use the XperGuide device from Philips, previously mentioned, to combine CBCT and live fluoroscopy with no need of registration in between the two modalities. However, similar to pre-operative data, this does not handle patient motion during the surgery. By-design registration works also include systems requiring only a one-time pre-operative registration to physically align the modalities such as the CamC [111], where the other modality (video camera) is attached to the C-arm in such a way that the X-ray image and video image overlay, whatever the C-arm position and surgical scene. To help the registration of the 2D projective intra-operative data, a mirror construction is used in order to merge virtually the optical centers and the axis of both modalities. The camera and mirror are positioned once during a one-time calibration, therefore, during surgery, no registration is required. Another example is provided by Beijst et al. [10] which registers through a one-time calibration a C-arm and a gamma camera. The gamma camera is placed above the X-ray source such as the C-arm source, intensifier, and the gamma camera are aligned.

Beyond the by-design registration, the remaining works are mainly about another widely used intra-operative modality: ultrasound. However, the fusion applications are limited due to the very different physical representation and characteristics between ultrasound and X-ray imaging, the X-ray image being a 2D projective image while the 2D ultrasound image is a slice of the 3D world. A mean of registration is to use an intermediate modality such as mechanical, electromagnetic tracking calibrated to the C-arm or a pre-operative modality calibrated with the techniques described in the previous sections. Jain et al. [63] use electromagnetic tracking and Ma et al. [98] a mechanical tracker to co-register trans-esophageal echocardiography (TEE, heart ultrasound) with X-ray images. TEE consists of inserting an ultrasound probe into the heart; both works, therefore, track the probe with respect to the C-arm calibrated pre-operatively with the tracking system. Wieczorek et al. [178] use an intermediate modality by the mean of a calibrated pre-operative CT to display modalities such as a 2D ultrasound slice or 3D tool model (tracked with optical tracking system intra-operatively with respect to the pre-operative CT) on the X-ray image.

The two categories proposed by Markelj et al. [102] to classify pre-operative to intra-operative and previously explained in this section can also be applied to classify the intra-operative modalities to X-ray image: extrinsic and intrinsic registration. Extrinsic registration can be found for the aforementioned TEE application. Lang et al. [83] attach fiducials on the TEE probe visible in the X-ray, which enables the registration. The placement of the fiducials on the probe are known thanks to a micro-CT. Intrinsic registration has also been investigated for the TEE case by Mountney et al. [109] and Gao et al. [46]. Mountney et al. compute the probe 6DOF pose thanks to its shape in the X-ray image using discriminative learning techniques combined with template matching. Of course, this work takes advantage of the visibility of the probe in the image. Gao et al. use a micro-CT of the probe to perform intensity-based registration by computing DRR of the probe that is registered to the X-ray image.

Visualization

The intra-operative nature of the second modality leads that most of the works use the raw data or a simple processing of it for the fusion. Navab et al. [111] use the raw video image to overlay along the X-ray image. The overlay is performed by alpha-blending of the X-ray image over the video image as shown in Figure 2.10, the alpha blending parameter can be tuned by the surgeon to observe the X-ray image at different transparency levels.

Similar to the visualization of Navab et al., uniform blending is the norm for visualization. For the fusion of TEE with X-ray, Gao et al. [46] project the 3D echo (ultrasound) on the X-ray image and display it via colored uniform transparency as shown in Figure 2.11. Beijst et al. [10] overlay by transparency the activity map of the nuclear probe on the X-ray image as shown on Figure 2.12, this a processed version of the raw data as only the locations of highest activity are displayed.

Finally, the visualization provided by Racadio et al. [129] and Leschka et al. [87] combining CBCT and X-ray image is similar to the visualization combining pre-operative CT with X-ray,

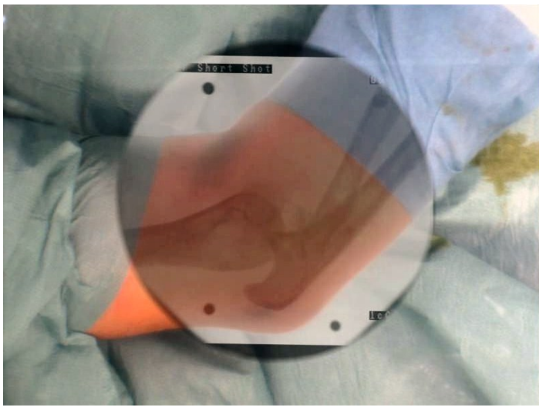


Fig. 2.10. Uniform overlay image of X-ray image over video image during an elbow surgery

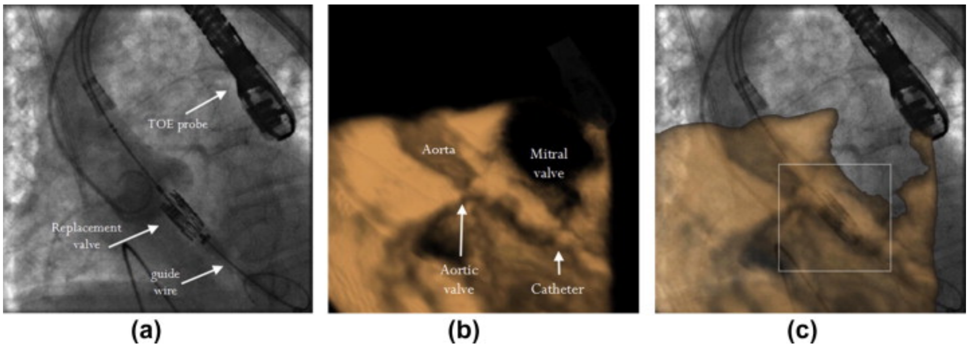


Fig. 2.11. Fusion of TEE with X-ray image, from Gao et al., reproduced with permission from [46], ©Elsevier

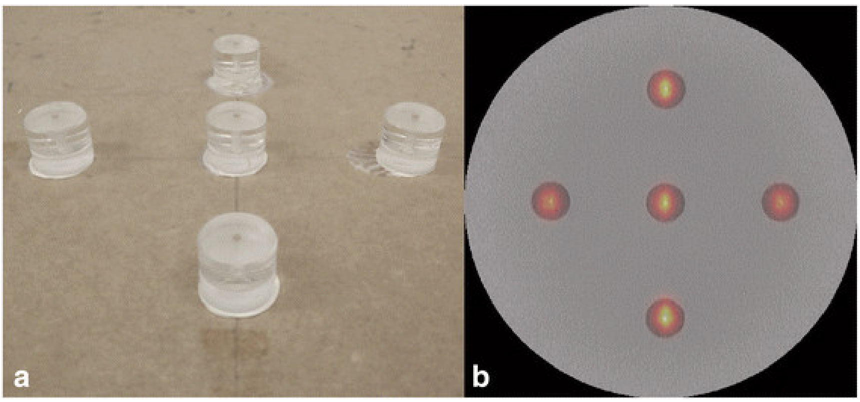


Fig. 2.12. On the right, overlay of X-ray image with nuclear activity of the phantoms presented on the left, from Beijst et al. [10] distributed under CC BY 4.0 license, Link to Image

they provide either the overlay of a CBCT slice over the X-ray image as already presented by Van et al. [18] and shown in Figure 2.13 or the overlay of the segmented 3D structure projected on the X-ray image.

The most advanced visualization is proposed by Wieczorek et al. [178] where they use information from the intermediate modality (pre-operative CT) to tune the blending value at each pixel between the additional modality and the X-ray image. The optical tracking registered with pre-operative CT allows to know the relative depth at every X-ray image pixel

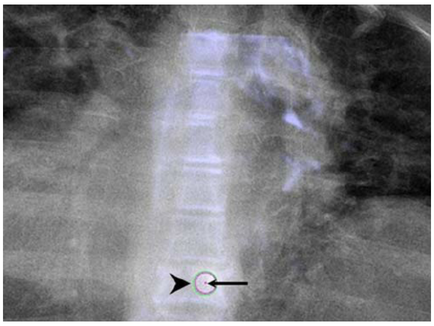


Fig. 2.13. CBCT slice overlaid over intra-operative X-ray image from Leschka et al. [87], entry point (purple) and target point (green) are shown as circles defining the 5-mm error margin, reproduced with permission from [87], ©Springer

of the objects in the additional modality, depth which is used to blend them accordingly to the X-ray image. Therefore, it improves the depth perception and allows to perceive the additional modality correctly over the X-ray image.

Figure 2.14 shows the difference in depth perception between their proposed visualization and the uniform blending visualization.

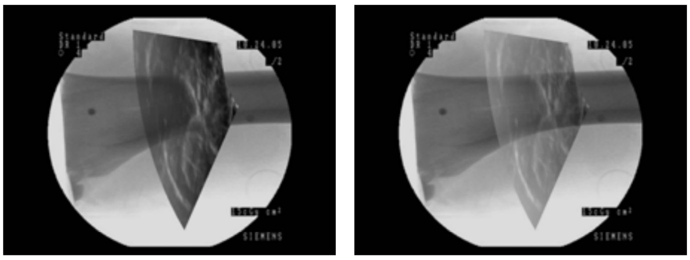


Fig. 2.14. Visualization from [178] to blend 2D ultrasound slice with X-ray image, (Left) Result the proposed visualization , (Right) Uniform blending, reproduced with permission from [178], ©Springer

Guidance and Navigation System

Navigation system builds on image-fusion works, which were already presented in this chapter. However, on top of providing the fusion of the image modalities, navigation systems also provide the tracking of surgical tools and integrate this tracking information to the visualization in the manner of a surgical GPS. This allows the surgeon to contextualize its current actions with the live X-ray image augmented with the pre-operative or intra-operative data visualization. The tracking can be external, based on external device such as optical tracking system often used in commercial navigation systems or electromagnetic tracking, or internal with the other modality used for the tracking of the surgical tool.

External Tracking

The surgical tool is tracked via an external device, calibrated along with the C-arm.

Optical Tracking

Optical tracking system is often found in commercial navigation systems, for this reason, the literature is quite extensive about this technology. However, the principle remains the same, an optical tracking system, composed of infrared cameras, is placed in the Operating Room. The C-arm, the surgical tool and potentially the patient (depending on the registration method used for the modalities data fusion) to be tracked are augmented with trackers composed of markers, which allows tracking them individually. The markers can be passive or active. Passive markers are spherical markers which reflect infrared light emitted by the LED light placed close to the camera and read again by those infrared cameras. Active markers emit infrared light activated by an electrical signal, and read by the infrared cameras [3]. Both technologies are very precise with a tracking accuracy around one millimeter [137], however, the main drawbacks of the technique are the line of sight issue, typical from outside-in tracking device as well as the cost of such systems. The goal of the navigation system is to track the surgical tool with respect to the pre-operative data and the intra-operative data already fused together in techniques explained earlier in this chapter. The C-arm motion is tracked with the help of the markers in order to place the tracked tool in the reference coordinate system of the data fusion with the ultimate goal to display the tool position on the data fusion visualization. The tool is represented through its 3D model, which is then projected on the 2D overlay view [87, 181] as shown in Figures 2.15 and 2.16.

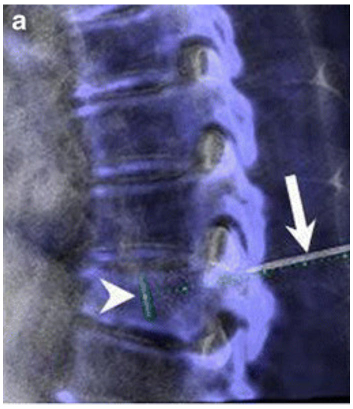


Fig. 2.15. CBCT slice overlaid over intra-operative X-ray image from Leschka et al. [87], with planned trajectory as green line, the needle model is also overlaid (arrow), reproduced with permission from [87], ©Springer

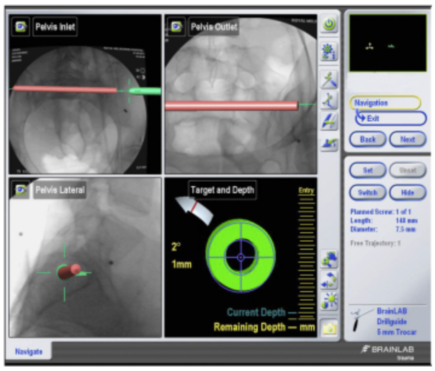


Fig. 2.16. Brainlab Vector Vision Navigation Interface with display of tracked screw (green) and planned screw path (red) from Wong et al., reproduced with permission from [181], ©Elsevier

Electromagnetic Tracking

Electromagnetic tracking consists of a magnetic field generator that induces a position-dependent current in sensor coils that are attached to the tracked objects. The main advantage of this technique is its absence of line-of-sight issue, which makes it ideal to track objects inside the human body such as catheters. The MediGuide commercial system [4] consists of a magnetic field generator attached to the C-arm detector and is dedicated to catheter tracking. Kircher et al. [75] presents the different visualization possible with this technology: catheter tip overlay as shown in Figure 2.17 or 3D model overlay of the heart and catheter.

3. https://www.ndigital.com/medical/products/polaris-family/
4. https://www.sjmglobal.com/en-int/professionals/featured-products/electrophysiology/navigation/advanced-navigation/mediguide-technology

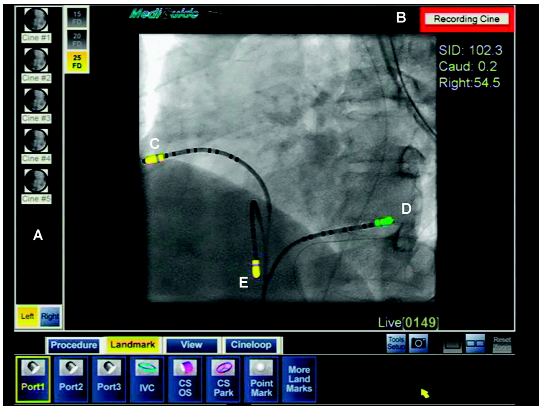


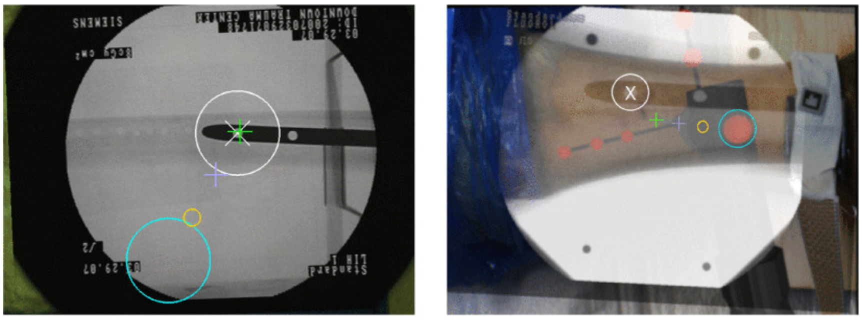
Fig. 2.17. Catheter tip tracked with EM (green and yellow dot) overlaid over X-ray image, reproduced with permission from [125], ©Wolters Kluwer Health, Inc.

Robotic Tracking

Robotic tracking goes further than the other types of external tracker and can replace the surgeon in the task of following the pre-operative planned path. Robots are calibrated to the C-arm using targets held by or attached to the robot and visible in the X-ray image. Then, 3D/2D extrinsic registration methods described earlier in this chapter can be applied to retrieve the relationship. Once calibrated to the C-arm and, therefore, placed in the coordinate system of the pre-operative data (using the X-ray image to pre-operative modality registration explained earlier), the robotic arm holding the tool can follow the path defined pre-operatively and reach the desired point on its own. Several types of robotic mount exist: to the bone [149], to the patient table [121] or to the ground [96]. Those works do not propose any fusion data visualization as the surgeon does not have to decide the navigation path. The advantage of robotic tracking is its accuracy while its drawback is its cumbersomeness in the OR and surgical area.

Internal Tracking

Without using any supplementary device, the additional modality can also be used for performing the tool tracking. Diotte et al. [35] use the video modality from the video augmented C-arm from Navab et al. [111] to provide, in addition of the video/X-ray overlay, the tracking of a drill used for distal locking of intramedullary nails. The tracking is possible thanks to a contraption whose cross-ratio are invariant in 2D, allowing tracking of the tip given the balls position tracking. The overlay is, therefore, enriched of the tool top and the tool tip as shown in Figure 2.18. This visualization is useful for the procedure as it requires down-the-beam alignment of the drill, meaning alignment of the tool axis to the X-ray source optical axis. With the proposed visualization, the surgeon only needs to align the tool top and tool tip, proving the good alignment.



(a) (b)

Fig. 2.18. (Left) Tool tracking visualization on X-ray image only, (Right) Tool tracking visualization on the video-X-ray overlay, the user must align the big circles to be in down-the-beam position from [35] ©[2015] IEEE

2.3 X-ray Image Interpretation

The interpretation of the X-ray image from the 3D real data to its representation into a 2D image is a difficult mental mapping for the surgeon and lifting the ambiguities due to the accumulative projection along the X-ray beam would help the perception of the X-ray image by the surgeon. The internal context is one part of the key for interpretation, therefore, we refer the reader to Section 2.2 for the details of the work targeting this cue. Internal context will help the surgeon in the 2D plane of the X-ray image, so, in this section, we will describe the works helping to recover information in the third dimension orthogonal to the X-ray image plane.

The variety of work in this category is very small compared to the variety of context works and they require an additional modality to recover the orthogonal dimension. Wang et al. [169] co-register intra-operative and intravascular ultrasound (IVUS). The IVUS transducer is tracked in the X-ray image using an image-based detection and tracking algorithm which provides at every position in the artery the cross-sectional artery ultrasound slice, orthogonal to the X-ray image plane. No work provides a fused visualization of both modalities, however Wang et al. [169] are the closest to that by linking together the modalities by showing in the X-ray image, the position of the current IVUS slice displayed alongside as shown in Figure 2.19. The combination of X-ray with IVUS help to recover information about the vessel wall thickness and tissues in the 3D dimensions while keeping orientation trackable thanks to the X-ray image.

While the IVUS-X-ray combination only gives information at one pixel at the time, Aichert et al. [3], as well as Wang et al. [170], propose a more global visualization. Using a co-registered CT to an X-ray image, they encode the depth of the structure inside the X-ray image using color mapping. This improves the relative depth perception of the different overlaying structures in the X-ray image. We show in Figure 2.20 the color schemes as well as the order of several structures that the color-encoded X-ray image helps to perceive.

Going further, Albarqouni et al. [5] predict the depth information from a single-view X-ray image with no co-registered CT required. A deep-learning based approach, with input an X-ray

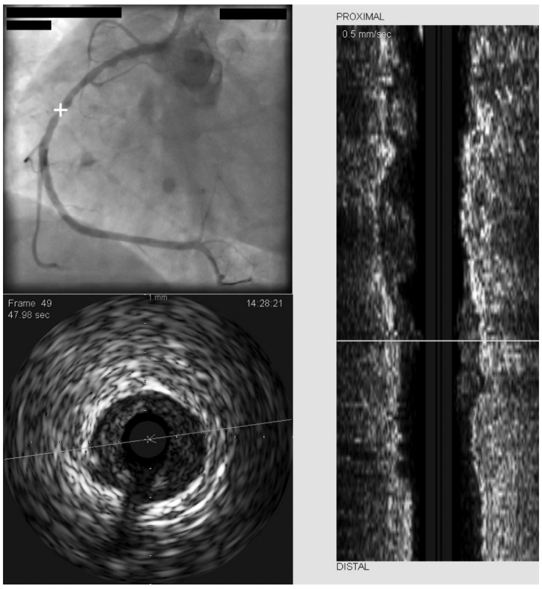


Fig. 2.19. (Left top) X-ray image with current position of the displayed IVUS slice as a cross (Left bottom) IVUS current slice (Right) Full IVUS scan along the artery from [169] ©[2013] IEEE

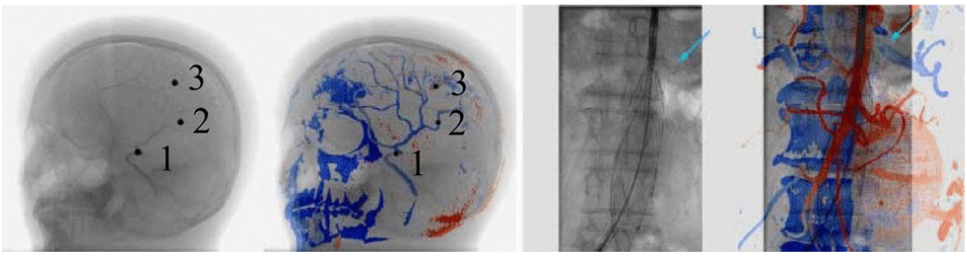


Fig. 2.20. Comparison of normal X-ray image with colored X-ray image for (Left) head vessel and (Right) the aorta from Wang et al. [170] ©[2014] IEEE

image and output the corresponding depth map, is used. The training data for such method is generated from available pre-operative CT (DRR images are generated for the input).

Also using a co-registered CT, Wang et al. [164] aims at visualizing the structures (such as aneurysms in the paper) occluded in the X-ray image due to the X-ray accumulation. For this purpose, they propose a mirror visualization that reflects back the occluded part of the arteries. The CT providing a 3D model of the artery as well as the mirror location are used to synthesize this mirror view. An example of this visualization is shown in Figure 2.21.

In the work of Navab et al. [111] that overlays an X-ray image over a video image of the surgical scene, the alpha-blending parameter is uniform all over the image, which can result in a non-natural result with the surgeon's hands being perceived as behind the patient's internal anatomy depicted in the X-ray image, even though it is physically the opposite as it is demonstrated on the left image of Figure 2.22. Later, Pauly et al. [122] replace the video camera by an RGBD camera and use the color and depth information in combination with Machine Learning to classify the different objects in the surgical scene such as the patient, the surgeon's hands, and the surgical tools. This knowledge allows defining specific blending values according to the class of a pixel in the image. The global overlay is then personalized

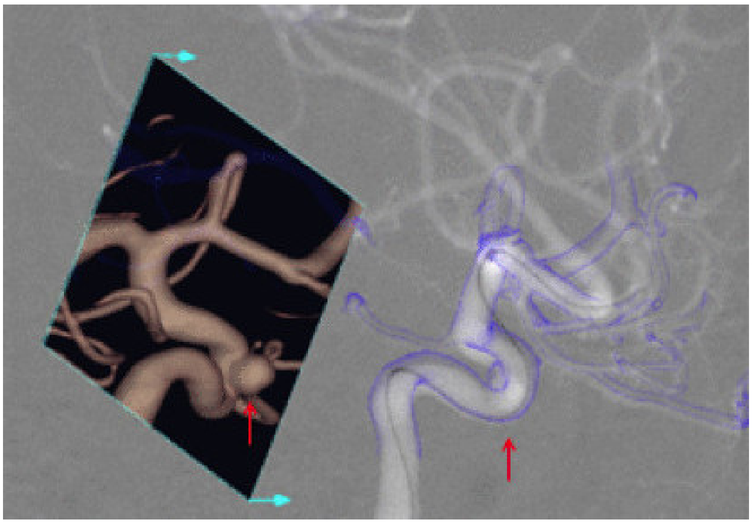


Fig. 2.21. Mirror visualization for displaying the occluded structures in the X-ray image such as aneurysm ©[2012] IEEE

according to the relevance of the different classes of objects in the image. In the end, this work provides a more natural overlay of X-ray image and video by allowing the surgeon's hands to be perceived over the patient by diminishing the blending value at the pixels representing the surgeon's hands, as displayed on the right image of Figure 2.22. Although the work does not provide depth information concerning the internal structure of the X-ray image, it provides depth perceptual information regarding the X-ray image by placing it perceptually at the correct relative order.

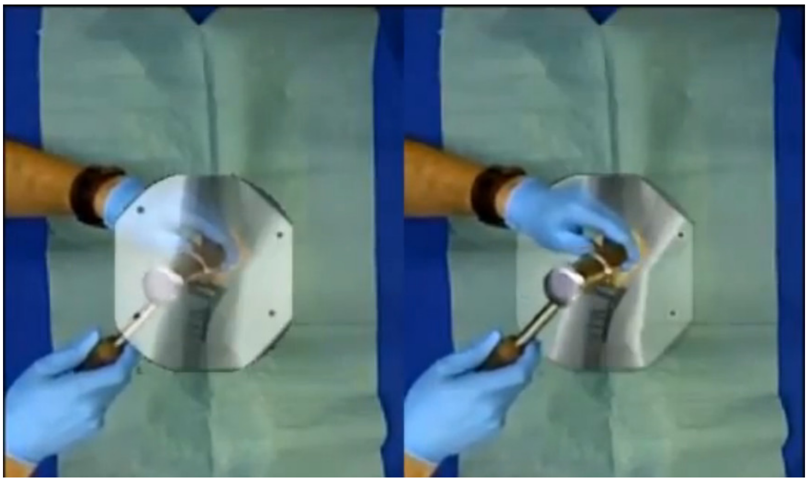


Fig. 2.22. (Left) Uniform blending of X-ray image over video, (Right) Relevance-based overlay

2.4 Safety in the OR

The Operating Room is a complex environment with numerous actors and devices present in the room. The C-arm is a critical device with multiple risks whose interactions with the surgical team and the environment must be monitored to avoid threats. The two main dangers of the C-arm are first the radiation that is emitted by the device and scattered through the OR to the full surgical staff, but also to the patient, and second, the C-arm collision to staff, to patient, or to surroundings.

2.4.1 Radiation Exposure Estimation

As explained in detail later in the Background chapter in Section 3.2.3, the radiation can be lethal in the long term. Several dispositions exist to reduce the exposure of the vital human organs such as lead aprons and lead glasses. However, the surgeon must often work with the hands in the beam, exposing heavily this unprotected anatomy. In the Section 2.2, we have previously explained how CAI works provide context to the X-ray image. Often, the additional information provided by augmenting the C-arm will reduce the number of X-ray images as the surgeon does not require additional X-ray images to understand the information inside the X-ray image and, therefore, reduce the radiation exposure. For example, navigation works explained in Section 2.2.2 will use either non-X-ray based trackers to track the surgical tool and as a consequence, the surgeon does not need to have hands in the beam while placing the surgical tool. Therefore, most of the works previously described in Section 2.2 have as a side effect (if not main target) to reduce radiation exposure. However, in this section, we will focus on the works that directly compute this radiation in order to sensibilize surgeons.

Multiple methods exist for the direct estimation of the physician's radiation exposure or the determination of the distribution of radiation combined with the physician's position in space. A pure mathematical method is presented in the work of Tsalafoutas et al. [162]. The distances and the scattering angles from the patient to various body parts of the surgeon and operating room personnel are estimated using a dosimeter affixed on the C-arm. However, this method loses precision as it does not take the actual position of the surgeon into account. An application by Bott et al. [22, 23] and Wagner et al. [163] deals with the radiation exposure estimation in a computer-based simulation program for C-arm training. The scattered radiation is simulated using Geant4[5], a toolkit for Monte Carlo Simulations, and visualized with pulsating spheres. Nevertheless, the actual radiation exposure of the surgeon cannot be evaluated as its position is not considered and the system is not designed for use in the ORs. In another work, the radiation exposure is estimated during a surgery [81]. Having 16 cameras mounted on the ceiling, the surgeon, and the patient can be tracked. The radiation is also simulated with the Geant4 toolkit using several detector spheres centered at the patient as the source of scatter radiation. Although this enables the surgeon to estimate their radiation exposure, the expensive setup with many cameras and several PCs makes it difficult to be directly deployed in an OR. More recently, Rodas et al. [133] combine the simulation of radiation propagation with the data obtained from wireless dosimeters. Three RGBD cameras mounted on the ceiling are used to generate a 3D point cloud representation of the scene. The radiation simulation is done with Geant4, where the radiation source, the flat panel detector, the table, and a water phantom, as well as eight dosimeters, are modeled. Due to the long computational time and the manual reading of the positions of the C-arm and the patient table, the system is not real-time. Moreover, the complex camera setup decreases its flexibility. Later, Rodas et al. [135] propose a mobile AR visualization of radiation exposure, shown in Figure 2.23 using a close setup from the one previously described. However, one of the RGBD

5. http://geant4.cern.ch/

cameras is placed on a hand-held screen. Using the RGBD data from the three cameras, the camera viewpoint is computed using ICP registration in a similar fashion to Kinect Fusion with the 3D reconstruction of the scene being updated and used for the next pose registration.

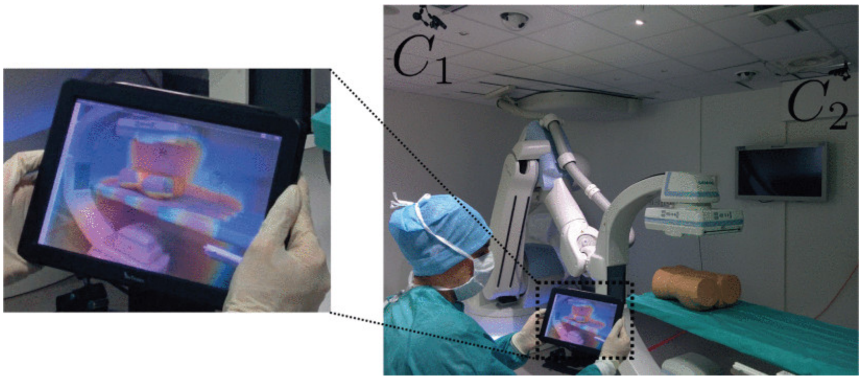


Fig. 2.23. (Left) AR Visualization of the radiation, (Right) System composed of two RGBD cameras fixed to the OR ceiling and a third one attached to the hand-held screen ©[2016] IEEE

2.4.2 C-arm Collision Avoidance

C-arm are cumbersome objects capable of ample motion at their five joints. In a cluttered environment as the OR where the patient along with its table, the surgeon, the surgical crew, the surgical lights, and the table holding surgical tools can be all present on the C-arm path and be hit by the C-arm in case of large C-arm motions. The C-arm motion is often left at the discretion of the surgical crew that must then take care that collisions will not happen when moving the C-arm. To help the C-arm maneuvering person, modern C-arms have built-in hardware for collision detectors such as pressure sensors on the source and detector housing [145]. Also, the C-arm kinematics software often presumes of the C-arm table and patient size to exclude those areas from the possible motion area of the C-arm [145]. In the literature, only one work can be found that includes the staff in the computation of collision avoidance. Ladikos et al. [82] perform a real-time 3D voxel-based reconstruction of the complete OR environment (surgical staff including) by mounting 16 cameras in the OR room. The bounding box of the different objects in the C-arm is computed from the 3D reconstruction and the intersection of the bounding boxes indicate the presence of collision. Figure 2.24 shows several examples of collision avoidance scenarios. For the third and fourth column, the person and the monitor are detected in the C-arm space and would potentially collide with the C-arm if this one is rotated.

2.5 Thesis Objectives

2.5.1 State of the Art Works in Light with Thesis Objectives

This chapter has shown the diversity of the works supplementing C-arm. We will now analyze this diversity in the scope of the theme developed in the Motivation chapter, where we discussed the necessity of introducing economical and minimally workflow disruptive

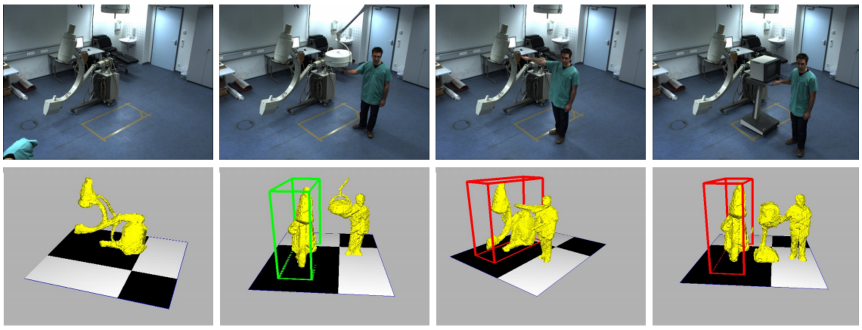


Fig. 2.24. Different scenarios of collision avoidance with a C-arm; first column with the reconstructed C-arm, the second column with the C-arm in a safe state (green bounding box), while the third and fourth columns with an object in the safety zone of the C-arm (red bounding box), reproduced with permission from [82], ©Springer

innovation at short term in the OR by supplementing C-arm in order to continuously improve the clinical impact. This chapter has exposed numerous directions to supplement C-arm, and we will extract which tracks are the most fitting the objectives: economical and minimal workflow disruption. We will perform this for several categories already used in this chapter: marker-based or markerless registration, pre-operative vs intra-operative additional modality, visualization paradigms, and single-purpose vs multiple-purpose technology. Based on the observations for each category, we will draw the perfect portrait of economical and minimally workflow disruptive hybrid supplement for C-arm devices.

Marker-based vs Markerless Registration

As marker-based registration works have shown, the use of markers is disruptive in the surgery. Fiducials need to be placed before or during surgery by the surgical staff and their appearance in the X-ray image, necessary for registration, is disturbing the content of the X-ray. Markerless registration relies only on the content of the imaging modalities to perform the registration. Reliability, robustness and accuracy are the strengths of marker-based registration, leading it to be still often used, such as in commercial navigation systems. It comes, however, at the cost of the disruption of the surgical workflow. The goal is to investigate markerless registration such as it gets as reliable, robust and accurate than marker-based registration without the burden of disturbing the workflow.

Pre-Operative vs Intra-Operative Additional Modality

Pre-operative data is often acquired prior to surgery. It is often of high-resolution and in 3D and can be of great use during surgery. However, this type of data and any of its derivative such as surgical planned path do not account for the current state of the surgery, where anatomies might move and even worse, deform. The surgeon risks to work with an outdated state of the anatomy, which puts the patient and the surgical outcomes at risk. Some work will use the intra-operative X-ray to update the pre-operative data to the current state of anatomy. Those works represent still a challenge in terms of registration, especially in the markerless case. Some works presented in this chapter which use an intra-operative additional modality (for example, the TEE works) still present the same drawbacks as pre-operative modality with the requirement to have the probe in the image to perform the fusion. In the end, the works providing the least disruption to the workflow are the ones based on intra-operative modality registered by design to the X-ray data where either no registration is required due to the C-arm construction or only one registration is necessary prior to surgery.

Single-Purpose vs Multiple-Purpose Technology

Numerous technology presented in this chapter can not be generalized to other purposes that the ones they were designed to, at most they can be used for different types of surgeries, close to the original target procedure. Optimizing C-arm for surgeries by very specialized supplement makes sense for surgeries performed at a high frequency. However, the search for cost-efficiency performed by hospitals drives them to rationalize the use of OR, which leads to a high turnover. Even inside a clinical field, the range of pathologies is wide and the OR must be ready to welcome this diversity. An economical supplement for C-arm devices must be as multi-purpose as OR is nowadays. If we take the example of tool tracking in navigation, multiple works require to use an external device for tool tracking. This single purpose device, even though it can be used for multiple procedures, clutters the OR even more, which can disturb the organization of the surgical team and affect their working conditions. Diotte et al. [35] reuse the additional intra-operative modality (i.e. the video image) to perform the tool tracking without overloading the OR even more for that purpose.

Visualization Paradigms

Almost all the visualization paradigms presented in this chapter propose the most elementary 2D multi-modal visualization, i.e. 2D uniform blending. Few considerations have been brought in the community to the validity of such overlays, which is as stated by Pauly et al. [122] often perceptually incorrect, not respecting the relative ordering of the different objects in the image. Uniform blending also overflows the image with all the available data, while maybe not useful as a whole. As a result, uniform blendings require an additional mental workload demanded to the surgeon to interpret those images correctly, which plays against this technology for a seamless integration in the OR. It is therefore important to follow the path started by Pauly et al. to investigate more complex and more perceptually accurate multi-modal visualization, using for example 3D or multiple layers.

2.5.2 Synthesis and Proposition

As analyzed in the previous section, the “perfect” economical and minimally workflow disruptive hybrid supplement for C-arm devices should preferably use as an additional modality a by-design registered intra-operative modality, should enable markerless registration, be multi-purpose and should have potential for enhanced multi-modal visualization. The Camera augmented mobile C-arm system [111] is one of the few system that can fulfill those criteria, strongly regarding the by-design registration and the multi-purpose criteria and softly for the 2 others. Indeed, this system possesses a few weaknesses regarding the markerless registration possibilities and the potential for visualization paradigms. Although the video can be used for markerless registration, all the Camera augmented mobile C-arm works performing tracking are using markers (AR markers for [168] or colored balls contraption for [35]). Video-based features used for registration are dependent of the scene content, especially its texture, to perform matching. This is a drawback for low-textured scenes such as surgical sites, especially in the case of minimally invasive surgery where the scene is mostly draped. Although the

Augmented Reality field has shown that multiple perception cues can be created from video images to perform a seamless multi-modal blending [94], they are still very scene dependent. For both cases, video images alone are not robust enough on low-textured scenes. RGBD cameras, that provide depth information along video imaging, allow overcoming this issue as depth-based features will rely on the scene geometry and can be used for low-textured scenes.

Therefore, this thesis investigates the use of RGBD cameras for augmenting C-arm and builds upon the Camera Augmented Mobile C-arm works performed by Navab et al. [111]. We will investigate how the depth modality can help the path towards economical and minimally workflow disruptive hybrid C-arm supplement.

2.5.3 Thesis Outline

This manuscript is organized as follows.

Part 1: Introduction We have already discussed the motivation of the works presented in this thesis as well as the state of the art works in the domain. In the remaining part, we describe the scientific foundations of the imaging modalities used in this thesis: X-ray imaging produced by mobile C-arm devices and RGBD sensing.

Part 2: RGBD Augmented C-arm Systems In this part, the mirror-based RGBD augmented C-arm is described in the first chapter. The feasibility of such setup is studied first through theoretical optics. Then, an empirical study to assess the validity of the RGBD data through mirror is conducted. In the second chapter, we focus on the mirror-less RGBD augmented C-arm, which is an alternative to the previous system providing similar output with minimal disruption on the C-arm housing. The new system is fully described, validated and also used in a pre-clinical study.

Part 3: Medical Applications of RGBD Augmented C-arm In this part, we describe the different medical applications which were explored during the span of the thesis and which are related to RGBD augmented C-arms. In the first chapter, we describe a new visualization paradigm overlaying the X-ray image over 3D reconstruction of the surgical scene using a depth-based C-arm to patient registration. In the second chapter, we present an augmented reality visualization overlaying the radiation exposure directly on the surgeon in an AR fashion, in order to sensibilize them to risks taken during such procedure. In the third chapter, we propose a multi-layer visualization paradigm building on the multi-camera background recovery capacity of the mirror-less RGBD augmented C-arm. In the fourth chapter, we use an RGBD augmented C-arm to provide an assistive tool during minimally-invasive scoliosis surgery by measuring the back deformation in real-time. Finally, in the last chapter, we present a mixed-reality setup for C-arm based surgery that mixes patient-based 3D printed anatomy and simulated X-ray image to create a realistic C-arm environment for training and new technology assessment such as the RGBD augmented C-arm.

Conclusion Finally, we draw the conclusions of this dissertation, and we propose perspectives for further developments.

In the appendix of this thesis, the reader can find the abstracts of the publications corresponding to other contributions made and not discussed herein, because they are out of the topic of this dissertation, as well as the publications list.

3 Background

In this chapter, we will describe extensively the scientific foundations of the imaging modalities used in this thesis: X-ray imaging produced by mobile C-arm devices and RGBD sensing. Both imaging modalities build on the same model for the image generation, namely the pinhole camera model, which we will describe first. Afterwards their specificities are described individually.

3.1 Pinhole Camera Model

The pinhole camera model finds its foundation in the oldest and simplest camera type: the pinhole camera. The pinhole camera is a simple light proof box represented in Figure 3.1 with a very small hole in the front, also called an aperture, and light-sensitive film paper laid inside the box on the side facing the aperture (image plane). When opened, the aperture lets the light rays which are reflected or emitted by the object and passing through it print on the film at the image plane. Ideally, per 3D point, only one ray should go through the aperture in order to obtain a focused image, which means the smaller the aperture is, the less blurred the image is. As a simpler representation than the physical image plane, the literature often refers to the virtual image plane which is the equivalent of the image plane but placed in between the object and the aperture, as shown in Figure 3.1 as the red plane.

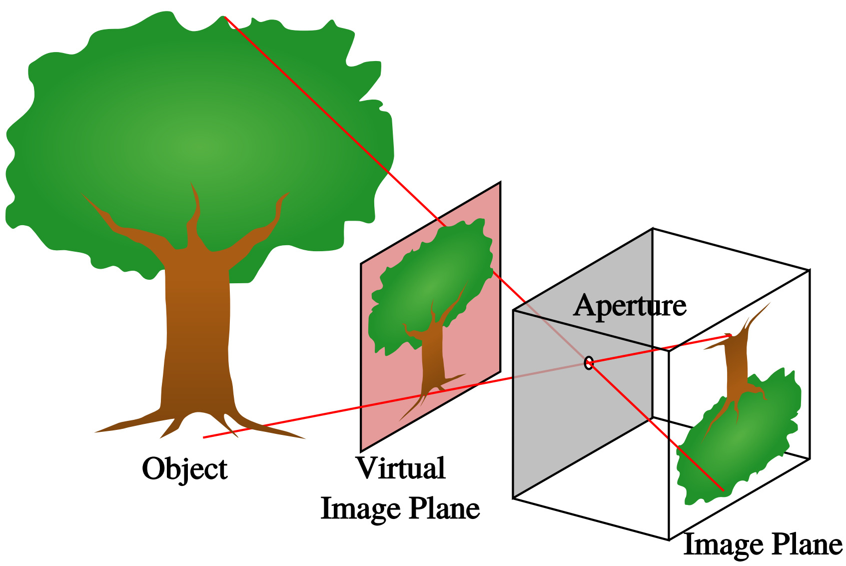


Fig. 3.1. Pinhole Camera concept

The pinhole camera model is the mathematical abstraction of this camera, where the aperture is left at its smallest possible size, a point. Therefore, only one ray by 3D point can go through the aperture. The pinhole camera model defines the geometric relationship between a 3D point $P = (X, Y, Z)$ and its unique 2D corresponding projection onto the virtual image plane $p = (u, v)$. The geometric mapping from 3D to 2D is then called a perspective projection. We denote the center of the perspective projection, the aperture, as the optical center or camera center C and the line perpendicular to the virtual image plane passing through the optical center, as the optical axis, which intersects the virtual image plane with the optical axis at the intersection point called the principal point $p_0 = (u_0, v_0)$. The distance between the virtual image plane and the optical center is called the focal length f.

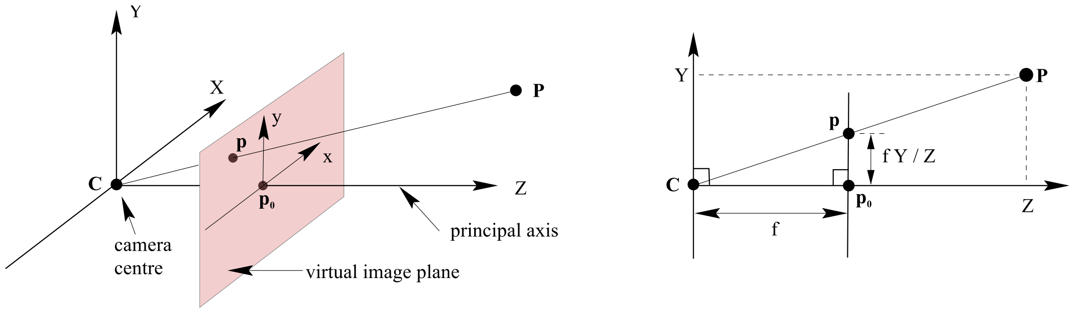


Fig. 3.2. Mathematical Representation of the Pinhole Camera Model, reproduced and modified with permission from [57], ©Cambridge University Press

In a perfect sensor, the projection from 3D to 2D can be mathematically represented by the equation 3.1.

$$\begin{pmatrix} uw \\ vw \\ w \end{pmatrix} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.1}$$

However, the image sensor can have imperfections such as skewed pixels or non-square pixels which are respectively described by the parameters s and m in the modified camera projection equation, giving Equation 3.2

$$\begin{pmatrix} uw \\ vw \\ w \end{pmatrix} = \begin{pmatrix} f & s & u_0 \\ 0 & mf & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.2}$$

In modern CCD sensor such as the one used for X-ray imaging on C-arm or for recent video camera sensor, those imperfections are negligible and we will suppose for the rest of the thesis that $s = 0$ and $m = 1$. The camera model, therefore, follows the Equation 3.1. The projection matrix is only depending on internal parameters and is, therefore, designed in the literature as the intrinsics parameter matrix K.

The 3D points imaged by the camera are usually not known in the camera coordinate system. The rigid transformation between the object coordinate system and the camera coordinate system is called the camera pose and is mathematically represented by the extrinsics parameters matrix T composed of the rotation R and the translation t. Therefore, the general relationship between a 3D point P and its 2D projection p in the image plane can be represented in the Equation 3.3.

$$\begin{pmatrix} uw \\ vw \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{K} \underbrace{\begin{pmatrix} R & t \end{pmatrix}}_{T} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.3}$$

Both RGBD sensing and X-ray imaging obey the pinhole camera model and, therefore, we can use the Equation 3.3 when dealing with those imaging modalities. However, X-ray imaging differs slightly from the video camera concerning where the rays are emitted from.

3.2 X-ray Imaging on Mobile C-arm

C-arm devices are radiography devices used during surgery to visualize the internal anatomical structure of the body. This internal imaging is possible thanks to X-ray electromagnetic radiation, that will describe first. Then, we will focus on its application on C-arm devices.

3.2.1 X-ray Radiation

The use of X-ray imaging in medicine has been rapidly investigated after the discovery of the X-ray radiation by Wilhelm Röntgen in 1895. Indeed, he imaged his wife's hand, taking the first medical radiography in history, shown in Figure 3.3.

Fig. 3.3. First X-ray image of Röntgen's wife hand, by Ulflund, public domain

The X-ray radiation is the electromagnetic radiation composed of high-energy photons with wavelength comprised between $10\,\mathrm{pm}$ and $10\,\mathrm{nm}$. The X-ray radiations used for medical imaging, referred as hard X-ray, are usually comprised between $10\,\mathrm{pm}$ and $100\,\mathrm{pm}$ as shown in Figure 3.4.

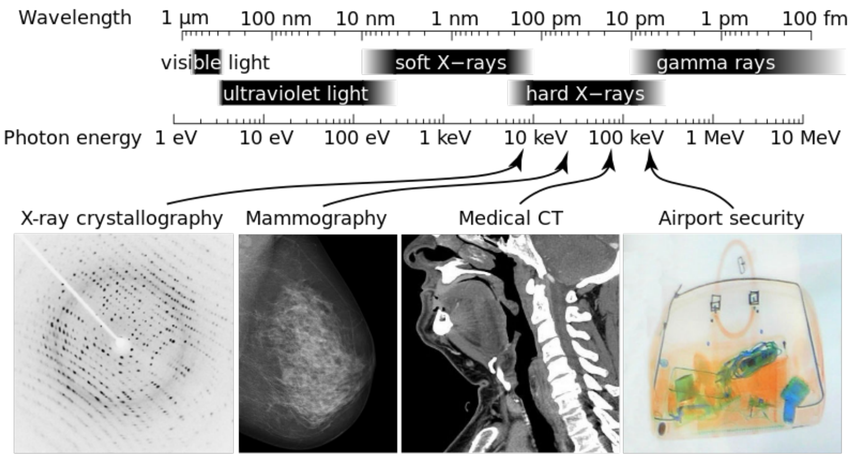


Fig. 3.4. X-ray radiation wavelength and its different applications, by Ulflund, distributed under CC BY-SA 3.0 license, Link to Image

At the level of energy of the hard X-rays, there are two major interaction effects of X-rays with tissue. The first is the photoelectric effect, where a photon uses up all its energy to eject an electron from an atom; while the electron will move around and ionize neighboring atoms, there are no scatter photons. The second major effect is Compton scattering, where a photon hits an atom and ionizes an electron but does not use up all of its energy. The photon then scatters in a different direction with a bit less energy, and the free electron ionizes neighboring atoms as in the photoelectric effect. Scattered photons can travel at any angle, no matter the original direction.

Due to both effects, the X-ray radiation is attenuated through its course in the matter. The model for this intensity attenuation along a beam I obeys the Beer–Lambert law shown in Equation 3.4, which shows an exponential decay of the X-ray radiation through the matter, depending on the depth of penetration x, the linear attenuation coefficient A and the initial intensity I_0.

$$I = I_0 e^{-Ax} \tag{3.4}$$

With a wavelength shorter than visible light, the energy of X-ray radiation is higher than the latter. This allows the radiation to penetrate denser and deeper structures than visible light such as bone, soft tissue, and skin. In Table 3.1, we show the linear attenuation coefficient for several matters of the human body. Bone is attenuating the most, followed by muscle, water, fat and finally air. The penetration distance to reach half the emitted radiation is in the order of centimeters, which makes it suitable for medical imaging.

Matter	Bone	Muscle	Water	Fat	Air
Linear Attenuation Coefficient (cm^{-1})	0.5727	0.2330	0.2245	0.1925	0.00025

Tab. 3.1. Linear Attenuation Coefficient at 50 keV (middle of hard X-ray spectrum) for several body matters

3.2.2 C-arm Device

A C-arm is a special type of radiography device designed to be used during surgery. In general, a radiography device is composed of a source that emits the X-ray radiation and of an intensifier (also, called detector) that will form an X-ray image from the radiation received from the source.

X-ray Image Formation

The patient is placed in between the source of radiation and the intensifier. As we explained earlier, the X-ray radiation can penetrate the human body, however its attenuation through matter differ according to the matter penetrated. An X-ray image is the resulting image of the X-ray radiation that has succeeded to penetrate the human body and exited with diminished intensity. This output intensity depends on the matter penetrated and will, therefore, infer about which structure has been crossed. Historically, the X-ray film placed at the intensifier was burnt with the incoming radiation. Depending on the film color, the dense matter will be then imaged as dark values (e.g. Röntgen's wife hand, Fig 3.3 and digital device) or as clear values (analog device). The intensifier is now digital sensor that will convert the radiation into a digital signal and provide a digital X-ray image.

The source of X-ray radiation is composed of X-ray photons, produced by an electron beam created from a heated cathode filament that is accelerated to a very high speed and strikes an anode target. The point where the electron beam strikes the anode target is called the focal spot. From this point, X-ray photons are emitted in all directions from the target surface. A collimator placed under the source selects the radiation into a conic beam directed to the intensifier. As the radiation is emitted from a single point (the focal spot) into a conic beam, the X-ray radiation on a radiography device, therefore, follows the pinhole model with the focal spot being the optical center. However, the difference to the usual video camera application is that the light is not reflected from the object but directly emitted by the optical center of the camera which is the X-ray source. The image plane is placed after the object at the image intensifier, as for the video application, we will consider a virtual image plane placed between the object and the optical center.

C-arm Specifications

A C-arm is a special radiography device on which the X-ray source and the image intensifier are rigidly attached on a C-shape on both extremities of the C (see Figure 3.5). This positions them directly opposite, aligned centrally to each other and mechanically dependent. Compared to screening radiography devices, C-arms are designed to be compact and lightweight to allow easy positioning with adequate space to work and a wide range of motion as shown in Figure 3.6.

Fig. 3.5. C-arm with the image intensifier on the top (red circle) and the X-ray source at the bottom (blue circle), by SteinsplitterBot, distributed under CC BY-SA 4.0 license, Link to Image

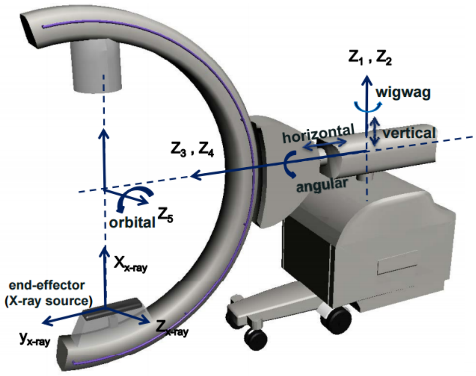


Fig. 3.6. C-arm degrees of freedom from [165]

In the Operating Room (OR), a C-arm can be used for any surgery that necessitates immediate feedback on the internal anatomy and the surgeon's actions. The C-arm can be used for guidance or control and is, therefore, particularly used for Minimally Invasive Surgery. The medical applications for C-arm are various, such as cardiac surgery, vascular surgery, gastroenterology, orthopedics, pain management, neurology procedures, brachytherapy, electrophysiology. We show in Figure 3.7 few examples of X-ray images acquired during those different types of procedure. The use of contrast agent injected in soft tissues can help to visualize them such as blood vessels for example.

(a) Pedicle Screw Placement (Spine Surgery) [108]

(b) Orthopedic Surgery

(c) Cardiac Surgery with Contrast Agent

Fig. 3.7. Images with C-arm devices for different types of procedure

3.2.3 Risks Related to X-ray Imaging

As explained earlier, the X-ray radiation has enough energy to penetrate the human body. This penetration is not without danger. As we have seen, the X-ray radiation interaction with matter produces ionized atoms (free radicals) that can damage the DNA of the penetrated cells and create mutations. Although the cells repair most of the damage and destroy most of the mutations, some remains. Those remaining DNA mutations might lead to cancer at the long term. However, cancer due to exposure to radiation occurs in a stochastic manner: there is no threshold point and risk increases in a linear-quadratic fashion with dose. Although the risk increases with dose, the severity of the effects does not; the person might develop cancer or maybe not. The patient and the surgical crew are exposed to those radiations during surgery. Except in the case of children where recurrent radiation must be avoided, the patient radiation is often considered an acceptable side effect due to its low frequency. The concern is more directed to the surgical crew which is operating several times a day and exposed daily to those radiations. Protections are taken in that regard such as lead apron, neck protection, and glasses. The best practice during surgery is to take the lowest amount of X-ray images necessary and to step back from the C-arm when an X-ray image is shot, as the radiation intensity is inverse to the distance to the source.

3.3 RGBD Sensing

RGBD cameras (RGB for the Red, Green and Blue channels of the video and D for the depth) is a type of sensor, that in addition to color data, offers depth information, i.e. the distance from the object to the plane containing the optical center parallel to the image plane. Depth imaging is often referred as 2.5D, as it gives 3D information but only from one viewpoint. The output image, called depth image, of such a camera, is shown in Fig 3.8. In this case, the different depth values are integers corresponding to the distance in mm to the camera plane of the scene.

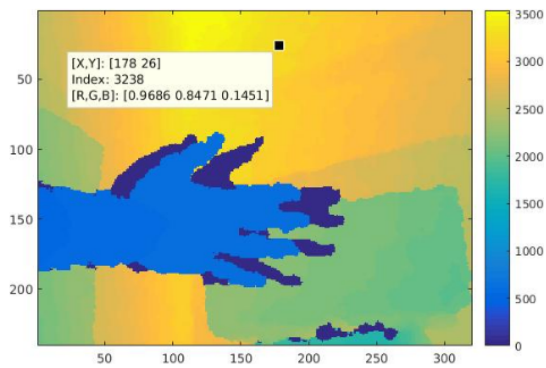


Fig. 3.8. Depth image acquired with Xtion Live Pro, scene of a hand in front of screen and further a wall. The colormap on the right shows the depth values range

From depth image to 3D

Without depth information, the 3D point corresponding to a 2D pixel is only known along the line passing the optical center and the pixel. In mathematical terms, in the perspective projection equation described earlier (Equation 3.3), the parameter w remains unknown. The depth information allows solving this unknown. The 3D point $P = (X, Y, Z)$ corresponding to a pixel $p = (u, v)$ in the depth camera virtual image plane of depth d can be retrieved using Equation 3.5.

$$X = \frac{d(u - u_0)}{f}; Y = \frac{d(v - v_0)}{f}; Z = d \tag{3.5}$$

3.3.1 Depth Sensing Technologies

Multiple technologies exist to compute the depth image. In the next sections, we are going to describe the three main technologies that can be found in RGBD cameras: stereo, structured light, and Time of Flight (ToF). First, we will describe the triangulation methodology behind the two first (stereo, structured light).

Depth Calculation Using Triangulation

Triangulation consists in finding the 3D point P based on two corresponding points in 2D p_1 and p_2 from different viewpoints (1 and 2). In each viewpoint i, a line passing by the optical center of the camera C_i to the 2D point p_i can be defined. The 3D point P is at the intersection of those two lines, in case they intersect which is not automatic in 3D space. The lines definition requires the knowledge of the intrinsic matrix of each camera K_i, the parametric equation, depending on the unknown scale factor w_i is presented in Equation 3.6 where $p_i = (u_i, v_i, 1)^T$ and P_i the 3D point along the line with coordinates $(X_i, Y_i, Z_i)^T$ in the camera coordinate system.

$$P_i = w_i K_i^{-1} p_i \tag{3.6}$$

Of course, the line intersection can only be computed in the same coordinate system, therefore, the pose T from the viewpoint 1 to viewpoint 2 (or vice versa T^{-1}) needs to be known too. Therefore, the line intersection is computed by searching the two variables w_1 and w_2 respecting the condition $P_2 = TP_1$. The 3D point $P = P_2 = TP_1$ is, therefore, known.

The triangulation principle is quite simple, however, we use as a prerequisite that we have two corresponding points in the two viewpoints. However, this is the bottleneck of the stereo geometry: how do we find matching points in two images in order to perform triangulation? The different technologies of depth cameras are actually based on the different techniques to perform this matching. First, we will discuss the stereo camera which uses passive matching where the correspondences are computed from two video cameras and the structured light cameras, using active matching, projecting a pattern in order to facilitate the correspondence problem.

Stereo Camera

Stereo cameras consist of two video cameras looking at the same scene from different viewpoints. As described before, to find the depth of one pixel in one image by the triangulation process, it is necessary to match it with a point from the other viewpoint. The triangulation brings some constraints to the location of the corresponding point. From one known 2D pixel point in one image to which we want to find its correspondent in the other image, one line can be created as explained in the triangulation paragraph. This line and the optical center of the other camera forms a plane, called epipolar plane and represented in red in Figure 3.9, in which the line of the corresponding point is included too. This plane intersects the other image as a line (red dashed line in Figure 3.9), limiting the corresponding point location to this line. This heavily decreases the search area for the correspondent.

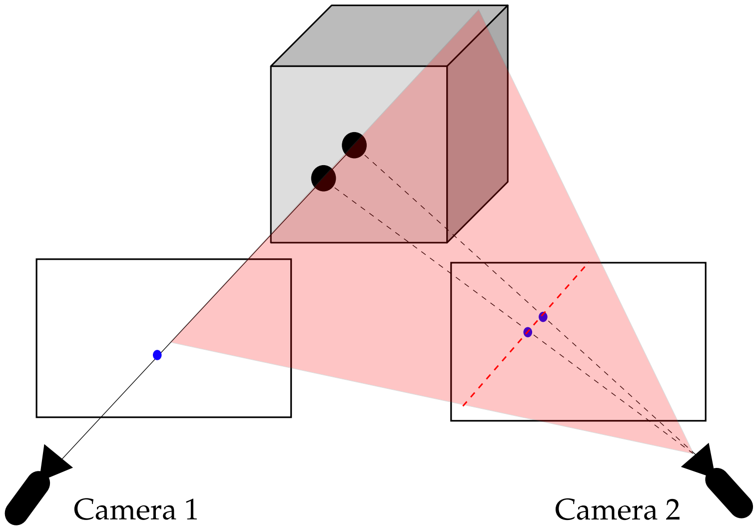


Fig. 3.9. Concept of epipolar plane (in red) for stereo camera system, on the red dotted line lies the correspondent

One very common algorithm to find the correspondent in stereo images is the Block Matching algorithm as implemented in OpenCV library. Along the line of search, it will compare patches (or blocks) –red and green boxes in Figure 3.10–to the patch surrounding the pairing pixel using a similarity metric such as SAD, NCC, etc. . . . The most similar patch along the line (green patch in Figure 3.10) will, therefore, give the corresponding pixel.

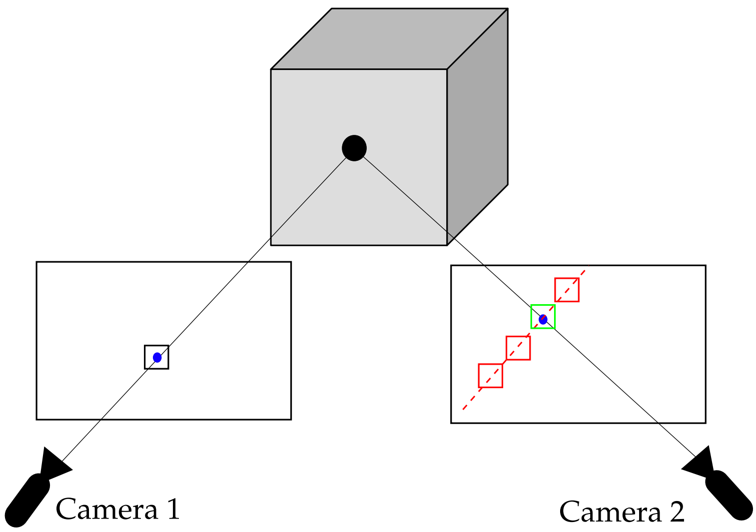


Fig. 3.10. Block-matching algorithm, each patch is compared using a similarity metric to the inquiry patch, in green is the corresponding one

Structured Light

In this case, the imaging device is composed of one projector projecting a known light pattern and one camera observing the pattern. As the emitted pattern is known, the search for the matching point in the observed image to the pattern point is a lot more simplified. As the scene is manipulated in order to solve the correspondence issue, this technique is called active triangulation. There is a wide diversity of light patterns and wavelengths used for Structured Light. At its simplest, a single point of light is projected on the object, its correspondent in the camera is straightforward to find as only one point appears too. However, this operation must be repeated for every pixel in the camera image. More complicated techniques have been developed to reduce the number of images necessary, reducing the acquisition time as well. However, as soon as the pattern gets more complicated such as a stripe of lights, the problem of correspondence occurs again. Therefore, temporally or spatially encoded patterns are the most common technique to overcome this issue. For an extensive overview and explanation of the different patterns, we refer the reader to the tutorial of Gang [48] and the paper of Pages et al. [117]. This thesis will only describe the most used patterns.

Temporal patterns

Temporal patterns consist of a temporal sequence of patterns of light stripes with different frequencies projected onto the object. This sequence of patterns is encoded in a way such that every projector pixel is uniquely identified by a temporal codeword in form of a binary sequence [126]. Its correspondent is the only point in the camera with the same codeword. Further works [79] encrypt the binary sequence with Gray-code, which will assign to adjacent stripes binary sequence differing of only one bit, allowing error correction and is more robust than binary encoding. In Figure 3.11, we roughly illustrate the temporal pattern coding. Two patterns of stripes with different widths are projected subsequently, coding the projector pixel with the sequence (1,0), the corresponding pixel in the camera image is the pixel with code (1,0) when looking at those subsequent projected patterns.

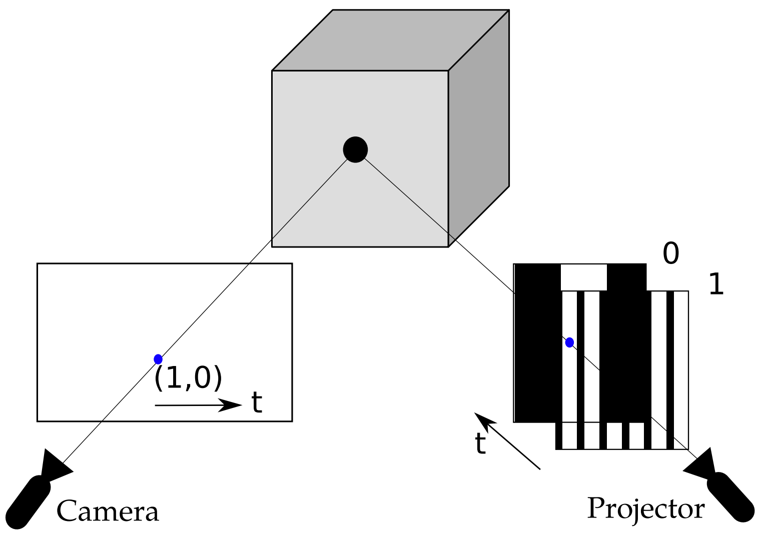


Fig. 3.11. Binary encoding temporal patterns, the projector sends subsequently different stripe patterns, coding temporally the pattern, the correspondent pixel is the pixel with same coding when reading the projection results.

The pattern used by the Intel RealSense F200 extends the binary temporal patterns to grayscale values as it can be seen in Figure 3.12. Patterns are sent at very high frequency (100Hz), however few artifacts due to the temporal nature of the system are reported by [183] such as ghosting, where previous actions appear in the current frame in case of fast motion. Although using fewer images than the single point pattern, the temporal acquisition still requires several frame acquisitions, which does not allow depth estimation of moving objects or only at the cost of artifacts.

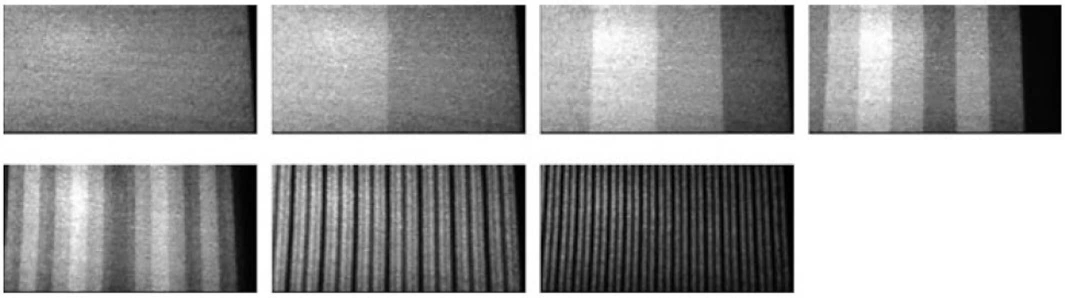


Fig. 3.12. Temporal patterns of the Intel RealSense F200, reproduced with permission from [183], ©Springer

Spatial neighborhood patterns

For spatial neighborhood patterns, the encoding is inside a unique pattern, therefore, it can be used in one frame and is suitable for moving objects. The codeword for each pixel is obtained from its pixel neighbors. The limitation is that if some neighbor pixels are occluded, then the codeword cannot be recovered and the depth not computed at this pixel. The encoding features can be intensities [104], colors [184] or neighborhood structure. This latest technique is the one used by the Kinect 1.0 camera in the form of a pseudo-random point pattern while the Intel RealSense R200 camera uses a combination of stripes pattern and grayscale values in its spectrum as visualized in Figure 3.13.

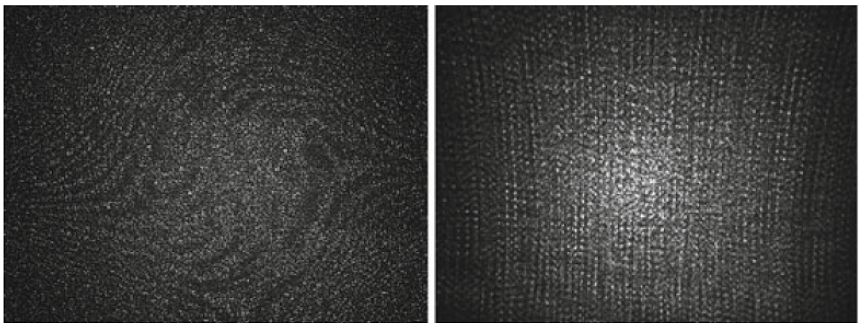


Fig. 3.13. Pseudo-random point pattern of Kinect 1.0 on the left and grayscale stripe pattern of the Intel RealSense R200 on the right, reproduced with permission from [183], ©Springer

Martinez et al. [103] explain that the Kinect 1.0 also uses an additional approach, known as projected texture [78], which uses the epipolar geometry to find that the correspondent lies on the epipolar line corresponding to the projector pixel in the camera image. The codeword must, therefore, be unique only along the epipolar lines, reducing the complexity of the pattern requirements. Ryan et al. [139] further discuss that a patch similarity search for every pattern pixel is performed along the corresponding epipolar line to find the matching pixel on the camera image using NCC. The codeword is, in this case, the patch, coding in the same manner as a QR code or an AR marker code. In Figure 3.14, we roughly illustrate this neighbor search. We search the patch in the camera image on the left that corresponds to the 3×3 patch in the pattern pixel (highlighted in red). Only one patch has the same neighbor pattern in the camera image, which allows finding the correspondent.

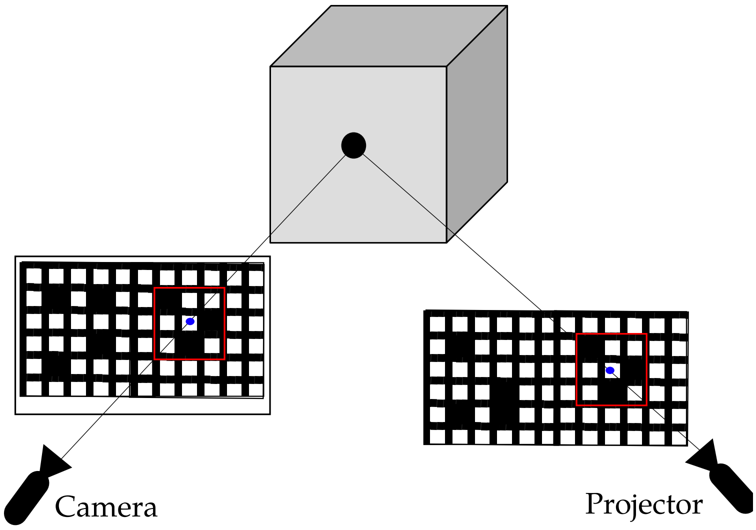


Fig. 3.14. Neighbor search for spatial neighborhood patterns, the correspondent pixel is the one with same neighbor patch (red square in both images)

Wavelength of Structured Light Devices

We have presented until now the main directions of Structured Light works without mentioning which wavelengths were used. The first works using Structured Light have been using visible light to project the patterns, however, this is destructive for the object texture. They can only recover the shape if the object is static, the texture can be acquired on a frame when no pattern is sent. For the systems aiming at imaging moving objects such as the Kinect 1.0, researchers have turned to other parts of the light spectrum, invisible to human eye. As most patterns are generated from a laser and the cheapest laser are emitting in infrared (IR), most of the invisible Structured Light devices use IR light. This is the case for the three aforementioned cameras Kinect 1.0, Intel RealSense F200 and R200. A large part of the Sun radiation reaching Earth is actually comprised in infrared, which unfortunately disturbs the visibility of the pattern by the IR camera in Structured Light devices. Therefore, those devices can only be used indoors, away from natural light.

Time of Flight

The last type of depth sensing technology is Time of Flight. This technology does not rely on triangulation and the distance is calculated directly or indirectly by calculating how long the light has traveled between its emission and its detection. There are two main types of technologies for Time of Flight cameras: pulsed modulation techniques and phase shift techniques [8]. For extensive details, we refer the reader to the book from Remondino [130].

Pulsed Modulation techniques

The pulsed modulation technique, which is part of LIDAR technology, consists of emitting one pulse of light and measuring its time τ to arrive at the camera as Figure 3.15 shows.

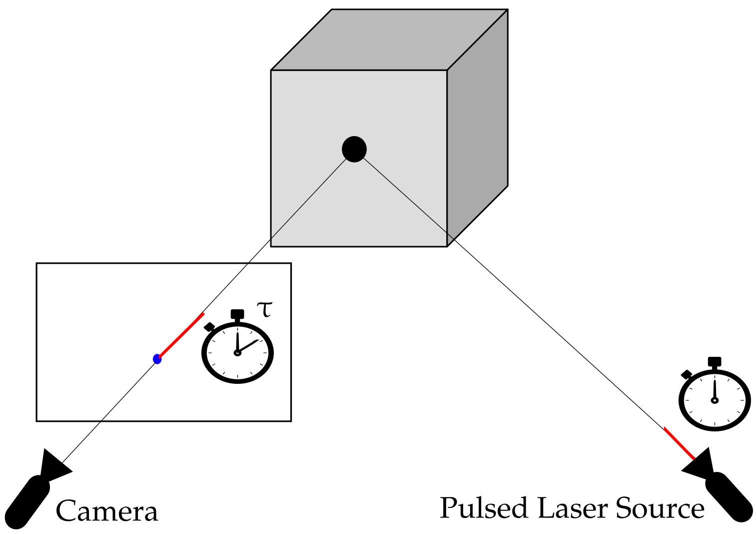


Fig. 3.15. Time measurement of the pulse of light (red line) flight

At every pixel in the camera array, the time when the photon hits the pixel is compared to the emission time. As we know the light speed ($c \approx 3e^8 m/s$), the depth d can be easily recovered for every pixel using Equation 3.7.

$$d = \frac{c\tau}{2} \tag{3.7}$$

The main drawback of those type of cameras is that the time measurement needs to be very accurate, which implies expensive hardware to do so. The advantage is however that it can calculate depth in the order of kilometers as used for example in airborne LIDAR.

Phase shift techniques

Another technology of Time of Flight cameras consists in indirectly calculating the light time of flight by comparing the phases of the signal between its emission and detection. A shift in phases φ appears due to the distance traveled by the light, as shown in Figure 3.16 and can inform about the depth d as Equation 3.8 demonstrates with f being the frequency of the signal. This technology is the one used in the Kinect v2.

$$d = \frac{c}{4\pi f}\varphi \tag{3.8}$$

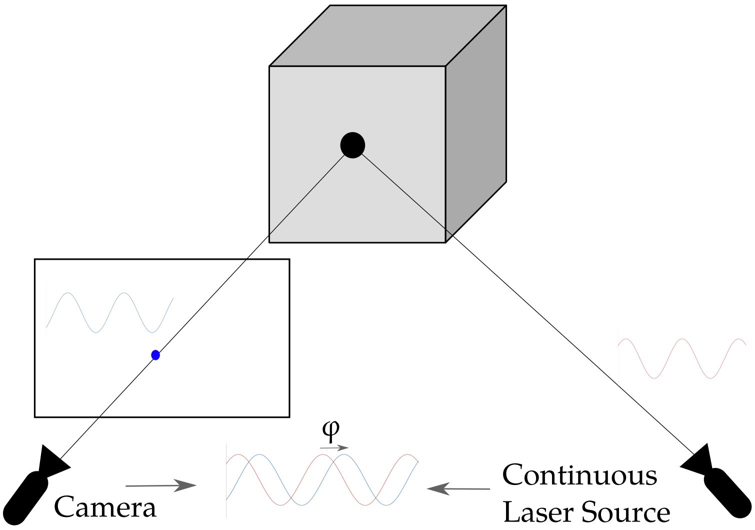


Fig. 3.16. Phase-shifting measurement, the phase shift φ between the emitted signal (pink wave) and the received signal (blue wave) is measured

3.3.2 Focus on Commodity RGBD Cameras

As shown in the previous paragraphs, the technology for depth sensing is quite diverse and has been widely implemented in commercial cameras during the last 20 years. However, for a long time, their prices remained high, keeping them into a niche market such as engineering and research. The release in 2011 of the Kinect 1.0 (see Figure 3.17) by Microsoft as a gaming device has allowed by the mass scale effect to heavily reduce the price of such RGBD device and to import such camera to people's home. With easier access to such technology, the number of applications using RGBD cameras has exploded in numerous fields starting with Computer Vision. In the path of the Kinect 1.0, numerous cameras have been released in the same price range: Kinect v2, Asus Xtion Pro Live, PrimeSense Carmine, Intel RealSense F200 and R200 (see Figure 3.17). Asus Xtion Pro Live and PrimeSense Carmine are produced by the manufacturer of Kinect 1.0 (PrimeSense) and present exactly the same characteristics regarding the depth sensors as the Kinect 1.0, while being lightweight.

Fig. 3.17. Kinect 1.0, Asus Xtion Live Pro, Intel RealSense F200 and R200 (reproduced with permission from [183], ©Springer) and Kinect v2 (public domain)

We have already explained the technology behind every camera in the previous paragraph and we are going to focus on the technical characteristics the Kinect v2 and Kinect 1.0 in the form of Asus Xtion Pro Live and PrimeSense Carmine, the cameras used in the works presented in this thesis. We show their main technical characteristics in Table 3.2.

Camera	Depth Resolution	Video Resolution	FPS	Working Range
Asus Xtion Live Pro	640×480	1280×960	30	0.5-5m
PrimeSense Carmine	640×480	1280×960	30	0.35-3m
Kinect v2	512×424	1920×1080	30	0.5-4.5 m

Tab. 3.2. Technical characteristics of RGBD cameras

Comparaison of the RGBD Cameras

As the specification section has shown, the depth images provided by all cameras have almost the same resolution. The Kinect v2 has, however, a much higher video image resolution. In a recent study, Wassenmuller et al. [175] have intensively compared the two Kinect versions. While the accuracy of Kinect 1.0 decreases quadratically with the distance, the Kinect v2 accuracy depends on the pixel location in the image (the center being more accurate than the corners). In general, the Kinect 1.0 presents fewer depth artifacts, especially for flat surfaces, where flying pixels can occur with the Kinect v2. The depth estimation of Kinect v2 is influenced by the scene color, as black will absorb the infrared, which does not happen with the Kinect 1.0. Due to bandwidth limitation, the IR image is not accessible at the same time as the video for Kinect 1.0, while it is possible with Kinect v2.

The type of technology used by the cameras has an impact when using several devices with overlapping fields of view. For the cameras using Structured Light, the patterns from different cameras overlays and, therefore, disturb the decoding of the pattern for all of them. As a result, the depth images present inaccurate values, mostly invalid in fact, making them unusable in this context. The Kinect v2 tune on a different signal frequency inside a defined range at every start. Therefore, when camera frequencies are different, multiple cameras do not disturb each other. By experience, it might happen very rarely that the cameras tune on the same frequency, creating, therefore, interference. Restarting them will allow to desynchronize them. Kinect v2 is, therefore, suitable for multi-camera setup with overlapping fields of view.

Part II

RGBD Augmented C-arm Systems

4 RGBD augmented mirror-based C-arm

The work presented in this chapter is an extended version of one part of the paper presented at ISMAR 2015 which is reproduced with permission from [53], ©IEEE.

4.1 Motivation & State of Art

In the State of the Art chapter, we have discussed the fusion of intra-operative data with X-ray image in order to contextualize the internal information of the X-ray image. Our first contribution builds on the Camera augmented mobile C-arm (CamC) setup described by Navab et al. [111]. CamC consists of displaying an alpha-blending overlay between X-ray and optical images in real-time. A video camera is placed close to the X-ray source, the overlay is possible thanks to a mirror construction placed under the X-ray source. This ensures that the optical axis and centers of both projective imaging systems are virtually aligned, after a one-time calibration. It supports surgeons in their understanding of the spatial relationships between anatomy, implants, and their surgical tools. Recently, the effective use of the video guidance from CamC has enabled the reduction of radiation exposure for the following clinical applications: the six degrees of freedom (DOF) kinematic modeling of a C-arm [166], the calibration of the C-arm [28], the robust pose estimation of a C-arm using AR markers [119], the parallax-free panorama generation for total knee arthroplasty surgery [167], the multimodal perceptual visualization of X-ray and optics [122] and the interlocking of intramedullary nailing [35, 95].

In this chapter, we build on those works and replace the video camera from Navab et al. setup by an RGBD camera of type infrared pattern emission (like the Kinect 1.0) whose characteristics are described in the Background Chapter in Section 3.3.2. The release of the Kinect 1.0 in 2011 opens a new range of possibilities by bringing depth information to the setup, with no additional calibration and additional hardware compared to Navab et al. [111] than a low-cost RGBD camera. Depth information has been studied very intensively in Computer Vision, the most famous applications being human body joint tracking [150] or 3D surface reconstruction [62], better known as Kinect Fusion. Those novel applications can be brought to the medical domain. This chapter investigates the feasibility and the design of the mirror-based RGBD augmented C-arm setup.

4.2 Setup

The setup is composed of one RGBD camera attached to the side of the X-ray source of the C-arm. A mirror construction is attached to the C-arm housing under the X-ray source in order to merge the optical axis and center of the video camera from the RGBD camera and the X-ray source. The setup is shown in Figure 4.1. At the year of the setup creation, only the cameras of type Kinect 1.0 existed. The camera used for this setup is the Asus Xtion Live Pro, lighter version of Kinect 1.0, with same technology but only powered through USB 2 or 3 and, therefore, more adapted to be mounted on the C-arm. The data from the camera is read through the OpenNI2 [1] library. The registration of the depth image to the video image is a feature offered by the library OpenNI2 based on manufacturer values.

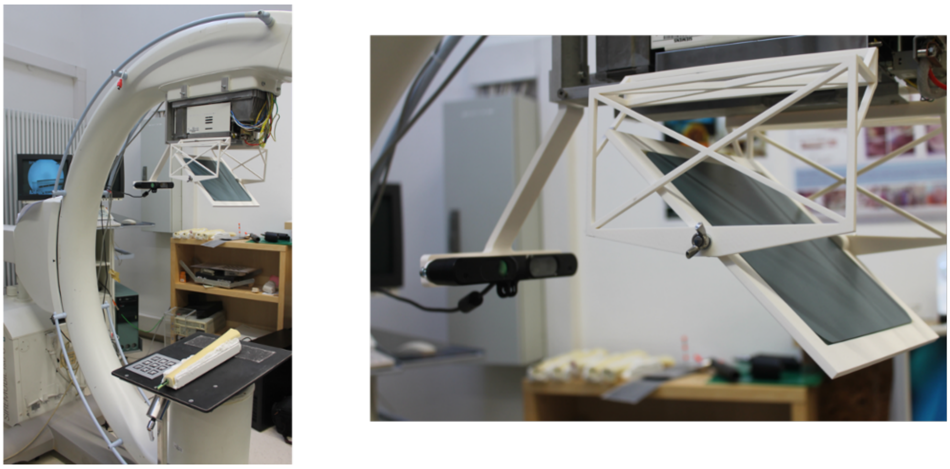


Fig. 4.1. Setup of RGBD augmented mirror-based C-arm, (Left) overview of setup, (Right) close-up on camera and mirror mounting

The mirror and cameras are mounted on the C-arm thanks to a 3D printed custom-made attachment. The mirror, placed at normal incidence with respect to the cameras optical axis, is of size $20\,\mathrm{cm} \times 20\,\mathrm{cm}$, which is slightly bigger than the mirror used by Navab et al [111]. Indeed, due to the RGBD cameras baseline between the infrared emitter and the infrared detector, a minimum size of the mirror is required such as it appears in both fields of view. However, working space under the C-arm is minimally impacted with a loss of $3\,\mathrm{cm}$ compared to the Camera Augmented C-arm [111]. The mirror is an ordinary silver back-coated glass mirror, the choice of the metal will be discussed in the next section.

Newer cameras have appeared since the realization of this setup. RealSense camera (Intel), of the same type as Kinect 1.0 and also describe in Chapter 3, possesses a shorter baseline which would allow the use of a smaller mirror. Smaller baseline implies shorter working range ($20\,\mathrm{cm}$ to $110\,\mathrm{cm}$) which would actually be more adapted to the C-arm than the working range of the Asus Xtion Live Pro ($50\,\mathrm{cm}$ to $10\,\mathrm{m}$). The use of this new camera is recommended for a future realization of this setup.

1. https://github.com/occipital/openni2

4.3 Infrared Pattern Emission Cameras with Mirror

A search of the literature has shown that no previous work has investigated if RGBD cameras of type Kinect 1.0, infrared pattern emission cameras, are working with mirrors. A short experiment shows that a depth image (with non-zero values) is returned when placing a mirror in the light path of the camera but with no insurance of depth accuracy. Few works have used right away infrared pattern emission cameras in combination with mirror without looking at the accuracy. Akay et al. [4] and Kim et al. [74] use a mirror in combination with RGBD cameras in order to get multiple viewpoints for 3D reconstruction, but without assuring if the depth returned through mirrors is accurate. To fill this gap in the literature, we investigate the accuracy of RGBD cameras in combination with mirrors, looking first which type of mirror is adapted to RGBD cameras.

4.3.1 Choice of Mirror

The most common mirrors are glass (called substrate) surface back coated with a thin metal layer reflecting the incoming light. Aluminum is the most used coating followed by silver, with aluminum being much cheaper than silver. Figure 4.2 shows the path of rays through a back-surface mirror, glass substrate deviates the rays (plain full line) until they reflect on the back surface and are deviated back when leaving the substrate. The mirror is equivalent to a front surface mirror which reflects the non-deviated rays (green dotted line) at the dotted red line. Without loss of generality, for the rest of the work, our back-surface mirror will be represented in our future drawings, for simplification, as a front surface mirror where the rays are not deviated by the substrate.

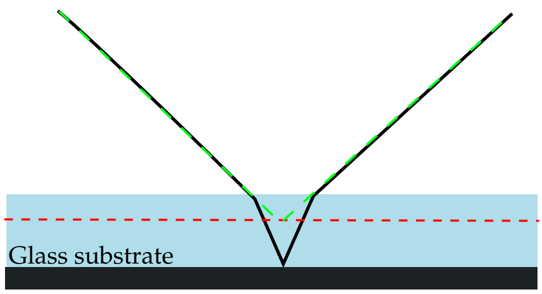


Fig. 4.2. Back surface mirror, glass substrate deviates the rays (plain full line) until they reflect on back surface and are deviated back when leaving the substrate, the mirror is equivalent to a front surface mirror which reflects at the dotted red line, the non-deviated rays (green dotted line)

Infrared pattern emission cameras use a near infrared wavelength of $825\,\mathrm{nm}$. For making the best coating choice, we look at the reflectance of the aluminum and silver at that wavelength. However, since the setup will also use a video camera, the mirror should also present good performances in the visible light. Figure 4.3 portrays the reflectance percentage of aluminum (Al), silver(Ag) and gold (Au) according to the wavelength. Visible light is comprised between $380\,\mathrm{nm}$ and $750\,\mathrm{nm}$. From this figure, we extract the reflectances at the interesting wavelengths and include them in Table 4.1.

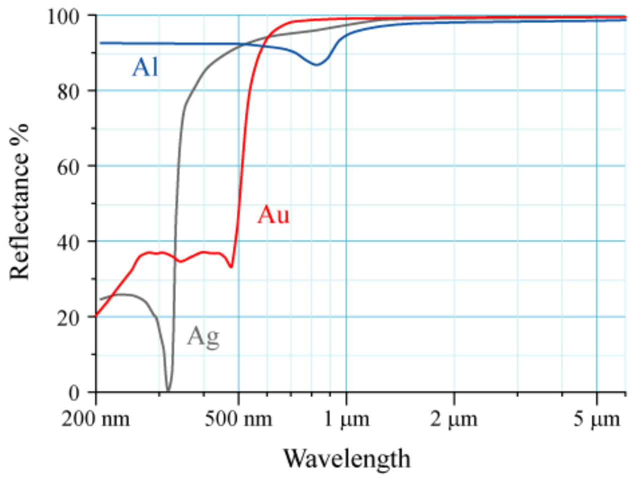


Fig. 4.3. Reflectance vs. wavelength curves for aluminum (Al), silver (Ag), and gold (Au) metal mirrors at normal incidence, by Bob Mellish, distributed under CC BY-SA 3.0 license, Link to Image

	Aluminum	Silver
Visible light Reflectance	90%	85% to 92%
Infrared Reflectance	85%	95%
Infrared Reflectance × Infrared Reflectance	72%	90%

Tab. 4.1. Reflectance rate extracted from Figure 4.3 for visible light and infrared light

While visible light is only reflected once (from the object to the video camera), the infrared light emitted and detected by the camera is reflected twice, first from the emitter to the object, then back from the object to the detector. To obtain the performance of the metal coating in combination with infrared pattern emission camera, the reflectance must be squared. From Figure 4.3, it can be observed that the aluminum has its lowest performance almost exactly at the wavelength of the RGBD infrared wavelength, with 85 % reflectance, while silver reflectance is 95%. By squaring those reflectances, the aluminum shows a high decrease in performance compared to silver (75% compared to 90%). In visible light, the silver reflectance is slightly lower than aluminum but the difference is much smaller than in infrared. Using those observations, a silver mirror is, therefore, the best choice as mirror coating for our setup.

4.3.2 Geometrical Optics Approach

In order to prove that the depth is optically calculated properly with the mirror, we will consider the geometrical optics principles applying to depth calculation. The geometrical optical laws are applied to verify if any modification is carried by the mirror in the path of the infrared rays. We consider the setup as described in Figure 4.4 where we compare two scenes at the same depth with or without a mirror.

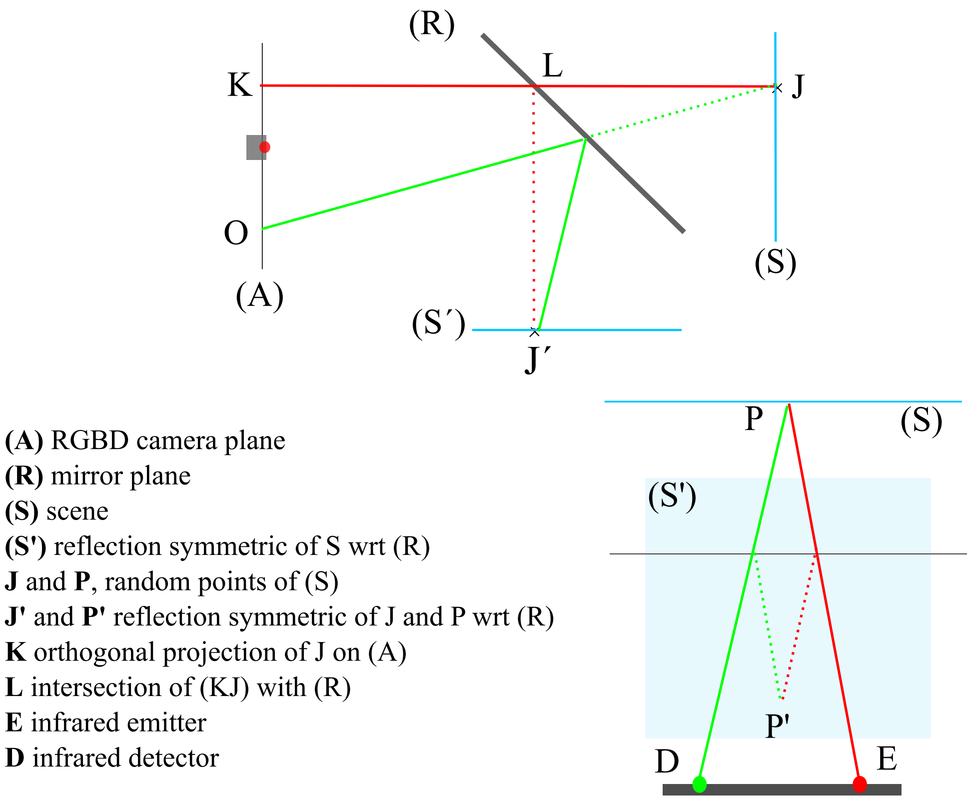


Fig. 4.4. Schema of the optical laws applied to the with mirror and without mirror scene for depth calculation

The first scene (S) is placed parallel to the RGBD camera, and the second scene (S') is the symmetric reflection of (S) by the mirror. In both scenes, we consider one point J in (S) and its symmetric J' in (S'). A reflective symmetry by a mirror is a reflectional symmetry and is, therefore, an affine isometric transformation preserving distances and angles after transformation, but not orientation.

Physical Depth and Reflection

We define the concept of *physical depth* of a point J (or $Phd(J)$), as seen by the RGBD camera, as the shortest distance between the plane of the camera (A) to a parallel plane to the camera containing J. Without the mirror, this *physical depth* is the depth value acquired by the camera. From Figure 4.4-top, without reflection on the mirror, this distance is equal to the distance between J and its orthogonal projection on (A) K. However, the mirror changes the definition: the physical depth seen through the mirror is, in fact, the shortest distance between two points of the camera plane (A) and the point J', considering one reflection on the mirror. In our case, the shortest path is the path beginning in K, reflecting on the mirror in L finishing on J'. This can be easily demonstrated by considering another random path with reflection on the mirror passing by J' (the green ray in Figure 4.4-top). When considering its equivalent reflection (dotted green path), the length of this path becomes then $||\vec{OJ}||$ which, by property of the orthogonal projection K on (A), is higher or equal (if $O = K$) than $||\vec{KJ}|| = Phd(S)$. The minimum is achieved for $O = K$, which confirms our affirmation. Therefore, since $Phd(S') = ||\vec{KL}|| + ||\vec{LJ'}||$ and $||\vec{LJ'}|| = ||\vec{LJ}||$ thanks to the mirror isometry, $Phd(S') = ||\vec{KJ}|| = Phd(S)$. As a consequence, this shows that (S') and (S) have the same *physical depth*, meaning that the depth camera if working properly through mirrors, should see the same value for both scenes. The scene (S') has then the same *physical depth* as (S) after one reflection on the mirror. We will show now that the depth camera should theoretically work properly through mirrors.

Deformation after Reflection

The depth calculated by the camera through the mirror is the same than without the mirror, if and only if the same infrared pattern deformation is observed by the infrared detector in both cases. In Figure 4.4-bottom, the optic configuration is represented. A parallel gray line represents the mirror, justified by the fact that the depth camera is facing the mirror. Any triangle derived by the emitter E, the detector D, and any point P of (S) intersects the mirror with a parallel line to the depth camera. A deformation induced by the mirror would imply that rays that are converging to a point P in (S) (represented by the red and green plain lines) would not converge to one point on (S') considering their equivalent reflected rays (in dashed lines). Since the mirror reflection is an isometric transformation conserving distances and angles, the reflected rays also converge in one point P' on (S'). This signifies that the depth calculation should not theoretically be disturbed by the mirror.

To confirm those demonstrations, we empirically verify by performing three experiences to test the depth acquisition through mirrors.

4.3.3 Experiments

In the previous section, we have shown that theoretically the mirror should not deform the pattern sent received by the RGBD camera, resulting in no change in depth image between a scene seen through a mirror or not.

With and without Mirror Comparison

Our first experiment is going to test this theory by reproducing the situation presented in Figure 4.4-top where at our scenes (S) and (S'), we placed a reference object, a flat box of size $32\,\mathrm{cm} \times 17\,\mathrm{cm}$, perpendicular to the RGBD camera optical axis. Between the scenes (S) and (S'), the box is symmetrically positioned with respect to the mirror. We performed the experiment at different *physical depth* along the RGBD optical axis, this *physical depth* being calculated through rigorous measurements with a precise ruler, set square and protractor. The range of *physical depth* is from $60\,\mathrm{cm}$ to $110\,\mathrm{cm}$, which is the C-arm working range with a step of $10\,\mathrm{cm}$. The mirror is itself placed at $30\,\mathrm{cm}$ from the camera. An example of the data acquired in this experiment is shown in Figure 4.5, on the first column is the data through the mirror and on the second column without the mirror.

For every pair of scenes, a data sequence lasting one minute is acquired, the depth at every pixel is then averaged temporally over the sequence over the valid values (depth > 0). This long sequence of images allows us to smooth the depth values which are noisy. Then, we perform a Sum of Squared Difference (SSD) operation between the two averaged depth image from the pair of symmetric scenes. Ideally, the resulting values should be zero, however the result of our experiment show that this is not the case as reported in Figure 4.6 and Table 4.2.

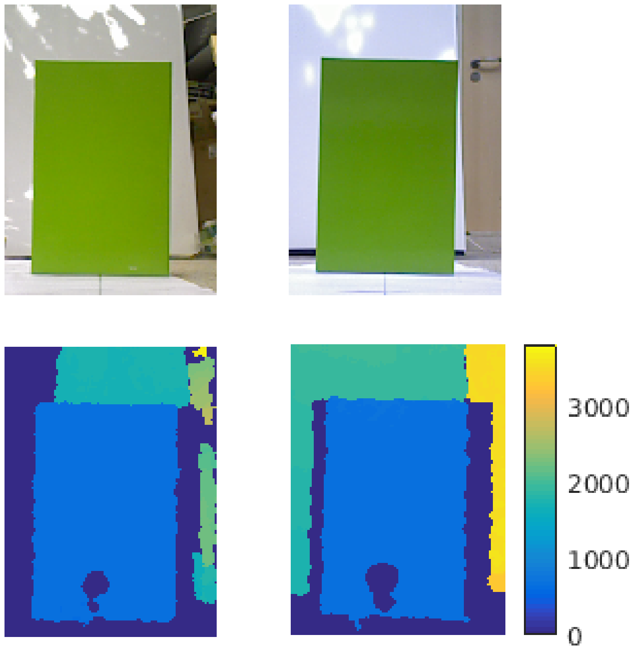


Fig. 4.5. Images acquired during the first experiment to compare with (Left) and without (Right) mirror situations

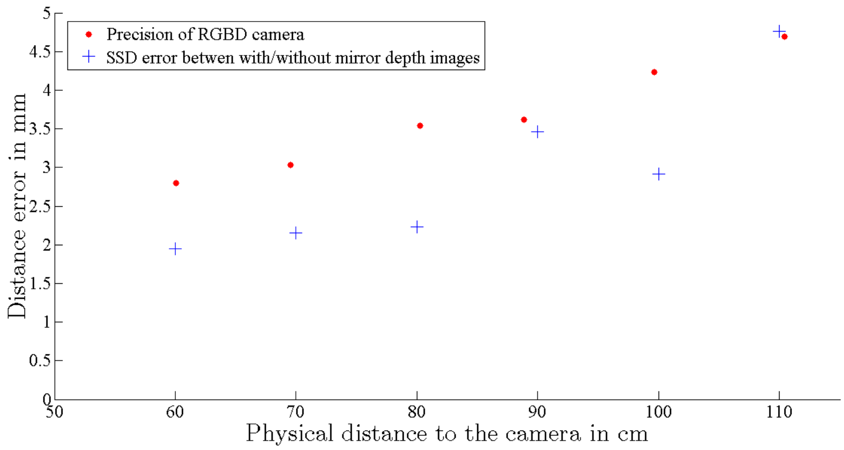


Fig. 4.6. SSD error between depth images at different distances

Physical Depth (in cm)	60	70	80	90	100	110
SSD error (in mm)	1.95	2.15	2.23	3.45	2.91	4.76
RGBD camera accuracy (in mm)	2.82	3.00	3.46	3.61	4.25	4.68

Tab. 4.2. SSD error between two averaged depth images from pairs of symmetric scenes at different *physical depth* compared to the RGBD camera accuracy

Our hypothesis is that the error is in the physical location of the reference object, even though very rigorous measurement has been done. However, this experiment needs to be put into context with other values to show interesting insights. We extracted from the work of Herrera

[61] the RGBD camera accuracy, also called axial error that we also reported in Figure 4.6 and Table 4.2. The depth difference error is for all the distance values under the inaccuracy value of the RGBD camera at that distance. While our experiment does not allow us to confirm the hypothesis that the mirror does not introduce deformation, our experiment shows us that even if some deformations appear, the depth error is of the same order as the depth inaccuracy, that is anyway present in all cases.

Two additional analyses are performed to study the behavior of the RGBD camera with a mirror. They were both performed using a precise measurement table. This device possesses three motors, one for each Euclidean axis, which can be controlled individually to access 3D point in the working range of the device. For our experiment, we used only one axis, placed parallel to the optical axis of our RGBD camera. This device allowed us very precise displacements within the range of $65\,\mathrm{cm}$ to $90\,\mathrm{cm}$ with step size of $1\,\mathrm{cm}$. The measurement table has a limited working range of $25\,\mathrm{cm}$, therefore, we had to choose a range included in the interval of $60\,\mathrm{cm}$-$110\,\mathrm{cm}$ used previously. This device could not be used for the previous experiment as the working space did not allow us to have at the same time scenes with and without a mirror. Therefore, in the following experiments, we only look at the scene through the mirror.

Relative Depth Comparison

Our second experiment tests if a variation of *physical depth* is exactly reproduced in the depth images values. For this experiment, the mirror is placed at $30\,\mathrm{cm}$ from the camera. When moving the reference object of one centimeter further using the measurement table, we acquire an average depth image (in a similar fashion as the previous experience) and compare it to the depth image at the previous position and calculate the SSD error. This is performed for the full range from $65\,\mathrm{cm}$ to $90\,\mathrm{cm}$. Our results are that for a variation of $10\,\mathrm{mm}$ in the real world, the average SSD error between two subsequent depth images is $9.94\,\mathrm{mm}{\pm}0.33\,\mathrm{mm}$ over our working range, showing that the mirror does not introduce relative error (0.006% error).

Spatially Located Error

For the previous experiments, the error metric consisted in averaging over the reference object, hence the error distribution on the mirror surface is unknown. Since the mirror is tilted at $45°$, our last experiment looks at the localization of error by assessing if any difference appears in depth when a scene is seen through the closer part of the mirror (in our experiment on the left) or further (right part). For this experiment, the camera is looking at a checkerboard located over the measurement table, which moves in the range $65\,\mathrm{cm}$ to $90\,\mathrm{cm}$ with a step size of $1\,\mathrm{cm}$. Using checkerboard corner tracking algorithms, we can detect the position in the video of the inside corners of the checkerboard, which are classified as being part of the close or further part of the mirror for all views. The checkerboard as seen through the mirror and the detected corners are shown in Figure 4.7.

For every corner pixel, we average the depth over a sequence in a similar fashion to the previous experiments. Then, for every corner inside a class, we average the depth values over the class members. We repeat this experiment for the full working range. The results are shown in Figure 4.8. By averaging over the working range, the depth difference between

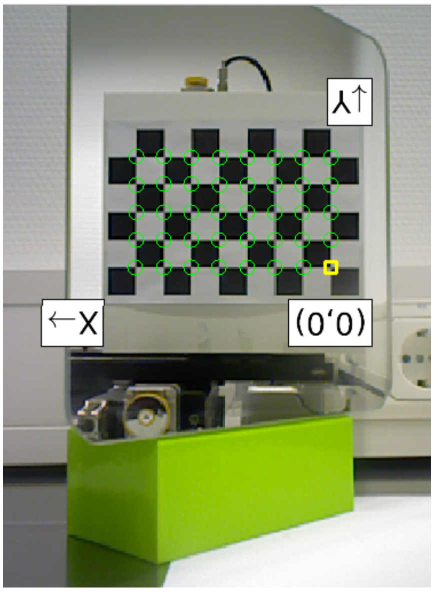


Fig. 4.7. Points used for the spatially located error detected using checkerboard corner detection

the closer and further part is very low: $0.7\,\mathrm{mm}$, which shows that no spatial located error is introduced by the mirror.

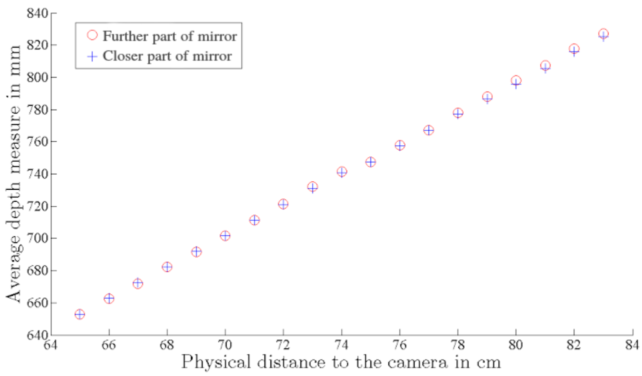


Fig. 4.8. Depth values at different distances for the further and closer parts of the mirror

Conclusion from the Experiments

From our experiments, we can conclude that no strong deformation is brought by the mirror. If one accepts to use the RGBD camera alone with its millimeter inaccuracy, one can also use it with a mirror with possibly a slightly higher inaccuracy, but still acceptable in regard to the initial RGBD camera inaccuracy. The inaccuracy of the RGBD camera with or without mirror also shows one limitation of the use of RGBD camera in a medical setting, with a depth inaccuracy comprised between $2.5\,\mathrm{mm}$ and $5\,\mathrm{mm}$, the depth data at one pixel cannot be used directly for 3D medical tool tracking. The inaccuracy is indeed superior to most allowed error range in image-guided surgery technologies [158]. Only post-processing or the use as 3D point cloud could overcome this limitation. Aware of the RGBD data accuracy limitation, we decided to push further with this setup as the depth data can be used for other purposes such as registration or visualization as proposed later in Chapter 6 and Chapter 7.

4.4 Calibration

Here, we describe the calibration process of our setup in order to get an aligned overlay of X-ray image over video image, similar to Navab et al [111]. Our RGBD camera is equipped with a video camera and we can, therefore, reproduce the same calibration process as Navab et al [111].

To align the video camera and X-ray source optical axes and centers, a two-level calibration object, shown in Figure 4.9, is used. The lower level includes five fixed X-ray markers, while the upper level possesses five movable rings of an inner diameter slightly higher than the markers diameter. The calibration process consists in aligning the upper rings and lower markers in both the video and the X-ray image. As shown in Figure 4.9-top, this ensures that optical axes and centers coincide. The X-ray source is not movable, therefore, the mirror and the RGBD camera must be moved in the process. First, we align the rings over the markers in the X-ray image. When achieved, the RGBD camera and mirror are moved in order to also see the alignment in the video image without touching the ring configuration anymore. Navab et al. [111] use X-ray image guidance for ring alignment while Chen et al. [29] have proposed a video guidance that allows reducing the number of X-ray images necessary for the procedure. The two cameras have different fields of view. Therefore, to obtain the overlay, a last step is necessary. Using an aluminum AR marker pattern visible in both video and X-ray images, we compute the homography H from X-ray to video image, which is used for the precise overlay.

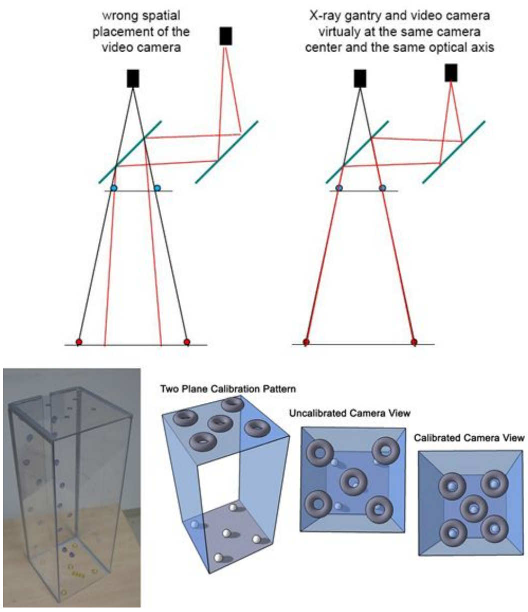


Fig. 4.9. (Top) Aligning the rings and markers in both modalities ensures virtual alignment of X-ray source and video optical center and axis (Bottom) Different views of the calibration object, from [111] ©[2010] IEEE

4.5 Visualization

Once the setup is calibrated, we obtain the overlay of X-ray image over the video image, the same overlay provided by the setup of Navab et al. [111]. However, for every pixel in the video image, we also have the depth information. We will show in Chapter 6 and 7 advanced visualizations using this information. However, we will already show a straightforward use based on the works of Pauly et al. [122]. Using the depth data, we can split between the background (acquired during an initialization sequence) and the foreground as shown in Figure 4.10, allowing a more perceptually correct overlay. Pauly et al. further use Machine Learning to split inside the foreground layer between tools and hands, we do not perform that in this example.

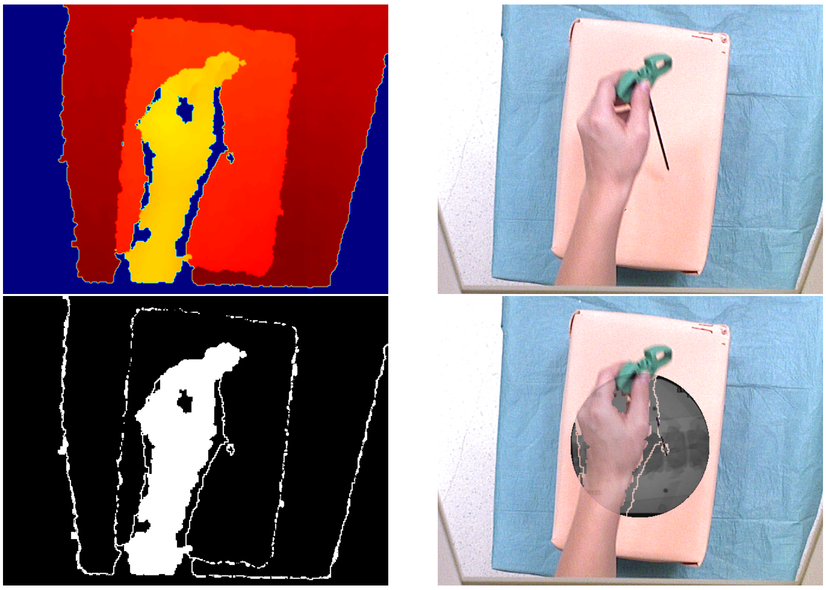


Fig. 4.10. (Top-left) Depth image from which is the foreground is segmented to obtain the (Bottom-left) mask, used for to personalize the (Bottom-right) overlay

4.6 Conclusion

In this chapter, we have described the mirror-based RGBD augmented C-arm. The feasibility of such setup was studied through theoretical optics and an empirical study to assess the validity of the RGBD data through was also conducted. We have shown that RGBD cameras of type pattern-emission can work with a mirror, making the first RGBD-augmented C-arm setup feasible and valid. We have also described its construction process from the design to the calibration. This work was an onset that led to the development of new medical applications such as 3D reconstruction along X-ray image visualization, presented in Chapter 6 but also an augmented reality radiation exposure sensibilization tool, presented in Chapter 7.

5 Stereo-RGBD mirrorless augmented C-arm

5.1 Motivation

The Camera augmented mobile C-arm from Navab et al. [111] or the RGBD augmented mirror-based C-arm presented in the previous chapter are both mirror-based setups with a mirror construction placed under the X-ray source in order to get the X-ray image over video image overlay. The mirror construction reduces the distance available for the surgeon between the intensifier and the source by approximately 15-20 %. For most surgeries, this is not an issue, but in the case of obese patients, the space reduction can be problematic. More importantly, the placement of the mirror construction under the X-ray source requires heavy modification of the C-arm structure in order to mount the mirror and the camera. This is restricting for an extensive integration of the setup in the Operating Room. Any inner modification of the C-arm invalidates its certification, requiring a long and difficult process of recertification to reuse it widely in the Operating Room with the new hardware addition. A setup requiring no inner modification could be used on any C-arm without recertification, which makes its cycle of development and evaluation shorter for research and cycle of integration easier for hospitals.

Those two constraints have led us to research a new system which does not include any modification of the inner structure of the C-arm and, therefore, does not need a mirror construction under the X-ray source, increasing the available working space. If the mirror construction was chosen by Navab et al. [111], this was due to the nature of the video and X-ray modality, which are both 3D to 2D projective imaging, following the pinhole model. To create an overlay with images from the same viewpoint, the video camera and X-rays source must hypothetically be placed at the same locations, which is physically impossible. The mirror allows deviating the rays from the video camera in such a way that they virtually come from the X-ray source. Placing the video camera in another viewpoint than the X-ray source without the use of mirror would be a solution, but would be associated with distortion artifacts. However, the use of RGBD camera lifts up this limitation, by allowing a change of viewpoint of the RGBD data without distortion since only the 3D rigid transformation between the camera pose and the target viewpoint needs to be computed, possible thanks to the 3D nature of the RGBD data. The cameras can then be placed in a more convenient location on the C-arm which requires less engineering on the C-arm and does not take space in the surgical workspace.

5.2 State of the Art

Several works in the Computer Vision field have used RGBD cameras to render another viewpoint than the one captured by the camera(s). Park et al. [118] first register and merge the point clouds generated from two RGBD cameras. They use the projective texture technique to seamlessly blend the point cloud color from the color images from the two viewpoints. The new viewpoint can be visualized by rotating the point cloud to the desired viewpoint. The main issue of point cloud technique for viewpoint change is the sparsity of points. If the new viewpoint has approximately the same characteristics in terms of distance or field of view, the result will look realistic, any change in distance or a smaller field of view will introduce holes as we zoom in on the point cloud. The next step in the use of RGBD data is to generate a mesh from the multiple cameras and then visualize it from the desired viewpoint such as performed in real-time by Alvaro et al. [31]. This solves the sparsity problem as the representation is dense. However, this type of work is reconstructing a full closed 360° model, on which mesh triangulation algorithm performs very well. However, with a fewer number of cameras, only providing a partial view of the entire scene, the mesh generation will lead to artifacts as the mesh is not closed properly. This leads to the last type of free-viewpoint rendering works using RGBD data which uses 3D volumetric reconstruction, also called Truncated Signed Distance Functions (TSDF). This technique, popularized by Kinect Fusion [62], is based on a voxel-based 3D volumetric representation of the space. This TSDF representation is filled at every voxel thanks to the RGBD data and implicitly represents the scene surface by means of the voxel subsets whose TSDF value is zero. Shen et al. [146] use this method to synthesize a mirror-like viewpoint, and apply it to render the background behind a person using raytracing on the volume. To render the foreground (i.e., the person), they use forward mapping of the point cloud to the desired mirror viewpoint. Jeya et al. [67] also use TSDF as a part of their pipeline for free-viewpoint video rendering for the aim of monitor animal behavior. They use multiple cameras at 360° around the animal and perform a TSDF reconstruction to then recreate a unified point cloud, that can be visualized from the desired viewpoint. Finally, Maimone et al. [100] completely use the TSDF reconstruction to recreate the color image at the desired viewpoint by performing raytracing on the TSDF for the application of telepresence. This is this last technique that we are going to apply for our setup. Indeed, performing raytracing on the TSDF provides a dense color image whatever the characteristics of the novel viewpoint in terms of field of view or distance.

5.3 Setup

5.3.1 Hardware Architecture

We present our setup with two RGBD cameras placed on each side of the X-ray source as shown in Figure 5.1. Placed at such locations, they do not impinge on the surgical workspace. As it can be seen on Figure 5.1, we offset the two cameras compared to the X-ray source thanks to a wood construction. This serves two goals: first, placed at this distance, their field of view is not partly occluded by the X-ray source; second, our research C-arm has a “naked” X-ray source due to previous experimentations, so offsetting the cameras is a more realistic placement similar to a placement on a “non-naked” X-ray source as on usual C-arm. The end goal of this setup is to be placed on “non-naked” X-ray source C-arm in a “headphone” analogy with no attachment done on the C-arm as shown in Figure 5.2, respecting the objective of minimal engineering on C-arm to avoid recertification.

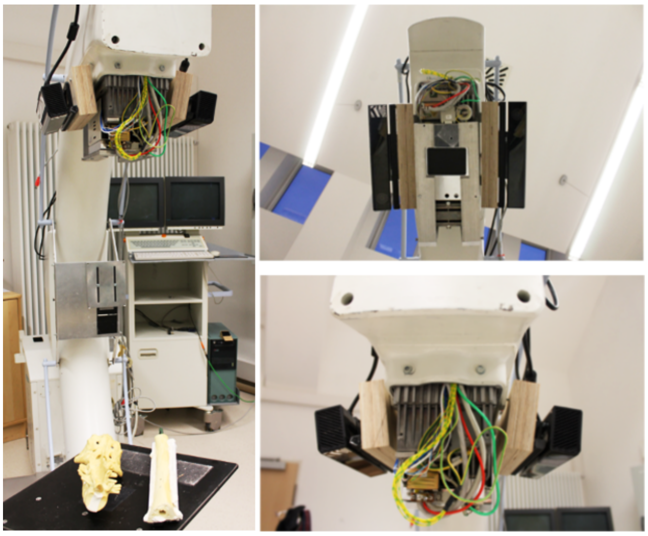


Fig. 5.1. Pictures of the setup with Kinect v2 placed on both side of the X-ray source and wood construction to offset the camera

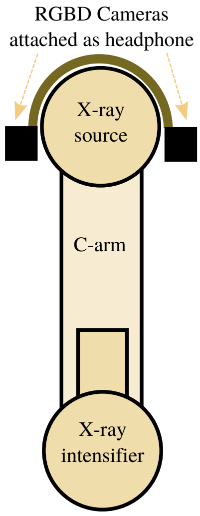


Fig. 5.2. End goal design with RGBD cameras placed as headphone

The C-arm used for this system is a Siremobil Iso-C 3-D from Siemens Medical Solutions. The two RGBD cameras used are Kinect v2. They are placed such as their fields of view are overlapping at the X-ray intensifier. Kinect v2 has been chosen because they do not interfere with each other compared to the Kinect 1.0 type of cameras when the fields of view of the RGBD cameras are overlapping as explained in the Background chapter in Section 3.3.2.

5.3.2 System Architecture

In the Background Chapter in Table 3.2, we have described the characteristics of the Kinect v2. To resume, it outputs video image of 1920× 1080 pixels at 30 FPS and depth images of 512× 424 pixels, same FPS. To access the data, the open-source Libfreenect2 library is used [38]. It gives access to the video image, the infrared image, and the depth image. The library also provides the registration of the depth image on the video image. Each RGBD sensor is connected via USB 3 to a different computer. The Microsoft SDK and the library we used (Libfreenect2) do not support multiple Kinect per computer due to bandwidth limitations, therefore, we need as many computers as cameras. The computers are connected via Ethernet communication, the user interface commands and the orders are emitted from the main processing computer, called PC_0 on which the camera 0 C_0 is attached. The known geometry of the scene, defined by the calibration and user defined parameters such as volume reconstruction dimensions and size, is sent via TCP protocol to the secondary computer, called PC_1 on which camera 1 C_1 is attached, in addition to the start order of the secondary camera (more details on this step in subsection 5.5.1). When the user starts the acquisition from the main processing computer, the data (video, depth and corresponding timestamp) issued by the camera 1 is sent via the PC_1 and TCP protocol to PC_0 that is equipped with a graphic card GeForce GTX 960 from Nvidia. On PC_0, the camera 0 is connected and run in a looping thread to send its latest data to the data processing thread. This data processing thread actually reads the latest data from both cameras and compare them using timestamps. If the time difference between timestamps is smaller than 30ms (which is the camera FPS), the pair of cameras data is sent to the rendering thread, that will create the synthesized color image. We schematize the hardware and software architecture of our framework in Figure 5.3.

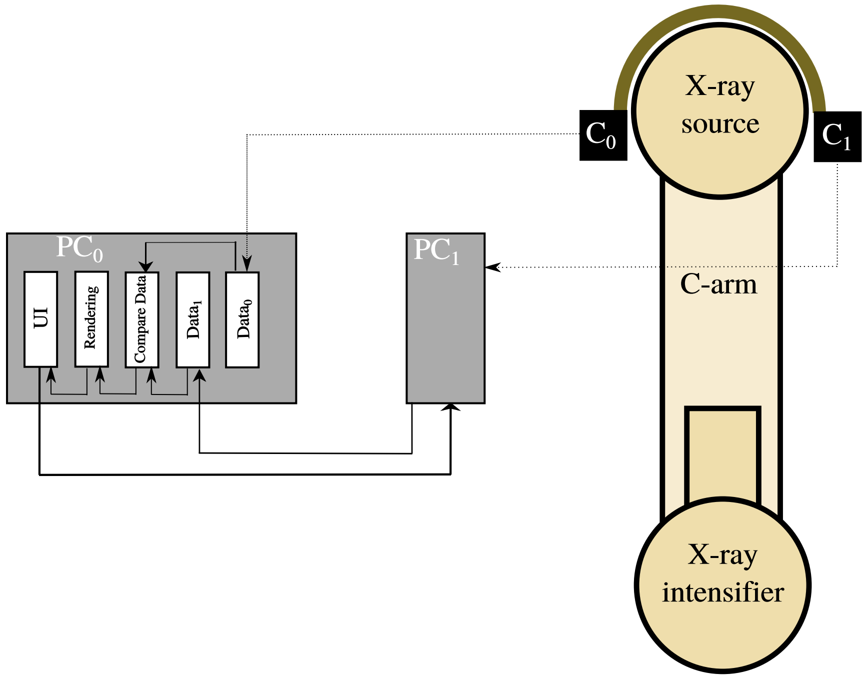


Fig. 5.3. Schematic representation of the hardware and software architecture of our setup

5.3.3 Notations

The X-ray image I_x is acquired from the X-ray source placed at the point $P_x \in \mathbb{R}^3$, I_x is defined on $\Omega_x = [1..480] \times [1..640]$ and maps to $[0..255]$, i.e $I_x : \Omega_x \mapsto [0..255]$. For every RGBD camera $i \in \{0, 1\}$, a color image I_c^i and a depth image I_d^i is defined. The color image I_c^i is acquired from a source placed at the point $P_c^i \in \mathbb{R}^3$, I_c^i is defined on $\Omega_c = [1..1080] \times [1..1920]$ and maps to $[0..255]^3$, i.e $I_c^i : \Omega_c \mapsto [0..255]^3$. The depth image I_d^i is acquired from a source placed at the point $P_d^i \in \mathbb{R}^3$, I_d^i is defined on Ω_d and maps to $\mathbb{N}^+$, i.e $I_d^i : \Omega_d \mapsto \mathbb{N}^+$. As explained earlier, the mapping from depth to video is given by the Libfreenect2 library, therefore, we suppose that $\Omega_d = \Omega_c$ and $P_d^i = P_c^i$ for $i \in \{0, 1\}$

5.4 Calibration

Our system is composed of multiple cameras and the X-ray source, therefore, the first step of our work is to calibrate this complex setup. The different steps are shown in Figure 5.4.

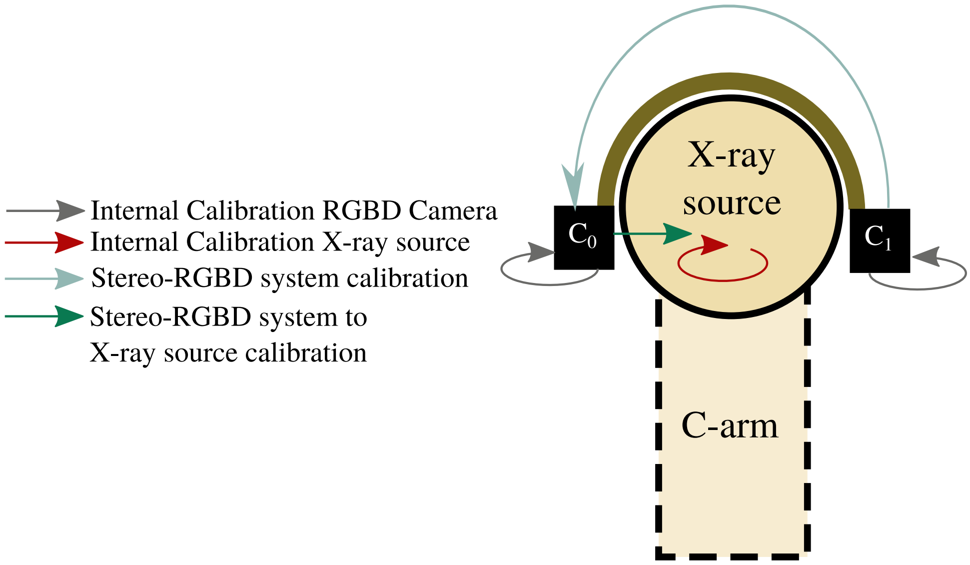


Fig. 5.4. Overview of the calibration process with the four calibrations performed on our setup

5.4.1 Internal X-ray Source Calibration

For internal calibration of the X-ray source, the methods described by Wang et al. [165] are used. A flat grid of X-ray markers is used for both the distortion correction and the intrinsic parameters calculation. The grid is built on a plastic board with metal beads of $3\,\mathrm{mm}$ diameter spaced uniformly every $10\,\mathrm{mm}$.

Distortion Correction

On old C-arms like ours (dating from before the introduction of flat panel detector), the generated X-ray image presents radial distortions following a pincushion shape that can be visualized as a stretching at the periphery of the X-ray image. For the distortion correction, the grid is placed directly on the intensifier. One X-ray image is then acquired. 3 points at the center of the X-ray image are clicked by the user and the distortion-free grid of points (green points in Figure 5.5, $(x_u, y_u) \in \Omega_x$) is created by building an uniform grid from the basis formed by the three clicked points. Using a blob detector algorithm and centroid calculation on the X-ray image, we can get the 2D coordinates of the grid points (pink points in Figure 5.5, $(x_d, y_d) \in \Omega_x$). Using the matching between the distortion-free points and the detected points, a bi-polynomial model of the distortion can be computed from Equation 5.1 by finding the coefficients $P_{i,j}$ and $Q_{i,j}$ for $(i,j) \in [0..M] \times [0..N]$ using the corresponding grid points (x_d, y_d) and (x_u, y_u).

$$
\begin{aligned}
x_d &= \sum_{i=0}^{M}\sum_{j=0}^{N} P_{i,j} x_u^i (1-x_u)^{M-i} y_u^j (1-y_u)^{N-j} \\
y_d &= \sum_{i=0}^{M}\sum_{j=0}^{N} Q_{i,j} y_u^i (1-y_u)^{M-i} x_u^j (1-x_u)^{N-j}
\end{aligned}
\tag{5.1}
$$

We choose M and N equal to 3. For the rest of the work, every X-ray image is undistorted before being processed by the rest of the algorithm.

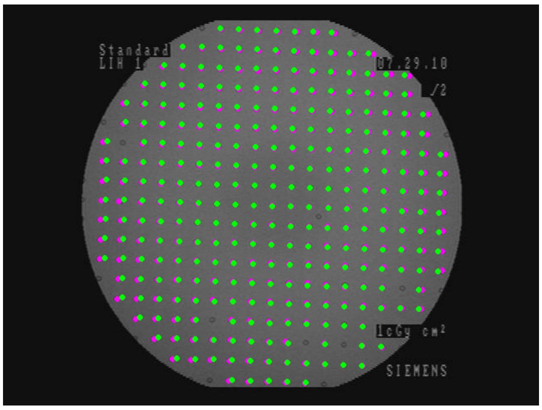


Fig. 5.5. Distortion-free grid in green and detected grid in pink

X-ray source Intrinsic Parameters

The intrinsic parameters of the X-ray source are obtained using the classical Zhang's method [185] from OpenCV on the X-ray marker grid. 15 poses of the grid are acquired, with various angulations and heights. For every image, the user clicks the same three points in order to create a common 3D grid and have a correct 2D image/3D grid matching. The Zhang's method is then applied to obtain the intrinsic matrix of the X-ray source.

5.4.2 Stereo-RGBD System Calibration

In this section, the stereo system composed of the two RGBD cameras is calibrated. The goal is to calculate the rigid transformation $T_{0\to 1} = (R_{0\to 1}|t_{0\to 1}) \in SE(3)$ from camera 0 C_0 to camera 1 C_1 as well as computing the internal parameters of the RGBD cameras. Our method consists simply of acquiring multiple views of a checkerboard seen in the two video cameras as shown in Figure 5.6 and apply Zhang's stereo calibration [185], this provides us with the intrinsics parameters of the RGBD cameras and the pose between the cameras.

Fig. 5.6. Checkerboard view from (Left) Camera 0 C_0 and (Right) Camera 1 C_1

Now that the two RGBD cameras are registered together, we will consider them as one system, called stereo-RGBD cameras, that needs to be registered as a whole to the X-ray source. For that, we place the coordinate system of the stereo-RGBD cameras setup in the coordinate system of the camera C_0 as the reference.

5.4.3 Stereo-RGBD System to X-ray Source Calibration

For calibration from the stereo-RGBD cameras system to the X-ray source, we employ the method from Wang et al. [173], except that we compute the transformation $T_{x \to 0} \in SE(3)$ between the X-ray source and one of the RGBD camera (C_0 in our case) instead of the projection matrix as done by Wang et al. This method uses a blue and white checkerboard pattern, rigidly attached to a grid of X-ray markers. The checkerboard square size is chosen as a multiple of the X-ray grid spacing. The design is made such that, given the 3D positions of the checkerboard corners, the 3D positions of the X-ray markers are easily retrievable, as they are a sub-multiple of the checkerboard spacing. The details of the calibration board can be seen in Figure 5.7 with the left image showing the checkerboard side and the right image showing the X-ray grid (which is, in fact, visible on both sides through X-ray).

Fig. 5.7. Images of the checkerboard used by Wang et al. [173] in the (Left) video and the (Right) X-ray image

Video images and depth images of several orientations of the checkerboard are acquired, as well as X-ray images. The fact that we use depth image has conditioned our choice of the blue color instead of black for the checkerboard, due to the fact that very dark surfaces have been shown by Lachat et al. [80] to have artifacts in the depth image with Kinect v2. The blue chosen is light enough to not provoke artifacts but contrasts enough for checkerboard detection.

Using the depth images of the checkerboard, point clouds are generated for every orientation of the checkerboard. To clean the checkerboard point clouds of noise, we perform a RANSAC plane fitting (from PCL library [1]) on every point cloud. This gives us for every point cloud a plane $\mathcal{M}$ of coefficients (a, b, c, d) defined as:

$$\mathcal{M} : aX + bY + cZ + d = 0 \tag{5.2}$$

For three extreme corners of the checkerboard detected in the video $p_i = (u_i, v_i)$ with $i \in [0..2]$ (those indexes are sorted such as each next point is the closest neighbor of the previous one), we extract their 3D coordinates $P_i = (X_i, Y_i, Z_i)$ in the RGBD camera coordinate system according to Equation 5.3 using the plane $\mathcal{M}$ and the knowledge of the intrinsic parameters of the RGBD camera.

$$X_i = \frac{u_i - u_0}{f}, Y_i = \frac{v_i - v_0}{f}, Z_i = -\frac{d + bY + aX}{c} \tag{5.3}$$

1. http://pointclouds.org/

with (u_0, v_0) the principal point and f the focal length of C_0. Therefore, we can retrieve the 3D coordinates of the X-ray markers which are only a sub-multiple of the 3D grid made of the three checkerboard corners by following Equation 5.4.

$$P_{xray} = \alpha\frac{(P_2 - P_1)}{s} + \beta\frac{(P_0 - P_1)}{s} \tag{5.4}$$

with s the sub-multiple order of the X-ray grid and $(\alpha, \beta) \in [0..s]^2$.

In the 2D X-ray images of the calibration board, a blob detector is applied in order to retrieve the center of every marker, then the user is asked to click the 3 markers on the X-ray images, corresponding to the three checkerboard extreme corners chosen previously (with the same order as the sorting performed). This allows pairing the 2D coordinates of the X-ray marker in the X-ray image p_{xray} to their 3D coordinates in the RGBD camera coordinate system P_{xray}. As we have calibrated earlier the intrinsics parameters of the X-ray source, we use the Ransac PnP method [17], which computes the extrinsic matrix, ie. $T_{x\to 0}$, using correspondence between 2D/3D points, instead of using the DLT algorithm for obtaining the projection matrix as performed by Wang et al.[173].

At this stage of the work, we know every relationship between the RGBD cameras and the X-ray system.

5.5 Space Efficient TSDF Generation

Once the full system geometry is known thanks to the calibration procedures, the image synthesization using volumetric reconstruction can be performed.

The 3D volumetric reconstruction is based on a 3D grid g, $g : [0..N]^3 \mapsto G \in \mathbb{R}^3$, composed of N voxels of size s_z in the three directions, placed orthogonally along the optical axis of the virtual viewpoint, ie. the calibrated X-ray source location in our work, at a distance d as it can be seen in Figure 5.8. Thanks to the calibration steps previously done between the virtual viewpoint and the RGBD cameras, the relationship between this 3D grid G and the RGBD sensors is known. We also know, thanks to user input at the start, the parameters such as d, N and s_z.

The default values used for this work are based on the characteristics of our setup. As we want to reconstruct the surgical scene around the intensifier, this sets $d \approx 90\,\mathrm{cm}$, as the intensifier is approximately $30\,\mathrm{cm}$ wide, this sets $Ns_z \approx 30\,\mathrm{cm}$. The grid maximum size N is ≈ 500 voxels due to computer memory (whether it is in CPU or GPU). Therefore, we use $s_z \approx 0.75\,\mathrm{mm}$. Those values can be changed in the user interface by the user.

5.5.1 Pre-Processing of RGBD Data

The calibration values are known before starting synthesization, as well as the user parameters (d, N, s_z). This gives upfront the full knowledge of the 3D grid location in space compared to the RGBD camera, which can be used to minimize the processing on the RGBD camera data to the minimum necessary. For our work, a grid big enough to visualize the surgical scene only projects itself into 10-20% of the video image due to the fact that the video camera is wide-angle and the X-ray source has a small field of view. Most of the RGBD data is, therefore, useless for our pipeline. To avoid pointless data transmission and processing, we compute the bounding box of the projection of the 3D grid G in the two RGBD cameras. First, we project the eight outer vertexes of G to the RGBD camera space and then compute the 2D convex hull of those eight 2D points (green polygon in Figure 5.9). As the perspective projection does not break convexity, all points inside the grid G will, therefore, also lie inside this convex hull. The bounding box of this convex hull is then computed to create an axis-aligned cropping (red ROI box in Figure 5.9). Only the data inside the bounding box is transmitted, for each camera, decreasing, therefore, the transmission latency of our system.

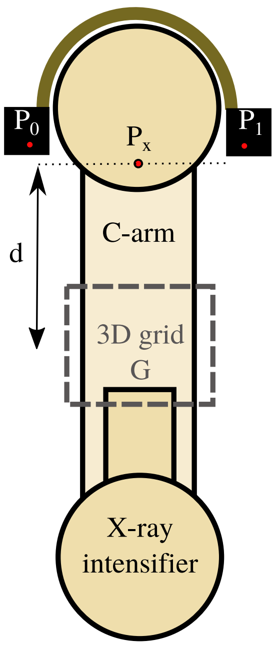


Fig. 5.8. Position of the grid G compared to the C-arm

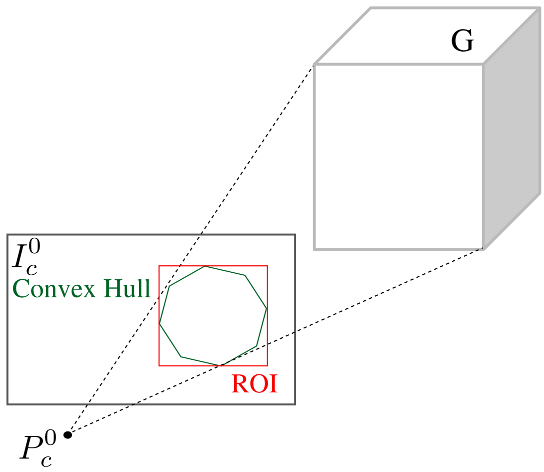


Fig. 5.9. Vertexes projections into the image, with Convex Hull (in green) and the ROI (in red) computed

5.5.2 Volumetric Reconstruction Using TSDF

The volumetric reconstruction consists of computing the function Truncated Signed Distance Field (TSDF) $f : G \mapsto \mathbb{R}$ which maps a 3D point $\mathbf{x} \in G$ to a truncated signed distance value. A graphical representation of the volumetric reconstruction step is shown in Figure 5.10. The TSDF value is the weighted average of truncated signed distance values $v_0(\mathbf{x})$ and $v_1(\mathbf{x})$ computed respectively in the two RGBD cameras. Therefore, the computation of f follows the Equation 5.5.

$$f(\mathbf{x}) = \begin{cases} \frac{w_0(\mathbf{x})v_0(\mathbf{x})+w_1(\mathbf{x})v_1(\mathbf{x})}{w_0(\mathbf{x})+w_1(\mathbf{x})} \text{ if } w_0(\mathbf{x}) + w_1(\mathbf{x}) \neq 0 \\ 1 \text{ else} \end{cases} \tag{5.5}$$

where w_0 and w_1 are the weights associated with each camera. The weights aim at discarding truncated signed values according to certain conditions (described in Equation 5.6) such as

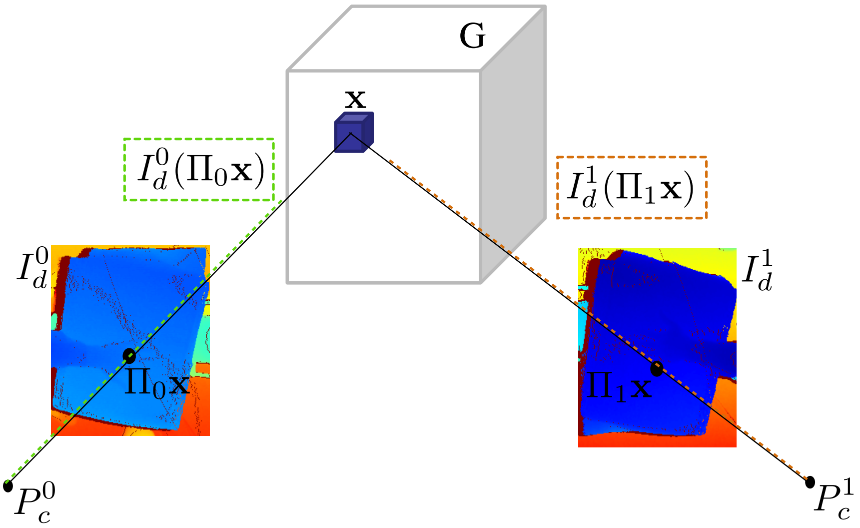


Fig. 5.10. Schematic representation of the volumetric reconstruction step

voxels too far from the surface according to a certain threshold. For each camera $i \in \{0, 1\}$, the weights $w_i(\mathbf{x})$ for each truncated signed value are computed using Equation 5.6.

$$w_i(\mathbf{x}) = \begin{cases} 1 \text{ if } I_d^i(\Pi_i\mathbf{x}) - ||\mathbf{x} - P_c^i|| < -\eta \\ 0 \text{ else} \end{cases} \tag{5.6}$$

where Π_i the projection matrix from G to RGBD camera space Ω_c^i. η is a tolerance on the visibility of $\mathbf{x}$ (we fixed η=1.5 cm). For each camera $i \in \{0, 1\}$, $v_i(\mathbf{x})$ geometrically represents the truncated value computed from the difference of the distance from $\mathbf{x}$ to P_c^i with the depth value obtained by projecting $\mathbf{x}$ into camera i. The truncated signed distances $v_i(\mathbf{x})$ are computed according to Equation 5.7.

$$v_i(\mathbf{x}) = \phi(I_d^i(\Pi_i\mathbf{x}) - ||\mathbf{x} - P_c^i||) \text{ with } \phi(s) = \begin{cases} sgn(s) \text{ if } \frac{|s|}{\delta} > 1 \\ \frac{s}{\delta} \text{ else} \end{cases} \tag{5.7}$$

with δ being a tolerance parameter to handle noise in depth measurements which depends on the specific characteristics of the RGBD sensor (we use here: δ=2 mm).

We illustrate in Figure 5.11 the evolution of the truncated signed distance according to the distance to the surface. The surface is represented in the TSDF grid with $f = 0$. The truncation allows focusing only the TSDF computation on the interesting values, the one close to zero.

Alongside with the TSDF f, we also create a volumetric color field $f_c : G \mapsto [0..255]^3$ following Equation 5.8.

$$f_c(\mathbf{x}) = \begin{cases} \frac{w_0(\mathbf{x})I_c^0(\Pi_0\mathbf{x}) + w_1(\mathbf{x})I_c^1(\Pi_1\mathbf{x})}{w_0(\mathbf{x}) + w_1(\mathbf{x})} \text{ if } w_0(\mathbf{x}) + w_1(\mathbf{x}) \neq 0 \\ (0, 0, 0) \text{ else} \end{cases} \tag{5.8}$$

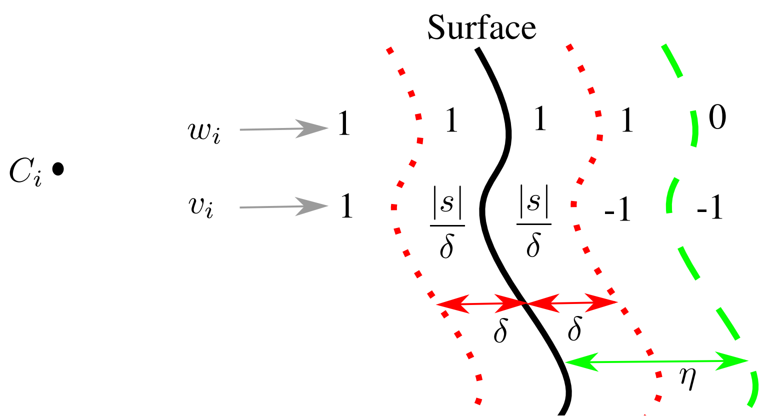


Fig. 5.11. Truncated Signed Distances Value and Weight according to distance to the surface in one camera

5.5.3 Occupancy Grid for Faster Reconstruction

To perform the volumetric reconstruction, we need to compute the TSDF values for every voxel. At this stage of the work, the entire grid needs to be visited voxel by voxel for performing so. We first implemented this on GPU using CUDA library, launching one kernel for the full grid reconstruction. However, the approach is not fast enough for a use with surgeons. Actually, only a few voxels represent the surface and are, therefore, of interest in the grid. Indeed most of the voxels in the grid are filled by default values $f = 1$ and $f_c = (0,0,0)$, showing that they are far from the surface, those voxels are either excluded by the weights criteria or respect the first line condition in the truncated value in Equation 5.7. Also, it is interesting to notice that the volumetric reconstruction respects continuity. If one voxel is far from the surface, then its neighbor will be too. Therefore, the "uninterested" space should be localized in block neighborhood that if detected before reconstruction would allow faster computation if they are skipping in the search, only focusing on the voxels close to the surface. To realize this, we use the idea of occupancy function f_o that maps the 3D grid $f_o : G \mapsto \{0,1\}$ to a binary map of occupancy. Before performing the TSDF computation, using the cropped depth and video images for both cameras and their registration, we can compute the point cloud $\mathcal{P}$ by merging the respective points clouds from the two cameras and transform it to the 3D grid space G, all in real-time. The point cloud is a sparse indication of where the surface should lie in and is, therefore, used to fill our occupancy function f_o. The grid G is subdivided in α^3 smaller subgrids $SG \subset G$ of size $n = \frac{N}{\alpha}$ voxels, with $\alpha > 1$. Then, using Equation 5.9, we can define the occupancy function f_0 at the voxel $\mathbf{x} \in SG$.

$$f_o(\mathbf{x}) = \begin{cases} 0 \text{ if } \mathcal{P} \cap SG = \varnothing \\ 1 \text{ else} \end{cases} \tag{5.9}$$

Then, for the 3D grid TSDF reconstruction, the computation is only done for the subgrids with $f_o = 1$, with one kernel launched by subgrid. The voxels inside subgrid with $f_o = 0$ are filled with the TSDF default values and, therefore, does not require GPU computation.

5.5.4 Raytracing

Raytracing can be efficiently used for rendering when dealing with implicit surfaces. This method searches for all 3D points $\mathbf{x}$ for which the condition $f(\mathbf{x}) = 0$ holds. For every pixel in our virtual view image I_s, we define a ray of unitary direction $\vec{r}$ from the optical center of the virtual camera P_x to the pixel center. The search starts by testing the TSDF grid at the point position $P = P_x + n\gamma\vec{r}$ with $n \in \mathbb{N}$, as we illustrate in Figure 5.12. To speed up the search, we use large values of γ ($\gamma = 10s_z$) and perform a binary search over several iterations. If during one n^{th} step $f(P_{curr}) < 0$ and $f(P_{prec}) > 0$, we refine this result by testing $P = \frac{P_{curr}+P_{prec}}{2}$. We then reiterate this process for the segment composed of P and the point (P_{prec} or P_{curr}) of opposite sign. At the end of the iterations, the last P is considered as the 3D point on the surface. The color at the pixel in the synthesized image I_s is the color at P in the color volumetric representation $f_c(P)$. A depth value d can also be computed as $d = ||P - P_x||$.

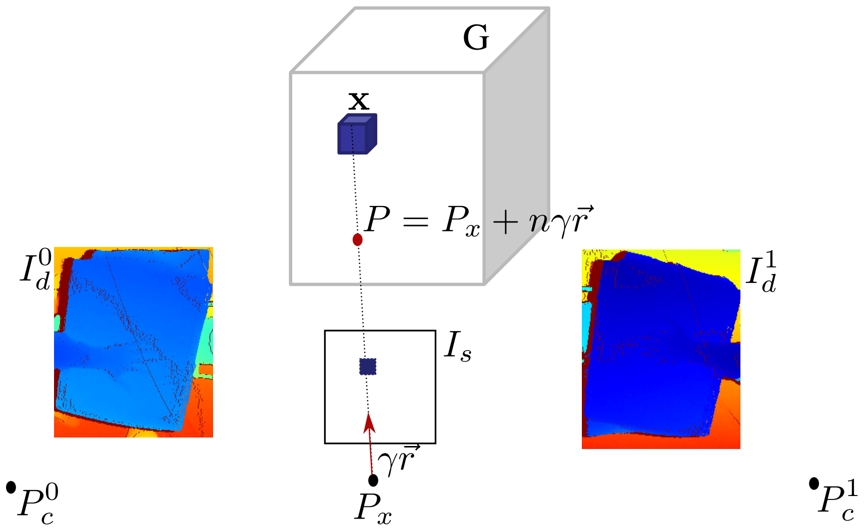


Fig. 5.12. Schematic representation of the raytracing steps

5.6 Results

We have evaluated our system on the technical side first and then, we performed a pre-clinical study.

5.6.1 Technical Evaluation

Calibration Errors

We first compute the RMS error for every part of our calibration chain. We compare our results with the RMS error obtained with similar components in the literature and report them in Table 5.1 The results show that every link of our calibration chain present comparable or better results than the literature.

	Intrinsic X-ray	Stereo-RGBD	Stereo-RGBD to X-ray
RMS Error (in pixel)	0.37	0.42	0.79
Reference RMS Error (in pixel)	0.48 [165]	< 0.5	1.23 [173]

Tab. 5.1. RMS error for every link of the calibration chain

Overlay Error

As the aim of our work is to provide an accurate overlay of X-ray image over synthesized video image, this accuracy is tested in the following paragraph. The influence of the C-arm rotation over this accuracy is also tested.

We attach a grid used for the internal X-ray calibration with markers visible in X-ray image and video image at the intensifier. For several angular and orbital rotations of the C-arm, the marker centers are detected in the X-ray image and the synthesized video image using blob detection and matched. The mean error distance between markers in video and X-ray image is then computed for every pose, as well as its standard deviation, are compiled in Figure 5.13.

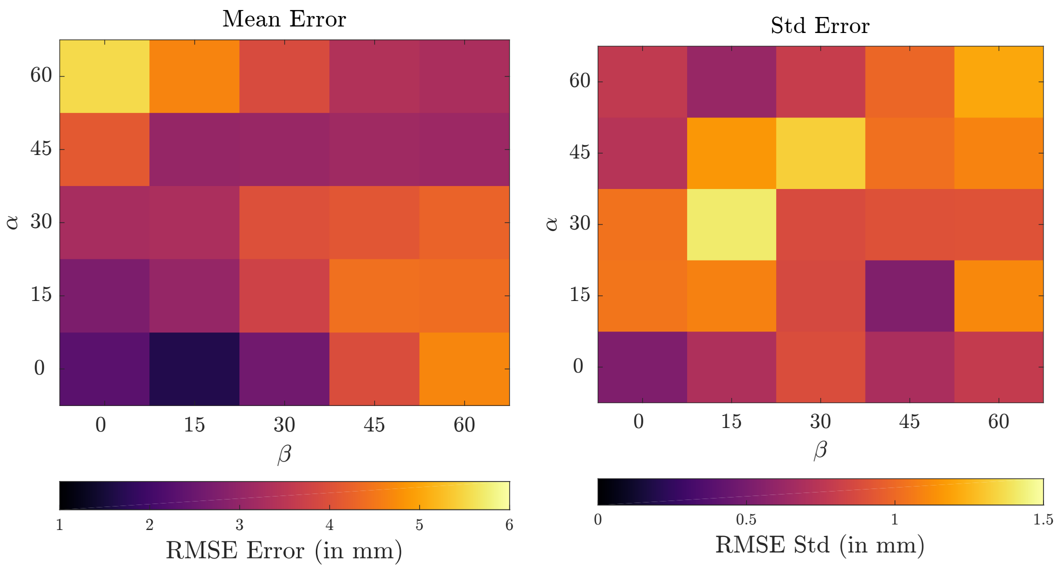


Fig. 5.13. Heatmap of (Left) mean overlay error and (Right) standard deviation of overlay error for several orbital α and angular β angulations

The mean overlay error is comprised between $2\,\mathrm{mm}$ and $6\,\mathrm{mm}$ over the poses. The error is minimal around the poses of small orbital and angular angles and increases with the angles increase. Our minimal error is higher than the minimal error of Navab et al. [111] which is $0.5\,\mathrm{mm}$. Our system is at best $2\,\mathrm{mm}$ precise.

Analysis about the Minimal Overlay Error

Our system is not as precise as Navab et al. which is not a surprise, as our overlay is possible through "software registration", which in our case includes six steps of calibration (four performed by us, plus the two RGBD camera internal calibrations). Each step comes with its imprecision, as shown in the previous subsection, even though we try to keep each step individually at the state of the art level. This should be compared to the calibration process from Navab et al. which consists of two steps ("hardware registration" of video camera to the X-ray source and homography calculation). However, our setup is a lot more flexible than Navab et al. as we calibrate it as it is built. Our calibration can be applied to any configuration of cameras placed on the side of the X-ray source while Navab et al. require the video camera to be placed at exactly one position which is very constraining for the C-arm design. Our long calibration pipeline is a consequence of this flexibility and, therefore, the accuracy of our setup suffers from it too. In general, we see our setup as an **alternative** to CamC and not a replacement. When it is not possible to modify the C-arm, we think it is still better to have a less precise overlay than no overlay at all.

Influence of Angles

For the increase of overlay error with the angles, the same increase of error is visible with CamC in the results provided by Navab et al. [111] and Wang [165]. Their overlay error for $(\alpha, \beta) = (60, 0)$ could go until $6\,\mathrm{mm}$, which is similar to our results. In their work, we can observe a stronger increase only with the orbital angles, we do not observe this in our setup as it increases sensibly the same for both angles. To explain their results, Navab et al. [111] describe that mechanical sagging, provoked for example by gravity, can happen in between the X-ray source and the intensifier when the C-arm is moved. This makes the intrinsics of the C-arm pose dependent. The same applies to our setup as we have calibrated our C-arm at the pose $(\alpha, \beta) = (0, 0)$, it explains we get our best results around this pose. To compensate those changes in intrinsics, they use the concept of Virtual Detector Plane described previously in [110]. They use X-ray markers placed close to the X-ray source on the housing (e.g. on the mirror in Wang et al. [165]) but still visible at the border of the X-ray image. Their positions in the 2D X-ray image at the X-ray intrinsics calibration C-arm pose is taken as reference and for any other pose, the X-ray image is wrapped such as those points are placed at the reference. They explain that this ensures fixed intrinsics compared to the calibration pose. Unfortunately, we could not use this method on our current setup as no support is possible for the X-ray markers that should be placed on the housing in the X-ray field of view. This is only due to the "naked" source nature of our C-arm. On a normal C-arm, as we target our setup to be used on, those X-ray markers could be simply glued on the plastic housing that is covering the X-ray source. We think that, with this improvement, the error should stabilize around the minimal error such as it occurs with Navab et al. This remains to be established in an additional study. Also, Navab et al. [111] explain that newer generation of C-arms (at the time of the paper in 2010) includes encoded projection matrices for every orientation of the C-arm, e.g. for reconstruction purposes. This information could be used by our system to adapt at every pose the synthesization of the video image. Indeed, our raytracing algorithm uses the X-ray source intrinsics and extrinsics which could be changed according to the C-arm pose and should reduce the overlay error.

Time Performance

To be usable during a medical procedure, we brought the system as close as possible to a real-time system, with low latency and as high as possible framerate. The latency is approximately around 40 ms, possible thanks to the data pre-processing step. From the data reception, the algorithm takes in between 80 ms to 200 ms to reconstruct a synthesized image (depending on the subdivision factor). While the framerate of 5 FPS results in a lagged visualization, a framerate of 8-10 FPS is enough to overcome this issue. We show in Figure 5.14 the details of computation time for each rendering step for the configuration with and without occupancy grid. We can observe the important gain in volumetric reconstruction when using the occupancy grid, justifying the extra cost of 5ms to create the occupancy grid.

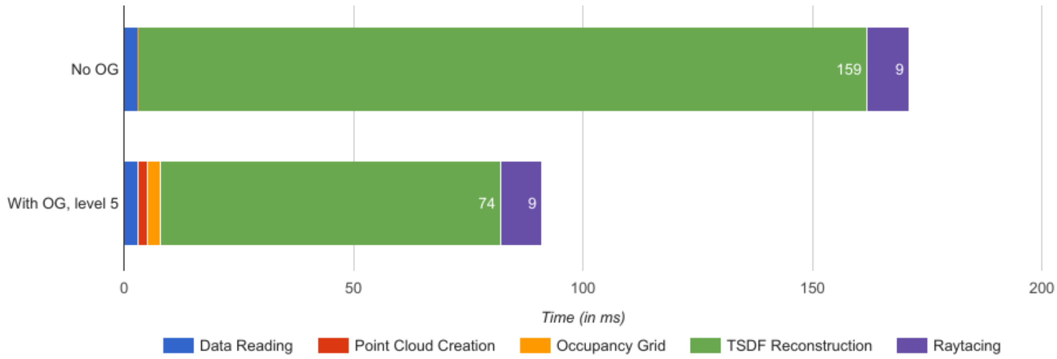


Fig. 5.14. Computation Time in ms for the different steps of the Rendering pipeline in the case with and without Occupancy Grid (OG)

Space Search Efficiency

We focus in this section on the TSDF reconstruction time reduction bought by the occupancy grid scheme. We show in Figure 5.15 the time of reconstruction according to the grid space percentage that went through TSDF reconstruction. This plot is realized from two experiments: 1) Experiment #1 with a scene made of a flat $5\,\mathrm{cm}$ thick object; 2) Experiment #2 same scene as Exp#1 but with a hand moving around the reconstruction volume. The first scene should have a low percentage of space reconstructed while the second one, the hand makes it more diverse in the percentage. For every scene, the image synthesization has been run for several subdivision grid level $\alpha \in \{1, 2, 3, 4, 5, 8, 10\}$. A colorbar allows finding the corresponding subdivision level for every acquired measurement.

For Experiment #1, an important part of the space is unoccupied and, therefore, less than 60% of the space is used in most trials of this experiment. For Experiment #2, the presence of the hand results in more space being occupied. For subdivision higher than 3, the occupied space is, however, lower than 70 %. Indeed the presence of the hand does not cover the full space and some void is still present which is more likely to be found as the subdivision level increases. Whatever the scene type, a subdivision of level 5 already allows a reduction of at least 30% of the reconstruction time. In a particularly empty scene, this can be achieved with a level 2, while a level 5 can bring a 60% reduction in time. We can clearly see from Figure 5.15 that the TSDF reconstruction time follows a linear relationship with the percentage of occupancy. We show as a dotted line the linear regression performed on the data, the

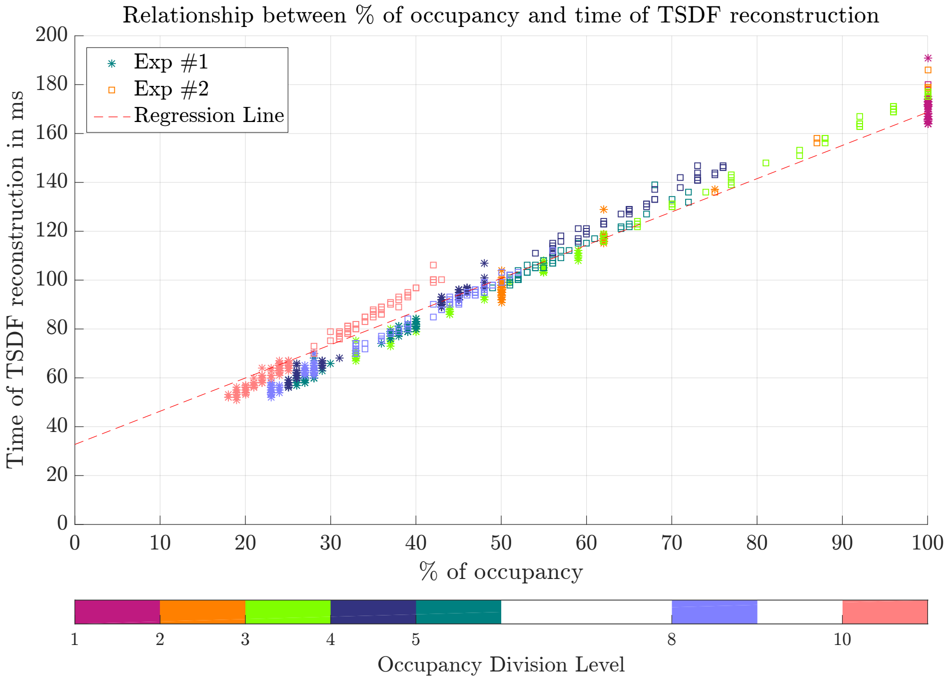


Fig. 5.15. Time of reconstruction according to the percentage of space used

correlation coefficient for that regression is $r = 0.9853$, which confirms this strong correlation. The type of scene does not matter in this relationship.

5.6.2 Pre-Clinical Study

We performed a pre-clinical study using our system. Like the Navab et al system [111], our system is ideal for surgeries requiring positioning down the beam the surgical tool such as interlocking of intra-medullary nails [35] or pedicle screw placements [111]. We use our system for another down the beam procedure, not tested yet with the system of Navab et al.: the lumbar facet joint injection.

Facet Joint Injection Procedure

Facet Joint Injection (FJI) is performed in the context of low back pain, which is an extremely common affliction whose lifetime prevalence is 60-70 % in industrialized countries [37]. Facets joints are suspected to be the source of chronic back pain for 10-15% of the cases [30]. Facets joints are a synovial joint that perform the articulation between adjacent vertebrae, restraining them in their relative motion. As visible in Figure 5.16, the joint is composed of the articular cartilage covering the adjacent articular processes, surrounded by synovium tissue, encapsulated inside the joint capsule.

The FJI is the common procedure to relieve patients of facet joint pain. This procedure consists of injecting through a needle a solution of local anesthetic and cortisone into the facet joint

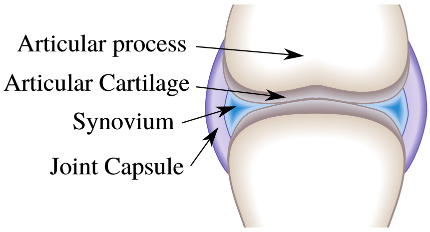


Fig. 5.16. Anatomy of Facet Joint, by Madhero88, distributed under CC BY-SA 3.0 license Link to Image

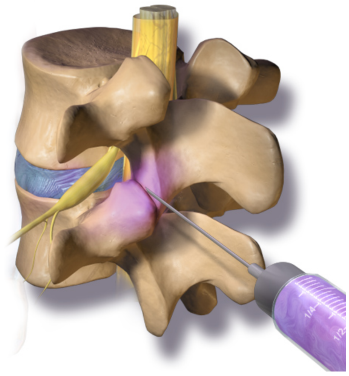


Fig. 5.17. Facet Joint Injection, distributed under CC BY 3.0 license, from [179]

as it can be seen on Figure 5.17. This procedure is performed ambulatory and the patient is only locally anesthetized. Although ambulatory, the procedure is most often performed under fluoroscopic guidance [127] to find the oblique view where the joint is parallel to the articular processes and then, to guide the needle to the facet joint. In parallel, during any fluoroscopic procedure, X-ray radiation exposure of the surgical staff is an important concern that should be addressed by following the ALARA principle: "As Low As Reasonably Achievable" as advised by the FDA [105]. During FJI, the main radiation exposure for surgeons is during the needle guidance step when they must often keep their hands in the beam during fluoroscopy while maneuvering the needle. C-arm positioning can be performed remotely, minimally exposing the surgeon. Therefore, we think our system can answer the radiation concern during needle insertion. The visibility of the target point in X-ray image with the surgical context of the video (hand and tool) should lead to reducing the needle insertion to one X-ray image. In comparison, Proschek et al. [127] use 4.53 images per joint for fluoroscopy-guided FJI. As the X-ray image is not used for needle navigation, but only for targeting, the same X-ray image can be re-used if several targets are present in the image. For lumbar facet joint, this assumption is true. On one spine side, three to four joints are visible and aligned in one image. Using our system, this reduces, even more, the number of X-ray images necessary for the procedure.

Pre-clinical Study Design

We perform a pre-clinical study performing FJI on a lumbar spine X-ray opaque phantom (LS01X model from Creaplast). In addition to the spine phantom, we also simulate the facet joint anatomy. We fill the facet joint "spaces" on the phantom with transparent silicone gel and create a bridge between the two articular processes on the lumbar spine phantom in order to resemble the facet joint anatomy as shown in Figure 5.18.

Silicone presents close stiffness property (in terms of Young's modulus) to facet joint [91]. The spine phantom with silicone facet joints is then covered with gelatin (at concentration 1:10 ratio to water) to simulate soft tissue and a plastic soft cover simulating the skin. Our partner expert surgeon has confirmed that the haptics of our phantom is close to the patient condition. The participant is asked to perform the procedure on three vertebrae levels (L1, L2, L3), i.e. on six facet joints using a G22 needle ($\oslash$ 0.7 mm). For each side, only one X-ray image should be necessary for target and, therefore, we take as a trial condition that only one X-ray image is

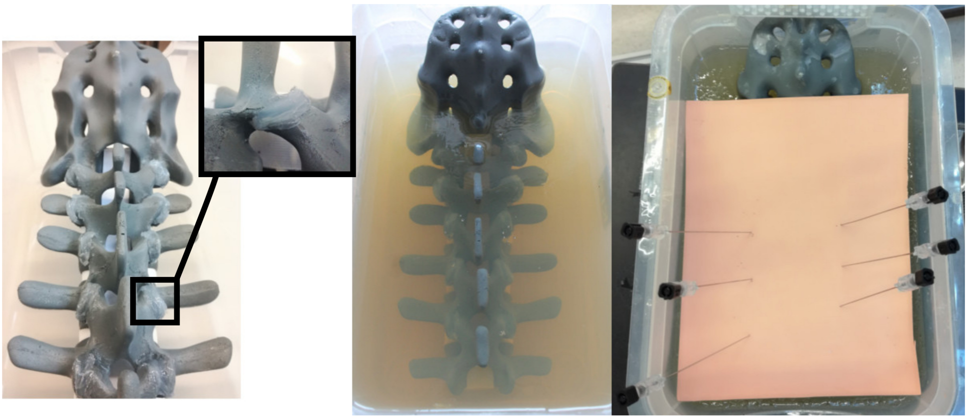


Fig. 5.18. Details on the phantom, (Left) Facet Joint simulated by silicone, (Center) Gelatin to simulate soft-tissue, (Right) Fabric cover to simulate skin

acquired per side, which means only two X-ray images are acquired by trial. For validating the final position of the needles, we keep the six needles placed by the participant on the phantom and perform a CT acquisition. First, an expert surgeon grades if the placement is a success. If the placement is failed, we compute the distance of the needle tip to the facet joint.

One expert surgeon performing daily facet injection has performed two full trials with our system, we also performed two control trials with X-ray images only to compare. We acquire the number of X-ray images acquired, the time for three needle placements per side.

Performance Results

Table 5.2 shows the results of the pre-clinical study. We can see that the trials have been successful at 84 %. Even when the trial fails the needle was close to the target at around $5\,\mathrm{mm}$. The control trials are successful at 100%, but at the cost of numerous X-rays as it can be seen on the fourth row of Figure 5.2 with, on average, 8.25 X-ray images acquired when only the X-ray image can be used for needle placement. The participant is faster using X-ray only (around 1:20 min for three needles placement compared to 2:09 min in average with our system). We can explain this result as any ambiguity in placement can be solved by shooting an X-ray image in the control test which was not possible with our system. The participant would then need longer to be sure of the placement using our system. We can conclude from this preliminary pre-clinical test that our system allows for high reduction of the number of X-ray images at th cost of a small loss in needle placement success.

Feedback & Discussion

This pre-clinical study is a preliminary test for a more extensive pre-clinical study. Based on the feedback of our participant and other surgical experts, the limit of one X-ray image per side is seen as too restrictive and they would like to acquire several check X-ray images. Knowing that those X-ray images can be re-used for needle target navigation, the number of X-ray images used for the procedure should then still be reduced compared to X-ray image only navigation. Our surgical experts believe that the success rate would then be close to

	Left Side				Right Side			
	T_1	T_2	T_1^c	T_2^c	T_1	T_2	T_1^c	T_2^c
L1	o	o	o	o	o	o	o	o
L2	x (6.5)	o	o	o	o	o	o	o
L3	o	x (4)	o	o	o	o	o	o
# X-ray	1	1	10	6	1	1	9	8
time (in min)	1:51	1:14	1:35	1:10	2:18	3:15	1:2	1:16

Tab. 5.2. Results of the pre-clinical study: o is a success, x is a fail with distance to the target in bracket. T_i is the i^{th} trial with our system, T_i^c is the i^{th} control trial

100%, while still reducing the number of X-ray images acquired. This remains to be proved with an extensive pre-clinical study.

Another lesson from this preliminary test is the change of behavior between the trials with our system and the control tests regarding the acquisition of X-ray image and the surgeon positioning. During the trials with our system, the surgeon would take the X-ray image while being away of the C-arm and then start the needle placement. While during the control trials, the participant would often acquire X-ray images with hands in the beam, himself being close to the beam. This change of behavior represents eventually, even more, radiation exposure reduction that the decrease of X-ray images number can bring as the radiation rapidly decrease with the distance to the source and no direct exposure happens. With the metrics acquired during our preliminary test, this change is not quantifiable. For a more extensive pre-clinical study, this change of behavior should be measured by tracking the participant compared to the C-arm using, for example, the skeleton tracking from the Kinect. The radiation exposure of the participant can also be measured using a dosimeter (in the limit of their precision). Both metrics should enlighten us strongly if our system reduces radiation exposure to the surgeon during facet joint injection procedure. Finally, the participant would appreciate if our system could provide a tool axis display or an indicator of down the beam alignment like provided by Diotte et al. [35] to facilitate the navigation from the time the needle enters the patient body.

5.7 Discussion

The visualization with our system is not sharp as a real image would be. For example, during our pre-clinical study, even though we could see a $0.7\,\mathrm{mm}$ needle with the system, it was sometimes difficult for the surgeon to navigate it. The image quality is linked to the RGBD cameras resolution and field of view. As mentioned before we only use 10% of the video image of the camera, in the end, we work with a low-resolution image to reconstruct to synthesize our view. Higher resolution cameras or camera with a smaller field of view of a similar resolution would allow us to have more information for the same surface, resulting in a sharper image. As another current limitation, the accuracy of the Kinect v2 depth estimation is low for medical standards with an axial error in the millimeter domain and inaccuracy at the edges. This has an influence on the image quality by potentially misplacing surface voxels in the TSDF, which can also influence the overlay accuracy by not placing the reconstructed object at the right place. We can note that even with a lower image quality, we could perform a study that captures successfully a $0.7\,\mathrm{mm}$ needle which is under our grid resolution $(0.75\,\mathrm{mm})$. The technology presented in this chapter is not mature for integration in the OR due to the current limitations of RGBD cameras. However, we think that the ideas disseminated in this chapter and the successful preliminary pre-clinical study show the full potential of the technology in the future.

5.8 Conclusion

In this chapter, we have presented the mirror-less RGBD augmented C-arm. This new setup, using two RGBD cameras affixed on the side of the C-arm source, can produce a similar overlay output as the mirror-based RGBD augmented C-arm setup but with minimal disruption on the C-arm housing which should facilitate the integration in the OR. The new system was fully described from software-hardware architecture to calibration, validated and also used for a pre-clinical study of Facet Joint Injection.

Part III

Medical Applications of RGBD Augmented C-arm

6 3D Visualization with 2D X-ray image

The work presented in this chapter is an extended version of one part of the paper presented at ISMAR 2015 which is reproduced with permission from [53], ©IEEE.

In Chapter 4, we have introduced an RGBD augmented C-arm that extends the video augmented C-arm from Navab et al. [111] with depth information. The presented setup has the same features as the video augmented C-arm regarding exact overlay of X-ray image over video image. However, it also provides the depth value of the imaged object at every color pixel. In this chapter, we will explain one application of the depth data for the RGBD augmented C-arm which is the overlay of the X-ray image over a 3D reconstruction of the surgical scene computed from RGBD data.

6.1 Introduction

The main advantage of the overlay of X-ray image over video from Navab et al. [111] is to give context about the external anatomy to the X-ray image. It allows precise localization and gives precious clues on where to target on the external patient surface to reach a desired internal point with the minimal amount of X-ray images. However, this contextualization is limited to the field of view of the video camera and only gives local information, which is enough for targeting. It lacks more global external information such as the C-arm position compared to the patient. The search for the next best X-ray view compared to the patient external surface is still a mental exercise that remains to be done. To solve this problem, a global 3D view of the scene is necessary. A 3D view would allow contextualizing in space altogether the C-arm position, the X-ray image, and the patient.

The pipeline to perform this global 3D visualization is shown in Figure 6.1. First, we reconstruct in 3D the surgical scene placed in the C-arm workspace, just before the surgery starts, by rotating the C-arm around the scene (Figure 6.1-step 1). During surgery, the surgeon positions the C-arm at the desired position and eventually acquires an X-ray image (Figure 6.1-step 2). Then, we can track the C-arm pose compared to the pre-operative 3D reconstruction using the current depth information via ICP registration algorithm (Figure 6.1-step 3). The C-arm pose knowledge allows computing the projection matrix (Figure 6.1-step 4), necessary to perform the overlay of the X-ray image over 3D reconstruction and also to visualize the current C-arm position with respect to the 3D reconstruction (Figure 6.1-step 5). We are now going to describe every step of this pipeline in detail.

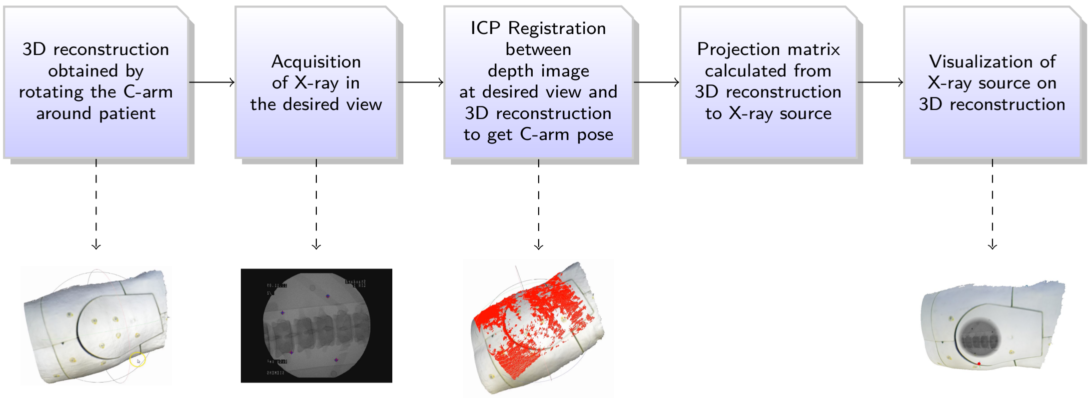


Fig. 6.1. Pipeline to visualize the X-ray image over 3D reconstruction of surgical scene

6.2 3D Surface Reconstruction from RGBD Data

The 3D surface reconstruction application, published by Izadi et al. [62], is the use of a temporal RGBD data sequence to reconstruct a more detailed, complete and accurate 3D point cloud or surface than a single depth image point cloud could provide. For every RGBD camera pose position, a single depth point cloud is computed and registered with the previous position's point cloud using ICP (Iterative Point Cloud), allowing us to compute the current camera pose compared to the previous one. Then, the RGBD data is used to reconstruct a volumetric reconstruction, called Truncated Signed Distance Field, which implicitly characterizes the 3D surface (more details on this technique in Chapter 5). Using implicit surface meshes generation algorithm such as Marching Cubes, the 3D mesh can be generated. The main advantage of this technique is that although a single depth can be very noisy, the temporal averaging inside the volumetric reconstruction allows reducing this noise and provides a more accurate reconstruction of objects than a single depth image.

In our application, we use an implementation of the Kinect Fusion, called RecFusion [1], to perform the 3D reconstruction. The multiple viewpoints are obtained by rotating the C-arm using its different joints around the object to reconstruct. The full acquisition takes less than two minutes. Figure 6.2 shows three 3D reconstructions of anatomical phantoms (bone on a table, thorax from 3D printed mannequin and spine phantom). In our defined workflow, this 3D reconstruction would be done when the patient is already under the C-arm before the surgery starts.

1. www.recfusion.net/

Fig. 6.2. 3D reconstruction of diverse anatomical phantoms

6.3 Automatic Depth-based C-arm Pose and Projection Matrix Estimation

During the surgery, X-ray images at different C-arm poses can be acquired. For each of them, we would like to show the C-arm position relatively to the 3D reconstruction. The depth image acquired at the same time as the X-ray image can be used to recover the pose. With a knowledge of the intrinsics parameters of the RGBD camera K, we can create a point cloud from the depth image as shown in Chapter 3.3. The pose (rotation R and translation t) of the RGBD camera compared to the 3D surface is then computed using the ICP algorithm between the point cloud of the depth image and the 3D reconstruction. The ICP algorithm is one of the most used 3D point cloud rigid registration algorithm. We use the PCL implementation of this algorithm [138]. The ICP algorithm registers one point cloud (the source) to the other (target). For each point in the source point cloud, the algorithm searches for the closest point in the target point cloud. The rigid transformation is then computed by minimizing on the cost function defined by the distance between closest points. The source point cloud is then transformed using the computed transformation and the process is repeated until convergence. The projection matrix M from the 3D reconstruction coordinate system to the depth image plane is obtained using the perspective projection equation (Equation 3.3 in the Background Chapter 3). The relationship between the depth image plane and the X-ray image plane is given by the homography H, that maps the X-ray image over the video image in the mirror-based RGBD augmented C-arm setup. Therefore, the projection matrix M from a 3D point of the 3D surface reconstruction to the X-ray image plane follows the Equation 6.1.

$$M = H^{-1}K(R|t) \tag{6.1}$$

Once the projection matrix is calculated, we can compute the C-arm X-ray source position compared to the 3D reconstruction and create the rendering paradigms of X-ray image along the 3D reconstruction.

6.4 Visualization

Efficiently merging 2D projective data and 3D surface information is an ill-posed problem. Indeed, the 3D position of 2D projective data has an infinite mathematical answer but only one right solution, unobtainable from the projection only. However, we propose two visualization paradigms to visualize the X-ray image along with the 3D reconstruction: texture mapping and virtual image plane visualization which are two particular cases of those infinite solutions. Both solutions present advantages and drawbacks that we are going to describe in detail. First, we explain how to display the C-arm X-ray source visualization with respect to the 3D reconstruction, before entering into the details of the two visualization paradigms.

6.4.1 C-arm X-ray Source Visualization

To easily correlate the X-ray source to the 3D reconstruction, the display of the 3D position of the X-ray source in the same 3D environment as the 3D reconstruction is essential. This C-arm X-ray source position C in the 3D reconstruction coordinate system can be easily computed from the C-arm pose T_M (of rotation R_M and translation t_M) with the Equation $C = -R_M^{-1}t_M$. This pose is obtained by decomposing the projection matrix M into its intrinsics K_M and extrinsics T_M matrices with QR decomposition such as $M = K_M T_M$. The source can be seen in Figure 6.3 as a red ball, the coordinate system axes of the X-ray source are also shown.

6.4.2 Texture Mapping

The first method of visualization consists of a texture mapping of the X-ray image on the 3D reconstruction. Using the projection matrix M, we project into the X-ray image plane the 3D reconstruction points. The color of the 3D points in the reconstruction is changed according to the position of their projections into the X-ray image. The simplest is to change to the X-ray color pixel on which a 3D point is projected. However, visualizing the full X-ray image can occlude a large part of the 3D reconstruction texture and impair the surgeon's perception. Therefore, we choose to show a window of the X-ray image in the 3D reconstruction on which we blend the X-ray image with the alpha-blending parameter α. We define this window as a disk in the X-ray image of center c and radius r. At the border of this window, we either blend progressively in a linear manner (on the distance m) the X-ray pixel color with the 3D reconstruction point color $C_{3D}(P)$ or make a sharp border ($m = 0$). Therefore, for a point P of the 3D reconstruction projecting into p in the X-ray image I_{xray}, its newly defined color $C(P)$ is defined by the Equation 6.2.

$$C(P) = \alpha\beta(p)I_{xray}(p) + (1 - \alpha\beta(p))C_{3D}(P) \tag{6.2}$$

The function $\beta(p)$ changes the blending value according to the pixel position in the window and its definition can be found in Equation 6.3.

$$\beta(p) = \begin{cases} 1 \text{ if } ||p - c|| < r \\ 1 - \dfrac{||p-c|| - r}{m} \text{ if } r < ||p - c|| < r + m \\ 0 \text{ if } ||p - c|| > r + m \end{cases} \qquad (6.3)$$

The window position, defined by c, can be moved interactively, as well as its size r, by the user. An example of this visualization paradigm is shown on the first row of Figure 6.3 with a blurry border (left) and a red sharp border (right).

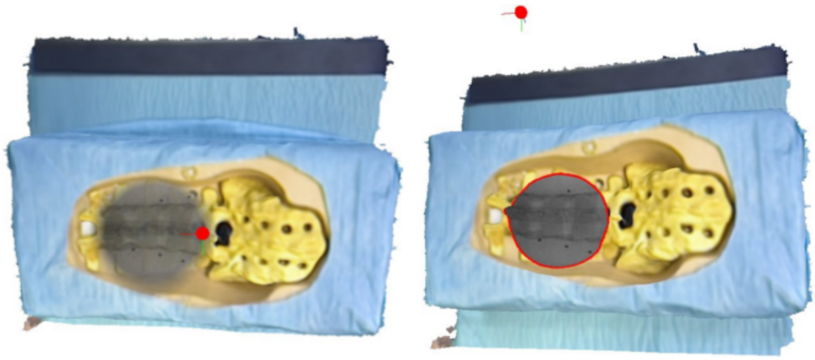


Fig. 6.3. X-ray image visualized with 3D reconstruction of a spine model with texture mapping (red dot is the source)

The texture mapping visualization paradigm allows observing the 3D reconstruction and the X-ray image on a single object, which is visually appealing but comes with two perception errors. The first error is the mapping distortion of the X-ray image onto the 3D reconstruction. Several points in the 3D reconstruction can correspond to the same projected point in the X-ray image. As we have stated before, due to the projective nature of the X-ray image, none of them is the right 3D point corresponding to the internal anatomy. In our visualization, all the points projecting into the same X-ray image pixel will get colorized by the X-ray value. This can lead to ghosting artifact where one structure in the X-ray image will appear several times in the 3D reconstruction. This can be seen in Figure 6.4 where one X-ray marker (red circle) can be seen twice because of the irregularity of the surface. Mapping distortion is only minimal when the display viewpoint direction is close to the mapping projection direction or the surface very smooth.

The second perception error is the misplacement of internal structures due to the projective nature of the X-ray imaging. As we explained earlier, knowing the real 3D position of the structure imaged by the X-ray image is an ill-problem which offers infinite mathematical answers, although only one solution is correct (the 3D structure imaged has a unique 3D position) but unobtainable from the projection matrix only. Using texture mapping, we choose one mathematical answer, which is to place the X-ray structure on the surface. However, this is obviously incorrect most of the time. When a structure is inside the patient, it will appear at the surface with our visualization which is not its proper localization. This can lead to an

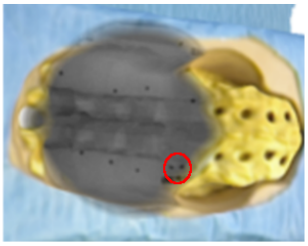


Fig. 6.4. Ghosting effect due to mapping distortion can be observed inside the red circle

erroneous perception of the structure's depth. An example is shown on Figure 6.5 where one marker is seen as being on the surface (left image) even if it is located inside the spine (right image).

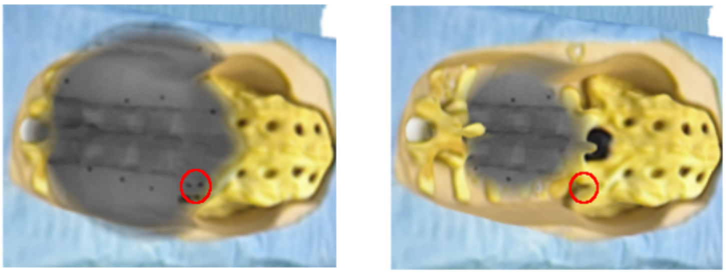


Fig. 6.5. Wrong structure perception, the marker seems on the surface on the left image while being in fact inside

This effect is also minimal when the display viewpoint corresponds to the mapping projection viewpoint. Because of the depth perception errors induced by the texture mapping of X-ray image over 3D reconstruction, this visualization paradigm is unsuitable for clinicians if the display viewpoint optical axis is far from the optical axis of the X-ray source, which happens as soon as the clinician is moving around the 3D reconstruction.

6.4.3 X-ray Image in the Virtual Image Plane

To overcome the drawbacks from the previous visualization paradigm, we propose to visualize the X-ray image in its virtual image plane. Therefore, the X-ray image is not mapped anymore on the 3D reconstruction but is visualized on a virtual image plane placed orthogonally to the X-ray optical axis at a given distance from the X-ray source. An example of this rendering is shown on Figure 6.6.

This visualization paradigm does not present the depth perception errors from the first paradigm, but when the display viewpoint direction is far to the mapping projection direction, the X-ray image is not visible anymore since it lies on a plane. As explained at the beginning of this section, it is impossible to retrieve the unique solution of the structure 3D position based only on the projection matrix. Therefore, both presented visualization paradigms are not perfect for X-ray image visualization on 3D reconstruction, especially when the display viewpoint is away from the X-ray projection axis. We either have incorrect information due to ghosting (for the second visualization paradigm) or no information at all about the X-ray image (for the second visualization paradigm). However, something that remains

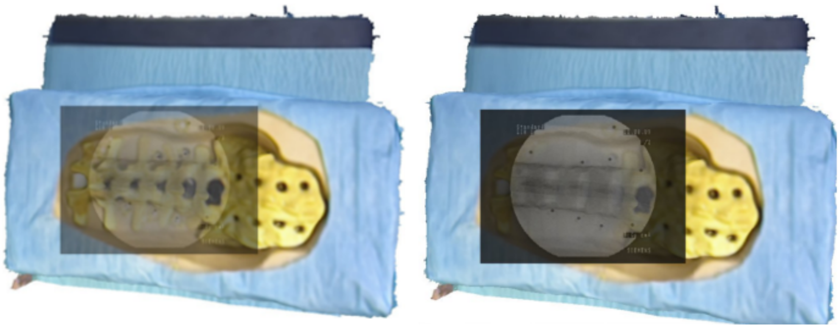


Fig. 6.6. X-ray image visualized with 3D reconstruction of a spine model in virtual image plane visualization paradigm, the two images represents different levels of blending

correct, independently of the display viewpoint, is the X-ray source and coordinate system visualization.

6.5 Results

6.5.1 Visualization Results

We have performed the 3D reconstruction of three anatomical phantoms (spine, thorax, and bone). We have shown along the section the visualization paradigms applied to the spine phantom. We show here the results for the bone phantom (Figure 6.7) and the thorax (Figure 6.8).

6.5.2 Projection Matrix Estimation Validation

We also validate the projection matrix calculation from the depth based C-arm pose estimation. On every anatomical phantom, we place X-ray markers, also visible on the 3D reconstruction (three for the bone, seven for the spine and four for the thorax). From the 3D reconstruction, we manually extract the 3D position of the X-ray markers, that we project into the X-ray image thanks to the estimated projection matrix from the depth-based C-arm pose estimation. Then, we compare the projected position of the markers to their real positions in the X-ray image.

We show the results of the projections quantitatively in Table 6.1 where we present the RMS error over the X-ray markers positions between the ground-truth (position in the X-ray image, blue in Figure 6.9) and their estimated positions (projection on X-ray image, red in in Figure 6.9).

We show the visual results in Figure 6.9 where we show the 2D real and estimated positions of the X-ray markers (top row of Figure 6.9). We also show, on the bottom row of Figure 6.9, using the virtual image plan paradigm how well the X-ray markers in the X-ray image match their positions in the 3D reconstruction.

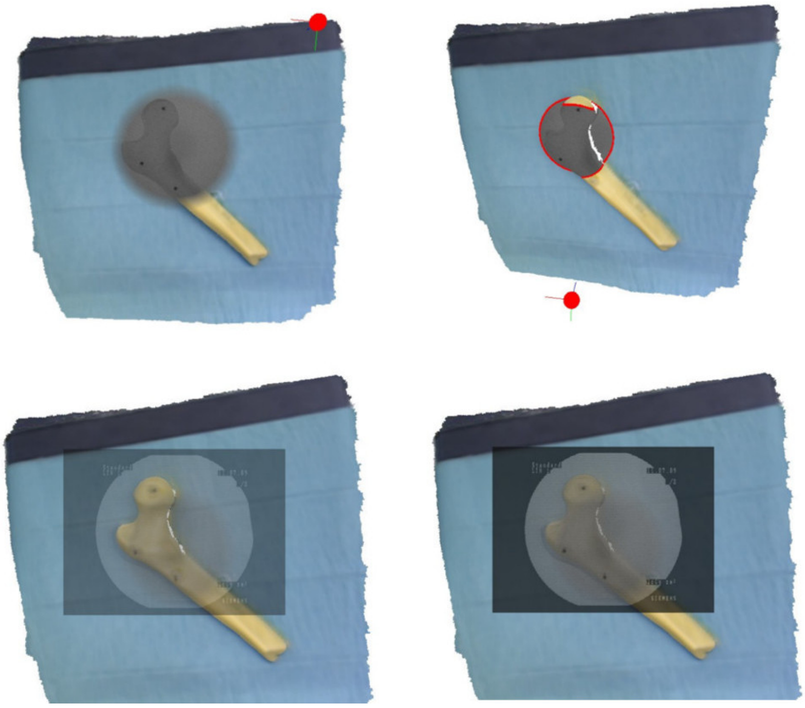


Fig. 6.7. X-ray image visualized with 3D reconstruction of a bone model, (Top) texture mapping, (Bottom) virtual image plane

Fig. 6.8. X-ray image visualized with 3D reconstruction of a thorax model, (Top) texture mapping, (Bottom) virtual image plane

The RMS projection error for the bone and the thorax model are under one millimeter. Although higher than the overlay error from Navab et al. [111] of $0.5\,\mathrm{mm}$, this error is still in the acceptable range of the precision (around $2\,\mathrm{mm}$) required for image-guided surgery technologies [158] For the spine, the error is higher, especially when considering all the markers but as it can be seen in Figure 6.9, only one marker is far from its projected position. This marker is placed at a different height than the others and is on the side. Our hypothesis

	Spine	Bone	Thorax
RMS Projection error (in pixels)	7.21 (3.02*)	1.62	2.09
RMS Projection error (in mm)	3.14 (1.31*)	0.71	0.91

* without outlier marker

Tab. 6.1. Projection error between the X-ray marker ground truth and real projections

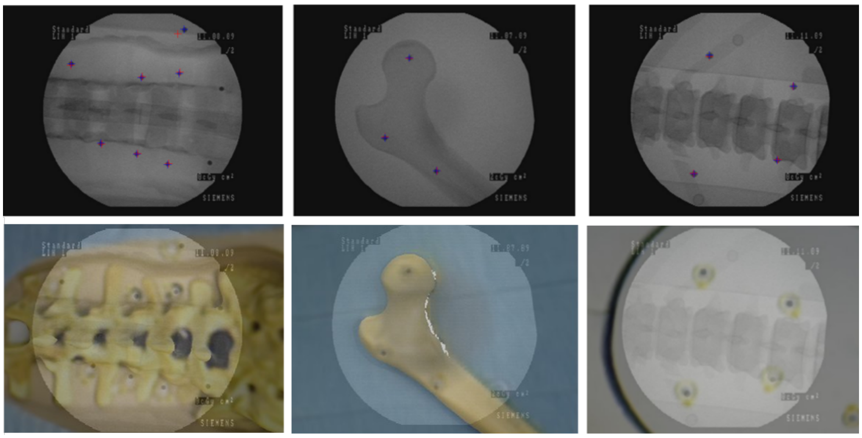


Fig. 6.9. (Top) Projections of the markers on the X-ray (blue - real positions and red - projected positions), (Bottom) Close-up on the markers in the virtual plane configuration

is that the spine model has a very challenging convex shape with sharp and thin structures. Then, the depth image acquired of the phantom is incomplete due to holes and edges which results in a depth-based ICP registration converging to a solution fitting well only a part of the phantom. Excluding this outlier marker from the RMS projection error brings the error down to $1.31\,\mathrm{mm}$.

6.6 Discussion

This work was the first work to propose RGBD data for C-arm pose estimation. Although the results are encouraging, we can see that depending on the object in the scene, the registration error can increase dramatically. This is the current limitation of our work. However, the registration algorithm used (ICP) is the most common for point cloud registration and is not state of the art (at least in its PCL version). The depth-based C-arm pose estimation could highly benefit from getting extended to the state of the art algorithms of point cloud registration in Computer Vision. Regarding the visualization paradigms exposed in this section, we saw that the limitation was due to the projective nature of the X-ray image, which makes the 2D/3D visualization challenging. As a follow-up on this work, Lee et al. [85] have explored the use of Cone Beam Computed Tomography (CBCT) imaging in combination with the 3D surface reconstruction. CBCT imaging uses a C-arm to create a CT-like volume of the internal structure, which is in this case registered with an RGBD camera attached at the intensifier. The main advantage of the CBCT volume is that a DRR image can be generated from any display viewpoint. When applied to the texture mapping visualization paradigm, this alleviates the depth perception issues, since the display viewpoint is always matching the source of DRR image. This system has been used by Fischer et al. [42] for a pre-clinical study of K-wire placement, we show an example of visualization in Figure 6.10. They have shown that the 3D visualization of patient, tool and DRR present advantages in terms of efficiency over X-ray imaging and provides intuitive feedback on how to place accurately and efficiently the medical tools during K-wire placement.

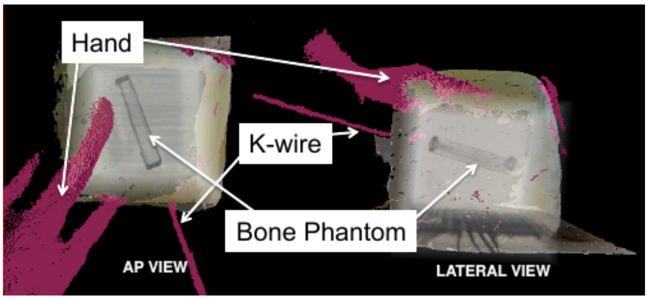


Fig. 6.10. 3D Visualization combined with DRR for the placement of K-wire, reproduced with permission from [42], ©Springer

6.7 Conclusion

In this chapter, we have presented the first application of a RGBD augmented C-arm, which is the visualization of the X-ray image over a 3D reconstruction of the surgical scene. This visualization is the extension to 3D of the 2D overlay provided by the video and RGBD augmented C-arm. The main goal of such visualization is to provide the external context of the X-ray image by showing its location compared to the patient as well as the localization of the X-ray source. We have also shown two types of rendering: texture mapping and virtual image plane. The first one provides the internal context of the X-ray image, however, the ill-posed nature of 3D X-ray image localization due to the projective nature of X-ray imaging might lead to incorrect depth perception in the internal context understanding. However, as we have shown, the future of this visualization is to be applied on CBCT where this limitation is lifted.

7 Radiation exposure estimation

The work presented in this chapter has been realized with the master student Nicola Leucht under my supervision and is an extended version of the paper presented at ISMAR 2015 which is reproduced with permission from [88], ©IEEE.

7.1 Introduction

As we explained in Chapter 3, a C-arm is a mobile intra-operative device, which is easy to operate and enables the surgeon to get immediate feedback of the patient's anatomy. It comes however at the cost of radiation exposure to the surgical staff with potential lethal consequences on their health as explained in Chapter 3 According to the Council Directive of the European Union [33], radiation exposure should always show a positive balance benefit versus amount of radiation. This implies that the dose is kept "As Low As Reasonably Achievable" (i.e. the ALARA principle). It means that the staff involved in C-arm based surgeries needs to have the adequate education and information, as well as, theoretical and practical training. The reason for this is that radiation emitted from the X-ray tube is scattered by the air and even more by the patient. The surgeon's position at a specific time during intervention becomes increasingly important, as the three main factors that influence the radiation exposure are time, distance, and shielding [26]. The first step to increase the distance and to minimize the time close to the source is to make surgeons aware of their proximity and to sensibilize them to the risks described in the Background Chapter in Section 3.2.3. To support this, it is important to compute in detail the level of radiation reaching a surgeon by knowing both the distribution of the scattered radiation and the surgeon's exact location.

7.2 State of the Art

We refer the reader to the Section 2.4.1 of the State of the Art Chapter for the literature review of the works computing radiation exposure for C-arm based surgeries.

7.2.1 Contributions

We present an inexpensive and flexible setup for the estimation of the final radiation exposure of the surgeon by attaching two depth cameras directly to the C-arm. We consider the surgeon's position as done previously [81, 133], but keep our mobility by not installing cameras on the OR ceiling. Another advantage over these works is that the scattering is simulated more precisely since the patient, as the source of scatter radiation, is modeled according to its real dimensions. Furthermore, no calibration with measured dosimeter values is required.

7.3 Methodology

7.3.1 Setup

The setup is based on the mirror-based RGBD augmented C-arm, described in Chapter 4, on which an additional RGBD camera (also an Asus Xtion Live Pro) is attached. The C-arm used is a Siremobile Iso-C 3D from Siemens Medical Solutions. While camera 1 (the camera from the mirror-based RGBD augmented C-arm) observes the scene through the mirror and, therefore, faces the image intensifier from above, the camera 2 (the additional camera) is attached horizontally at the middle of the C-arm and watches the scene directly. Figure 7.1 depicts the current RGBD augmented C-arm setup.

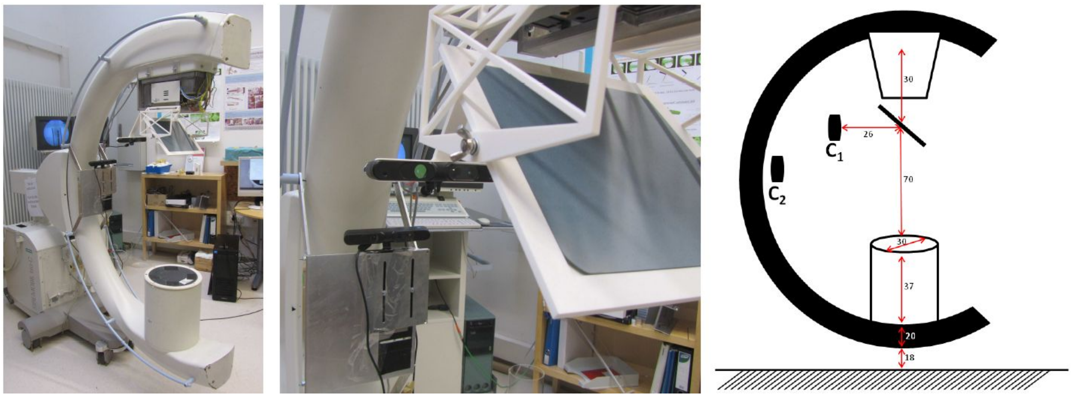


Fig. 7.1. (Left, Middle) The setup with the mirror-based RGBD augmented C-arm and an additional RGBD camera, (Right) The setup's sketch including the measured dimensions in centimeters

The patient, patient table, and surgeon also belong to the scene. The camera calibration is performed with a checkerboard pattern using the Zhang's calibration from OpenCV. To find the coordinate systems' rotation difference, the point clouds of the two cameras are demeaned first in order to be centered around the origin. Then, the covariance matrix of all the points is computed. Using singular value decomposition of the covariance matrix, the rotation between the coordinate systems is determined. The remaining translation is then computed using the previously calculated mean values of the point clouds. For the calibration process, ten images from each camera are acquired with the checkerboard pattern at various positions and rotations. Applying the resulting rotation matrix to the point cloud of camera 2 transforms the points into the coordinate system of the camera 1 and then allows a correct visualization of the complete observed scene.

7.3.2 Scene Reconstruction

The C-arm is modeled according to its true fabricated size, having camera 1 attached at a fixed position and camera 2 positioned relatively according to the calibration result. For the accurate modeling of the patient, the data of camera 1 is used. Taking a previously captured background image into account, the patient's shape and size is determined and a triangulated volume is created. Since camera 2 has a better view of the surgeon and is still close enough in order to capture their surface well, its depth data is the one used for the surgeons' triangulation. The depth images of the previously generated patient scene as the background and the current scene containing the surgeon are used to determine the surgeon's dimensions by subtracting the background information. After a clustering algorithm for noise elimination, the points from the depth image are converted into world coordinates and a triangular representation is generated.

7.3.3 Radiation Simulation

The characteristic photon energy spectrum for a certain voltage and filtering is determined with SPEKTR [151]. The obtained energy level probabilities are used for the particle generation in Geant4 [1], which is the physical simulation toolkit used in this work. It enables the user to create a scene containing different components and uses a Monte Carlo algorithm for simulating the passage of particles through matter. The simulation starts as soon as the background and patient scenes are available. In order to estimate the dose distribution in space, a voxel grid is used for counting the hits and then determines the energy deposited at each position. Knowing the surgeon's surface, the dose is projected on it. For a given voltage, current, time, filtering, and field size, the number of particles needed for the simulation is calculated. In order to simulate the radiation in mGy/h for our validation study based on the Siremobile Iso-C 3D, the following parameter values are defined using the C-arm user manual: a voltage of U= $110\,\mathrm{kV}$, a current of I= $3\,\mathrm{mA}$, and a time t= $3600\,\mathrm{s}$. The energy then amounts to $W_0 = U \times I \times t = 1.188 \times 10^6 J$ which is equivalent to $6.03 \times 10^{22} eV$. Using the spectrum generated with SPEKTR, the mean energy of the photons is computed. By dividing the total energy W through the mean energy $W_m = 55.42\,\mathrm{kV}$ of the created photons, the number N_0 of the photons being emitted is estimated as $W/W_m = 1.09 \times 10^{18}$. Knowing the total number of photons being created, the influence of the collimator has to be considered. Since the rays are spread equally in all directions from the target, they form a hemisphere. Then, the number of photons that pass through the collimator and form the desired field at the image intensifier is computed by determining the fracture of the hemisphere which corresponds to the opening of the collimator. We calculate for the case of a square field of $18\,\mathrm{cm}$ x $18\,\mathrm{cm}$ at the image intensifier, at a distance of one meter from the X-ray source, and arrive at $N = 5.5 \times 10^{18}$ photons to be simulated.

7.3.4 Dose Computation & Visualization

A C-arm can be utilized in two different ways, depending on the situation. One possible application is single acquisitions, which means that multiple single X-ray images are taken. The other application is fluoroscopy and refers to the constant irradiation during a surgery. In single image mode, the scene is captured and the current surgeon triangulation is augmented with a radiation color map. For the final dose estimation, the dose area product [153] for the scene is computed and the results of the individual X-ray acquisitions are summed. For the

1. http://geant4.cern.ch/

fluoroscopy mode, the skeleton tracking algorithm of NiTE2 is used. The depth image of the camera is acquired and a machine learning algorithm, as introduced by Shotton et al. [150] classifies the body parts and tracks the joint positions. At these positions, the dose obtained from the Geant4 simulation is accumulated over time. By interpolating between the joints, the triangulated representation of the surgeon is assigned a radiation color map and the final dose area product is calculated directly. The scene visualization created in OpenGL is shown in Figure 7.2 as well as the predefined colormap.

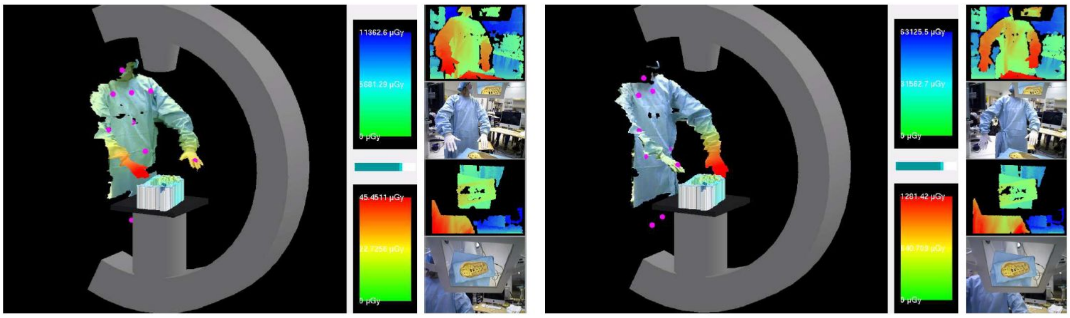


Fig. 7.2. (Left) X-ray single shot mode with color-coded radiation exposure, (Right) Color-coded radiation map augmentation on the surgeon and the skeleton joints visualized with magenta spheres

7.4 Evaluation & Results

Two different approaches are used to validate the results of the proposed approach: user-manual validation and dosimeter validation.

7.4.1 User-Manuel

The user manual of our C-arm provides a statement of the scatter radiation in the important sojourn areas A1, A2, and B. For the constant irradiation of a water phantom of $25\,\mathrm{cm}$ x $25\,\mathrm{cm}$ x $15\,\mathrm{cm}$ on top of the image intensifier with a voltage of $110\,\mathrm{kV}$, a current of 3mA, a filtering of $3\,\mathrm{mm}$ Al equivalent, and a field size of $18\,\mathrm{cm}$ x $18\,\mathrm{cm}$ at the image intensifier, dose values in specified areas are given. The same scene is modeled in Geant4 and a simulation is run five times with 10 million particles per round. Since the mean error of the user manual values compared to the simulated results over the three areas is 16.46%, this shows that the radiation simulation is in the correct order of magnitude. A reason for the difference could be that the values in the manual are the "maximum interference radiation" and, therefore, the values of the simulation are mostly below the data stated there. The setup and dose values are shown in Figure 7.3 and in Table 7.1.

7.4.2 Dosimeter

For the validation in a real setup, measurements with a dosimeter at different positions are performed. Those dose values are compared to the Geant4 simulation results at the dosimeters'

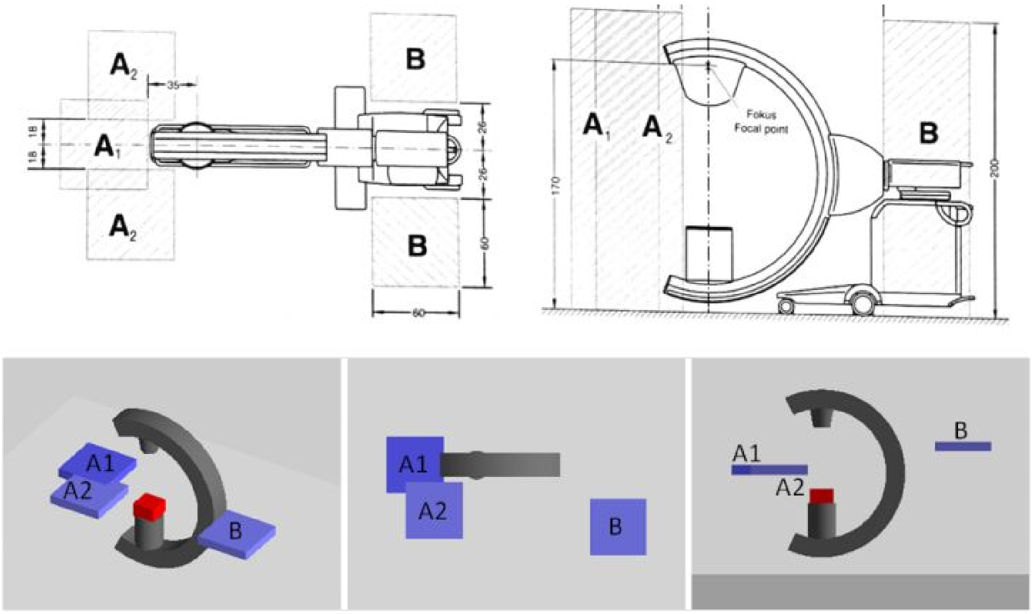


Fig. 7.3. (Top) The C-arm setup as used for the dose measurements in the user-manual of a Siremobile Iso-C 3D, (Bottom) The same scene created in Geant4

	User Manual (in mGy/h)	Geant 4 (in mGy/h)	Error (%)
A1	7.25	5.79 ± 0.39	20.1
A2	6.83	5.31 ± 0.84	22.3
A3	0.84	0.90 ± 0.18	6.9

Tab. 7.1. Dose measurements in the areas A1, A2, and B

positions. Instead of a patient, a cylindrical water phantom with a height of $20\,\mathrm{cm}$ and a diameter of $8\,\mathrm{cm}$ is placed under the X-ray source as shown in Figure 7.4.

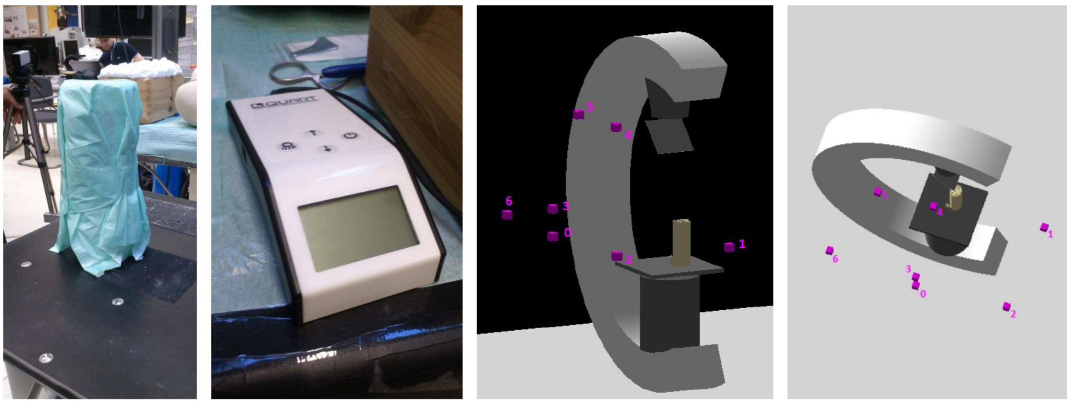


Fig. 7.4. The components of the experiment: (Left) Water phantom on the table, (Second Left) Dosimeter and (Right Images) Scene modeled in Geant4 with the seven dosimeter positions

After acquisition by the depth camera, the simulation is started with this configuration, modeling the scatter radiation of the phantom and creating the respective dose distribution. The dose values are accumulated in a grid with a voxel size of $5\,\mathrm{cm}$ x$5\,\mathrm{cm}$ x $5\,\mathrm{cm}$. In reality, the dose is measured with a QUART didoSVM dosimeter, which is "currently the most compact and light weight high-performance survey meter in its class"[2], at seven different positions around the phantom. Each time after one X-ray image acquisition, taken with a voltage of $110\,\mathrm{kV}$ and a current of $8\,\mathrm{mA}$, the value of the dosimeter is read out. The duration of one X-ray shot is measured as $210\,\mathrm{ms}$. Two separate dose measurements are made at each dosimeter position (positions on the two right images of Figure 7.4). The simulation is run ten times with 100 million photons per round and evaluated at the dosimeter positions. Table 7.2 shows the results of the simulation and the measured values. The mean error of the simulation results compared to the measured dosimeter results is 16.39%, which is, in contrast to an error of 30% despite a dosimeter calibration in the work of Rodas et al. [133], an improvement. The best results are achieved for the positions 2 and 3 and the worst are observed at positions 5 and 6 with values up to 32.88%. At these positions, possible sources of errors are attributed to the plastic construction holding the mirror, which is not modeled in the Geant4 simulation. Also, the metal construction holding camera 2 next to it could also be responsible due to a change in scattering. We note that the high standard deviation values of at these positions could be improved with a larger number of simulation runs.

Position	Dosimeter (in mGy/h)	Geant 4 (in mGy/h)	Error (%)
0	627 ± 25.5	693.4 ± 169	10.6
1	468 ± 0	385.9 ± 89	17.7
2	257.5 ± 0.7	253.7 ± 105	1.46
3	698 ± 7.1	697.7 ± 142	0.03
4	1074 ± 31.1	841.3 ± 222	21.6
5	587 ± 1.4	780 ± 451	32.8
6	326.5 ± 3.5	425.7 ± 61.2	30.4

Tab. 7.2. The average values over either the two measurements or the ten simulation runs $\pm$ the standard deviation

7.5 Discussion

The setup used for this work is the mirror-based RGBD augmented C-arm, on which an additional RGBD camera is attached to track the surgeon's skeleton. We have shown that the estimation of radiation exposure with this setup is equivalent to the state of the art. However, the future of this project is to go away from the restrictive framework of the mirror-based RGBD augmented C-arm which is very specific. The full radiation exposure estimation framework is not bound to this setup and could be generalized to any RGBD augmented C-arm. Indeed, as RGBD data allows to place the cameras anywhere on the C-arm to obtain 3D reconstruction, any lighter setup placing RGBD cameras at more convenient places on

2. http://quart.de/en.html

the C-arm (e.g. at the intensifier) would be more suitable for better integration into the OR. The current implementation is also too slow to be used intra-operatively, however, a GPU version of the library used for our work has recently been released [11], reducing the radiation estimation computational time by a factor of ninety.

7.6 Conclusion

In this chapter, a new system for radiation exposure estimation simulation composed of a setup based on 2 RGBD cameras affixed to the C-arm, a framework simulation and an augmented reality visualization of the accumulated radiation on the surgeon was presented. We have shown that the estimation of radiation exposure with this setup is equivalent to the state of the art. The end goal was to provide an augmented reality visualization overlaying on the surgeons their radiation exposure during a surgical procedure in order to sensibilize them to the risks inherent to it. We have proposed this visualization in a heatmap manner overlaid on the surgeon.

Multi-layer Visualization for Medical Mixed Reality

8

8.1 Introduction

The term "Surgery" comes from the Greek "Kheirourgia" which means handy work. Despite numerous technological improvements in the last centuries, surgery remains a manual work where surgeons perform complex tasks using their hands and surgical instrumentation. As it is yet not possible to retrieve the view as seen directly by the surgeon, numerous works are using video cameras to record the entire surgical scene. Such a solution is applicable for training medical students using "first-person" view cameras [15], or more commonly for Medical Augmented Reality where another modality (intra-operative or pre-operative) is overlaid over the video to give context to the medical data. Having the hands and instrumentation positioned in the field of action inherently signifies the occlusion of the surgical scene and the anatomy being treated. This is true both from the surgeon viewpoint or any imaging modality viewpoint. It would be advantageous if there was a solution to display to the surgeon any occluded region of interest without losing the information about the action that is given by the hands and surgical instrument positions. Introducing transparency on the occluding regions links the problem to the Diminished Reality field of study. Combining Diminished Reality and Augmented Reality has been referred in the literature to Mediated Reality [101], which is the action of altering the reality by subtraction and/or addition of elements.

8.2 Related Work

Making the occluding layer transparent or even disappear in order to visualize what is beyond has been studied in Diminished Reality (DR). In contrast to Augmented Reality where graphics are overlaid on a real scene, DR withdraws or attenuates real elements from a scene. The works in DR can be divided into three categories according to the background recovering method: multi-viewpoint, temporal, and inpainting. The temporal methods [25, 32, 147] suppose that the camera has seen the scene previously without the occluder (or with the occluder at another position) and use this previous information to recover the current occluded pixels. The inpainting methods recover the occluded part of an image with information from its non-occluded part using patch-based methods [58, 71] or combined pixels methods [59]. The multi-viewpoint techniques use additional cameras placed at other viewpoints that can observe the occluded background totally, or partially in order to recover it from the occluded viewpoint. Jarusirisawad and Saitoo [66] use perspective wrapping from the non-occluded cameras to the occluded camera to recover background pixels. More recently, using RGBD cameras, several works [106, 140] have generated surface mesh models of the background from one or multiple side cameras. Observing the mesh from the occluded viewpoint requires only a rigid transformation, avoiding distortions due to wrapping. Sugimoto et al. [156] use the 3D geometry to back-project the occluded pixels to the side views and, therefore, recover it. By design, the multi-viewpoint recovery can be used for the stereo-RGBD augmented C-arm described in Chapter 5 composed of two RGBD cameras, placed on the side of the X-ray source viewpoint. However, instead of using a mesh, the volumetric field can be used. However, no work in literature has used volumetric field such as TSDF to recover background information to the best of our knowledge. Concerning the visualization of the foreground layer in combination with the background layer, the most used technique is transparency [25, 156]. As explained by Livingston et al. in their review of depth cues for "X-ray" vision augmented reality [94], transparency is indeed the most natural depth cues as it can be experienced in the real world with transparent objects.

8.3 Background Recovery using the TSDF Methodology

For this work, we use the setup and the TSDF algorithm described in Chapter 5 and we refer the reader to this chapter for the details concerning the setup and the TSDF volumetric reconstruction.

After the first raytracing step described in Chapter 5, the synthesized video image I_s as seen by the X-ray source viewpoint, as well as its corresponding depth image I_{sd}, are generated. The volumetric TSDF field is a dense representation which contains information about the full 3D space around the C-arm detector whereas the raytracing step only stops at the first found surface voxel (corresponding to $f(\mathbf{x}) = 0$) from the raytracing viewpoint origin. Therefore, the TSDF field contains more information than is actually used for our application in Chapter 5. In a typical surgical setting such as described in the introduction of this chapter with hands or tools placed above the patient, beyond the foreground (hand or tools) synthesized by the first raytracing step, other surface voxels can be present along the ray. This is especially true since the two RGBD cameras are placed on the side of the C-arm, giving additional information from another viewpoint. This situation is illustrated in Figure 8.1 where the background occluded by a hand from the X-ray source viewpoint (the blue point) can be seen by at least one of the two cameras (red and green points). In a TSDF representation, this means those occluded surface background voxels also have a TSDF value equal to 0. To find those additional surface voxels, a modified "second run" raytracing must be performed from the pixels in I_s detected as foreground (e.g. surgeon's hands or surgical tools).

8.3.1 Foreground Segmentation

As a first step, the foreground needs to be segmented using the synthesized depth image. A background model is computed from an initialization sequence of N depth images $I_{sdi} : \Omega_x \mapsto \mathbb{N}^+$ where no hands or surgical instruments are introduced yet. An average depth image

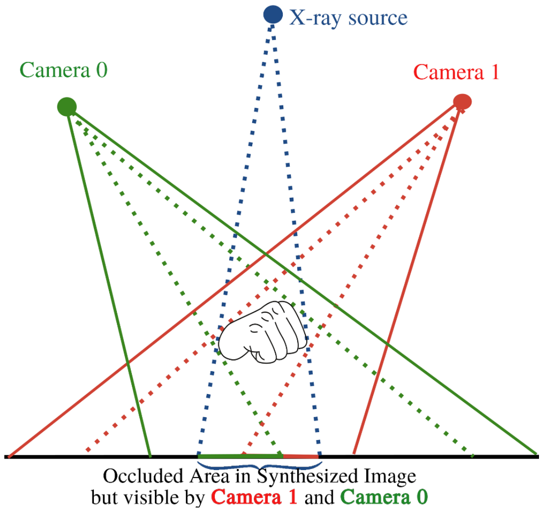


Fig. 8.1. Background surface occluded by the foreground can be seen by the side RGBD cameras

$I_a : \Omega_x \mapsto \mathbb{R}^+$ is created by averaging the depth at every pixel along the initialization sequence as shown in Equation 8.1.

$$I_a(p) = \frac{\sum_{i=1}^{N} I_{sdi}(p)}{N} \text{ with } p \in \Omega_x \tag{8.1}$$

Then, for every new image (with potential hands or surgical instruments present), the incoming depth image I_{sd} is compared to the mean image in order to create a binary mask image I_m using Equation 8.2.

$$I_m(p) = \begin{cases} 1 \text{ if } |I_{sd}(p) - I_a(p)| > \delta \\ 0 \text{ else} \end{cases} \tag{8.2}$$

where δ is a margin on the foreground detection (δ=3 cm). The method is rudimentary compared to state of the art background subtraction methods, however, the margin allows the background to change shape (in the limit of the margin). A noise removal step is added using morphological opening on the mask image. An example of scaled depth image I_{sd} and its corresponding mask I_m are shown on Figure 8.2.

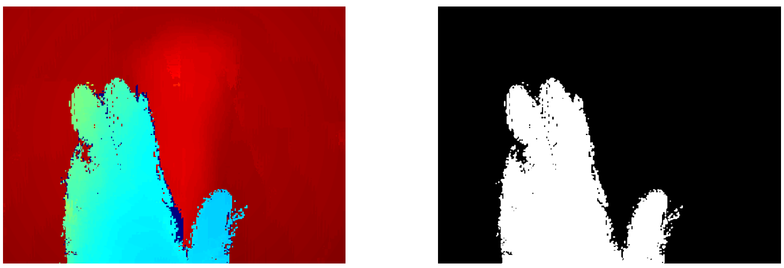


Fig. 8.2. (Left) The synthesized depth image I_{sd} and (Right) its corresponding segmented mask

We call $\Omega_f \in \Omega_x$ the pixels domain that include the pixels classified as foreground and $\Omega_b = \Omega_f^c$ the pixels domain of background pixels.

8.3.2 Second-run Raytracing

Once the foreground has been segmented, a second raytracing can be performed on the pixels of Ω_f. Instead of beginning the raytracing from the X-ray source viewpoint, the ray search starts at the voxel $\mathbf{y}$ found at the first raytracing run plus a margin μ (that we choose equal to $4\,\mathrm{cm}$ in our setup). This margin is the insurance to not find a surface voxel still related to the foreground. The starting voxel $\mathbf{y}$ for the pixel p in Ω_f can be easily retrieved using the depth image I_{sd} resulting from the first raytracing, such as $\mathbf{y} = P_x + I_{sd}(p)\vec{r}$. The raytracing is then performed forward using binary search in a similar fashion to the first run of raytracing. This second raytracing results in the creation of the image $I_r : \Omega_f \mapsto [0..255]^3$ which comprises the recovered background on Ω_f. As a result, the full background (visible and occluded) $I_b : \Omega_x \mapsto [0..255]^3$ can be created and is defined by $I_b(\Omega_f) = I_r(\Omega_f)$ and $I_b(\Omega_b) = I_s(\Omega_b)$. The foreground image $I_f : \Omega_f \mapsto [0..255]^3$ can also be created from I_s on the foreground segmented pixels Ω_f such $I_f(\Omega_f) = I_s(\Omega_f)$.

8.3.3 Multi-Layer Visualization

On top of the background image I_b, the foreground layer I_f can be overlaid with transparency as well as the X-ray image I_{xray}. A multi-layer image $I_l : \Omega_x \mapsto [0..255]^3$ can then be created by blending all the layers according to Equation 8.3.

$$I_l(p) = \begin{cases} \alpha I_f(p) + \beta I_b(p) + \gamma I_{xray}(p) \text{ if } p \in \Omega_f \\ (1-\delta) I_b(p) + \delta I_{xray}(p) \text{ if } p \in \Omega_b \end{cases} \tag{8.3}$$

with $(\alpha, \beta, \gamma, \delta) \in [0,1]^4$ with $\alpha + \beta + \gamma = 1$ are the blending parameters associated with each level. They can also be seen as specific weight values which emphasize a specific layer during the blending process. The visualization scheme we propose allows us then to observe three layers of structures (displayed in Figure 8.3) according to those parameters.

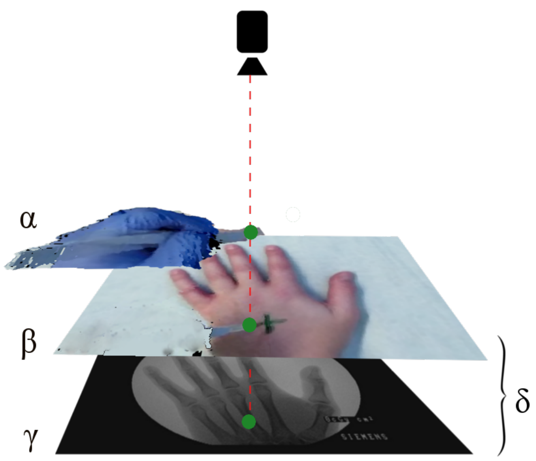


Fig. 8.3. The different layers in the multi-layer visualization, all can be observed depending on the chosen blending values $\alpha, \beta, \gamma, \delta$

The furthest layer is the X-ray image, which can be observed in its totality in the image I_l with $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$. As we get closer to the camera, another layer is the background recovered using the TSDF volumetric reconstruction. It can be observed in its entirety in I_l with $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$. Finally, the foreground layer comprising the surgeon's hands and medical instruments can be observed in I_l using $(\alpha, \beta, \gamma) = (1, 0, 0)$. The foreground layer only exists on Ω_f, therefore, we are free to choose any δ values in $[0, 1]$. The aforementioned modes are the basic modes where only one weight is activated at the time. Obviously, more complex visualization can also be proposed where the different layers (X-ray image, background, foreground) can be seen by transparency by choosing blending parameters $(\alpha, \beta, \gamma, \delta)$ not equal to 0 and 1. The choice of blending values depends on multiple parameters such as surgeon preferences, step in the surgical workflow, type of instrument used. It can be changed on the fly during surgery according to such parameters.

As an example, for the entry point scenario in down the beam surgery, where the hand will occlude the entry point as defined on the X-ray image, the configuration $(\alpha, \beta, \gamma, \delta) = (0.25, 0.25, 0.5, 0.5)$ would allow seeing the target point in the X-ray through the surgeon's hands still present for context but half-transparent and through which the entry point could be targeted on the background layer. Without the multi-layer visualization, the surgeon would only see the X-ray image overlaid over its hands and it would be difficult to target the entry point without moving its hands/tools. Then, as another configuration example, once an instrument has penetrated the skin, the background is not necessary to visualize. The transparent hands can be overlaid directly on the X-ray image, skipping the background layer. This scenario corresponds to blending parameters $(\alpha, \beta, \gamma, \delta) = (1 - \gamma, 0, \gamma, 1)$ with $0 < \gamma < 1$. With the configuration $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 1)$, the visualization consists of fully opaque hands or surgical tools on the X-ray image, giving a similar output as [122] which aimed at obtaining a natural ordering of hands over X-ray image compared to a uniform blending of X-ray image over video which corresponds to the configuration $((\alpha, \beta, \gamma, \delta) = (1 - \gamma, 0, \gamma, \gamma))$.

After the image synthesization step, every layer is known, however, the multi-layer configuration is not fixed and can be changed according to the preferences on the fly. For example, for the case of medical students re-watching a surgery, the multi-layer visualization can be replayed to medical students and residents with other blending configuration than the one used in surgery. They can have full control for the observation of the layers having the choice to emphasize particular layers of interest for their learning.

8.4 Results

8.4.1 Experimental Protocol

Six sequences of around 60 images have been recorded depicting example scenarios which include both surgeon's hands and surgical tools. Both a realistic hand model phantom and a real patient hand are used and positioned on a surgical table. A clinician wearing purple examination gloves introduces partial occlusions randomly to the scene. Sequences 1 and 3 contain the motion of the clinician's hand above the hand model phantom at $20\,\mathrm{cm}$ and $30\,\mathrm{cm}$ respectively. Sequences 2 and 4 contain the motion of a clinician's hand closed and above the hand model phantom at $20\,\mathrm{cm}$ and $30\,\mathrm{cm}$ respectively. Sequences 3 and 4 also contain incision lines drawn using a marker on the hand model phantom. Finally, Sequences 5 and 6 are recorded with surgical tools above a real patient hand. Sequence 5 includes actions using a surgical hammer aiming for a cross target drawn on the patient hand. Sequence 6 includes a scalpel targeting the same cross. The heights of the surgical instruments to the patient hand vary up from $5\,\mathrm{cm}$ to $30\,\mathrm{cm}$.

8.4.2 Background Recovery

For every sequence, the mean value of the recovered pixels percentage is calculated and indicated in Table 8.1. The natural observation in Table 8.1 is that the closer the surgeon's hand and surgical tools are to the anatomy the larger the occlusion in both side cameras will be. This signifies a lower percentage of recovered pixels by our algorithm which is demonstrated in Table 8.1.

Sequences	1	2	3	4	5	6
Pixels recovered (in %)	69.3	65.2	88.2	97.4	84.1	45.2

Tab. 8.1. Background recovery results

Sequences 1 and 2 were recorded with a surgeon's hand open (69.3%) and closed (65.2%) Fewer pixels are recovered for the close hand scenario as mainly the palm is present in the scene. The palm cannot be recovered in the other scenario but the fingers are also occluding, which are easier to recover from (due to their thin shape), in percentage, the open hand scenario recovers more, even if occluding more. Sequences 3 and 4 resulted in larger recovery percentages (88.2% and 97.4% respectively) because the surgeon's hand was farther away from the hand model. This implies that there is a greater probability for the background voxels to be seen by the RGBD sensors. Sequence 6 with a scalpel confirms that the height strongly influences the recovery. The scalpel scenario which includes numerous images with hands and instruments close to the background (less than $10\,\mathrm{cm}$) shows a low recovery result as expected. Due to the hammer's shape, the sequence 5 shows, however, a higher recovery percentage.

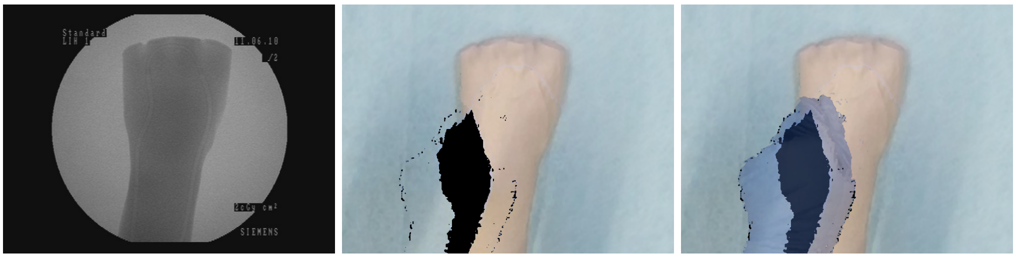
8.4.3 Visualization Results

In the Figures 8.4–8.9, for each scenario, one selected image I_l in the sequence can observed with different values of α, β, γ and δ. For every scenario, from left to right (on the two rows), the layer visualized in I_l is getting closer to the X-ray source viewpoint. In the column (a), the furthest layer (the X-ray image) is displayed. In the column (b), the second layer (the background), in the column (c), the blending of the front layer with the background, in the column (d), the blending of the three layers and finally, in the column (e), the closest layer is shown. Despite the fact that the background cannot be ideally recovered, a manual post processing step involving inpainting is applied and displayed in the column (f) of the Figures 8.4–8.9. We believe that the multi-layer visualization concept is an interesting and profound solution offering numerous possibilities in the surgical areas as well as the mixed reality communities.

(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ **(b)** $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ **(c)** $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ **(e)** $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ **(f)** Inpainting

Fig. 8.4. Multi-layer image I_l of one selected frame from Sequence 1 with different blending parameters $(\alpha, \beta, \gamma, \delta)$



(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ **(b)** $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ **(c)** $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ **(e)** $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ **(f)** Inpainting

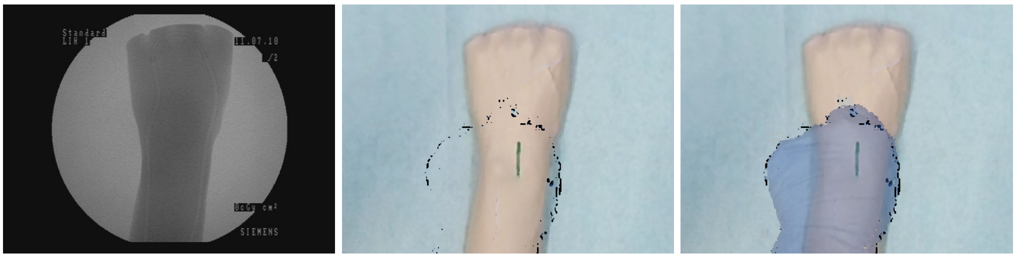
Fig. 8.5. Multi-layer image I_l of one selected frame from Sequence 2 with different blending parameters $(\alpha, \beta, \gamma, \delta)$

Similar to results presented for the mirror-less setup in Chapter 5, the images resulting from synthesization are not as sharp as a real video image. The area synthesized by our algorithm is approximately $20\,\mathrm{cm} \times 20\,\mathrm{cm}$ (C-arm detector size), which is small compared to the wide-angle field of view from the Kinect v2. Reduced to the area of synthesization, the video and depth from the RGBD camera are not of high resolution enough for sharper results. More

(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ (b) $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ (c) $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ (e) $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ (f) Inpainting

Fig. 8.6. Multi-layer image I_l of one selected frame from Sequence 3 with different blending parameters $(\alpha, \beta, \gamma, \delta)$

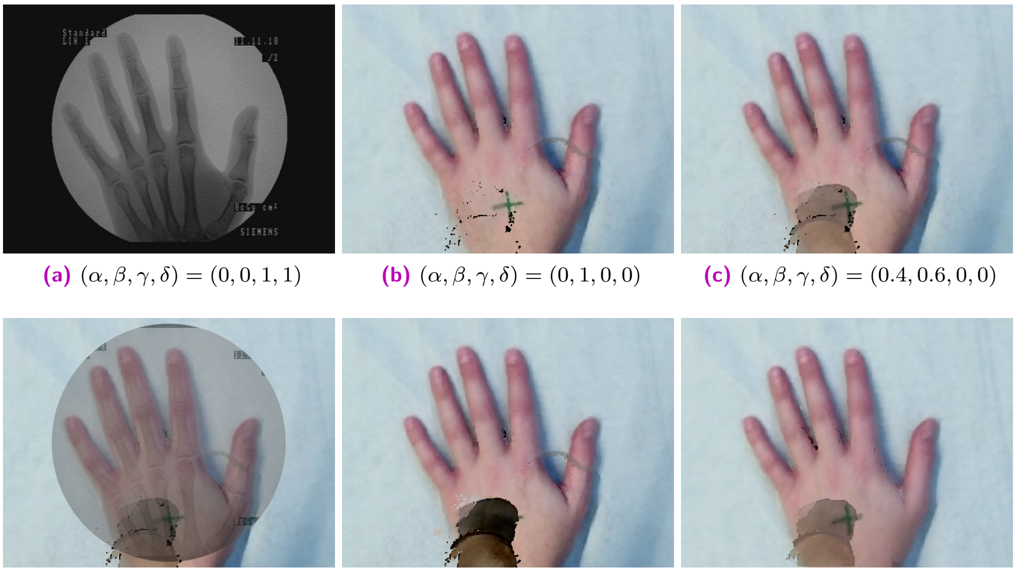


(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ (b) $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ (c) $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ (e) $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ (f) Inpainting

Fig. 8.7. Multi-layer image I_l of one selected frame from Sequence 4 with different blending parameters $(\alpha, \beta, \gamma, \delta)$

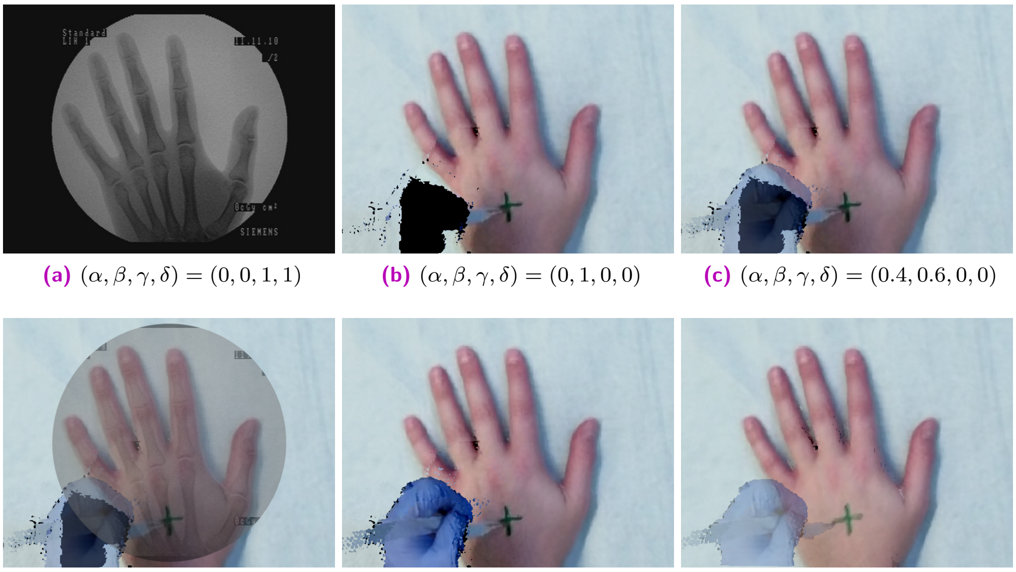
specialized hardware with a smaller field of view and higher resolution RGBD data would solve this problem. Moreover, several artifacts can be seen around the hand and surgical instruments in the synthesized image due to high difference and noise in depth in the RGBD data from the two cameras. However, our results demonstrate that our method is working well, since the incision line and cross drawn on the hand model and patient hand are perfectly visible in the recovered background image and can be seen in transparency through the hands and surgical tools in the images of Figure 8.6–8.9-column (c) and (d). In the scalpel sequence



(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ (b) $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ (c) $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ (e) $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ (f) Inpainting

Fig. 8.8. Multi-layer image I_l of one selected frame from Sequence 5 with different blending parameters $(\alpha, \beta, \gamma, \delta)$



(a) $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$ (b) $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$ (c) $(\alpha, \beta, \gamma, \delta) = (0.4, 0.6, 0, 0)$

(d) $(\alpha, \beta, \gamma, \delta) = (0.2, 0.3, 0.5, 0.5)$ (e) $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$ (f) Inpainting

Fig. 8.9. Multi-layer image I_l of one selected frame from Sequence 6 with different blending parameters $(\alpha, \beta, \gamma, \delta)$

(sequence 6) in Figure 8.9-column (b), it can be seen that the tip of the scalpel is considered as background, this is due to the margin of few centimeters used for background segmentation. In this image, the scalpel is actually touching the skin.

8.5 Discussion

Inferring temporal priors could help alleviate occlusion. Methods involving volumetric fields [114] use temporal information as the field is sequentially updating with new information, instead of being fully reinitialized as with our method. The percentage of pixels recovered is also dependent on the side cameras configuration. In our clinical case, the camera setup is constrained by the C-arm design and the disparity between the X-ray source and the two RGBD cameras is low. A higher disparity would lead to less occlusion in at least one of the cameras. Even with our constrained and difficult clinical setup, the results are promising and we are convinced the work could also be easily extended to less restrictive settings. A potential application is Industrial Mixed Reality where workers, wearing a Head Mounted Display with two cameras placed on their side (with a higher disparity than our setup), could see their viewpoint synthesized with their hands in transparency.

8.6 Conclusion

In this chapter, we have presented a novel visualization paradigm combining Diminished and Augmented Reality in the medical domain. Our visualization scheme proposes a user-adjustable multiple layer visualization where each layer can be blended with others. The multiple layers comprise the anatomy within the X-ray image, the patient background, and the surgeon's hand(s) and surgical instruments. The result of our visualization scheme offers the clinician to choose which layers are to become transparent depending on the surgical scenario or workflow step. Beyond the medical domain, this work is the first use of the volumetric field for background recovery in Diminished Reality and Mixed Reality. Future works should involve adding additional layers, by disassociating the surgeon's hand(s) layer from the surgical instruments layer, in order to adjust further the visualization to the user preferences.

9 Assistive system for Minimally Invasive Scoliosis Surgery

The work presented in this chapter is an extended version of the paper presented at AECAI 2017 [52] whose content is distributed under CC BY-NC 3.0 license.

9.1 Introduction

9.1.1 Idiopathic Scoliosis

The term “scoliosis” derives from the ancient Greek word σκολίωσις “skoliosis” (a bending) and was first established by Galen (130–201 AD). Adolescent idiopathic scoliosis is a lifetime condition of unknown cause, resulting in a three-dimensional deviation of a person’s spine of more than ten degrees. Its incidence is about 2% in the population [70] and is the most common spinal disorder in children and adolescents. 90 % of idiopathic scoliosis cases start at the teenage age [77]. The prevalence of scoliosis is higher in girls than boys with a ratio of 1.5-3 for 1, this ratio increases with the severity of the disease [77]. The most common diagnostic tool for scoliosis is the radiographies in the sagittal and the coronal plane, which will respectively give the sagittal and coronal Cobb Angles, i.e. the maximal angles of deviation. We show in Figure 9.1 the definitions of the body planes (sagittal, coronal and transverse), as well as the illustration of the coronal Cobb angle in Figure 9.2.

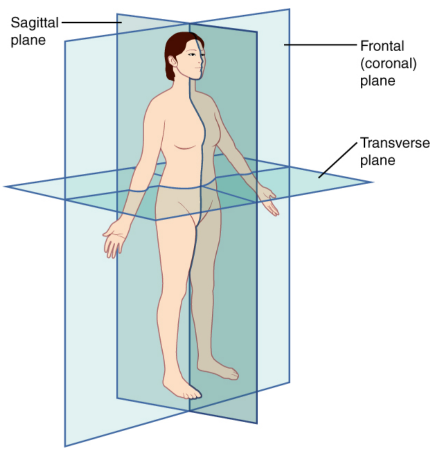


Fig. 9.1. Planes of the body, by Connexions, distributed under CC BY 4.0 license, Link to Image

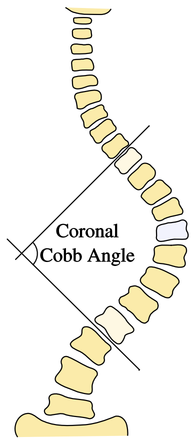


Fig. 9.2. Coronal Cobb Angle, by Wdwdbot, distributed under CC BY-SA 3.0 license, Link to Image

In Figure 9.3, we show images from a scoliotic patient in the 3D different planes, either through photography (transverse plane) or radiography (sagittal and coronal planes). The scoliosis correction depends on its severity and patient growth. If the patient is still growing and the Cobb Angles are higher than $20°$, the patient can wear a bracing that forces the spine to straighten. The growth of the patient will then naturally correct the deformities. We show in Figure 9.4 an example of bracing straightening the spine. If the scoliosis is very severe (more than $45°$), the surgery is recommended. We are going to explain in the next section the specificities of the scoliosis surgery.

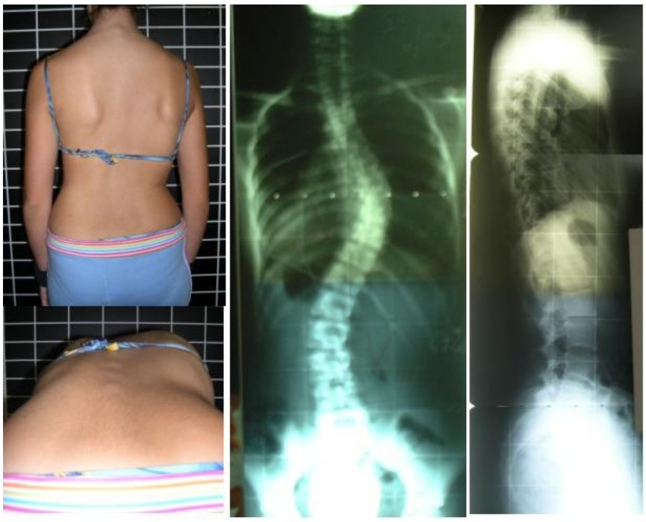


Fig. 9.3. (Left-top) Scoliosis in the coronal plane (body surface) - (Left-bottom) Scoliosis in the transverse plane (body surface) - (Center) Radiography in the coronal plane - (Right) Radiography in the sagittal plane, by Rigo M., Negrini S., Weiss HR., Grivas TB., Maruyama T., Kotwicki T., distributed under CC BY 2.0 license, Link to Image

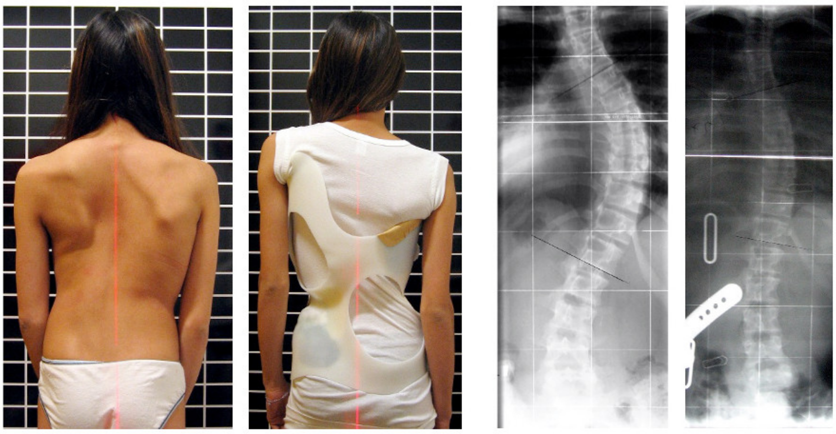


Fig. 9.4. The use of a brace allows straightening the spine, as shown in the right radiography by Weiss HR, distributed under CC BY 2.0 license, Link to Image

9.1.2 Scoliosis Surgery

Scoliosis surgery consists of permanently fusing two or more adjacent vertebrae such as they form a solid bone that no longer moves. Modern surgical approaches include tools such as rods, screws, hooks, and/or wires placed in the spine in order to straighten it. Scoliosis has been, for a long time, corrected using open surgery, however, with the development in the last years of Minimally Invasive Surgeries (MIS), the latter has become a popular alternative. It allows reducing the scar anesthetics, the patient pain and discomfort, inducing a faster recovery time, resulting in fine in an overall treatment costs reduction. MIS, however, come with new challenges for the surgeons. The surgical action is only performed through tools and endoscopic images, which can result in a loss of tactile feedback and dexterity as well as a loss in the perception of the surgical site. One surgical technique used for MIS scoliosis surgery is Video-Assisted Thoracoscopic Surgery (VATS) [115], shown in Figure 9.5, that consists of introducing the tools and endoscope camera through the patient side and ribs. It prevents from unaesthetic scar on the patient back, minimize trauma to the patient muscles and allows faster recovery. Fluoroscopic images are also acquired to guide the pedicle screw placement

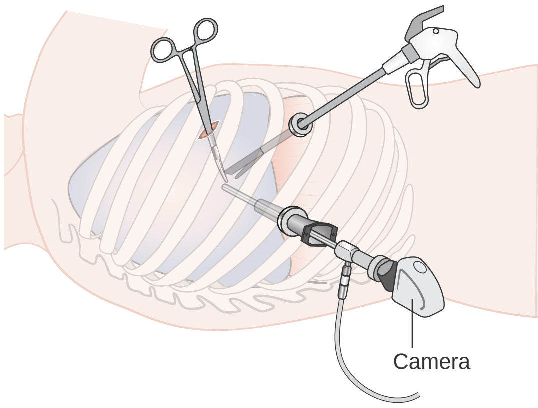


Fig. 9.5. VATS surgery, by Cancer Research UK, distributed under CC BY-SA 4.0 license, Link to Image

and to visualize the effect of the intervention on the curvature of the spine, however, it only shows the effect on the spine in the coronal and sagittal plane. Cosmetic external appearance is the main concern of the patients, while the correction of the spine by achieving coronal and sagittal trunk balance, visible through fluoroscopy, is the priority for surgeons [24]. The main issue is that the two concerns are not necessarily correlated, as a correction of the spine deformations in the coronal and sagittal planes do not automatically reduce the back deformities or can even increase them. In the images shown in Figure 9.6, we show the results of a scoliosis surgery that is very satisfying regarding the correction of the spine (second image from the left) as the spine is almost straight, however, the cosmetic results are unsatisfactory to the patient due to the visual prominence of the shoulder blade [177].

This non-automatic correlation between the patient surface deformities and the correction of the spine has led the scoliosis community to invest systems and metrics to assess those back deformities.

9.1.3 Patient Surface Acquisition

Since the seventies, the community has investigated patient trunk surface reconstruction in order to assess pre- and post-operatively the scoliosis inferred deformations [93] using visible structured light. With the Kinect 1.0 release, the research has lately moved towards RGBD

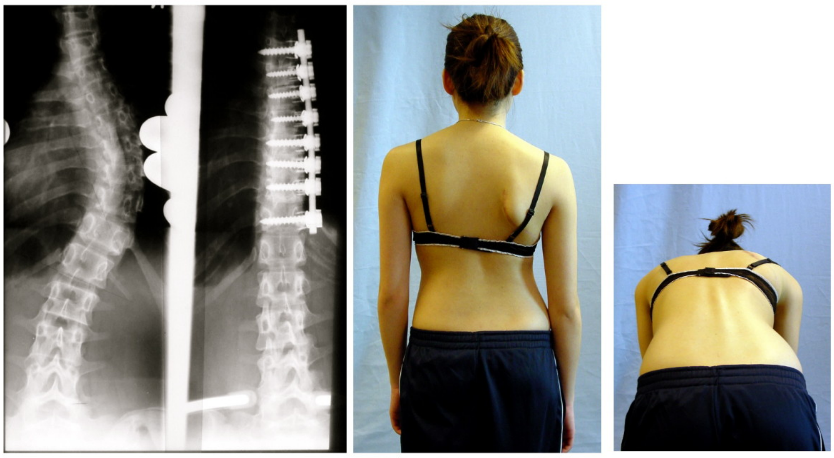


Fig. 9.6. Severe scoliosis with back deformities remaining after surgery, by Weiss HR [177], distributed under CC BY 2.0 license, Link to Image

cameras to perform pre- and post-operative patient surface reconstruction. Three techniques can be found in the literature:

- Fusion of point cloud from multiple cameras [20]
- Kinect Fusion 3D surface reconstruction [27]
- Point cloud from single RGBD camera [131]

The first and third items could already be found before with visible structured light [144]. We show in Figure 9.7 a patient surface reconstruction from visible structured light INSPECK system. However, those devices are costly, in opposition to RGBD cameras that the mass scale effect made low cost (< 200\$).

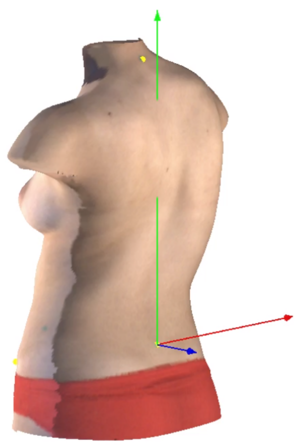


Fig. 9.7. Patient Surface Acquisition by visible light structured light system Inspeck

9.1.4 Metric of Patient Surface Deformities

From the patient surface reconstruction, the researchers have aimed to create indexes and metrics to assess the scoliosis deformities in the three planes. We refer the reader to the extensive review of those metrics by Patias et al. [120]. For the sagittal and coronal planes, the main target is to replace the control radiography and, therefore, the sagittal and coronal Cobb Angles by non-invasive measurements in order to reduce repeated exposure to radiation. Lately, Seoud et al. proposed a novel index to assess patient surface deformities in the three planes [144]. This new index allows a description of the deformations at all trunk levels and not only at the deformity apex and includes complementary measurements taken in the three planes: the Back Surface axial Rotation (BSR), the trunk deviations in the coronal plane and in the sagittal plane. This index was already used to evaluate quantitatively the patient surface deformities post-operatively to compare it with spinal measurements from radiographies [143]. As a result of the study, it was shown that current surgical techniques perform well in realigning the patient surface in the coronal and sagittal plane, however, the deformity correction in the transverse plane, measured by the BSR, is more challenging. Indeed, the latest cannot be measured with radiographies during surgery in opposition to the spinal deformities in the two other planes and in scoliosis surgery, often a C-arm is the only assessment tool available. The surgeon must, therefore, rely on experience and naked eye view during surgery to assess the deformations in the transverse plane.

9.1.5 Intra-operative Assessment of Scoliosis

C-arm devices are often the only available tool for spinal deformities measurement in the Operating Room. At the best of our knowledge, no patient surface reconstruction work in the literature tried to bring the external deformities measurement intraoperatively. To be used as an assistive tool during surgery, the assessment of trunk surface deformation must respect several constraints in order to bring minimal perturbation to the surgical workflow such as:

- Real-time acquisition
- Minimal setup
- Automated process (minimum human intervention)
- Legible visualization of the metrics

Most of the trunk reconstruction works discussed in Section 9.1.3 consist of spatial or temporal multi-camera acquisition. Multi-camera acquisition during surgery is feasible and can be done in real-time, however, it is cumbersome on a C-arm or prone to occlusion if mounted on the surgical room ceiling. Using only one camera is preferable, but with a real-time reconstruction, which was a criterion not met by the only work proposing a single camera setup [131].

9.1.6 Proposed Solution

We propose the first intraoperative assistive system for scoliosis VATS surgery composed of a single RGBD camera affixed on a C-arm, that reconstructs the patient back surface and provides the real-time visualization of the surgery effects on the patient surface in the transverse plane using the BSR index calculated from RGBD data. This comes as a complement of the coronal and sagittal deviations metric measurements performed via X-ray images during surgery. The proposed setup enables the surgeon to adapt the strategy dynamically according to the patient response.

9.2 Methodology

9.2.1 Setup

The setup of the assistive system consists of an RGBD camera (PrimeSense Carmine Short Range) attached on a C-arm. We attach the camera at the middle of the C-curve as it can be seen on Figure 9.8, due to the configuration of the C-arm during VATS surgery. Indeed, during such operations, the surgeon usually acquires LAT images (vertical position of the C-arm) to assess the sagittal plane deformities and AP images (horizontal position of the C-arm) for the coronal deviation. We assume that the C-arm resting position (when not used) is the vertical position as it is less cumbersome and takes less space and we, therefore, place the camera at the middle of the C-curve. If the surgeon keeps the C-arm in the horizontal position, the camera could be placed at the intensifier and our work would still be valid. In both cases, the goal of the camera position is to be in normal incidence to the patient back in order to visualize the whole back of the patient. Our setup is also justified by the description of VATS scoliosis surgery reported by Newton et al. [115]. Indeed, they describe that the surgeon is placed anterior to the patient that is laying on its side, with its back free of any draping making it visible by the camera that will be placed posterior.

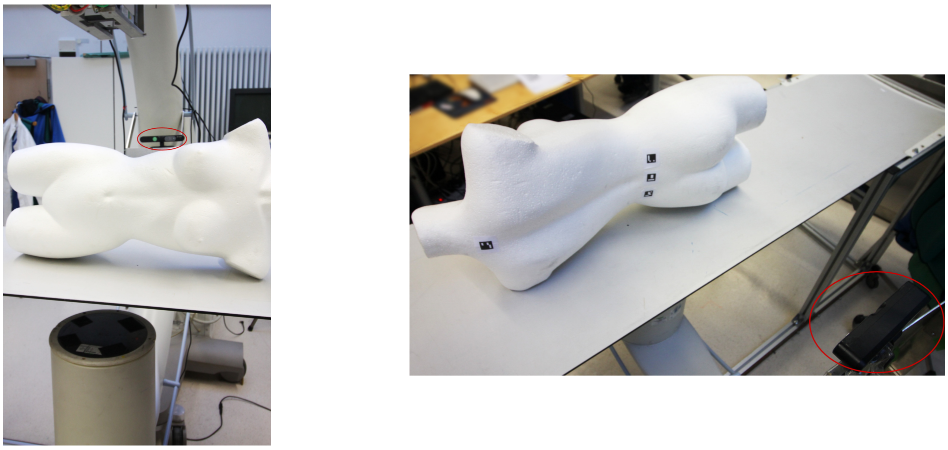


Fig. 9.8. Setup with RGBD camera (red circle) placed at middle of the C-arm curve

Using the data acquired from the RGBD camera, the Back Surface Rotation metric is computed following the different steps explained in the pipeline figure 9.9. We explain further those different steps.

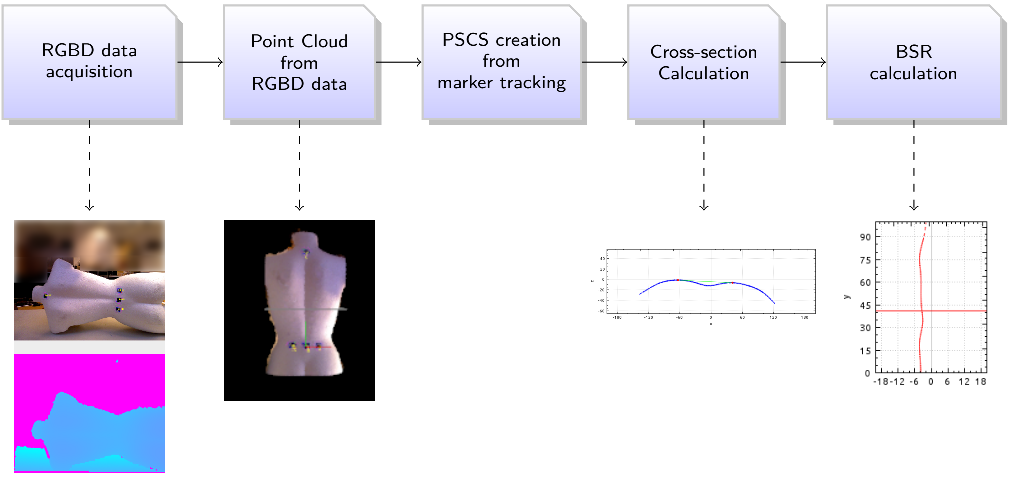


Fig. 9.9. Pipeline of the assistive tool

9.2.2 Patient Point Cloud from RGBD Data

Using the OpenNI library [1], the video image and the depth image of the RGBD camera can be acquired at the frame rate of 30FPS, the video image size is 1280×1024 pixels while the depth image is 640×480 pixels. The OpenNI library also offers the mapping Ω from the depth image to the video image. As explained in the background chapter with the Equation 3.5, for every incoming pair of video/depth images, we can reconstruct the corresponding colored point cloud in the depth camera coordinate system as shown in Fig 9.10.

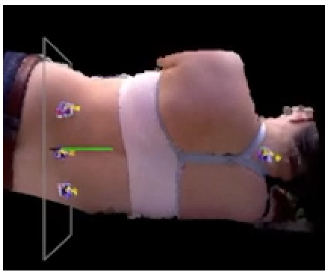


Fig. 9.10. Point cloud representing the subject

9.2.3 Patient-Specific Coordinate System

Patient-Specific Coordinate System (PSCS) are commonly used [120] for scoliosis measurement in order to provide comparable intra- and inter- patient metrics. The PSCS is defined by anatomical landmarks that are easily traceable, ours are defined by the four anatomical landmarks (L_1, L_2, L_3, L_4), respectively the C7 Vertebral Prominence (VP), the posterior-superior iliac spines' midpoint (MPSIS) and the left and right superior iliac spines (LPSIS, RPSIS). The landmarks location is shown in Figure 9.11.

1. https://github.com/occipital/openni2

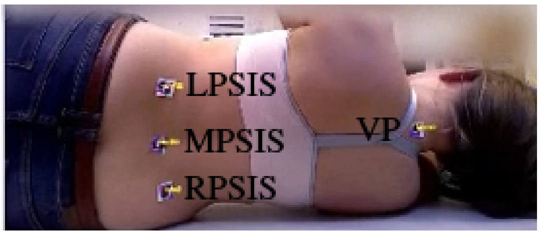


Fig. 9.11. Anatomical landmarks location for the PSCS

The anatomical landmarks are identified by the surgeon at the beginning of the procedure. Once identified, trackable markers (in our case, AR markers) are placed at those landmarks allowing them to be tracked all along the surgery automatically without any human intervention. Using the ARUCO library[2], we detect the AR markers centers in the color image (u_{cm}, v_{cm}) and compute their coordinates in the depth image $(u_{dm}, v_{dm}) = \Omega^{-1}(u_{cm}, v_{cm})$. As depth images are noisy, we use as depth value the average of the valid depth values around a local neighborhood of the marker center instead of relying on a single depth value. Thus, we finally compute the 3D points for the four anatomical landmarks using the depth to 3D point equation: $L_i = (\frac{d(u_{dm}-u_0)}{f}, \frac{d(v_{dm}-v_0)}{f}, d)^T$. Using the four 3D anatomical landmarks positions, we compute the transformation $T_{PSCS \to D}$ from the PSCS to the depth camera coordinate system D using Equation 9.1. We can then transform the patient point cloud generated in D to the PSCS.

$$T_{PSCS \to D} = (\vec{n_x}, \vec{n_y}, \vec{n_x} \wedge \vec{n_y} | L_2) \text{ with } \vec{n_x} = \frac{L_4 - L_3}{||L_4 - L_3||}, \vec{n_y} = \frac{L_1 - L_2}{||L_1 - L_2||} \tag{9.1}$$

As the patient is lying on the side, the table can also appear in the point cloud and, therefore, disturb the BSR metric calculation. However, as the patient is centered in the PSCS, we can create a cropping box aligned to the PSCS axis, that will exclude every point with $|X| > X_T$. This threshold X_T can be dynamically changed by the user in the User Interface.

9.2.4 Cross-section Computation and BSR Metric

In the PSCS, we compute the BSR metric as defined by Seoud et al. [143] by creating N cross-sections of thickness t along the PSCS Y-axis with $t = \frac{||L_1 - L_2||}{N}$. In this work, we fixed the number of cross-sections to $N = 100$ to make our measurements comparable. Every point (X_p, Y_p, Z_p) in the point cloud is assigned into its n-th cross-section with $n = \lfloor \frac{Y_p}{t} \rfloor$ if $n \in [0, N-1]$. If $n < 0$, then the point is under the waistline and is not of interest, same for $n \geq N$ above the VP landmark. In every cross-section, the 3D points are mapped orthogonally to the 2D (X, Z) plane. Therefore, for every cross-section, we obtain a 2D curve representing the outline of the back as shown in Figure 9.12-top. Then, we perform a cubic spline regression on the curve to smooth the outline. The double tangent line to the outline (the line touching the outline at only two points) is then computed (green line in Figure 9.12-bottom). Its angle to the X-axis (the horizontal axis in our graph) is the BSR value at this cross-section. The BSR curve is, therefore, obtained by computing the BSR value at every cross-section.

2. https://www.uco.es/investiga/grupos/ava/node/26

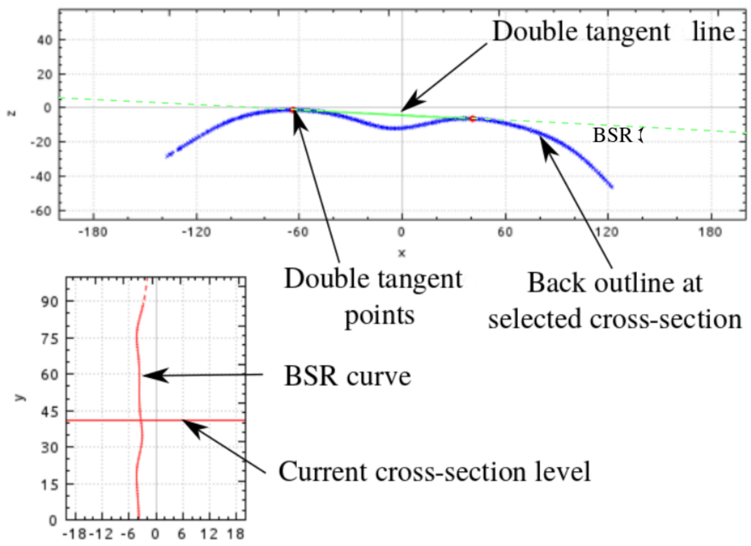


Fig. 9.12. (Top) Outline of the back curve at one cross-section, (Bottom) BSR curve along the cross-sections, the red line shows the selected cross-section

9.3 Experiment and Results

We perform several experiments to validate the feasibility of our approach, first using simulated depth images from real scoliotic patient models, then using real acquisition on a non-scoliotic mannequin that is compared to 3D reconstruction, and finally, we show qualitative results on a non-scoliotic moving person.

9.3.1 Evaluation Using Simulated Point Cloud from Real Scoliotic Patient 3D Models

The first experiment aims at quantifying the error induced by the use of RGBD data on the BSR metric calculation. Complete (back and front) 3D models of scoliotic patients are used, acquired by the INSPECK system at the Sainte-Justine Hospital in Montréal before and after surgery. Each model is already placed in its respective PSCS. To measure the error induced by the RGBD data regardless of tracking error on the PSCS, we simulate depth images from the 3D models (which is our ground truth) and reconstruct the 3D point cloud corresponding to the simulated depth image, which is then a degraded partial view of the original 3D model. Both are placed in the same PSCS. We show the pipeline of our experiment in Figure 9.13. The simulated depth image is computed by constructing an octree on the full 3D model point cloud and by performing raytracing from the simulated viewpoint. The first intersection with the octree of a ray originating from the simulated viewpoint for each simulated depth image pixel gives the depth at this pixel. For more realism, we add noise on the obtained depth value following the gaussian model of the axial error on Kinect 1.0 type of camera given by Nguyen et al. [116].

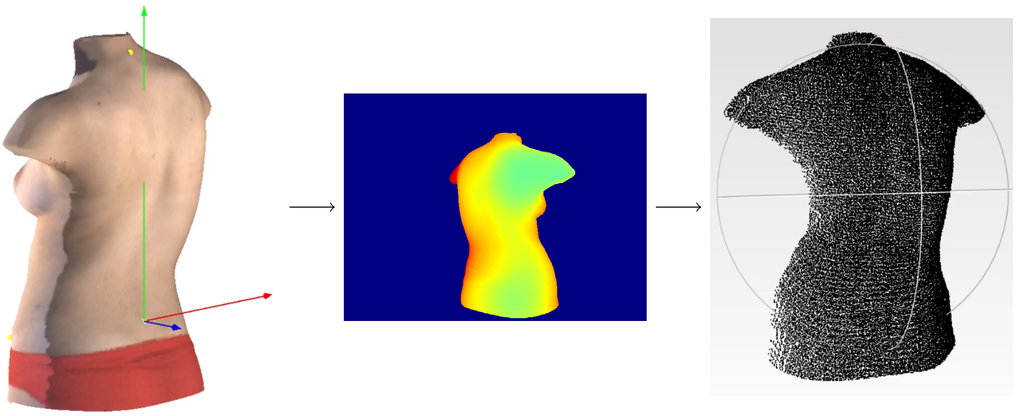


Fig. 9.13. Process of simulated point cloud from scoliotic patient data : (Left) 3D model of scoliotic patient, (Center) simulated depth image, (Right) point cloud generated from simulated depth image, distributed under CC BY-NC 3.0 license

For each six 3D patient models, we generate simulated depth images and, therefore, point clouds from different poses. Thoses poses are defined by a rotation R computed with the Euler chained rotations around the PSCS X-axis and Y-axis, leading to $R = R_x(\alpha)R_y(\beta)$ with $(\alpha, \beta) \in \{-30, -20, -10, 0, 10, 20, 30\}^2$ and a translation on the Z-axis from $70\,\mathrm{cm}$ to 1 meter by step of $10\,\mathrm{cm}$. In total, 196 poses are used per model. For every class of poses C_i, we compute $RMSE_{C_i}$, according to Equation 9.2, it represents the RMSE error on the sum of residuals on each pose p in the class, which, in fact, consists of calculating the RMSE error between the BSR measurement $BSR_{p,j}$ at all cross-sections j and pose p compared to ground truth GT.

$$RMSE_{C_i} = \sqrt{\frac{\sum_{p \in C_i} \sum_{j=0}^{N} (BSR_{p,j} - GT_j)^2}{N|C_i|}} \tag{9.2}$$

The results for poses classified by angles (α, β) per model are presented in Figure 9.14 under heatmap plots with α variating on the vertical axis and β on the horizontal axis. Each class (one heatmap cell) is in this case composed of four poses only variating by their distance to the patient. The results for poses classified by the distance to the patient per model are shown in Figure 9.15. Each class is in this case composed of 49 poses.

In the two graphs, the RMSE error is under $1.3°$, which is under the BSR Typical Error of Measurement of $1.75°$ reported by state of the art work from Seoud et al. [144] using the INSPECK system. In Figure 9.14, for $\alpha > 10$, we can observe that the RMSE error increases for all patients, indeed those poses are leaning towards the patient head and parts of the back surface are then occluded by the posterior rib hump, more prominent than the waistline even in a normal subject. The observation of the RMSE error range for each patient shows that the angle influence is patient-dependent, as the error range is small for Patient 1 (within a range of $0.1°$) while the range is 10 times higher for Patient 3. Classifying the poses by the distance to the patient as in Figure 9.15, we observe an increase with the distance, consistent with the axial depth noise property. This experiment shows that a system using RGBD data from a

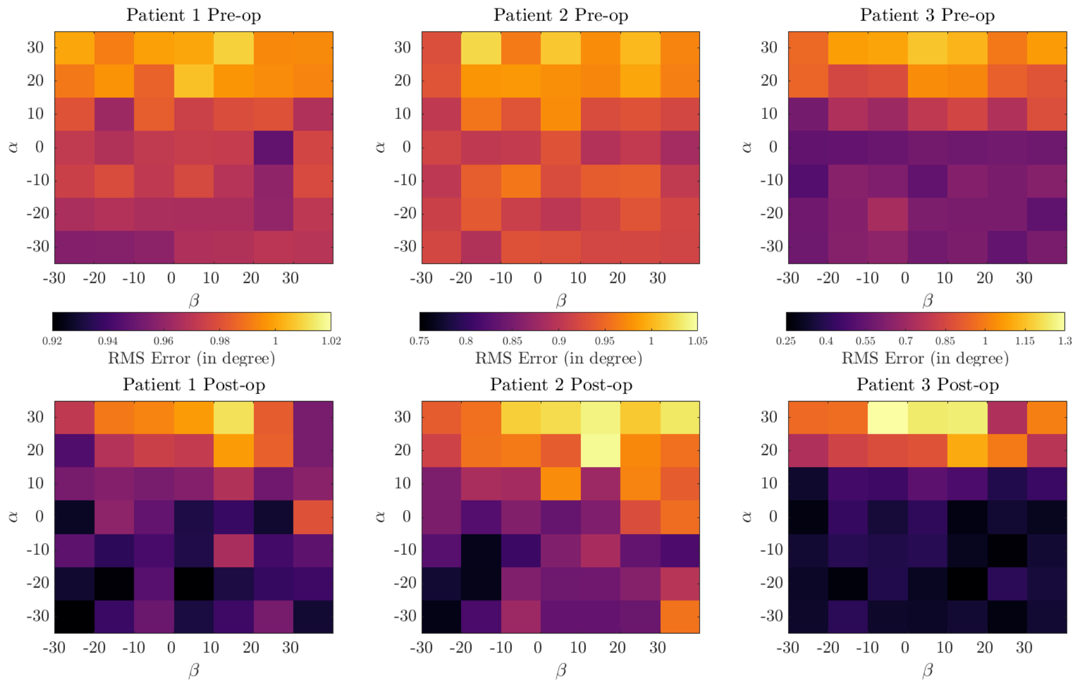


Fig. 9.14. RMSE error for (α, β) angles at all distances, one colorbar by patient

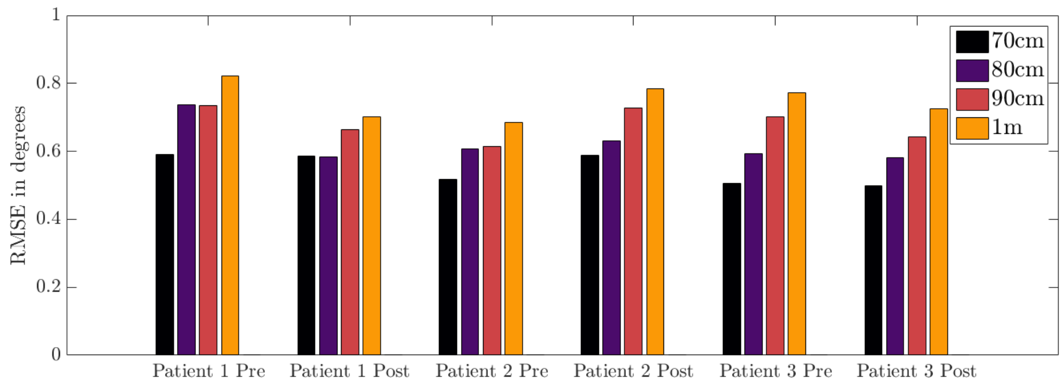


Fig. 9.15. RMSE error for poses classified by distance

single viewpoint has an acceptable accuracy for BSR calculation, assuming, however, no error on the PSCS tracking. The latter is taken into account in the next experiment.

9.3.2 Evaluation on Real Acquisition of Non-Scoliotic Mannequin

In opposition to the first experiment, this experiment is using live RGBD data from our framework. For evaluation, we use a static polystyrene female mannequin with no scoliosis. Four AR markers are placed at the anatomical landmarks described in Section 9.2.3. Then, we position the mannequin at four different poses that can be tracked by our system with various in angulations. The mannequin is approximately placed at 80 cm from the camera. First, the reliability of our measurement is tested by recording for each pose multiple BSR measurements (100 for our experiment). We compute the absolute agreement score between all the measurements for each pose using the IntraClass Correlation $ICC_{3,1}$. The reliability score for each pose is shown in Table 9.1. Then, we quantify the full error induced by our system including the PSCS tracking error by reconstructing in 3D the back of the static mannequin using the non-real-time Kinect Fusion algorithm from RecFusion[3] and shown in Figure 9.17.

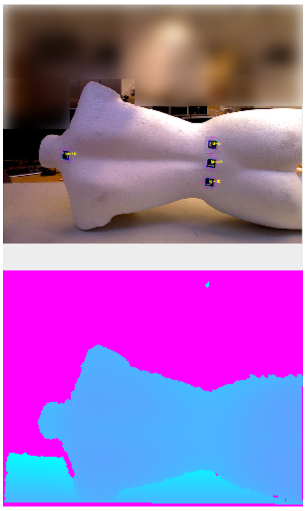


Fig. 9.16. Live Acquisition of the Mannequin

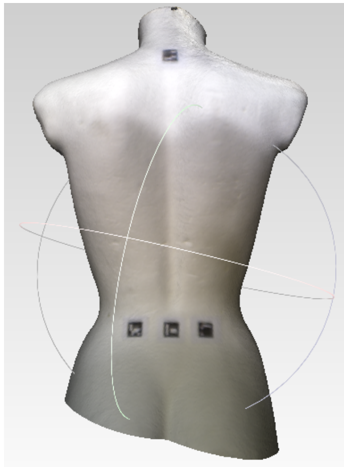


Fig. 9.17. 3D Reconstruction of non-scoliotic mannequin

We manually segment the 3D position of the marker centers on the 3D mannequin reconstruction and compute the PSCS for the 3D reconstructed mannequin. Then, we compare the BSR measurements between the 3D mannequin and live acquisitions with our system at the different poses performed during the reliability test by computing the RMSE error on the sum of residuals from all BSR measurements and the 3D reconstruction BSR measurements, using the Equation 9.2 with only one pose per class. The RMSE error is shown in Table 9.1.

Pose	$R_x(30)$	0	$R_y(30)$	$R_y(-10)$
Reliability Score	0.885	0.953	0.881	0.971
RMSE error (°)	0.507	0.674	0.508	0.879

Tab. 9.1. Reliability Score and RMSE error to 3D reconstruction for several poses

The reliability score is above 0.85 for the four poses, considered as a good reliability [144]. This implies that for one given pose, we should always get the same measurement. The RMSE error is under one degree for the four poses, which is under the BSR Typical Error of Measurement of $1.75°$ previously reported. As this experiment tests the full framework pipeline, it shows that our system can be used for BSR calculation.

As mentioned earlier, we reported the RMSE error for poses when markers could be tracked, we can see in the poses used that they are not symmetric in the angles used. Indeed, the AR marker detection is heavily sensitive to the inclination to the camera. If the marker is too

3. www.recfusion.net

inclined, the tracking is lost and the PSCS calculation cannot be done. In our experiment, the detection of the VP marker is in some poses complicated, e.g. for poses where $\alpha < 0$ corresponding to the neck further than the pelvis. The neck curve makes the marker, very inclined and, therefore, not detectable. Robust to perspective change markers such as proposed by Birdal et al. [14] would help to overcome this issue.

9.3.3 Qualitative Results with Real Acquisition of a Non-Scoliotic Person

For the last experiment, we use our system on a non-scoliotic person (the author of this thesis, a 30-year-old female) deforming her back during the acquisition. As a result, we show that our system can track the deformations progression in real-time. The processing time per RGBD frame is less than 25 milliseconds. We show in Figure 9.18 two frames of this acquisition; the left image is taken as a reference (green line on the BSR graph). On the right image, the person is increasing the posterior rib hump by moving backward the left shoulder, increasing the BSR angle at the spinal top level in the negative values as it can be seen on the red curve of the BSR graph.

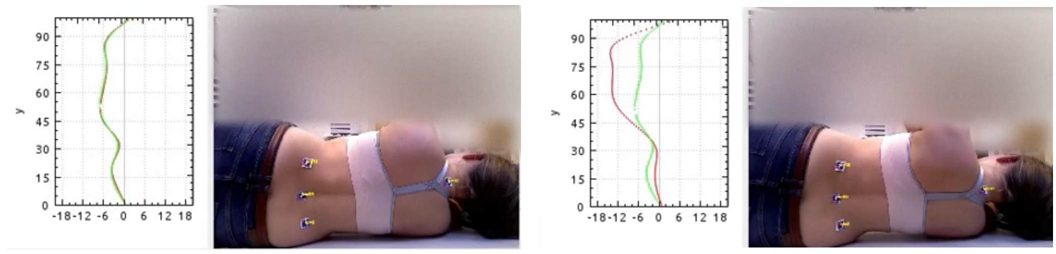


Fig. 9.18. Left image position as reference, right with person increasing rib hump

9.4 Discussion and Perspective

The last experiment brings the question about the reference position. The subject of this experiment is not scoliotic, but her BSR curve is, however, not straight. This is due to the choice of the reference position. For the pre-operative and post-operative BSR measurements, the acquisitions are done with the patient standing in a position that is going to be similar because natural for most of the subjects. However, there is no standard position when lying on the side and a position similar to the standing one (where the patient is all straight from tip to toe while lying on the side) is very difficult to hold and will be hard to achieve during surgery without the help of instrumentation. Two opposite directions can be investigated to solve this issue. First, we can enforce a standard position by bringing the patient the closest to a straight position while lying with the help of instrumentation. A pillow between the waist and the table would, for example, help to keep the spine straight but would appear in the depth image, however, this is a direction to explore. The second direction is to adapt to the patient, by correlating the intra-operative changes with a standing pre-operative model. This could be done at the beginning of the surgery, by registering rigidly or elastically the patient's back point cloud when lying on the side with the standing position. During surgery, the system can then measure the differences compared to the reference at the beginning of surgery and then, induce them to the BSR values measured while being in standing position.

Finally, in this work, we did not exploit the fact that the RGBD camera is attached on the C-arm and, therefore, the X-ray images should also be fully integrated into the framework in combination with the 3D point cloud acquired from the RGBD camera. To perform so, the RGBD camera needs to be registered to the X-ray source using, for example, the works of Wang et al. [173], also used in Chapter 5. Then, the X-ray image could be used in the framework in combination with 3D point cloud for visualization such as proposed in Chapter 6 and later, to retrieve a 3D spine reconstruction from single X-ray image using the articulated model such as proposed by Boisvert et al. [16]. The combination of both 3D representations would finally allow exploring their mechanical interaction to suggest intra-operatively strategies of spine corrections based on trunk deformations.

Through our experiments, we have demonstrated that the proposed system shows promising results regarding accuracy, showing a better accuracy than the state of the art. The next step is to validate our system on scoliotic patients by bringing it into the Operating Room, first as a non-interfering system to study the validity of the BSR metric within the clinical context as well as the hypothesis regarding the surgical context described in Section 9.2.1 on which our system is built. Existing works in the literature such as Navab et al. [113] have proven that camera integration on C-arm in the OR is feasible. However, several challenges remain to be answered and investigated before the full integration in the scoliosis surgery workflow.

9.5 Conclusion

To conclude, we have shown in this chapter the feasibility of the first intra-operative assistive system for scoliosis VATS surgery composed of a single RGBD camera affixed on a C-arm through multiple experiments from simulated data to the real-time acquisition on a person. Our proposed system can measure the changes on the BSR index in the accuracy range defined by state of the art of scoliotic patient reconstruction.

10 C-arm based Surgery Simulation for Training and new Technology Assessment

The work presented in this chapter was co-authored with Philipp Stefan and is an extended version of the paper presented at MICCAI 2017 which is reproduced with permission from [154], ©Springer.

10.1 Introduction

Despite the progress in image-guided interventions over the last 25 years [124] and a widespread use of navigation systems in North America and Europe [56], conventional fluoroscopy remains the most frequently used intra-operative imaging modality in surgery. In spine surgery, 87% of surgeons worldwide use routinely fluoroscopy compared to 11% using navigation systems [56]. The primary challenge in minimally invasive image-guided interventions is the surgeons' ability to mentally recreate the 3D surgical scene from intra-operative images [124], as surgeons do not have a direct view of the surgical area. During C-arm based procedures this ability directly depends on the correct manipulation of the C-arm performed by an operator, usually a nurse [21], based on the communication with the surgeon. Perfection in surgery requires extensive and immersive experiences to acquire the relevant surgical skills [2]. However, due to several obligatory working-hour restrictions [2], increasing costs of operating room time and ethical concerns regarding patient-safety, clinical training opportunities are continuously decreasing while the complexity of interventions is continuously increasing. At the same time, the integration in the Operating Room of new C-arm based technologies such as the RGBD augmented C-arm presented in the previous chapters, that could increase the surgery efficiency in terms of time, outcome, and radiation exposure reduction by facilitating the mental mapping, is difficult due to the extensive validation required before entering the OR in addition to the limitations already explained before for the training case. If not happening in the OR, training and assessment of new technology both require as realistic as possible setups for being efficient and valid [65].

10.2 State of Art

In that prospect, training/technology assessment models have been proposed. While the use of animal or human cadaver for training/technology assessment provides adequate haptic feedback and fluoroscopic images, it requires X-ray radiation, is costly, ethically problematic, and pathologies relevant to the trained procedure are, in general, not present in the specimen.

For the case of new technology assessment, the lack of reproducibility due to the unique nature of every animal and human cadaver is an obstacle to statistical significance, necessary to obtain strong evidence of improvement compared to the state of the art. Commercially available synthetic training models offer only a very limited range of pathologies and typically do not show realistic images under X-ray imaging. To keep the realism without the exposure to X-ray radiation, computer-based simulation has been developed [21, 23, 49, 55, 182]. Most simulators including X-ray imaging target the spine, due to its complex anatomy and proximity and interlacing to critical structures. Most described works on C-arm simulators use the principle of Digitally Reconstructed Radiographs (DRR) to create X-ray images without radiation from Computed Tomography (CT) [21, 23, 49, 182]. The representation of a C-arm and its control has been realized in simulators at different degrees of realism from virtual to real C-arms. Gong et al. [49] mount a webcam next to the C-arm source to track it relative to a virtual patient represented by an empty cardboard box with AR markers. A DRR image is generated from clinical 4D CT data using the tracked position. Bott et al. [23] use an electromagnetic tracking system to track a physical C-arm, the operating table and a mannequin representing the patient to generate the DRR images. The two latest systems, however, are not suited for interventional surgical training, as no physical anatomy model matching the image data is present.

To add realism to the simulation, several works aim at presenting patient-based anatomy in a tangible manner. Despite a high cost, haptic devices are broadly used in surgical simulators for force feedback generation depending on the anatomy represented in the CT, in a few cases combined with a real C-arm. Wucherer et al. [182] place a physical C-arm as part of an operating room decor, but its functions are not linked to the simulator. Rudarakanchana et al. [136] combine a C-arm simulator with an endovascular simulator, whether both systems are spatially registered is not stated. Patient anatomy can also be represented physically by 3D printing, which is already commonly used for procedure planning and training [174]. Harrop et al. [55] reproduce the equivalent of multiplanar display used in navigation using 3D printed models from CT scans. To resume, several works simulating C-arm operation and replicating patient anatomy from medical imaging data exist but none of them bring both in an accurately registered spatial relation.

10.2.1 Contributions

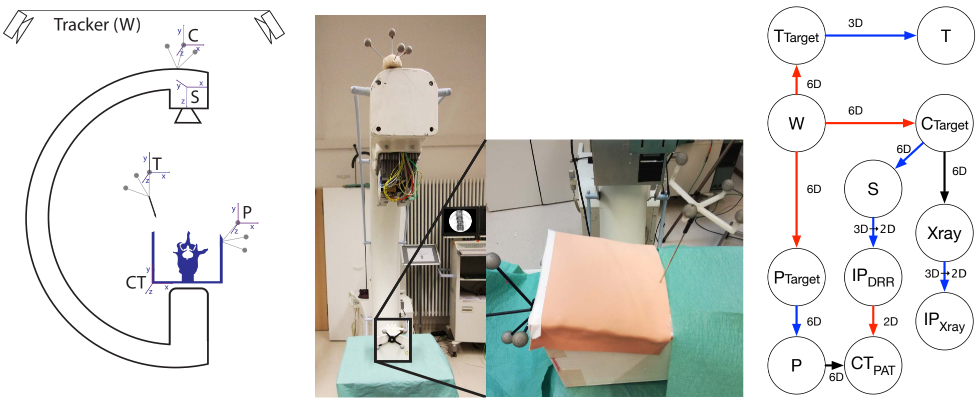
The proposed mixed-reality approach that combines patient-based 3D printed anatomy and simulated X-ray imaging with a real C-arm is a complement to the traditional training and C-arm based new technology assessment pipeline. To the best of our knowledge, no other simulation environment places a radiation-free physically present C-arm in an accurate spatial relation to simulated patient anatomy. This allows the use of real instruments and accurately aligns C-arm images with a physical patient model, which is important for the training of hand-eye coordination and mental mapping of projection images to the surgical scene and patient anatomy, as well for the realism necessary to the validity of new technology assessment. The patient-based models are created from CT data using a 3D printer and can be replicated as often as needed at low cost. The printed models contain the pathology present in the underlying CT data, in contrast to cadaver specimens that most often do not contain a relevant pathology. A further contribution is the transfer of the Spatial Relationship

Graphs (SRG) concept from Industrial AR [128] to Computer Assisted Interventions (CAI). An SRG is a directed graph in which the nodes represent coordinate systems, edges represent transformations between adjacent coordinates systems. Along this work, we use SRGs to provide an intuitive and visual description of the complex, dynamic transformations chain of tracked objects and calibrations implicated in the mixed-reality system.

10.3 Methodology

10.3.1 Setup

In the proposed system, both the C-arm (C), the 3D printed patient model (P) and the tool (T) are physical objects tracked using a ARTTRACK2 four optical cameras outside-in tracking system (W). A schematic representation of the setup is shown in Figure 10.1a, whereas the real setup is depicted in Figure 10.1b. In order to simulate an X-ray acquisition, the position of the virtual camera (S) in the CT coordinate system needs to be computed. Figure 10.1c shows the SRG of this simulation system, detailing on the transformations spatially linking all components. Edges are labeled with the type of transformation: $6D$ for 3D rigid transformations, $2D$ and $3D$ for 2D and 3D translations, $3D \rightarrow 2D$ for projective transformations. Edges not varying over time are labeled *static*, edges that do vary are labeled *dynamic*. Edges that need to be *calibrated* are static by definition throughout this work. The following colors are used in figures: blue: *calibrated*, black: *static*, red: *dynamic*.



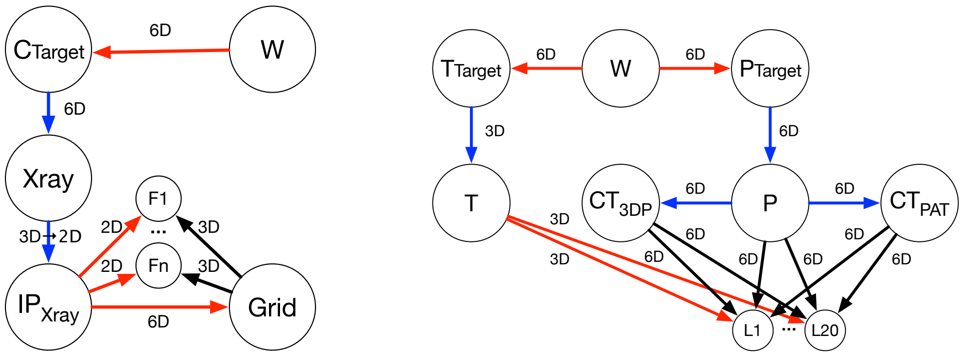
(a) Schematic Setup (b) C-arm and 3D Print (c) SRG

Fig. 10.1. (a) and (b) Overview of the proposed system with C-arm, 3D print, and optical marker targets, (c) Spatial Relationship Graph (SRG) of the simulation system.

10.3.2 Synthetic Patient Model

From a patient CT dataset, a segmentation of the spine is created and four walls of a box are added around it. On the surface of these walls, twenty artificial landmark holes L_i are placed for the registration of the printed patient model to the CT data. From the segmentation, a surface mesh is created, which is then smoothed and printed in PLA on an Ultimaker2+ 3D printer. To this printed model (P), an optical tracking target (P_{Target}) is rigidly attached. For evaluation purposes, CT markers (shown in Figure 10.3b as yellow circles) are attached to the printed model and a CT scan of it is acquired (CT_{3DP}).

10.3.3 System Calibration



(a) $T_{C_{Target} \rightarrow Xray}$ calibration (b) $T_{P_{Target} \rightarrow CT_{PAT}}$ and $T_{P_{Target} \rightarrow CT_{3DP}}$ calibration

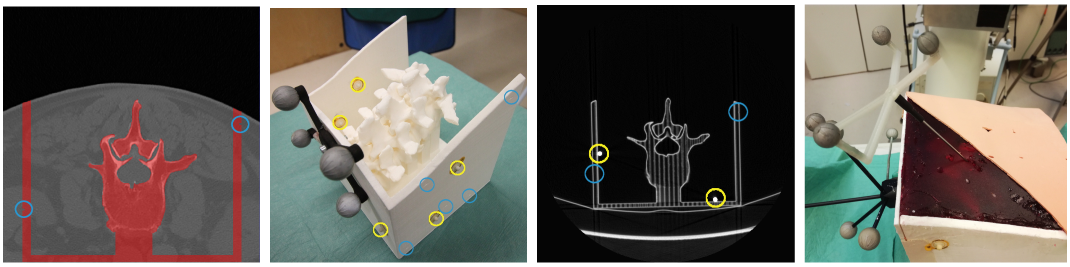
Fig. 10.2. Spatial Relationship Graphs: (a) C-arm Target to X-ray Source, (b) Print Target to patient CT and CT of printed model.

Calibration of Simulated X-ray Source S to the C-arm Target C_{Target}

To place the simulated X-ray source S at the C-arm real X-ray source, the calibrated transformation $T_{C_{Target} \rightarrow Xray}$ (Figure 10.2a) is required to obtain the dynamic transformation $T_{C_{Target} \rightarrow S}$. The calculation of this transformation can be solved using the hand-eye calibration algorithm [161] known in robotics and augmented reality. A planar grid of X-ray visible markers is placed on a fixed surface between the real X-ray source and the C-arm image intensifier. Multiple X-ray images of the grid are acquired from different poses of the C-arm and, based on the grid of markers, the C-arm camera pose with respect to the grid is computed using the Ransac PnP algorithm [17] implemented in OpenCV. For every X-ray image acquired, a pair of poses, composed of a C-arm tracking target pose $T_{W \rightarrow C_{Target}}$ and a C-arm camera pose in the grid coordinate system $T_{Xray \rightarrow Grid}$, is computed. From those pose pairs, the hand-eye calibration algorithm estimates $T_{C_{Target} \rightarrow S}$.

Calibration of Printed Model P to CT_{PAT} and CT_{3DP}

To render the DRR image spatially aligned with the printed model, we also need to obtain the transformation from the printed model tracking target to the patient CT coordinate system $T_{P_{Target} \rightarrow CT_{PAT}}$ (Figure 10.2b). For evaluation purposes, we also want to obtain the transformation $T_{P_{Target} \rightarrow CT_{3DP}}$ from the printed model tracking target to the CT of the printed model. For registration, 20 artificial landmarks L_i are placed in the segmentation of the patient CT and thus are observable in the printed model and its corresponding CT (CT_{3DP}) (Figure 10.3, blue circles). Using a pointer tool, the 3D position of every landmark L_i in the printed model is located in the coordinate system of the printed model tracking target P_{Target}. The same landmark positions are also extracted manually from the CT of the printed model CT_{3DP}. Using the corresponding 3D points sequence, the transformations $T_{P_{Target} \rightarrow CT_{3DP}}$ and $T_{P_{Target} \rightarrow CT_{PAT}}$ are estimated using the least mean square minimization on the distances between corresponding points [12].



(a) (b) (c) (d)

Fig. 10.3. Artificial landmarks (blue) and CT markers (yellow) in (a) patient CT and segmentation, (b) 3D print, (c) 3D print CT. (d) Synthetic patient print filled with red-colored wax used during the user-study.

10.3.4 Full-chain Transformation

Knowing $T_{C_{Target} \to S}$ and $T_{P_{Target} \to CT_{PAT}}$, we compute the transformation from the patient CT to the simulated X-ray source $T_{CT_{PAT} \to S}$ for any C-arm and printed model pose with Equation 10.1.

$$T_{CT_{PAT} \to S} = T^{-1}_{P_{Target} \to CT_{PAT}} T^{-1}_{W \to P_{Target}} T_{W \to C_{Target}} T_{C_{Target} \to S} \tag{10.1}$$

The pose of the simulated X-ray source in the patient CT coordinate system $T^{-1}_{CT_{PAT} \to S}$ is used to position a virtual camera to compute the DRR image. The intrinsics of the X-ray imaging are derived by a standard camera calibration method, already explained in Chapter 5.4.1. The generated DRR image can be used straight away as replacement of X-ray image in the training scenario but it can also be integrated to more complex systems for the case of new C-arm based technologies. We demonstrate this by integrating the simulated X-ray image into the RGBD augmented C-arm system described in Chapter 4.

10.4 System Evaluation

10.4.1 3D Print to Patient CT and Printed Patient CT Calibration Quality

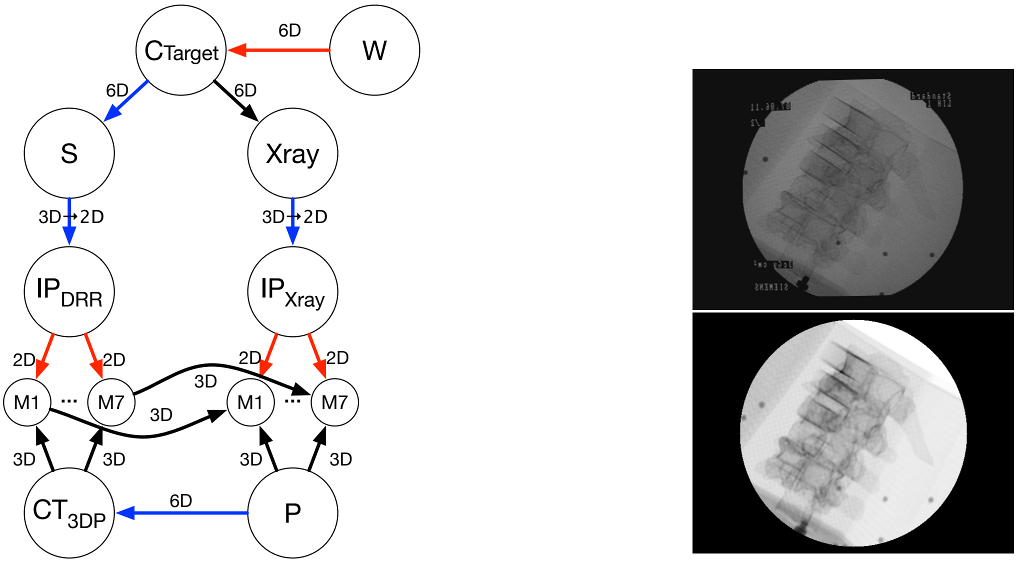
First, we evaluated the errors: a) of the printing process, i.e. the registration of CT_{PAT} to CT_{3DP}, b) the registration of P to CT_{PAT} used to visualize the DRR image spatially aligned with patient model in the user study and c) the registration P to CT_{3DP} used in the evaluation of the error between the DRR and real X-ray images. The respective rigid transformations describing those spatial relationships $T_{CT_{PAT} \to CT_{3DP}}$, $T_{P \to CT_{PAT}}$ and $T_{P \to CT_{3DP}}$ are calculated based on a least mean square error minimization of the distances between corresponding artificial landmarks. The root-mean-square error (RSME) on the distance residuals is given in Table 10.1.

	3D print to patient CT	3D print to CT of 3D print	patient CT to CT of 3D print
RMSE (in mm)	0.752	0.844	0.575

Tab. 10.1. RSME on the distance residuals for the different transformations

10.4.2 Evaluation of Tracking Full-chain

Then, we evaluated the full-chain accuracy of tracking, answering the question to what extent the simulated X-ray image matches the real X-ray image. We compare the 2D positions of CT markers placed on the 3D print (see Figure 10.3, yellow circles) in DRR images generated from the CT of the printed model (CT_{3DP}) and in real X-ray images of the printed model (P). This evaluation step is represented as an SRG graph along with an exemplary X-ray image and DRR image pair used in the evaluation in Figure 10.4. The RMSE error over 7 C-arm poses, combining C-arm angulations between $\approx \pm$ 20° Wigwag and $\approx \pm$ 20° Orbital, including AP, is 4.85±2.37 pixels (1.85±0.90 mm).



(a) SRG for evaluation **(b)** (Top) X-ray image and (Bottom) DRR image

Fig. 10.4. Full-chain tracking evaluation: (a) Spatial Relationship Graph, (b) example pair of real X-ray image and DRR image acquired during evaluation

10.4.3 Qualitative Results with Patient Data

In Figure 10.5, DRR images generated from the 3D print CT (first row) and from the patient CT (second row) for three poses of the C-arm (AP, $\approx$ -20° Wigwag, $\approx$ -20° Orbital) are shown. In Figure 10.6, we show the result of the integration of the simulated C-arm into the CamC framework. Instead of showing a real X-ray image, the system overlays the DRR image over the video.

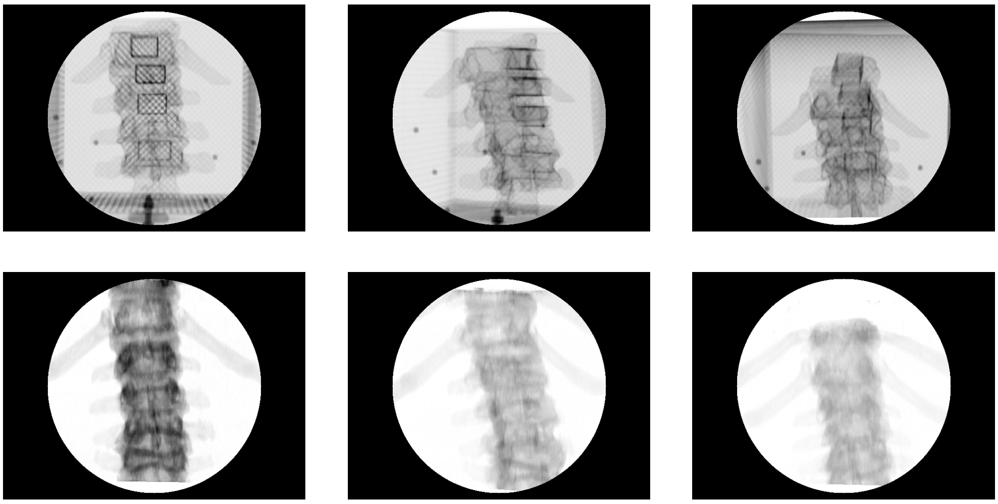


Fig. 10.5. 1st row: DRR images from 3D print CT, 2nd row: DRR images from patient CT, left to right: AP, ≈ -20° Wigwag, ≈ -20° Orbital

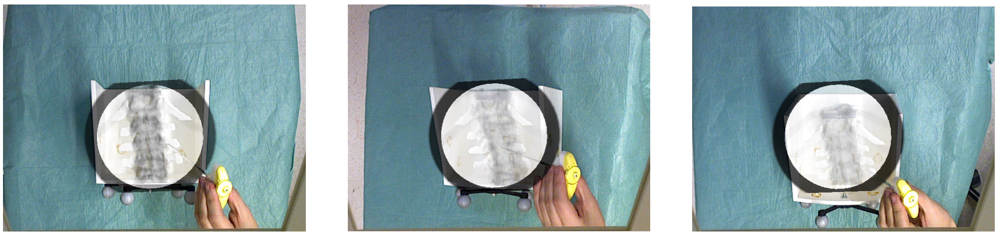


Fig. 10.6. Overlay of DRR images from patient CT with video, left to right: AP, ≈ -20° Wigwag, ≈ -20° Orbital

10.4.4 User Study

We perform a user study to test our simulation environment in the training scenario. For the user study, a synthetic patient print is filled with red-colored gel candle wax, using a print of the segmented skin as a mold to exactly recreate the patient's body shape, then covered with a skin-colored foam rubber sheet to imitate skin. This model, shown in Figure 10.3d, is placed in between a mannequin phantom, to indicate where head and feet of the patient are located, then positioned on an operating table and finally draped as visible in Figure 10.7. The surgeons participating in the study are presented with a patient case suggesting a FJI and asked to perform four injections into L1/L2 and L2/L3 on both sides using the simulated C-arm operated by a standardized nurse following the surgeons' instructions. After the performance, the participants are asked to answer a questionnaire. A total of $N = 6$ surgeons (5 trauma and 1 orthopedic surgeons), mean age 40 (SD 10.7, range 32-61), with prior experience in general spine surgery of mean 6.8 years (SD 6.6, range 2-20) and experience in FJI of mean 4.2 years (SD 4.4, range 0-10), 3 participants with teaching experience in both image guided surgery and FJI, 2 participants with ≥ 1000 procedures performed, the rest with ≤ 60, participate in the study. All, except one, participants had prior experience with surgical simulators, 2 participants had used this simulator before. We show in Figure 10.7 a participant using our simulation environment in action, the picture is taken at the beginning of the procedure where the participant palpates the simulated patient in order to find the right vertebral level where to place the needle.

Fig. 10.7. Participant palpating the phantom to find the vertebral level to target

The result of the questionnaire is summarized in Figure 10.8. Participants express agreement with the overall realism of the simulation (Q1) and strong agreement with the usefulness of the system for training of novices (Q12) and experts (Q13). The participants strongly agree that an integration into medical education would be useful (Q15). Free-text areas for improvements in the questionnaire reflect the positive reception of the participants: "[Replicate] facet joint capsule (haptic sensation when feeling around)", "Improve haptics of the soft-tissue and ligaments surrounding the vertebrae", "Current state very good, possibly further develop for more spine procedures".

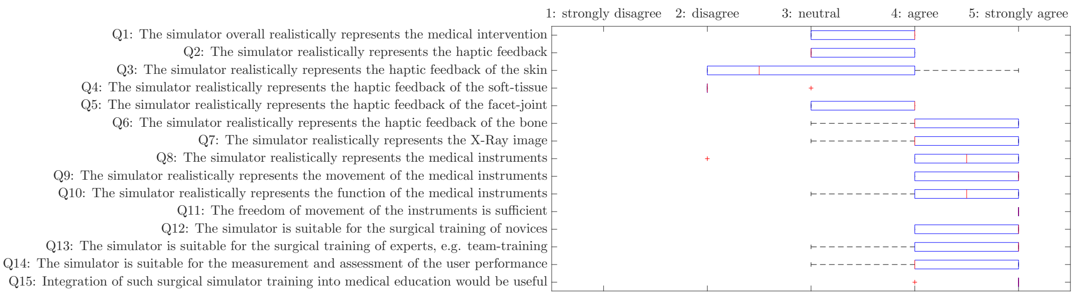


Fig. 10.8. Box plot of the 5-point Likert scale questionnaire results from the user study.

The user study has validated the use of our system for training. In the future, a similar user study should be conducted with the setup integrated inside the RGBD augmented C-arm system in order to validate the use of our simulation environment for technology assessment.

10.5 Discussion

The current training of teams of surgeons and operators in C-arm based procedures in general and the assessment of new C-arm based technology involve X-ray radiation for the full length of cadaver/phantoms procedure. The proposed mixed-reality system has the potential to complement or even replace large parts of cadaver use. 3D printing enables the accurate replication of patient anatomy. With the presented methodology, these can be correctly spatially aligned to the C-arm and surgical instruments, allowing training institutions to include any available patient case with its specific pathologies in a training scenario and research centers to reproduce at the infinity the same patient-based case, decisive to statistical significance, in new C-arm based technology assessment. The SRG methodology used throughout this chapter proved to be a versatile tool in providing an intuitive description of the spatial relations involved in the simulation system, in identifying the required transformations and in modeling appropriate calibrations. Therefore, we suggest the general usage of SRGs for high-level descriptions of complex, dynamic real-world spatial relations in CAI applications. The simulated X-ray images generated by our system result in an accuracy within the tolerable range of $\leq 2\,\mathrm{mm}$ for image-guided spine surgery [158]. The system is thus well suited for training of technical skills, e.g. the hand-eye coordination in surgical tool usage or the mental mapping from 2D projective images to the 3D surgical scene as well as for new C-arm based technology assessment. Additionally, it can potentially be used for the training of non-technical skills such as communication between surgeon and C-arm operator.

Future Work

To improve the model fidelity, e.g. replication of ligaments and capsule tissue, the latest generation 3D printers could be used, which supports materials with varying consistency and density [174], could be used. We have presented two main directions for our mixed-reality C-arm based surgery simulation system. The training scenario has been investigated deeper with a qualitative user-study. It remains however to prove quantitatively the benefit of our setup for training in a randomized controlled study where our system is compared to the cadaver training. During this study, metrics such as surgical performance (accuracy, time) on different cases (some pathological), level of radiation exposure could be measured. The direction regarding new C-arm based technology assessment has been less investigated in this thesis, a qualitative user study (for the specific case of RGBD augmented C-arm, but also for any C-arm based surgery) remains to be performed as well as a randomized controlled study later on where the level of radiation is the main metric to be considered in this case. The potential of SRG must also be exploited outside of the work presented in this chapter. It can be used for any CAI setup requiring multiple sensors and tracking devices. The SRG methodology permits to realize that an mixed-reality ultrasound training system is easy to realize based on our current setup, the topology of the graph presented in Figure 10.1c remains identical. The X-ray source is replaced by the ultrasound source and the image plane by the ultrasound slice. The edge between the ultrasound slice to CT remains as the simulation of the ultrasound slice is created from CT based on the work of Salehi et al. [141]. The rest of the graph remains identical, showing the powerfulness of the SRG to prototype new CAI setup.

10.6 Conclusion

In this chapter, we have proposed a C-arm based surgery simulation system that accurately simulates patient anatomy and X-ray imaging. We have shown the feasibility of using the system to simulate a surgical procedure with a fidelity sufficient for the training of novices and experts and integration in medical education, according to surgical experts that evaluated the system in a user study. We have also shown that this system can be used for the assessment of novel technology before their introduction in the OR for the case of RGBD augmented C-arm.

Part IV

Conclusion

Discussion & Conclusion

11

Within this thesis, we have presented two setups supplementing C-arm devices by affixing RGBD devices. We have also explored what those setups can offer as novel medical applications in terms of visualization as well as a mean to validate them without radiation involved. This chapter will summarize and discuss these works and provide ideas and possible directions for future work and improvements.

11.1 Summary and Perspective

11.1.1 RGBD Augmented C-arm Systems

In this part, two different setups supplementing C-arm devices with RGBD camera have been introduced. First, the mirror-based RGBD augmented C-arm was presented. The feasibility of such setup was first studied through theoretical optics, followed by an empirical study to assess the validity of the RGBD data. We have demonstrated that RGBD cameras of type pattern-emission can work with a mirror, making the first RGBD-augmented C-arm setup feasible and valid. The construction process of the setup from the design to the calibration is also described in this chapter. This was the pioneering work for investigating RGBD-augmented C-arm, however, the 3D capabilities of the RGBD camera and the cumbersomeness of the mirror construction have led very fast to focus on the setup presented in the second chapter of this part: the mirror-less RGBD augmented C-arm. This setup, using two RGBD cameras placed on the side of the C-arm source, can produce a similar overlay output as the previous setup but with only minimum disruption on the C-arm housing. The new system was fully described, validated and also used during a pre-clinical study. During this study, we could show that our system could decrease significantly the number of X-ray images acquired. However, in its current state, the setup has limited image quality (due to hardware limitations) which leads to a small discrepancy in needle navigation success. The next step for the mirror-less setup is to replace the two RGBD cameras with cameras of higher resolutions and/or smaller field of view and higher depth accuracy, limited factors in the synthesized video image quality. The design of a no-screw contraption to hold the cameras on the C-arm would also facilitate its rapid use and integration in the OR.

11.1.2 Medical Applications

In this thesis, we have also described in this thesis different medical applications made possible thanks to the development of RGBD augmented C-arms.

3D Visualization with 2D X-ray Image

A new visualization paradigm overlaying the X-ray image over the 3D reconstruction of the surgical scene is presented along a depth-based C-arm to patient registration. The main goal of such visualization is to provide the external context of the X-ray image by showing its location compared to the patient as well as the localization of the X-ray source. We have also developed two types of rendering: texture mapping and virtual image plane. The first one provides the internal context of the X-ray image, however, the ill-posed nature of X-ray image 3D localization due to its projective nature might lead to wrong depth perception in the internal context understanding. Nevertheless, the future of this visualization is to be applied on CBCT where this limitation is lifted [42, 85].

Radiation Exposure Estimation

In this chapter, an augmented reality visualization of the surgeons radiation exposure during a surgical procedure is introduced in order to sensibilize them to the risks inherent to the exposure. The radiation exposure is visualized via an heatmap overlay on the surgeon in an augmented reality fashion. The estimation of radiation exposure with our setup is equivalent to the state of the art. The setup for the project is the mirror-based RGBD augmented C-arm, on which an additional RGBD camera is attached to track the surgeon's skeleton. However, the future of this project is to go away from the restrictive framework of the mirror-based RGBD augmented C-arm which is very specific. The full radiation exposure estimation framework is not bound to this setup and could be generalized to any RGBD augmented C-arm. Indeed, as RGBD data allows to place the cameras anywhere on the C-arm to obtain 3D reconstruction, a more lightweight setup placing RGBD cameras at more convenient places on the C-arm (e.g. at the intensifier) would be more suitable for better integration into the OR. The current implementation is also too slow to be used intra-operatively. However, a GPU version of the library used for our work has recently been released [11], reducing the radiation estimation computational time by a factor of ninety. The integration of this library is the next step for a feasible future integration inside a clinical workflow.

Multi-layer Visualization for Medical Mixed Reality

In this chapter, a novel visualization paradigm combining Diminished and Augmented Reality in the medical domain is presented. Our visualization scheme proposes a user-adjustable multiple layer visualization where each layer of the surgical scene can be blended with others. The multiple layers comprise the anatomy with the X-ray image, the patient background recovered thanks to the mirror-less RGBD augmented C-arm setup, and the surgeons hand and surgical instruments. The result of our visualization scheme offers the clinician to choose which layer(s) can become transparent depending on the surgical scenario or workflow step. Beyond the medical domain, this work is the first use of a volumetric field for background recovery in Diminished Reality and Mixed Reality. Future works should split the scene into even more additional layers, by disassociating the surgeon hand layer from the surgical instruments layer, in order to further adjust the visualization to the user preferences. Also, our visualization paradigm could also be applied out of the medical domain for Industrial Mixed Reality where workers, wearing a Head Mounted Display with two cameras placed on their side (with a higher disparity than our setup), see their viewpoint synthesized with their hands in transparency.

Assistive System for Minimally Invasive Scoliosis Surgery

In this chapter, the first intra-operative assistive system for scoliosis VATS surgery is presented, this setup is composed of a single RGBD camera affixed on a C-arm and its feasibility and validity are demonstrated through multiple experiments from simulated data to a real-time acquisition on a person. The proposed system is able to measure the changes in the back surface deformities in the transverse plane measured by the Back Surface Rotation (BSR) index in the accuracy range defined by the state of the art of scoliotic patient reconstruction. The future of this work is to further investigate its total integration into the scoliosis surgery workflow such as correlation of intra-operative metric to pre-operative metric and combination of the BSR metric with the internal scoliosis information from the X-ray image.

C-arm based Surgery Simulation for Training and new Technology Assessment

In this chapter, a C-arm based surgery simulation system that accurately simulates patient anatomy and X-ray imaging is presented. We have shown the feasibility of using the system to simulate a surgical procedure with a fidelity sufficient for the training of novices and experts and integration in medical education, according to surgical experts that evaluated the system in a user study. We have also presented the possibility for this system to be used for the assessment of novel technology before their introduction in the OR such as RGBD augmented C-arm that has been described extensively in this thesis. The next step is to perform a user-study for the new technology assessment application, similar to the one performed for the training application. The potential of the Spatial Relationship Graph must also be applied to other CAI applications requiring multiples devices such as tracking, sensors.

11.2 General Conclusion

11.2.1 Works in Light with Thesis Objectives

In the State of the Art Chapter in Section 2.5.2, we have made the proposition that augmenting C-arm with RGBD cameras could fit the description of an economical and minimally workflow disruptive hybrid C-arm supplement. Among the directions that we have extracted from the literature, we have described that such setup(s) should use i) as an additional modality a by-design registered intra-operative modality, ii) should enable markerless registration, iii) be multi-purpose, and iv) should have potential for enhanced multi-modal visualization. The two setups presented in this work, both augmenting the C-arm with calibrated RGBD camera, both requires only a one-time calibration procedure prior to surgery. They both enter the category of by-design registration. We also use a setup where a RGBD camera is not calibrated to the C-arm for the assistive tool for scoliosis surgery application. As described in the future works regarding this work, the calibration of the RGBD camera to the C-arm would also be a by-design registration performed one time prior to surgery using Wang et al. works [173]. The setups presented in this thesis, therefore, all fit the first direction. We have also presented in this thesis the first work performing depth-based C-arm pose estimation to the surgical scene. The use of depth data allows alleviating the need for markers, using the 3D geometry as distinctive feature. This shows the potential of RGBD camera for markerless pose estimation, and could replace the use of markers in the Camera Augmented Mobile C-arm (CamC [111]) works performing C-arm pose estimation for X-ray images panorama stitching [168]. The recently published work by Lee et al. [86] has already shown the potential of RGBD data for markerless tool pose estimation for RGBD augmented C-arm. RGBD augmented C-arm have therefore proven their ability to enable markerless registration, which can be used for C-arm or tool pose estimation. Our setups build upon the work of CamC, which has already proven to be multi-purpose as CamC can be retrieved at numerous occurrences along the state of the art chapter. Our setups can at least perform as well, in addition to the applications enabled thanks to the RGBD data that we have presented in this thesis. Finally, we have presented several visualization paradigms, two for the multi-modal X-ray/video fusion. The first proposition that performs X-ray image overlay over 3D reconstruction does not solve the perception issue of the 2D uniform blending, but allows the external contextualization of the X-ray image in its 3D environment. The second proposition that decomposes the image into multiple layers thanks to the RGBD data is a step towards more perceptually correct visualizations. The multi-layer framework embraces the visualization proposed by Pauly et al. [122] and brings it further as the user can fully personalize the visualization. The main goal is that the surgeon uses the personalized visualization leading to minimal mental workload. Therefore, we conclude that the thesis proposes multi-modal enhanced visualizations that are a step forward perceptual multi-modal visualization. As a consequence, we can claim that the works presented in this thesis have fulfilled the directions extracted from the literature and that RGBD augmented C-arm works is a viable idea for economical and minimally workflow disruptive hybrid C-arm supplement. We will now discuss the potential clinical impact of the different works presented in this thesis and draw insights from this discussion. We will then discuss the future on the RGBD augmented C-arm technology and its potential to become at short term a hybrid C-arm supplement.

11.2.2 Insights on the Medical Applications

We have explored different directions for visualization paradigms based on the data provided by RGBD augmented C-arm as well as a mean to validate them without radiation. Insights can be drawn from those different works by looking at their direct clinical impact. Two works (the radiation exposure estimation and the simulation of C-arm based surgeries) propose a general framework that can apply to any C-arm based surgery. Beyond the technical details that must be improved for both works, their potential clinical impacts are promising and we could see their integration into clinical workflow, training environment and new technology assessment protocol realized even at a short term. The same fast integration could also happen to the assistive system for scoliosis MIS, even though in this case, the work was already conducted with a precise clinical workflow in mind. Two works conducted during the thesis span do not have a specific and clear surgical application: 3D reconstruction visualization with 2D X-ray image and the multi-layer visualization. For the first work, this is due to the wrong depth perception that the most appealing visualization (texture mapping) leads to. As we have seen in the corresponding chapter of this work, the proposed visualization when applied to CBCT has led to clinical applications such as K-wire placement [42]. However, if limited to a normal 2D C-arm, this work is not suitable for clinical purpose. This is the risk of research that sometimes leads to dead-ends due to the current research conditions. But it is actually worth exploring these directions so that other research can build on later if some research conditions have changed. Unfortunately, such change of research conditions has not happened yet to the work of multi-layer visualization. Although deeply convinced that such visualization can be used for down the beam procedures where the surgeon's hands occlude the target area, the work has not yet found its surgical application where it could be of benefit, even after multiple discussions with clinical experts. This difficulty in finding the surgical application questions both the suitability of the multi-layer visualization work for clinical purposes in its current state and the mental barrier from surgeons to use a open-choice visualization.

11.2.3 Trends for the Future

Although the presented setups are in an early technical stage, with RGBD cameras of limited capacities and accuracy, we have shown the feasibility of such setups as well as their potentials, while also exploring several medical applications. We believe that the different directions explored could lead to future clinical impact at a short and long term. During the course of the thesis work, Philips has released the first commercially-available video augmented CBCT C-arm showing the interest and the potential of such technology in the Operating Room. On the other side, RGBD cameras are starting to replace video cameras in common electronic devices (phone, laptop) while the computer vision community currently investigates the generation of depth image based on single video image using Deep Learning [39]. Therefore, it is only a question of time and technology advancement (for RGBD cameras and/or RGBD data generation from video data) before the junction happens and that the introduction on RGBD augmented C-arm in the OR occurs. As we have discussed in the Motivation Chapter, the innovation should happen in a short time in the OR for constantly improving the clinical impact. Taken chronologically, the works described in this thesis show a clear trend towards minimalistic out-of-C-arm-housing setup with minimal engineering required to affix the camera along the C-arm. Minimalistic setups can be integrated easily in the OR as they do not require the C-arm to pass recertification again. The radiation-free mixed-reality environment for C-arm based surgeries, with the assessment of new technologies being one of its main goals (such as the presented RGBD augmented C-arm), can demonstrate strong evidence of the innovation usefulness, leading to cross the bridge to OR integration faster than with a traditional technology assessment pipeline. Those two findings of our work are steps forward the aforementioned goal of short term integration of innovation (here via the RGBD augmented C-arm) in the OR. Only time may see this possibility happen, perhaps through the setups and/or medical applications explored in this thesis but also via other minimalist designs and medical applications not yet explored.

Part V

Appendix

A List of Authored and Co-authored Publications

The works with * indicate a first co-authorship.

2017

[154] Philipp Stefan* and **Séverine Habert*** and Alexander Winkler and Marc Lazarovici and Julian Fürtmetz and Ulrich Eck and Nassir Navab. "A Mixed-Reality Approach to Radiation-free Training of C-arm based Surgery". *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2017, Québec, CND*.

[52] **Séverine Habert** and Ulrich Eck and Stefan Parent and Pascal Fallavollita and Nassir Navab and Farida Cheriet. "A Novel Application of an RGBD augmented C-arm for Minimally Invasive Scoliosis Surgery Assistance". *Augmented Environments for Computer-Assisted Interventions (AECAI), 2017, Québec, CND*.

2016

[171] Xiang Wang and **Séverine Habert** and Christian Schulte zu Berge and Pascal Fallavollita and Nassir Navab. "Inverse visualization concept for RGB-D augmented C-arms". *Computers in Biology and Medicine*.

[42] Marius Fischer and Bernhard Fuerst and Sing Chun Lee and Javad Fotouhi and **Séverine Habert** and Simon Weidert and Ekkehard Euler and Greg Osgood and Nassir Navab. "Preclinical usability study of multiple augmented reality concepts for K-wire placement". *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*.

[97] Ma Meng and Pascal Fallavollita and **Séverine Habert** and Simon Weidert and Nassir Navab "Device-and system-independent personal touchless user interface for operating rooms". *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*.

[173] Xiang Wang and **Séverine Habert** and Ma Meng and Chun-Hao Huang and Pascal Fallavollita and Nassir Navab. "Precise 3D/2D calibration between a RGB-D sensor and a C-arm fluoroscope". *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*.

2015

[53] **Séverine Habert** and José Gardiazabal and Pascal Fallavollita and Nassir Navab. "RGBDX: first design and experimental validation of a mirror-based RGBD Xray imaging system". *IEEE Symposium on Mixed and Augmented Reality (ISMAR), 2015, Fukuoka, JPN*.

[51] **Séverine Habert*** and Ma Meng* and Wadim Kehl and Xiang Wang and Federico Tombari and Pascal Fallavollita and Nassir Navab. "Augmenting mobile C-arm fluoroscopes via Stereo-RGBD sensors for multimodal visualization". *IEEE Symposium on Mixed and Augmented Reality (ISMAR), 2015, Fukuoka, JPN*.

[172] Xiang Wang and **Séverine Habert** and Ma Meng and Xiang Wang and Chun-Hao Huang and Pascal Fallavollita and Nassir Navab. RGB-D/C-arm Calibration and Application in Medical Augmented Reality *IEEE Symposium on Mixed and Augmented Reality (ISMAR), 2015, Fukuoka, JPN*.

[88] Nicola Leucht and **Séverine Habert** and Patrick Wucherer and Simon Weidert and Pascal Fallavollita and Nassir Navab. "Augmented Reality for Radiation Awareness". *IEEE Symposium on Mixed and Augmented Reality (ISMAR), 2015, Fukuoka, JPN*.

2014

[41] Pascal Fallavollita and Alexander Winkler and **Séverine Habert** and Patrick Wucherer and Philipp Stefan and Riad Mansour and Reza Ghotbi and Nassir Navab. "Desired-view controlled positioning of angiographic C-arms". *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2014, Boston, USA*.

[122] Olivier Pauly and Benoit Diotte and **Séverine Habert** and Simon Weidert and Ekkehard Euler and Pascal Fallavollita and Nassir Navab. "Relevance-based visualization to improve surgeon perception" *Information Processing in Computer-Assisted Interventions (IPCAI), 2014, Fukuoka, JPN*.

B Abstracts of Publications not Discussed in this Thesis

Inverse visualization concept for RGB-D augmented C-arms

X. Wang, S. Habert, C. Schulte zu Berge, P. Fallavollita, N. Navab

X-ray is still the essential imaging for many minimally-invasive interventions. Overlaying X-ray images with an optical view of the surgery scene has been demonstrated to be an efficient way to reduce radiation exposure and surgery time. However, clinicians are recommended to place the X-ray source under the patient table while the optical view of the real scene must be captured from the top in order to see the patient, surgical tools, and the surgical site. With the help of a RGB-D (red-green-blue-depth) camera, which can measure depth in addition to color, the 3D model of the real scene is registered to the X-ray image. However, fusing two opposing viewpoints and visualizing them in the context of medical applications has never been attempted. In this paper, we propose first experiences of a novel inverse visualization technique for RGB-D augmented C-arms. A user study consisting of 16 participants demonstrated that our method shows a meaningful visualization with potential in providing clinicians multi-modal fused data in real-time during surgery.

Computers in Biology and Medicine, Volume 77, 1 October 2016, Pages 135-147

Device and System Independent Personal Touchless User Interface for Operating Rooms

M. Meng, P. Fallavollita, S. Habert, S. Weidert, N. Navab

Introduction In the modern day operating room, the surgeon performs surgeries with the support of different medical systems that showcase patient information, physiological data, and medical images. It is generally accepted that numerous interactions must be performed by the surgical team to control the corresponding medical system to retrieve the desired information. Joysticks and physical keys are still present in the operating room due to the disadvantages of mouses, and surgeons often communicate instructions to the surgical team when requiring information from a specific medical system. In this paper, a novel user interface is developed that allows the surgeon to personally perform touchless interaction with the various medical systems, switch effortlessly among them, all of this without modifying the systems' software and hardware.

Methods To achieve this, a wearable RGB-D sensor is mounted on the surgeon's head for inside-out tracking of his/her finger with any of the medical systems' displays. Android devices with a special application are connected to the computers on which the medical systems are running, simulating a normal USB mouse and keyboard. When the surgeon performs interaction using pointing gestures, the desired cursor position in the targeted medical system display, and gestures, are transformed into general events and then sent to the corresponding Android device. Finally, the application running on the Android devices generates the corresponding mouse or keyboard events according to the targeted medical system.

Results and conclusion To simulate an operating room setting, our unique user interface was tested by seven medical participants who performed several interactions with the visualization of CT, MRI, and fluoroscopy images at varying distances from them. Results from the system usability scale and NASA-TLX workload index indicated a strong acceptance of our proposed user interface.

International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), 2016

Preclinical usability study of multiple augmented reality concepts for K-wire placement

M. Fischer, B. Fuerst, S.C. Lee, J. Fotouhi, S. Habert, S. Weidert, E. Euler, G. Osgood, N. Navab

Introduction In many orthopedic surgeries, there is a demand for correctly placing medical instruments (e.g., K-wire or drill) to perform bone fracture repairs. The main challenge is the mental alignment of X-ray images acquired using a C-arm, the medical instruments, and the patient, which dramatically increases in complexity during pelvic surgeries. Current solutions include the continuous acquisition of many intra-operative X-ray images from various views, which will result in high radiation exposure, long surgical durations, and significant effort and frustration for the surgical staff. This work conducts a preclinical usability study to test and evaluate mixed reality visualization techniques using intra-operative X-ray, optical, and RGBD imaging to augment the surgeon's view to assist accurate placement of tools.

Methods We design and perform a usability study to compare the performance of surgeons and their task load using three different mixed reality systems during K-wire placements. The three systems are interventional X-ray imaging, X-ray augmentation on 2D video, and 3D surface reconstruction augmented by digitally reconstructed radiographs and live tool visualization.

Results The evaluation criteria include duration, number of X-ray images acquired, placement accuracy, and the surgical task load, which are observed during 21 clinically relevant interventions performed by surgeons on phantoms. Finally, we test for statistically significant improvements and show that the mixed reality visualization leads to a significantly improved efficiency.

Conclusion The 3D visualization of patient, tool, and DRR shows clear advantages over the conventional X-ray imaging and provides intuitive feedback to place the medical tools correctly and efficiently.

International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), 2016

Desired-View Controlled Positioning of Angiographic C-arms

P. Fallavollita, A. Winkler, S. Habert, P. Wucherer, P. Stefan, R. Mansour, R. Ghotbi, N. Navab

We present the idea of a user interface concept, which resolves the challenges involved in the control of angiographic C-arms for their constant repositioning during interventions by either the surgeons or the surgical staff. Our aim is to shift the paradigm of interventional image acquisition workflow from the traditional control device interfaces to 'desired-view' control. This allows the physicians to only communicate the desired outcome of imaging, based on simulated X-rays from pre-operative CT or CTA data, while the system takes care of computing the positioning of the imaging device relative to the patient's anatomy through inverse kinematics and CT to patient registration. Together with our clinical partners, we evaluate the new technique using 5 patient CTA and their corresponding intraoperative X-ray angiography datasets.

Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2014

Relevance-Based Visualization to Improve Surgeon Perception

O. Pauly, B. Diotte, S. Habert, S. Weidert, E. Euler, P. Fallavollita, N. Navab

In computer-aided interventions, the visual feedback of the doctor is vital. Enhancing the relevant object will help for the perception of this feedback. In this paper, we present a learning-based labeling of the surgical scene using a depth camera (comprised of RGB and depth range sensors). The depth sensor is used for background extraction and Random Forests are used for segmenting color images. The end result is a labeled scene consisting of surgeon hands, surgical instruments and background labels. We evaluated the method by conducting 10 simulated surgeries with 5 clinicians and demonstrated that the approach provides surgeons a dissected surgical scene, enhanced visualization, and upgraded depth perception.

International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), 2014

Bibliography

[1] N. Abi-Jaoudeh, P. Mielekamp, N. Noordhoek, A. M. Venkatesan, C. Millo, A. Radaelli, B. Carelsen, and B. J. Wood. "Cone-beam CT Fusion/navigation for real-time PET guided biopsies and ablations: a feasibility study". In: *Journal of Vascular and Interventional Radiology* 23.6 (2012), p. 737 (cit. on p. 20).

[2] N. Ahmed, K. S. Devitt, I. Keshet, J. Spicer, K. Imrie, L. Feldman, J. Cools-Lartigue, A. Kayssi, N. Lipsman, and M. Elmi. "A Systematic Review of the Effects of Resident Duty Hour Restrictions in Surgery: Impact on Resident Wellness, Training, and Patient Outcomes". en. In: *Annals of Surgery* 259.6 (2014), pp. 1041–1053 (cit. on p. 131).

[3] A. Aichert, M. Wieczorek, J. Wang, M. Kreiser, L. Wang, P. Fallavollita, and N. Navab. "The colored x-rays". In: *Workshop on Augmented Environments for Computer-Assisted Interventions*. Springer. 2012, pp. 45–54 (cit. on p. 27).

[4] A. Akay and Y. S. Akgul. "3D reconstruction with mirrors and RGB-D cameras". In: *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*. Vol. 3. IEEE. 2014, pp. 325–334 (cit. on p. 57).

[5] S. Albarqouni, U. Konrad, L. Wang, N. Navab, and S. Demirci. "Single-view X-ray depth recovery: toward a novel concept for image-guided interventions". In: *International journal of computer assisted radiology and surgery* 11.6 (2016), pp. 873–880 (cit. on p. 27).

[6] M. Amini, T. Ngo, R. McCormack, and S. Amiri. "An Imaged-Based Technique for Tracking C-Arm Fluoroscopes with an Example Use in High Tibial Osteotomy". In: *Bone Joint J* 98.SUPP 5 (2016), pp. 27–27 (cit. on p. 16).

[7] S. Amiri, D. R. Wilson, B. A. Masri, and C. Anglin. "A low-cost tracked C-arm (TC-arm) upgrade system for versatile quantitative intraoperative imaging". In: *International journal of computer assisted radiology and surgery* 9.4 (2014), pp. 695–711 (cit. on p. 17).

[8] C. S. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, et al. "A 0.13 μm CMOS system-on-chip for a 512$\times$ 424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 MHz and 2 GS/s ADC". In: *IEEE Journal of Solid-State Circuits* 50.1 (2015), pp. 303–319 (cit. on p. 48).

[9] J. M. Behar, P. Mountney, D. Toth, S. Reiml, M. Panayiotou, A. Brost, B. Fahn, R. Karim, S. Claridge, T. Jackson, et al. "Real-Time X-MRI-Guided Left Ventricular Lead Implantation for Targeted Delivery of Cardiac Resynchronization Therapy". In: *JACC: Clinical Electrophysiology* (2017) (cit. on pp. 20, 21).

[10] C. Beijst, M. Elschot, S. van der Velden, and H. W. de Jong. "Multimodality calibration for simultaneous fluoroscopic and nuclear imaging". In: *EJNMMI physics* 3.1 (2016), p. 20 (cit. on pp. 21–23).

[11] J. Bert, H. Perez-Ponce, Z. El Bitar, S. Jan, Y. Boursier, D. Vintache, A. Bonissent, C. Morel, D. Brasse, and D. Visvikis. "Geant4-based Monte Carlo simulations on GPU for medical applications". In: *Physics in medicine and biology* 58.16 (2013), p. 5593 (cit. on pp. 105, 144).

[12] P. Besl and N. McKay. “A Method for Registration of 3-D Shapes”. In: *Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 239–256 (cit. on p. 134).

[13] J Bijhold. “Three-dimensional verification of patient placement during radiotherapy using portal images”. In: *Medical physics* 20.2 (1993), pp. 347–356 (cit. on p. 19).

[14] T. Birdal, I. Dobryden, and S. Ilic. “X-Tag: A Fiducial Tag for Flexible and Accurate Bundle Adjustment”. In: *3D Vision (3DV)*. IEEE. 2016, pp. 556–564 (cit. on p. 129).

[15] N. Bizzotto, A. Sandri, F. Lavini, C. Dalloca, and D. Regis. “Video in Operating Room GoPro HERO3 Camera on Surgeon’s Head to Film Operations—A Test”. In: *Surgical innovation* (2013), p. 1553350613513514 (cit. on p. 107).

[16] J. Boisvert, F. Cheriet, X. Pennec, and N. Ayache. “3D reconstruction of the human spine from radiograph (s) using a multi-body statistical model”. In: *SPIE Medical Imaging*. 2009, pp. 72612D–72612D (cit. on p. 130).

[17] R. C. Bolles and M. A. Fischler. “A RANSAC-Based Approach to Model Fitting and Its Application to Finding Cylinders in Range Data.” In: *IJCAI*. Vol. 1981. 1981, pp. 637–643 (cit. on pp. 74, 134).

[18] I. M. van der Bom, R. P. Moser, G. Gao, M. Sena-Esteves, N. Aronin, and M. J. Gounis. “Frameless multimodal image guidance of localized convection-enhanced delivery of therapeutics in the brain”. In: *Journal of neurointerventional surgery* () (cit. on pp. 20, 21, 23).

[19] M. Van der Bom, J. Pluim, M. Gounis, E. van de Kraats, S. Sprinkhuizen, J Timmer, R Homan, and L. Bartels. “Registration of 2D x-ray images to 3D MRI by generating pseudo-CT data”. In: *Physics in medicine and biology* 56.4 (2011), p. 1031 (cit. on p. 19).

[20] V. Bonnet, T. Yamaguchi, A. Dupeyron, S. Andary, A. Seilles, P. Fraisse, and G. Venture. “Automatic estimate of back anatomical landmarks and 3D spine curve from a Kinect sensor”. In: *Biomedical Robotics and Biomechatronics (BioRob)*. 2016, pp. 924–929 (cit. on p. 120).

[21] O. J. Bott, M. Teistler, C. Duwenkamp, M. Wagner, M. Marschollek, M. Plischke, B. W. Raab, K. M. Stürmer, D. P. Pretschner, and K. Dresing. “virtX - Evaluation of a Computer-Based Training System for Mobile C-arm Systems in Trauma and Orthopedic Surgery”. ENG. In: *Methods of Information in Medicine* 47.3 (2008), pp. 270–278 (cit. on pp. 131, 132).

[22] O. J. Bott, M. Wagner, C. Duwenkamp, N. Hellrung, and K. Dresing. “Improving education on C-arm operation and radiation protection with a computer-based training and simulation system”. In: *International journal of computer assisted radiology and surgery* 4.4 (2009), pp. 399–407 (cit. on p. 30).

[23] O. J. Bott, K. Dresing, M. Wagner, B.-W. Raab, and M. Teistler. “Informatics in radiology: use of a C-arm fluoroscopy simulator to support training in intraoperative radiography”. In: *Radiographics* 31.3 (2011), E65–E75 (cit. on pp. 30, 132).

[24] R. Buchanan, J. G. Birch, A. A. Morton, and R. H. Browne. “Do you see what I see? Looking at scoliosis surgical outcomes through orthopedists’ eyes”. In: *Spine* 28.24 (2003), pp. 2700–2704 (cit. on p. 119).

[25] V. Buchmann, T. Nilsen, and M. Billinghurst. “Interaction with partially transparent hands and objects”. In: *Proceedings of the Sixth Australasian conference on User interface-Volume 40*. Australian Computer Society, Inc. 2005, pp. 17–20 (cit. on pp. 107, 108).

[26] F. Campeau and J. Fleitz. *Limited radiography*. Cengage Learning, 2016 (cit. on p. 99).

[27] A. Castro, J. Pacheco, C Lourenço, S Queirós, A. Moreira, N. Rodrigues, and J. Vilaça. “Evaluation of spinal posture using Microsoft Kinect™: A preliminary case-study with 98 volunteers”. In: *Porto Biomedical Journal* (2016) (cit. on p. 120).

[28] X. Chen, L. Wang, P. Fallavollita, and N. Navab. “Precise X-ray and video overlay for augmented reality fluoroscopy”. In: *International journal of computer assisted radiology and surgery* 8.1 (2013), pp. 29–38 (cit. on p. 55).

[29] X. Chen, H. Naik, L. Wang, N. Navab, and P. Fallavollita. “Video-guided calibration of an augmented reality mobile C-arm”. In: *International journal of computer assisted radiology and surgery* 9.6 (2014), pp. 987–996 (cit. on p. 64).

[30] S. P. Cohen and S. N. Raja. “Pathogenesis, diagnosis, and treatment of lumbar zygapophysial (facet) joint pain”. In: *The Journal of the American Society of Anesthesiologists* 106.3 (2007), pp. 591–614 (cit. on p. 82).

[31] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan. “High-quality Streamable Free-viewpoint Video”. In: *ACM Trans. Graph.* 34.4 (July 2015), 69:1–69:13 (cit. on p. 68).

[32] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy. “Augmented touch without visual obtrusion”. In: *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE. 2009, pp. 99–102 (cit. on p. 107).

[33] “COUNCIL DIRECTIVE 2013/59/EURATOM of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122”. In: *Official Journal of the European Union* L 13 (2014-01-17), pp. 1–73 (cit. on p. 99).

[34] L. Delrue, R. Gosselin, B. Ilsen, A. Van Landeghem, J. de Mey, and P. Duyck. “Difficulties in the interpretation of chest radiography”. In: *Comparative Interpretation of CT and Standard Radiography of the Chest*. Springer, 2011, pp. 27–49 (cit. on p. 10).

[35] B. Diotte, P. Fallavollita, L. Wang, S. Weidert, E. Euler, P. Thaller, and N. Navab. “Multi-modal intra-operative navigation during distal locking of intramedullary nails”. In: *Medical Imaging, IEEE Transactions on* 34.2 (2015), pp. 487–495 (cit. on pp. 26, 27, 33, 55, 82, 85).

[36] L. Dubinsky. “Mobile C-arms: Is the era of flat panel detectors and 3-D technology upon us?” In: *March 2016 issue of DOTmed HealthCare Business News magazine* (2016) (cit. on p. 8).

[37] B. Duthey. *World Health Organization - Background Paper (6.24) - Low back pain*. 2013 (cit. on p. 82).

[38] F. Echtler and al. *libfreenect2: Release 0.1.1*. Jan. 2016 (cit. on p. 69).

[39] D. Eigen, C. Puhrsch, and R. Fergus. “Depth map prediction from a single image using a multi-scale deep network”. In: *Advances in neural information processing systems*. 2014, pp. 2366–2374 (cit. on p. 147).

[40] P. Fallavollita, C. Burdette, D. Y. Song, P. Abolmaesumi, and G. Fichtinger. “C-arm tracking by intensity-based registration of a fiducial in prostate brachytherapy”. In: *International Conference on Information Processing in Computer-Assisted Interventions*. Springer. 2010, pp. 45–55 (cit. on p. 16).

[41] P. Fallavollita, A. Winkler, S. Habert, P. Wucherer, P. Stefan, R. Mansour, R. Ghotbi, and N. Navab. “Desired-view controlled positioning of angiographic c-arms”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, pp. 659–666 (cit. on pp. 15, 152).

[42] M. Fischer, B. Fuerst, S. C. Lee, J. Fotouhi, S. Habert, S. Weidert, E. Euler, G. Osgood, and N. Navab. “Preclinical usability study of multiple augmented reality concepts for K-wire placement”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.6 (2016), pp. 1007–1014 (cit. on pp. 98, 144, 146, 151).

[43] D. Fu and G. Kuduvalli. “A fast, accurate, and automatic 2D–3D image registration for image-guided cranial radiosurgery”. In: *Medical physics* 35.5 (2008), pp. 2180–2194 (cit. on p. 19).

[44] B. Fuerst, J. Fotouhi, and N. Navab. “Vision-based intraoperative cone-beam ct stitching for non-overlapping volumes”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 387–395 (cit. on p. 16).

[45] K. P. Gall, A. F. Thornton, J. Munzenrider, and S. Rosenthal. “Experience using radiopaque fiducial points for patient alignment during radiotherapy”. In: *International Journal of Radiation Oncology Biology Physics* 27 (1993), pp. 161–162 (cit. on p. 18).

[46] G. Gao, G. Penney, Y. Ma, N. Gogin, P. Cathier, A. Arujuna, G. Morton, D. Caulfield, J. Gill, C. A. Rinaldi, et al. “Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking”. In: *Medical image analysis* 16.1 (2012), pp. 38–49 (cit. on pp. 22, 23).

[47] A. Gawande. “Two hundred years of surgery”. In: *New England Journal of Medicine* 366.18 (2012), pp. 1716–1723 (cit. on pp. 4, 5).

[48] J. Geng. “Structured-light 3D surface imaging: a tutorial”. In: *Advances in Optics and Photonics* 3.2 (2011), pp. 128–160 (cit. on p. 46).

[49] R. H. Gong, B. Jenkins, R. W. Sze, and Z. Yaniv. “A Cost Effective and High Fidelity Fluoroscopy Simulator Using the Image-Guided Surgery Toolkit (IGSTK)”. In: *SPIE Medical Imaging*. 2014, pp. 903618–903618 (cit. on p. 132).

[50] V. Grzeda and G. Fichtinger. “C-arm rotation encoding with accelerometers”. In: *International journal of computer assisted radiology and surgery* 5.4 (2010), pp. 385–391 (cit. on p. 17).

[51] S. Habert, M. Meng, W. Kehl, X. Wang, F. Tombari, P. Fallavollita, and N. Navab. “Augmenting Mobile C-arm Fluoroscopes via Stereo-RGBD Sensors for Multimodal Visualization”. In: *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. 2015, pp. 72–75 (cit. on p. 152).

[52] S. Habert, U. Eck, P. Fallavollita, S. Parent, N. Navab, and F. Cheriet. “A Novel Application of an RGBD augmented C-arm for Minimally Invasive Scoliosis Surgery Assistance”. In: *Healthcare Technology Letters* (2017) (cit. on pp. 117, 151).

[53] S. Habert, J. Gardiazabal, P. Fallavollita, and N. Navab. “RGBDX: First Design and Experimental Validation of a Mirror-Based RGBD X-ray Imaging System”. In: *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 13–18 (cit. on pp. 55, 89, 152).

[54] G. Haller, T. Laroche, and F. Clergue. “Morbidity in anaesthesia: today and tomorrow”. In: *Best Practice & Research Clinical Anaesthesiology* 25.2 (2011), pp. 123–132 (cit. on p. 6).

[55] J. Harrop, A. R. Rezai, D. J. Hoh, G. M. Ghobrial, and A. Sharan. “Neurosurgical Training With a Novel Cervical Spine Simulator: Posterior Foraminotomy and Laminectomy”. en. In: *Neurosurgery* 73 (2013), S94–S99 (cit. on p. 132).

[56] R. Härtl, K. S. Lam, J. Wang, A. Korge, F. Kandziora, and L. Audige. “Worldwide Survey on the Use of Navigation in Spine Surgery”. In: *World Neurosurgery* 79.1 (2013), pp. 162–172 (cit. on p. 131).

[57] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, ISBN: 0521540518, 2004 (cit. on p. 38).

[58] J. Herling and W. Broll. “Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments”. In: *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE. 2010, pp. 207–212 (cit. on p. 107).

[59] J. Herling and W. Broll. “Pixmix: A real-time approach to high-quality diminished reality”. In: *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE. 2012, pp. 141–150 (cit. on p. 107).

[60] A. Herr. “Cost and technical efficiency of German hospitals: does ownership matter?” In: *Health Economics* 17.9 (2008), pp. 1057–1071 (cit. on p. 6).

[61] D. Herrera, J. Kannala, and J. Heikkilä. "Joint depth and color camera calibration with distortion correction". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.10 (2012), pp. 2058–2064 (cit. on p. 62).

[62] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera". In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM. 2011, pp. 559–568 (cit. on pp. 55, 68, 90).

[63] A. Jain, L. Gutierrez, and D. Stanton. "3D TEE registration with X-ray fluoroscopy for interventional cardiac applications". In: *Functional Imaging and Modeling of the Heart* (2009), pp. 321–329 (cit. on p. 22).

[64] A. K. Jain, T. Mustufa, Y. Zhou, E. Burdette, G. S. Chirikjian, and G. Fichtinger. "A robust fluoroscope tracking (FTRAC) fiducial". In: *Medical Imaging*. International Society for Optics and Photonics. 2005, pp. 798–809 (cit. on p. 16).

[65] P. Jannin and W. Korb. "Assessment of image-guided interventions". In: *Image Guided Interventions Technology and Applications. Springer* (2008), pp. 531–549 (cit. on p. 131).

[66] S. Jarusirisawad and H. Saito. "Diminished Reality Via Multiple Hand-Held CAMERAS". In: *Distributed Smart Cameras, 2007. ICDSC '07. First ACM/IEEE International Conference on*. 2007, pp. 251–258 (cit. on p. 107).

[67] L. Jeya and J. Zelek. "KinectScenes: Robust Real-Time RGB-D Fusion for Animal Behavioural Monitoring". In: *Computer and Robot Vision (CRV), 2016 13th Conference on*. IEEE. 2016, pp. 383–390 (cit. on p. 68).

[68] J.-Y. Jin, S. Ryu, K. Faber, T. Mikkelsen, Q. Chen, S. Li, and B. Movsas. "2D/3D image fusion for accurate target localization and evaluation of a mask based stereotactic system in fractionated stereotactic radiotherapy of cranial lesions". In: *Medical physics* 33.12 (2006), pp. 4557–4566 (cit. on p. 18).

[69] B. Kainz, M. Grabner, and M. Rüther. "Fast marker based C-arm pose estimation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008* (2008), pp. 652–659 (cit. on p. 16).

[70] W. J. Kane. "Scoliosis prevalence: a call for a statement of terms." In: *Clinical orthopaedics and related research* 126 (1977), pp. 43–46 (cit. on p. 117).

[71] N. Kawai, T. Sato, and N. Yokoya. "Diminished reality considering background structures". In: *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*. IEEE. 2013, pp. 259–260 (cit. on p. 107).

[72] E. Kerrien, M. Berger, E. Maurincomme, L. Launay, R. Vaillant, and L. Picard. "Fully automatic 3D/2D subtracted angiography registration". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99*. Springer. 1999, pp. 664–671 (cit. on p. 19).

[73] J. Kim, F.-F. Yin, Y. Zhao, and J. H. Kim. "Effects of x-ray and CT image enhancements on the robustness and accuracy of a rigid 3D/2D image registration". In: *Medical physics* 32.4 (2005), pp. 866–873 (cit. on p. 19).

[74] M.-H. Kim, J. Park, and I. S. Kweon. "RGBD sensor and mirrors: A practical setup for 3D reconstruction of dynamic objects". In: *Ubiquitous Robots and Ambient Intelligence (URAI), 2014 11th International Conference on*. IEEE. 2014, pp. 651–652 (cit. on p. 57).

[75] S. Kircher, S. Rolf, G. Hindricks, and P. Sommer. "Ablation of typical atrial flutter using a novel non-fluoroscopic electromagnetic catheter tracking system". In: *Interventional Cardiology* 6.2 (2014), p. 149 (cit. on p. 25).

[76] R. M. Klevens, J. R. Edwards, C. L. Richards Jr, T. C. Horan, R. P. Gaynes, D. A. Pollock, and D. M. Cardo. "Estimating health care-associated infections and deaths in US hospitals, 2002". In: *Public health reports* 122.2 (2007), pp. 160–166 (cit. on p. 6).

[77] M. R. Konieczny, H. Senyurt, and R. Krauspe. “Epidemiology of adolescent idiopathic scoliosis”. In: *Journal of children’s orthopaedics* 7.1 (2013), pp. 3–9 (cit. on p. 117).

[78] K. Konolige. “Projected texture stereo”. In: *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE. 2010, pp. 148–155 (cit. on p. 47).

[79] W Krattenthaler, K. Mayer, and H. Duwe. “3D-surface measurement with coded light approach”. In: *Proceedings of the 17th meeting of the Austrian Association for pattern recognition on Image analysis and synthesis*. R. Oldenbourg Verlag GmbH. 1994, pp. 103–114 (cit. on p. 46).

[80] E. Lachat, H. Macher, T. Landes, and P. Grussenmeyer. “Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling”. In: *Remote Sensing* 7.10 (2015), pp. 13070–13097 (cit. on p. 73).

[81] A. Ladikos, C. Cagniart, R. Ghotbi, M. Reiser, and N. Navab. “Estimating radiation exposure in interventional environments”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010* (2010), pp. 237–244 (cit. on pp. 30, 99).

[82] A. Ladikos, S. Benhimane, and N. Navab. “Real-time 3D reconstruction for collision avoidance in interventional environments”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2008, pp. 526–534 (cit. on p. 31).

[83] P. Lang, P. Seslija, M. W. Chu, D. Bainbridge, G. M. Guiraudon, D. L. Jones, and T. M. Peters. “US–fluoroscopy registration for transcatheter aortic valve implantation”. In: *IEEE Transactions on Biomedical Engineering* 59.5 (2012), pp. 1444–1453 (cit. on p. 22).

[84] S. Lavallée and R. Szeliski. “Recovering the position and orientation of free-form objects from image contours using 3D distance maps”. In: *IEEE Transactions on pattern analysis and machine intelligence* 17.4 (1995), pp. 378–390 (cit. on p. 19).

[85] S. C. Lee, B. Fuerst, J. Fotouhi, M. Fischer, G. Osgood, and N. Navab. “Calibration of RGBD camera and cone-beam CT for 3D intra-operative mixed reality visualization”. In: *International journal of computer assisted radiology and surgery* 11.6 (2016), pp. 967–975 (cit. on pp. 97, 144).

[86] S. C. Lee, B. Fuerst, K. Tateno, A. Johnson, J. Fotouhi, G. Osgood, F. Tombari, and N. Navab. “Multi-modal Imaging, Model-based Tracking and Mixed Reality Visualisation for Orthopaedic Surgery”. In: *Healthcare Technology Letters* (2017) (cit. on p. 146).

[87] S. C. Leschka, D. Babic, S. El Shikh, C. Wossmann, M. Schumacher, and C. A. Taschner. “C-arm cone beam computed tomography needle path overlay for image-guided procedures of the spine and pelvis”. In: *Neuroradiology* 54.3 (2012), pp. 215–223 (cit. on pp. 21, 22, 24, 25).

[88] N. Leucht, S. Habert, P. Wucherer, S. Weidert, N. Navab, and P. Fallavollita. “[POSTER] Augmented Reality for Radiation Awareness”. In: *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 60–63 (cit. on pp. 99, 152).

[89] M. R. Levitt, S. S. Vaidya, D. K. Su, K. S. Moe, L. J. Kim, L. N. Sekhar, D. K. Hallam, and B. V. Ghodke. “The “triple-overlay” technique for percutaneous diagnosis and treatment of lesions of the head and neck: combined three-dimensional guidance with magnetic resonance imaging, cone-beam computed tomography, and fluoroscopy”. In: *World neurosurgery* 79.3 (2013), pp. 509–514 (cit. on p. 20).

[90] J. Lister. “Antiseptic principle in the practice of surgery.” In: *British Medical Journal* 2.5543 (1967), pp. 9–12 (cit. on p. 4).

[91] J. S. Little and P. S. Khalsa. “Material properties of the human lumbar facet joint capsule”. In: *Journal of biomechanical engineering* 127.1 (2005), pp. 15–24 (cit. on p. 83).

[92] D. Litzenberg, L. A. Dawson, H. Sandler, M. G. Sanda, D. L. McShan, R. K. Ten Haken, K. L. Lam, K. K. Brock, and J. M. Balter. “Daily prostate targeting using implanted radiopaque markers”. In: *International Journal of Radiation Oncology Biology Physics* 52.3 (2002), pp. 699–703 (cit. on p. 19).

[93] X. Liu, J. Thometz, J. Tassone, L. Paulsen, and R. Lyon. “Historical review and experience with the use of surface topographic systems in children with idiopathic scoliosis”. In: *ISIS* 2 (2003), p. 8 (cit. on p. 119).

[94] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. *Pursuit of “X-ray vision” for augmented reality*. Springer, 2013 (cit. on pp. 33, 108).

[95] R. Londei, M. Esposito, B. Diotte, S. Weidert, E. Euler, P. Thaller, N. Navab, and P. Fallavollita. “The ‘augmented’circles: a video-guided solution for the down-the-beam positioning of IM nail holes”. In: *Information Processing in Computer-Assisted Interventions*. Springer, 2014, pp. 100–107 (cit. on p. 55).

[96] N. Lonjon, E. Chan-Seng, V. Costalat, B. Bonnafoux, M. Vassal, and J. Boetto. “Robot-assisted spine surgery: feasibility study through a prospective case-matched analysis”. In: *European Spine Journal* 25.3 (2016), pp. 947–955 (cit. on p. 26).

[97] M. Ma, P. Fallavollita, S. Habert, S. Weidert, and N. Navab. “Device-and system-independent personal touchless user interface for operating rooms: One personal UI to control all displays in an operating room.” In: *International journal of computer assisted radiology and surgery* 11.6 (2016), pp. 853–861 (cit. on p. 151).

[98] Y. Ma, G. P. Penney, D. Bos, P. Frissen, C. A. Rinaldi, R. Razavi, and K. S. Rhode. “Hybrid echo and x-ray image guidance for cardiac catheterization procedures by using a robotic arm: a feasibility study”. In: *Physics in medicine and biology* 55.13 (2010), N371 (cit. on p. 22).

[99] Y. Ma, S. Duckett, P. Chinchapatnam, G. Gao, A. Shetty, C. A. Rinaldi, T. Schaeffter, and K. S. Rhode. “MRI to X-ray fluoroscopy overlay for guidance of cardiac resynchronization therapy procedures”. In: *Computing in Cardiology, 2010*. IEEE. 2010, pp. 229–232 (cit. on pp. 19, 20).

[100] A. Maimone and H. Fuchs. “Real-time volumetric 3D capture of room-sized scenes for telepresence”. In: *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*. IEEE. 2012, pp. 1–4 (cit. on p. 68).

[101] S. Mann and S. M. Nnlf. “Mediated reality”. In: (1994) (cit. on p. 107).

[102] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš. “A review of 3D/2D registration methods for image-guided interventions”. In: *Medical image analysis* 16.3 (2012), pp. 642–661 (cit. on pp. 18, 19, 22).

[103] M. Martinez and R. Stiefelhagen. “Kinect Unleashed: Getting Control over High Resolution Depth Maps.” In: *MVA*. 2013, pp. 247–250 (cit. on p. 47).

[104] M. Maruyama and S. Abe. “Range sensing by projecting multiple slits with random cuts”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.6 (1993), pp. 647–651 (cit. on p. 47).

[105] *Medical X-ray Imaging*. https://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/MedicalX-Rays/default.htm. Accessed: 2017-05-02 (cit. on pp. 9, 83).

[106] S. Meerits and H. Saito. “Real-Time Diminished Reality for Dynamic Scenes”. In: *Mixed and Augmented Reality Workshops (ISMARW), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 53–59 (cit. on p. 107).

[107] A. Moataz, A. Soliman, A. M. Ghanem, M. al Shatouri, A. Atia, and E. A. Rashed. “Three-dimensional angiography using mobile C-arm with IMU sensor attached: Initial study”. In: *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015 IEEE*. IEEE. 2015, pp. 1–3 (cit. on p. 17).

[108] R. J. Mobbs, P. Sivabalan, and J. Li. “Technique, challenges and indications for percutaneous pedicle screw fixation”. In: *Journal of Clinical Neuroscience* 18.6 (2011), pp. 741–749 (cit. on p. 42).

[109] P. Mountney, R. Ionasec, M. Kaizer, S. Mamaghani, W. Wu, T. Chen, M. John, J. Boese, and D. Comaniciu. “Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012* (2012), pp. 544–551 (cit. on p. 22).

[110] N. Navab, M. Mitschke, and O. Schütz. “Camera-augmented mobile C-arm (CAMC) application: 3D reconstruction using a low-cost mobile C-arm”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI’99*. Springer. 1999, pp. 688–697 (cit. on p. 80).

[111] N. Navab, S.-M. Heining, and J. Traub. “Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications”. In: *Medical Imaging, IEEE Transactions on* 29.7 (2010), pp. 1412–1423 (cit. on pp. 16, 21, 22, 26, 28, 33, 34, 55, 56, 64, 65, 67, 79, 80, 82, 89, 96, 146).

[112] N. Navab, A. R. Bani-Hashemi, M. M. Mitschke, D. W. Holdsworth, R. Fahrig, A. J. Fox, and R Graumann. “Dynamic geometrical calibration for 3D cerebral angiography”. In: *Medical Imaging 1996*. International Society for Optics and Photonics. 1996, pp. 361–370 (cit. on p. 16).

[113] N. Navab, T. Blum, L. Wang, A. Okur, and T. Wendler. “First Deployments of Augmented Reality in Operating Rooms”. In: *IEEE Computer* 45.7 (2012), pp. 48–55 (cit. on p. 130).

[114] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. “KinectFusion: Real-time dense surface mapping and tracking”. In: *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*. IEEE. 2011, pp. 127–136 (cit. on p. 116).

[115] P. O. Newton and A. Perry. “Thoracoscopic Deformity Correction”. In: *Minimally Invasive Spine Surgery*. Springer, 2009, pp. 77–86 (cit. on pp. 119, 122).

[116] C. V. Nguyen, S. Izadi, and D. Lovell. “Modeling kinect sensor noise for improved 3d reconstruction and tracking”. In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*. IEEE. 2012, pp. 524–530 (cit. on p. 125).

[117] J. Pages, J. Salvi, R. Garcia, and C. Matabosch. “Overview of coded light projection techniques for automatic 3D profiling”. In: *Robotics and Automation, 2003. Proceedings. ICRA’03. IEEE International Conference on*. Vol. 1. IEEE. 2003, pp. 133–138 (cit. on p. 46).

[118] A. Park and J. Kim. “GPU accelerated view synthesis from multiple RGB-D images”. In: *2012 19th IEEE International Conference on Image Processing*. IEEE. 2012, pp. 573–576 (cit. on p. 68).

[119] S. Pati, O. Erat, L. Wang, S. Weidert, E. Euler, N. Navab, and P. Fallavollita. “Accurate pose estimation using single marker single camera calibration system”. In: *SPIE medical imaging*. International Society for Optics and Photonics. 2013, pp. 867126–867126 (cit. on p. 55).

[120] P. Patias, T. B. Grivas, A. Kaspiris, C. Aggouris, and E. Drakoutos. “A review of the trunk surface metrics used as Scoliosis and other deformities evaluation indices”. In: *Scoliosis* 5.1 (2010), p. 12 (cit. on pp. 121, 123).

[121] A. Patriciu, D. Stoianovici, L. L. Whitcomb, T. Jarrett, D. Mazilu, A. Stanimir, I. Iordachita, J. Anderson, R. Taylor, and L. R. Kavoussi. “Motion-based robotic instrument targeting under C-Arm fluoroscopy”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2000, pp. 988–998 (cit. on p. 26).

[122] O. Pauly, B. Diotte, S. Habert, S. Weidert, E. Euler, P. Fallavollita, and N. Navab. “Relevance-based visualization to improve surgeon perception”. In: *International Conference on Information Processing in Computer-Assisted Interventions*. Springer. 2014, pp. 178–185 (cit. on pp. 28, 33, 55, 65, 111, 146, 152).

[123] *Pediatric X-ray Imaging*. https://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/ucm298899.htm. Accessed: 2017-05-02 (cit. on p. 9).

[124] T. M. Peters and C. A. Linte. "Image-Guided Interventions and Computer-Integrated Therapy: Quo Vadis?" In: *Medical Image Analysis*. 20th anniversary of the Medical Image Analysis journal (MedIA) 33 (2016), pp. 56–63 (cit. on p. 131).

[125] C. Piorkowski and G. Hindricks. "Nonfluoroscopic sensor-guided navigation of intracardiac electrophysiology catheters within prerecorded cine loops". In: *Circulation: Arrhythmia and Electrophysiology* 4.4 (2011), e36–e38 (cit. on p. 26).

[126] J. Posdamer and M. Altschuler. "Surface measurement by space-encoded projected beam systems". In: *Computer graphics and image processing* 18.1 (1982), pp. 1–17 (cit. on p. 46).

[127] D. Proschek, K Kafchitsas, M. Rauschmann, A. Kurth, T. Vogl, and F. Geiger. "Reduction of radiation dose during facet joint injection using the new image guidance system SabreSource™: a prospective study in 60 patients". In: *European spine journal* 18.4 (2009), p. 546 (cit. on p. 83).

[128] D. Pustka, M. Huber, M. Bauer, and G. Klinker. "Spatial Relationship Patterns: Elements of Reusable Tracking and Calibration Systems". In: *International Symposium on Mixed and Augmented Reality*. 2006, pp. 88–97 (cit. on p. 133).

[129] J. M. Racadio, D. Babic, R. Homan, J. W. Rampton, M. N. Patel, J. M. Racadio, and N. D. Johnson. "Live 3D guidance in the interventional radiology suite". In: *American Journal of Roentgenology* 189.6 (2007), W357–W364 (cit. on pp. 21, 22).

[130] F. Remondino and D. Stoppa. *TOF range-imaging cameras*. Vol. 68121. Springer, 2013 (cit. on p. 48).

[131] M. Reyes, A. Clapés, J. Ramírez, J. R. Revilla, and S. Escalera. "Automatic digital biometry analysis based on depth maps". In: *Computers in Industry* 64.9 (2013), pp. 1316–1325 (cit. on pp. 120, 121).

[132] D. J. Riskin, M. T. Longaker, M. Gertner, and T. M. Krummel. "Innovation in surgery: a historical perspective". In: *Annals of surgery* 244.5 (2006), p. 686 (cit. on p. 5).

[133] N. L. Rodas and N. Padoy. "3D global estimation and augmented reality visualization of intra-operative X-ray dose". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, pp. 415–422 (cit. on pp. 30, 99, 104).

[134] N. L. Rodas, J. Bert, D. Visvikis, M. de Mathelin, and N. Padoy. "Pose optimization of a C-arm imaging device to reduce intraoperative radiation exposure of staff and patient during interventional procedures". In: *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE. 2017, pp. 4200–4207 (cit. on p. 15).

[135] N. L. Rodas, F. Barrera, and N. Padoy. "See It With Your Own Eyes: Markerless Mobile Augmented Reality for Radiation Awareness in the Hybrid Room". In: *IEEE Transactions on Biomedical Engineering* 64.2 (2017), pp. 429–440 (cit. on p. 30).

[136] N. Rudarakanchana, I. V. Herzeele, C. D. Bicknell, C. V. Riga, A. Rolls, N. J. W. Cheshire, and M. S. Hamady. "Endovascular Repair of Ruptured Abdominal Aortic Aneurysm: Technical and Team Training in an Immersive Virtual Reality Environment". en. In: *CardioVascular and Interventional Radiology* 37.4 (2013), pp. 920–927 (cit. on p. 132).

[137] T. Rudolph, L. Ebert, and J. Kowal. "Comparison of three optical tracking systems in a complex navigation scenario". In: *Computer aided surgery* 15.4-6 (2010), pp. 104–109 (cit. on p. 25).

[138] R. B. Rusu and S. Cousins. "3D is here: Point Cloud Library (PCL)". In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China, 2011 (cit. on p. 91).

[139] S. Ryan Fanello, C. Rhemann, V. Tankovich, A. Kowdle, S. Orts Escolano, D. Kim, and S. Izadi. "Hyperdepth: Learning depth from structured light without matching". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 5441–5450 (cit. on p. 47).

[140] H. Saito, T. Honda, Y. Nakayama, and F. De Sorbier. “Camera pose estimation for mixed and diminished reality in FTV”. In: *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2014*. IEEE. 2014, pp. 1–4 (cit. on p. 107).

[141] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, and W. Wein. “Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 510–518 (cit. on p. 139).

[142] C. Schaller, C. Rohkohl, J. Penne, M. Stürmer, and J. Hornegger. “Inverse C-arm positioning for interventional procedures using real-time body part detection”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009* (2009), pp. 549–556 (cit. on p. 17).

[143] L. Seoud, F. Cheriet, H. Labelle, and S. Parent. “Changes in Trunk Appearance After Scoliosis Spinal Surgery and Their Relation to Changes in Spinal Measurements”. In: *Spine Deformity* 3.6 (2015), pp. 595–603 (cit. on pp. 121, 124).

[144] L. Seoud, J. Dansereau, H. Labelle, and F. Cheriet. “Multilevel analysis of trunk surface measurements for noninvasive assessment of scoliosis deformities”. In: *Spine* 37.17 (2012), E1045–E1053 (cit. on pp. 120, 121, 126, 128).

[145] C. C. Shaw. *Cone beam computed tomography*. Taylor & Francis, 2014 (cit. on p. 31).

[146] J. Shen, P.-C. Su, S.-c. S. Cheung, and J. Zhao. “Virtual mirror rendering with stationary rgb-d cameras and stored 3-d background”. In: *Image Processing, IEEE Transactions on* 22.9 (2013), pp. 3433–3448 (cit. on p. 68).

[147] Y. Shen, F. Lu, X. Cao, and H. Foroosh. “Video completion for perspective camera under constrained motion”. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 3. IEEE. 2006, pp. 63–66 (cit. on p. 107).

[148] H. Shirato, S. Shimizu, K. Kitamura, T. Nishioka, K. Kagei, S. Hashimoto, H. Aoyama, T. Kunieda, N. Shinohara, H. Dosaka-Akita, et al. “Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor”. In: *International Journal of Radiation Oncology Biology Physics* 48.2 (2000), pp. 435–442 (cit. on pp. 18, 19).

[149] M. Shoham, S. Brink-Danan, A. Friedlander, and N. Knoller. “Bone-mounted miniature robotic system for spine surgery”. In: *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*. IEEE. 2006, pp. 917–920 (cit. on p. 26).

[150] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. “Real-time human pose recognition in parts from single depth images”. In: *Communications of the ACM* 56.1 (2013), pp. 116–124 (cit. on pp. 55, 102).

[151] J. Siewerdsen, A. Waese, D. Moseley, S Richard, and D. Jaffray. “Spektr: A computational tool for x-ray spectral analysis and imaging system optimization”. In: *Medical physics* 31.11 (2004), pp. 3057–3067 (cit. on p. 101).

[152] G. Soete, J. Van de Steene, D. Verellen, V. Vinh-Hung, D. Van den Berge, D. Michielsen, F. Keuppens, P. De Roover, and G. Storme. “Initial clinical experience with infrared-reflecting skin markers in the positioning of patients treated by conformal radiotherapy for prostate cancer”. In: *International Journal of Radiation Oncology Biology Physics* 52.3 (2002), pp. 694–698 (cit. on p. 18).

[153] P. Sprawls. *Physical principles of medical imaging*. Aspen Publishers, 1987 (cit. on p. 101).

[154] P. Stefan, S. Habert, A. Winkler, M. Lazarovici, J. Fürmetz, U. Eck, and N. Navab. “A Mixed-Reality Approach to Radiation-Free Training of C-arm Based Surgery”. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II*. Springer International Publishing, 2017, pp. 540–547 (cit. on pp. 131, 151).

[155] T. Steger, M. Hoßbach, and S. Wesarg. “Marker detection evaluation by phantom and cadaver experiments for C-arm pose estimation pattern”. In: *SPIE medical imaging*. International Society for Optics and Photonics. 2013, pp. 86711V–86711V (cit. on p. 16).

[156] K. Sugimoto, H. Fujii, A. Yamashita, and H. Asama. “Half-diminished reality image using three rgb-d sensors for remote control robots”. In: *Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on*. IEEE. 2014, pp. 1–6 (cit. on p. 108).

[157] T. S. Tang, N. MacIntyre, H. Gill, R. Fellows, N. Hill, D. Wilson, and R. E. Ellis. “Accurate assessment of patellar tracking using fiducial and intensity-based fluoroscopic techniques”. In: *Medical image analysis* 8.3 (2004), pp. 343–351 (cit. on p. 19).

[158] T. Tjardes, S. Shafizadeh, D. Rixen, T. Paffrath, B. Bouillon, E. S. Steinhausen, and H. Baethis. “Image-Guided Spine Surgery: State of the Art and Future Directions”. en. In: *European Spine Journal* 19.1 (2009), pp. 25–45 (cit. on pp. 63, 96, 139).

[159] D. Tomazevic, B. Likar, and F. Pernus. “3-D/2-D registration by integrating 2-D information in 3-D”. In: *IEEE transactions on medical imaging* 25.1 (2006), pp. 17–27 (cit. on p. 19).

[160] D. Tomazevic, B. Likar, T. Slivnik, and F. Pernus. “3-D/2-D registration of CT and MR to X-ray images”. In: *IEEE transactions on medical imaging* 22.11 (2003), pp. 1407–1416 (cit. on p. 19).

[161] R. Y. Tsai and R. K. Lenz. “Real Time Versatile Robotics Hand/Eye Calibration Using 3D Machine Vision”. In: *IEEE International Conference on Robotics and Automation*. 1988, pp. 554–561 (cit. on p. 134).

[162] I. A. Tsalafoutas, V. Tsapaki, A. Kaliakmanis, S. Pneumaticos, F. Tsoronis, E. D. Koulentianos, and G. Papachristou. “Estimation of radiation doses to patients and surgeons from various fluoroscopically guided orthopaedic surgeries”. In: *Radiation protection dosimetry* 128.1 (2007), pp. 112–119 (cit. on p. 30).

[163] M. Wagner, C. Duwenkamp, K. Dresing, and O. J. Bott. “An approach to calculate and visualize intraoperative scattered radiation exposure.” In: *MIE*. 2009, pp. 831–835 (cit. on p. 30).

[164] J. Wang, P. Fallavollita, L. Wang, M. Kreiser, and N. Navab. “Augmented reality during angiography: integration of a virtual mirror for improved 2d/3d visualization”. In: *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE. 2012, pp. 257–264 (cit. on p. 27).

[165] L. Wang. “Novel techniques for integrating video augmented X-ray imaging into orthopedic and trauma surgery”. Dissertation. München: Technische Universität München, 2012 (cit. on pp. 42, 71, 79, 80).

[166] L. Wang, P. Fallavollita, R. Zou, X. Chen, S. Weidert, and N. Navab. “Closed-form inverse kinematics for interventional C-arm X-ray imaging with six degrees of freedom: modeling and application”. In: *Medical Imaging, IEEE Transactions on* 31.5 (2012), pp. 1086–1099 (cit. on p. 55).

[167] L. Wang, P. Fallavollita, A. Brand, O. Erat, S. Weidert, P. Thaller, E. Euler, and N. Navab. “Intra-op measurement of the mechanical axis deviation: an evaluation study on 19 human cadaver legs”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 609–616 (cit. on p. 55).

[168] L. Wang, J. Traub, S. Weidert, S. Heining, E. Euler, and N. Navab. “Parallax-free long bone X-ray image stitching”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009* (2009), pp. 173–180 (cit. on pp. 16, 33, 146).

[169] P. Wang, O. Ecabert, T. Chen, M. Wels, J. Rieber, M. Ostermeier, and D. Comaniciu. “Image-based co-registration of angiography and intravascular ultrasound images”. In: *IEEE transactions on medical imaging* 32.12 (2013), pp. 2238–2249 (cit. on pp. 27, 28).

[170] X. Wang, C. S. zu Berge, S. Demirci, P. Fallavollita, and N. Navab. “Improved interventional X-ray appearance”. In: *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*. IEEE. 2014, pp. 237–242 (cit. on pp. 27, 28).

[171] X. Wang, S. Habert, C. S. zu Berge, P. Fallavollita, and N. Navab. "Inverse visualization concept for RGB-D augmented C-arms". In: *Computers in biology and medicine* 77 (2016), pp. 135–147 (cit. on p. 151).

[172] X. Wang, S. Habert, M. Ma, C.-H. Huang, P. Fallavollita, and N. Navab. "[POSTER] RGB-D/C-arm Calibration and Application in Medical Augmented Reality". In: *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 100–103 (cit. on p. 152).

[173] X. Wang, S. Habert, M. Ma, C.-H. Huang, P. Fallavollita, and N. Navab. "Precise 3D/2D calibration between a RGB-D sensor and a C-arm fluoroscope". In: *International journal of computer assisted radiology and surgery* 11.8 (2016), pp. 1385–1395 (cit. on pp. 73, 74, 79, 130, 145, 151).

[174] V. Waran, V. Narayanan, R. Karuppiah, S. L. Owen, and T. Aziz. "Utility of Multimaterial 3D Printers in Creating Models with Pathological Entities to Enhance the Training Experience of Neurosurgeons". In: *Journal of Neurosurgery* 120.2 (2014), pp. 489–492 (cit. on pp. 132, 139).

[175] O. Wasenmüller and D. Stricker. "Comparison of Kinect v1 and v2 Depth Images in Terms of Accuracy and Precision". In: *Asian Conference on Computer Vision Workshop (ACCV workshop), Springer*. 2016 (cit. on p. 51).

[176] W. Wein, B. Röper, and N. Navab. "2D/3D registration based on volume gradients". In: *SPIE Medical Imaging*. Vol. 5747. 2005, pp. 144–150 (cit. on p. 19).

[177] H.-R. Weiss and D. Goodall. "Rate of complications in scoliosis surgery–a systematic review of the Pub Med literature". In: *Scoliosis* 3.1 (2008), p. 9 (cit. on pp. 119, 120).

[178] M. Wieczorek, A. Aichert, P. Fallavollita, O. Kutter, A. Ahmadi, L. Wang, and N. Navab. "Interactive 3D visualization of a single-view X-Ray image". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2011, pp. 73–80 (cit. on pp. 22–24).

[179] Wikiversity. *WikiJournal of Medicine/Medical gallery of Blausen Medical 2014 — Wikiversity,* (cit. on p. 83).

[180] T. Wolff, A. Lasso, M. Eblenkamp, E. Wintermantel, and G. Fichtinger. "C-arm angle measurement with accelerometer for brachytherapy: an accuracy study". In: *International journal of computer assisted radiology and surgery* 9.1 (2014), pp. 137–144 (cit. on p. 17).

[181] J. M.-L. Wong, S. Bewsher, J. Yew, A. Bucknill, and R. de Steiger. "Fluoroscopically assisted computer navigation enables accurate percutaneous screw placement for pelvic and acetabular fracture fixation". In: *Injury* 46.6 (2015), pp. 1064–1068 (cit. on p. 25).

[182] P. Wucherer, P. Stefan, K. Abhari, et al. "Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis". In: *IEEE Transactions on Medical Imaging* 34.8 (2015), pp. 1730–1737 (cit. on p. 132).

[183] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo. "Operating Principles of Structured Light Depth Cameras". In: *Time-of-Flight and Structured Light Depth Cameras: Technology and Applications*. Cham: Springer International Publishing, 2016, pp. 43–79 (cit. on pp. 47, 50).

[184] L. Zhang, B. Curless, and S. M. Seitz. "Rapid shape acquisition using color structured light and multi-pass dynamic programming". In: *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*. IEEE. 2002, pp. 24–36 (cit. on p. 47).

[185] Z. Zhang. "A flexible new technique for camera calibration". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.11 (2000), pp. 1330–1334 (cit. on p. 72).

[186] Z. Zhang. "Flexible camera calibration by viewing a plane from unknown orientations". In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Vol. 1. Ieee. 1999, pp. 666–673 (cit. on p. 16).

[187] D. Zikic, M. Groher, A. Khamene, and N. Navab. “Deformable registration of 3D vessel structures to a single projection image”. In: *Proc. SPIE (Medical Imaging: Image Processing)*. 2008 (cit. on p. 19).

List of Figures

List of Tables