AUGMENTED REALITY AND ARTIFICIAL INTELLIGENCE IN IMAGE-GUIDED AND ROBOT-ASSISTED INTERVENTIONS

by

Javad Fotouhi

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

October, 2020

Abstract

In minimally invasive orthopedic procedures, the surgeon places wires, screws, and surgical implants through the muscles and bony structures under image guidance. These interventions require alignment of the pre- and intra-operative patient data, the intra-operative scanner, surgical instruments, and the patient. Suboptimal interaction with patient data and challenges in mastering 3D anatomy based on ill-posed 2D interventional images are essential concerns in image-guided therapies.

State of the art approaches often support the surgeon by using external navigation systems or ill-conditioned image-based registration methods that both have certain drawbacks. Augmented reality (AR) has been introduced in the operating rooms in the last decade; however, in image-guided interventions, it has often only been considered as a visualization device improving traditional workflows. Consequently, the technology is gaining minimum maturity that it requires to redefine new procedures, user interfaces, and interactions.

This dissertation investigates the applications of AR, artificial intelligence, and robotics in interventional medicine. Our solutions were applied in a broad spectrum

of problems for various tasks, namely improving imaging and acquisition, image computing and analytics for registration and image understanding, and enhancing the interventional visualization. The benefits of these approaches were also discovered in robot-assisted interventions.

We revealed how exemplary workflows are redefined via AR by taking full advantage of head-mounted displays when entirely co-registered with the imaging systems and the environment at all times. The proposed AR landscape is enabled by co-localizing the users and the imaging devices via the operating room environment and exploiting all involved frustums to move spatial information between different bodies. The system's awareness of the geometric and physical characteristics of X-ray imaging allows the exploration of different human-machine interfaces. We also leveraged the principles governing image formation and combined it with deep learning and RGBD sensing to fuse images and reconstruct interventional data.

We hope that our holistic approaches towards improving the interface of surgery and enhancing the usability of interventional imaging, not only augments the surgeon's capabilities but also augments the surgical team's experience in carrying out an effective intervention with reduced complications.

Primary Reader and Advisor: Nassir Navab (Johns Hopkins University)

Co-Advisor: Mehran Armand (Johns Hopkins University)

Secondary Readers: Mathias Unberath (Johns Hopkins University), Greg Osgood (Johns Hopkins Hospital)

Acknowledgments

I would like to thank here the numerous people who supported me through this journey, contributed directly or indirectly to this thesis, and made these years of research such an exciting time.

I owe enormous thanks to my advisor, Nassir Navab, that his enthusiasm for research is infectious and his inspirations radiate. I learned not only technical skills from him, but also took lessons from how he leads people with humility and interacts with students and colleagues with respect. Despite that we were often thousands of miles away, he was always available to quickly jump on a Skype call for a scientific or even personal conversation. I am grateful to Nassir that he encouraged me to have fun with research, ask more questions, think loudly in the lab, and get as creative as possible. I am also thankful to him for granting me all the liberty to pursue ideas that have prepared me to make better decisions and shaped me as a scientist. I will undoubtedly miss the range he brings to projects and his ability to apply knowledge to many diverse domains.

I cannot find enough appraisal words for the immense support and cordial guid-

ance from my co-advisor, Mehran Armand. Mehran gave me the benefit of his vast experience in designing computer-assisted surgical solutions. He challenged me to step back, see the big picture, and identify the impact our contributions may make on reshaping the future of clinical care. Time and again, he shared his fair judgment and unbiased perspective of research with me and encouraged me to think about different dimensions of my work that was essential to my progress.

Sincere thanks to my two friends, colleagues, and mentors Mathias Unberath and Bernhard Fuerst, who played a crucial role in shaping my research character. Their contributions are evident everywhere in this thesis and beyond. I greatly enjoyed all our fruitful and stimulating brainstorming conversations during quad-walks and learned a great deal from their critical input.

Johns Hopkins offered me a rare and unique experience to build strong partnerships with many talented and world-renowned clinicians and surgeons, and enjoy their clinical support towards designing translatable clinical solutions. I wish to warmly thank my medical advisor, Greg Osgood, who patiently explained clinical challenges to me, even during his surgeries. Observing patient experiences and understanding his enthusiasm for improving patient safety made me even more passionate about empowering clinicians with technology and improving clinical outcomes. I want to also express my gratitude to my other clinical champion, Alex Johnson, that brought substantial momentum to our pre-clinical studies.

I want to express my gratitude to Russell Taylor for his input on my Ph.D. projects

and his leadership at LCSR. I also thank all who make this place such a unique and vibrant environment. It was a great pleasure to spend these years here and interact with experts working at the intersection of engineering and medicine.

I want to thank the great people from the LCSR and CS family at JHU, colleagues and friends who supported me in scientific and non-scientific matters: Sing Chun Lee, Daniil Pakhomov, Alejandro Martin Gomez, Arian Mehrfard, Giacomo Taylor, Qiaochu Wang, Tianyu Song, Wenhao Gu, Emerson Tucker, Jonas Hajek, Yordanka Velikova, Bastian Bier, Jan-Nico Zaech, Marius Fischer, Risto Kojcev, Nikita Ivkin, Oliver Zettinig, Sebastian Andress, Alexander Winkler, Clayton Alexander, Florian Goldmann, Laura Fink, Nima Befrui, Alexander Barthel, Ashkan Khakzar, Christoph Graumann, Erin Sutton, Felix Bork, Francisco Pinto, Kevin Yu, Leila Barmaki, Mareike Thies, Michael Sommersperger, Philipp Nikuta, Shahriar Sefati, Farshid Alambeigi, Cong Gao, Robert Grupp, Amirhossein Farvardin, Mahsan Bakhtiarijenad, Alireza Chamani, Joshua Liu, Rachel Hegeman, Robert Di Pietro, Ayushi Sinha, Xingtong Liu, Jie Ying Wu, Long Qian, Hasan Sen, Ali Ebrahimi, Berk Gonenc, Chloe Audigier, Anand Malpani, Princy Parsana, Nicola Fuerst, Iris Unberath, Fabian Prada, Yasamin Nazari, and Anton Deguet. I hope these friendships will last for many years to come.

A very special thank you is dedicated to the friendly and supportive administrative staff at LCSR and CS that made everything so smooth: Alison Morrow, Lorrie Dodd, Ashley Moriarty, Jordan Card, Zack Burwell, Debbie DeFord, Rose Chase,

Kim Franklin, and Cathy Thornton.

I want to thank friends who became family to me when I was away from my family: Hani, Nikta, Shahriar, Parastoo, Sahar, and Lida. They made my life beautiful and kept me sane. The laughs we shared will remain unforgettable. I am thankful to them for their constant affection, and I am glad that they are all part of my life.

To my parents, Reza and Mina, I express my deep love and gratitude for bringing me up to this path; my dad, who always encouraged me to think critically and shaped me into a planner, and my mom, who showed me to love with all my heart. I am greatly indebted that they instilled in me the importance of education and put their trust in my choices. I also express my love and gratitude to my sister Faranak and my brothers Hossein and Ali, that I always looked up to them. I am incredibly thankful to my extended family, Farrah, Hassan, and Yashar that welcomed me in the US and made me feel at home. I am grateful for their love and unending support through this process.

My deepest gratitude to my best friend and my constant cheerleader - my beautiful wife, Gilda. Her everlasting smile and steady stream of love accompanied me throughout this exciting journey. She stood by me every step of the way and pushed me to deliver my best. She showed me that with grit, I get the ability to exceed my limits. Gilda shared with me her knowledge about leadership and gave me insightful feedback regarding interpersonal interactions. It was always my pleasure to discuss research ideas with her and see her enthusiastically, inspiring me to continue pushing

and making an impact on human lives. She shared with me the good and bad of this journey and helped me become the person I am today. I owe her everything.

Dedication

To my loving parents, *Reza* and *Mina*, and to my beautiful wife *Gilda*, who have been great sources of inspiration and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Clinical Background

Surgical care is a fundamental part of the healthcare system that has emerged as an essential concern in global health. Due to the cost and complexity of delivering surgical care, yet a considerable population in the world do not have access to safe and affordable surgery (Fig. 1.1). During the global COVID-19 pandemic and public health crises, the challenges in delivering surgical procedures were even further noticed, as expediting (elective) surgeries with fast recoveries such that the patients spend the minimum time in the hospital environment became a paramount concern. Long open surgeries with extreme collateral damage will worsen the backlog and put an additional burden on the healthcare system.

In total hip arthroplasty (THA), also referred to as total hip replacement, the

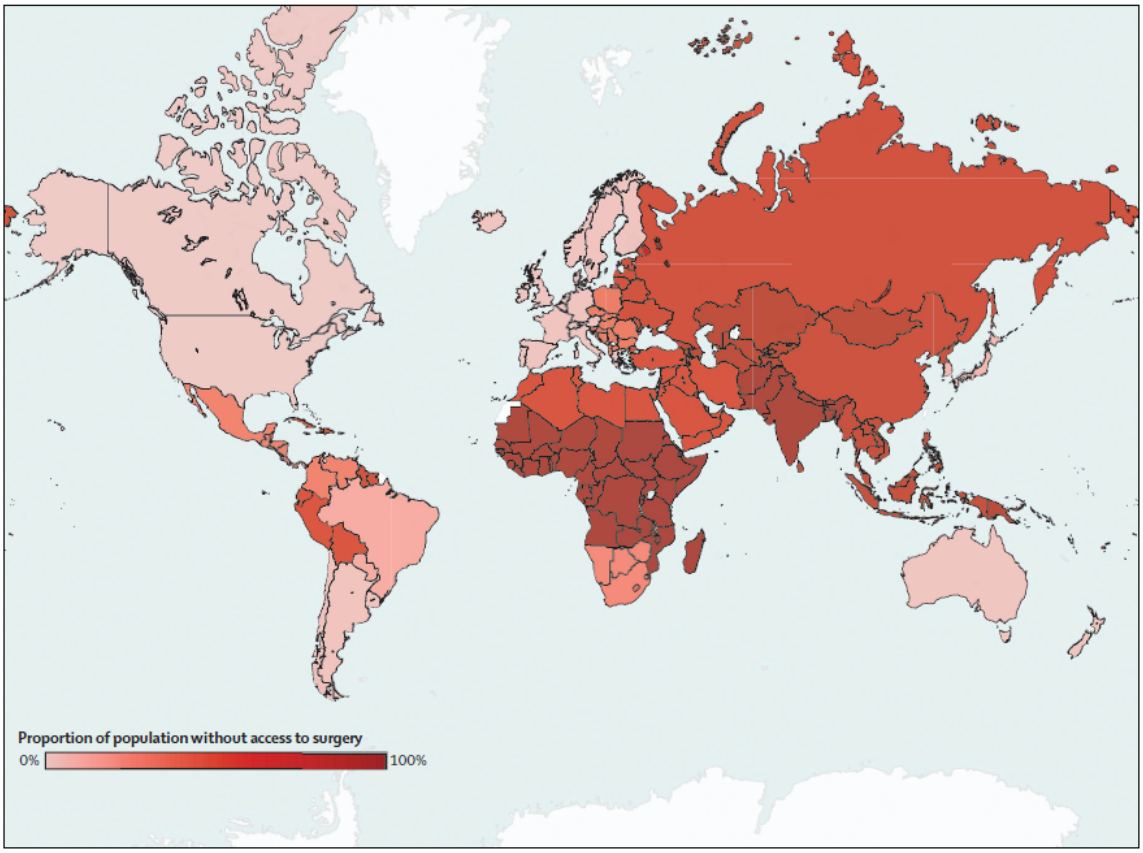

Figure 1.1: The heatmap shows the percentage of world population at each region without access to safe and affordable surgical care [1].

damaged bone and cartilage are replaced with prosthetic components. The procedure relives pain and disability with a high success rate. In 2010, there were approximately 330,000 THAs performed in the US. This is projected to increase to 570,000 THAs by 2030 [2] as younger patients are considered for THA. Together with a prolonged life expectancy, the consideration of younger, more active patients for THA suggests that implant longevity is of increasing importance as it is associated with the time to revision arthroplasty [2]. The time to repeat surgery is affected by the wear of the

implants that is correlated with their physical properties as well as acetabular component positioning. Sub-optimal surgical outcomes that demand revision procedures would risk the patient safety and use additional resources of our healthcare system.

In image-guided orthopedic interventions, X-ray images are acquired to guide the procedure. These surgeries usually yield high X-ray exposure for both patient and clinical staff, and may increase fatigue and frustration due to the difficulty in continuous re-positioning of the mobile X-ray machine (C-arm). To perform a complex 3D procedure, the surgeon relies on these X-ray images which are 2D projective representations of the anatomy. Therefore, guidance and visualization both become very challenging for the surgeon.

Surgeries performed under image guidance heavily depend on surgeons' mental alignment and spatial mapping between patient and the medical data. For instance, in a minimally invasive K-wire placement procedure (Fig. 1.2), the surgeon would need to verify the K-wire 3D pose relative to patient for every small advancement of the K-wire using several X-ray images acquired from various angles. This image-guided procedure is considered to be delicate as small misplacement of the K-wire could cause severe damage to the external iliac artery and vein, obturator nerve, or to structures such as the inguinal canal and intra-articular hip joint.

A reliable solution for direct view into patient anatomy as well as an intuitive visualization to observe the pre- and intra-interventional data within the surgical site is not available yet (Fig. 1.3). In addition to visualization and perceptual challenges,

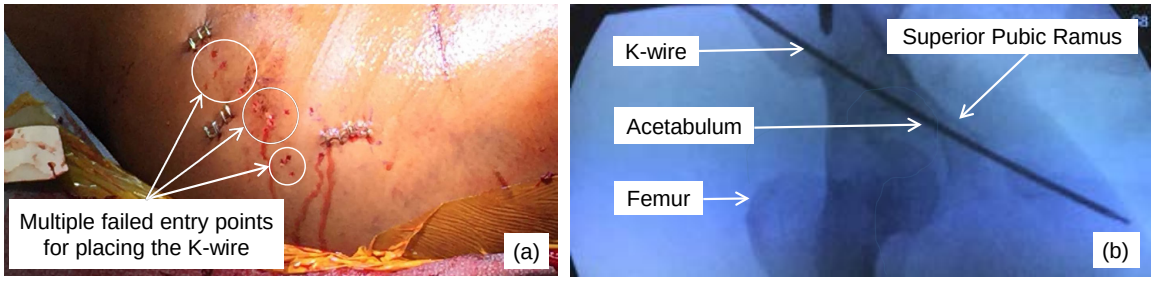

Figure 1.2: **(a)** During a fracture reduction surgery, the surgeon may attempt and fail frequently in accurately placing the wire within a safe zone inside the bony anatomy. **(b)** The narrow area between the acetabulum and the superior pubic ramus is shown as an example of a typical safe zone for K-wire placement in a trauma procedure.

registration and guidance may often be delivered using external trackers and invasive fiducials which dramatically increase the setup complexity and require invasive procedures for their implantation into the bone [3, 4].

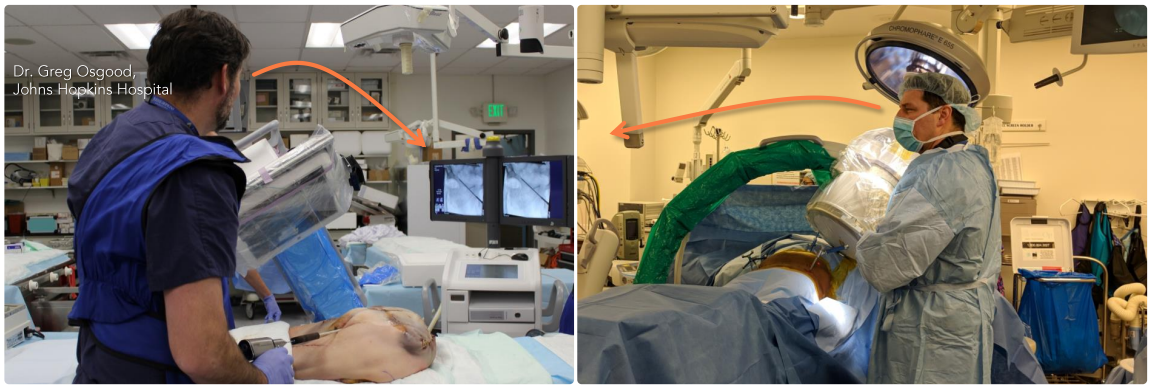

Figure 1.3: The interventional data are commonly shown off-axis and indirectly to the surgeon. Consequently, the surgeons have to take their gaze away from the patient site to visualize each image update.

C-arm 2D fluoroscopy is the crucial imaging modality for several image-guided interventions. During these procedures, the C-arm is frequently re-positioned to acquire images from various perspectives of a target anatomy. Performing surgery solely

under 2D fluoroscopy guidance is a challenging task as a single point of view lacks the information required to navigate complex 3D structures, making the intervention and procedure extremely difficult.

For orthopedic traumatologists, restoring the correct length, alignment, and rotation of the affected extremity is the goal of any fracture management strategy regardless of the fixation technique. This can be difficult with the use of conventional fluoroscopy with limited field of view and lack of 3D cues. For instance, it is estimated that malalignment ($> 5^\circ$ in the coronal or sagittal plane) is seen in approximately 10%, and malrotation ($> 15^\circ$) in up to approximately 30% of femoral nailing cases [5, 6].

Complex and frequent K-wire placements occur after ilio-sacaral joint and superior pubic ramus fractures. As shown in Fig. 1.4, the common treatment for both injuries is the placement of screws to stabilize the fracture. During the minimally invasive approach, the surgeon places the K-wire through the muscles and all bone fragments under image-guidance, requiring a good mental alignment of the intra-operative X-ray imaging, the medical instrument, and the patient [7, 8]. Due to the 2D nature of the X-ray images, this results in frequent re–positioning of the C–arm [9]. For instance, in pelvic acetabulum fractures, the surgeon needs to find the correct trajectory of the K-wire through the superior pubic ramus. The misplacement of the K-wire could cause severe damage to the external iliac artery and vein, obturator nerve, or to structures such as the inguinal canal and intra-articular hip joint [10]. It is not unusual that a

single K-wire placement for one screw takes up to ten minutes [11].

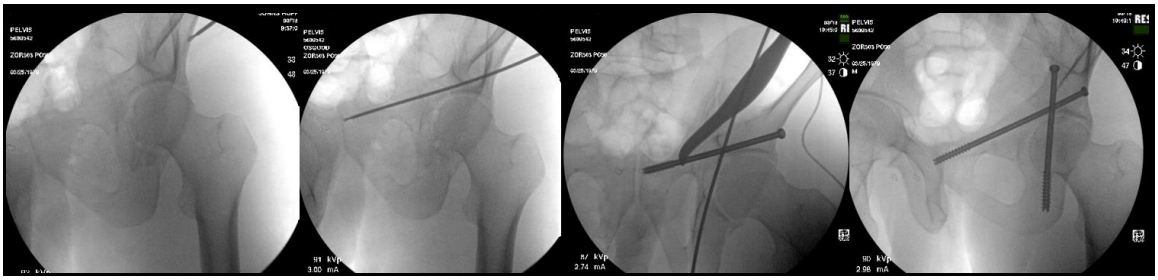

Figure 1.4: Common treatment for sacroiliac joint and superior pubic ramus fractures is the placement of K-wires under fluoroscopic guidance.

In fracture care surgery, it is of utmost importance to achieve proper alignment between the ends of extremities and reduce discrepancies to the contralateral side. It is commonly acknowledged that the verification mechanisms from 2D images are mentally challenging and cumbersome. The unmet clinical needs for intra-operative measurement of different biomechanical axes and bone lengths are also important concerns in other fields of orthopedic surgery, including high tibial osteotomy and total hip arthroplasty [12–14].

Ensuring adequate fracture reduction in pelvis and acetabulum surgery can be difficult. Surgeons rely on known radiographic densities and parameters to guide fracture reduction. In acetabulum surgery, for example, surgeons acquire specific radiographic views to visualize the ilioischial and/or iliopectineal lines to assess reduction of the posterior and anterior pelvic columns respectively [15, 16]. This method can become unreliable when the bone is comminuted and the anatomic area of interest is distorted.

Additionally, surgeons may obtain radiographic views of the healthy contralateral

side for comparison when attempting to reconstruct the comminuted operative side. Yet, this method has its own limitations. For instance, it is rare to obtain a radiographic view that possesses clear and symmetric views of both sides of the pelvis at one time. When the image intensifier is very close to the patient this may be achieved for the anterior/posterior (AP), inlet, and outlet views; however, many more radiographic views are typically used during a pelvis or acetabulum fracture case that does not possess this property. Additionally, even when both sides are visualized, the surgeon must imagine a mirrored version of the healthy side superimposed on the operative side, which increases the level of mental task load during the surgery. Studies have shown that these 2D fluoroscopic methods can be unreliable [17], which has led surgeons to seek other methods for reduction assessment.

Compositions of X-ray images were investigated to assist with quantifying the total length and angular measurements in pre- or post-operative settings for patients that underwent osteotomy, endoprosthesis, or fracture reduction procedures [18, 19]. These works do not address the image stitching problem and instead focus on disambiguating relative poses between multiple 2D images by using radiopaque scales that are placed approximately parallel to the extremity. Other approaches use bi-planar X-ray scanners with orthogonal planes and recover 3D anatomical landmarks through stereo 3D reconstruction [20].

In orthopedic sports-related and adult reconstruction procedures, high tibial and distal femoral osteotomies are utilized to shift contact forces in the knee in patients

with unilateral knee osteoarthritis. These osteotomies rely on precise correction of the mechanical axis to achieve positive clinical results.

Acquiring Cone-Beam Computed Tomographic (CBCT) volumes which provides 3D information of the anatomy is a solution to this problem. However, CBCT acquisition does not yield real-time feedback. A considerable number of scans would need to be performed at every step of the procedure, requiring the surgical site to be prepared for each scan. Pre-interventional 3D patient data can be fused to the C-arm images and augment the intra-interventional flouroscopy images with 3D information. This multi-modal image fusion is used in several different interventions, particularly in radiation therapy [21], cardiac and endovascular procedures [22], and orthopedics and trauma interventions [23]. Additionally, other image-guided orthopedic procedures such as femoroplasty [24], (robot-assisted) bone augmentation, K-wire and screw placement in pelvic fractures, cup placement for hips, and surgeries for patients with osteonecrosis can benefit from the registration of pre-interventional data with real-time intra-interventional C-arm images.

Intra-operative CBCT has the potential to provide a navigation system for osteotomies about the knee while integrating well with the conventional surgical workflow. Another promising use for intraoperative CBCT in orthopedics is for comminuted fractures of the mid femur. Intraoperative 3D CBCT has the potential to verify length, alignment, and rotation and to reduce the need for revision surgery due to malreduction [25]. In Fig. 1.5 the difficulty in addressing rotational alignment

in mid-shaft comminuted femur fractures and the clinical impact of misalignment is demonstrated. Fig. 1.6 demonstrates the anatomical landmarks used to estimate the 3D position of the bone. Traditionally, to ensure proper femoral rotation, the contralateral leg is used as a reference: First, an AP radiograph of the contralateral hip is acquired, and particular attention is paid to anatomical landmarks such as how much of the lesser trochanter is visible along the medial side of the femur. Second, the C-arm is translated distally to the knee and then rotated $\sim 90^\circ$ to obtain a lateral radiograph of the healthy knee with the posterior condyles overlapping. These two images, the AP of the hip and lateral of the knee, determine the rotational alignment of the healthy side. To ensure correct rotational alignment of the injured side, an AP of the hip (on the injured side) is obtained, attempting to reproduce the AP radiograph acquired of the contralateral side (a similar amount of lesser trochanter visible along the medial side of the femur). This ensures that the position of both hips is similar. The C-arm is then moved distally to the knee of the injured femur and rotated $\sim 90^\circ$ to a lateral view. This lateral image should match that of the healthy side. If they do not match, rotational correction of the femur can be performed, attempting to obtain a lateral radiograph of the knee on the injured side similar to that of the contralateral side. This procedure motivates the need for intraoperative 3D imaging with large field of view, where leg length discrepancy and malrotation can be quantified intraoperatively and compared with the geometric measurements from the preoperative CT scan of the contralateral side.

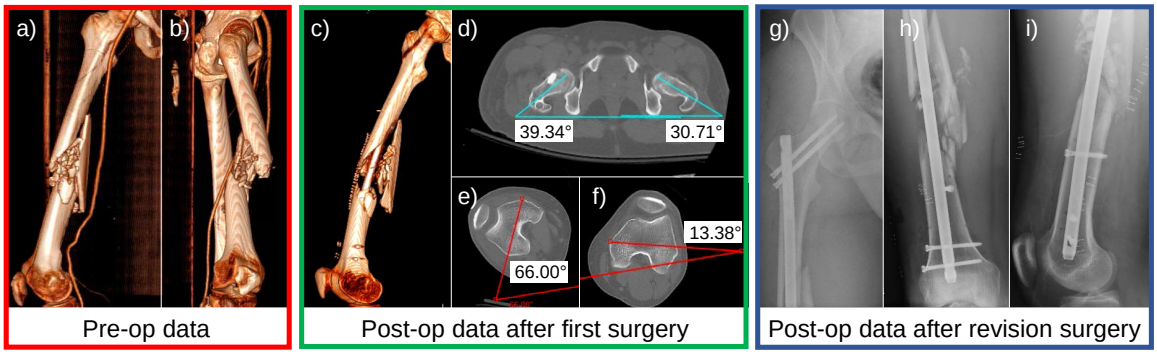

Figure 1.5: Difficulties arise in addressing rotational alignment in long bone fractures - The 3D preoperative CT scan of the right femur of a patient with a ballistic fracture of the femoral shaft is shown in **(a-b)**. As seen in these images, due to the significant comminution, there are few anatomical cues as to the correct rotational alignment of the bone. **(c)** shows the postoperative CT of the same femur after reduction and placement of a cephallomedullary nail. The varus/valgus alignment appears to be restored (see Fig. 1.6); however, significant rotational malalignment is present with excessive external rotation of the distal aspect of the femur. Axial cuts from the postoperative CT scan are shown in **(d-f)**. As shown in **(d)**, the hips are in relatively similar position (right hip $\sim 10^\circ$ externally rotated vs. the left). However, in **(e)**, the operative right knee is over 40° more externally rotated than the healthy contralateral side in **(f)**. Figures **(g-i)** show the anteroposterior (AP) view of the right hip, AP view of the right femur, and the lateral postoperative radiographs after revision cephalomedullary nailing with correction of the rotational deformity. The revision surgery includes removal and correct replacement of the intramedullary nail.

We recognize the need for more intuitive and practical mechanisms to employ pre- and intra-interventional data that can yield lower frustration for the surgical team, improve patient treatment, and reduce the overall cost for care delivery.

1.2 Navigation and Robotic Assistance

Surgical navigation and robotic systems are developed to support surgery with localization and execution of well-defined tasks [26–29]. Though these systems increase

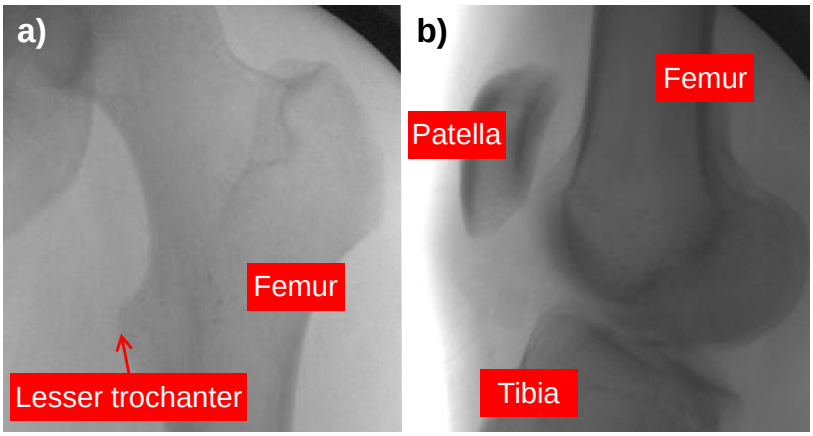

Figure 1.6: Contralateral images for guidance in rotational alignment - **(a)** and **(b)** are intraoperative fluoroscopic images from the revision surgery; AP view of the contralateral hip and lateral view of the contralateral knee. These images were utilized to guide rotational alignment of the fractured femur. By visualizing landmarks on these radiographs and understanding the change in angulation of the C-arm, the surgeon can estimate the rotational alignment of the healthy femur and attempt to recreate this alignment on the operative side.

the accuracy, their complex setup and explicit tracking nature may overburden the surgical workflow and consequently impede their acceptance in clinical routines [30]. Image-based navigation alleviates the requirements for external tracking, though depends strongly on pre-operative data which become outdated when the anatomy is altered during the surgery [31, 32].

During computer-assisted interventions, surgical navigation systems are utilized to support the mental alignment and guide the surgeon 1.7. Most current systems use pre-operative X-ray Computed Tomography (CT) volumes, infrared optical tracking systems, and tracking targets on tools and the patient, to provide a visualization of the tool relative to the patient's anatomy. Some systems include the use of a C-arm and enable tracking of tools relative to intra-operatively acquired X-ray images [33]. After

the guided procedure, the validation of the placement is performed using conventional X-ray imaging.

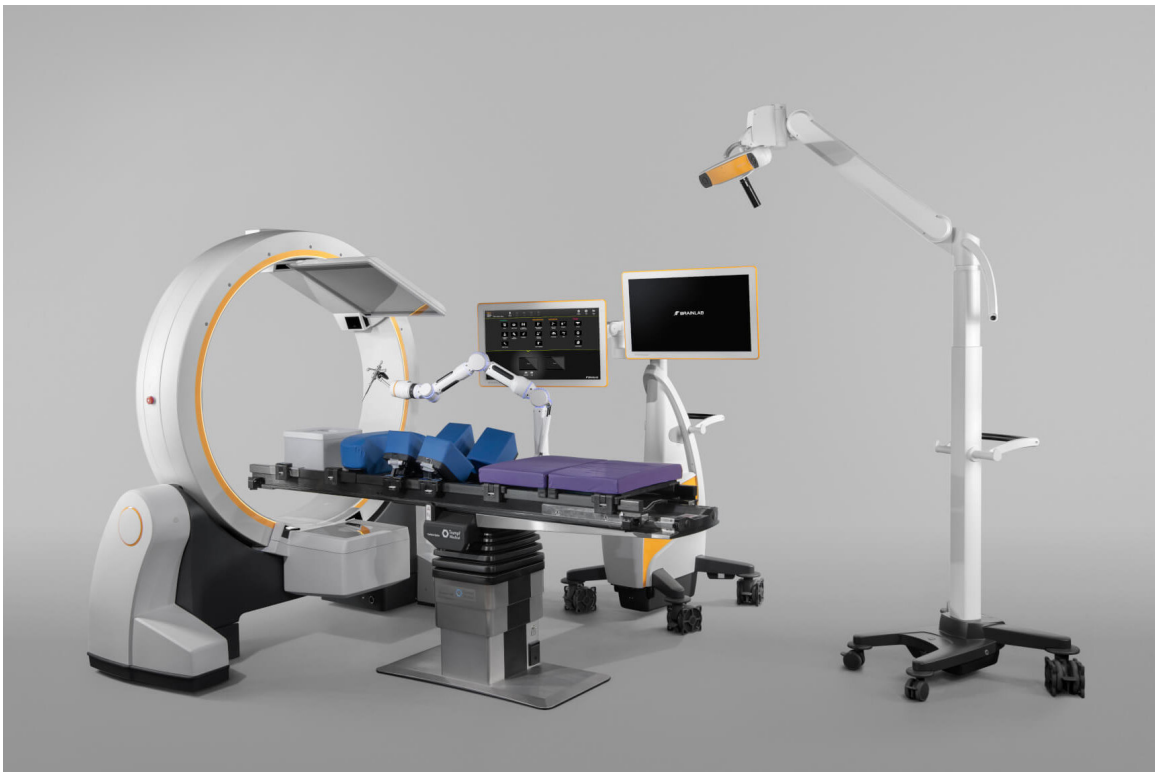

Figure 1.7: Surgical navigation systems provide guidance with respect to patient and the imaging data. Image source: BrainLab

While some publications indicate that surgical navigation systems reduce the radiation dose [34] and increase accuracy [35, 36], a more recent study shows no clear advantage of using navigation systems in some procedures [37]. After two decades of navigation systems, we can see that these systems may have failed to provide the promised advantages as they do not reduce the required time in the operating room, show no systematic, significant influence on the patient outcome, and do not reduce

the frustration of the surgeon and staff.

External navigation systems commonly use certain points on the anatomy of interest, as decided by the surgeon, and conform to a "map" of the known morphology of the anatomy of interest. Despite the fact that THA frequently uses X-ray images for navigation and pre-operative patient CT may not be available, several computer assisted THA solutions suggest planning the desired pose of the acetabular component pre-operatively on a CT scan of the patient [38, 39]. Pre-operative CT imaging allows planning of the implants in 3D, automatically estimating the orientation of the natural acetabular opening, and predicting the appropriate size of the cup implant [40].

Navigation-based THA with external trackers are performed based on pre-operative patient CT, or image-less computer assisted approaches. The planning outcome in a CT-based navigation approach is used intra-operatively with external optical navigation systems to estimate the relative pose of the implant with respect to the patient anatomy during the procedure. Tracking of the patient is commonly performed using fiducials that are drilled into the patient's bones. Registration of the pre-operative CT data to the patient on the surgical bed is performed by manually touching anatomical landmarks on the surface of the patient using a tracked tool [38]. In addition to the paired-point transformation estimated by matching the few anatomical landmarks, several points are sampled on the surface of the pelvis and matched to the segmentation of the pelvis in the CT data [41]. CT-based navigation showed statistically

significant improvement in orienting the acetabular component and eliminating malpositioning, while resulting in increased blood loss, cost, and time for surgery [42, 43]. Combined simultaneous navigation of the acetabulum and the femur was used in 10 clinical tests where the surgical outcome based on post-operative imaging showed 2.98 mm and 4.25° error in cup position and orientation, respectively [44].

Image-less navigation systems do not require any pre-operatively acquired radiology data. In this method, the pelvic plane is located in 3D by only identifying anatomical landmarks on the surface of the patient using a tracked pointer reference tool and optically visible markers attached to the patient [45]. This approach showed improvement in terms of cup positioning [46]. However, few number of samples points for registration as well as pelvis tilts resulted in unreliable registration [47].

Robotic systems are developed to provide additional confidence to the surgical team, reduce human error, increase precision, and ensure reproducibility [48–53]. In a robotic system, pins are implanted into the patient's femur prior to acquiring a pre-operative CT scan. After the surgeon has performed the planning on the CT data, the robot is introduced into the operating room. To close the registration loop between patient, robot, and CT volume, each pre-operatively implanted pin is touched by the robot with manual support. To eliminate the need for fiducial implantation, registration is either achieved by selecting several points on the surface of the bone using a digitizer and using an iterative closest point algorithm to perform registration to patient CT data [54], or by using intra-operative C-arm fluoroscopy

and performing 2D/3D registration between the X-ray image and CT volume [55]. After registering the pre-operative CT data to patient, the robot assists the surgeon in placing the femoral stem and the acetabular component according to the planning. The outcome of 97 robot-assisted THA procedures indicates performance similar to the conventional technique [52]; However, in some cases additional complications such as nerve damage, post-operative knee effusion, incorrect orientation of the acetabular component, and deep reaming resulting in leg length discrepancy were reported when the robotic system was used. These studies and their respective outcomes all attest to the need for computer-assisted solutions that satisfy both the requirements concerning accuracy and ergonomics.

1.3 Augmented Reality in Orthopedic Surgery

Augmented Reality (AR) can support the surgeon by intuitive augmentation of medical information. AR refers to the real world augmented with virtual information, as opposed to Virtual Reality (VR), in which the user is confronted with a completely virtual setting [56, 57]. The user's view is augmented either via monitor-based display system, optical see-through system or video see-through system [58]. With recent commercial products such as Google Glass (Google Inc., Mountain View, California, USA) and Microsoft HoloLens (Microsoft, Redmond, WA), optical see-through systems have gained broad availability. Such head-mounted-displays (HMD) allow a high

degree of flexibility by enabling the user to visualize virtual content that is directly overlaid onto the present reality.

An important component of AR is the underlying tracking-system. Tracking is essential when placing virtual objects into the real world in correct relations and positions. Most systems are based on external markers, where a particular pattern in the real world is tracked as a reference [59]. Visual markers are widely used for this purpose, where unique and high contrast patterns are detected by optical cameras [60]. On the other hand, modern systems act independently from such predefined patterns and are referred to as marker-less systems. The marker-less tracking technology is enabled by using several Inertial Measurement Unit (IMU), Red-Green-Blue (RGB) and Infrared sensors on the HMD which allow creating a spatial map of the room and performing real-time inside-out tracking and localization with respect to the environment. Therefore, they are capable of orienting themselves on already present objects, without additional markers [61].

In interventional medicine, AR is already introduced in several specialties, namely, neuro- [62] and visceral-surgeries [63, 64]. The growing interest for AR in orthopedics and trauma is not surprising, since the surgical procedures in orthopedic surgery frequently use *i)* visual data such as medical images acquired both pre- and intra-operatively and *ii)* often include mechanical steps such as screw or implant insertions, osteotomies and correction of deformities that can be visualizing the rigid relations in AR environments. Hence, such technical tasks seem predisposed to applications

of AR. As AR-assistance in orthopedic surgery is a strong focus in this dissertation, a comprehensive overview of the state-of-the-art for AR-supported systems in orthopedic interventions is presented in this section.

1.3.1 Instrument and Implant Placement

The orthopedic surgeon often relies on his three-dimensional (3D) orientation to place instruments or implants. Intra-operative fluoroscopy provides two-dimensional (2D) information. Therefore, the surgeon has to perform the mental task of mapping the 2D radiographs to the 3D anatomy. AR solutions can potentially reduce the dependence of the outcome on the surgeon's parameters by providing pre-operative planning in the field of view of the surgeon, or even showing correct trajectories for placing implants with overlays. Jing-Ren Wu et al. [65] used a camera-projector AR system to project the spinal bony anatomy on the back of a patient with entry points for vertebroplasty, based on pre-operative CT data. For registration, markers were attached to the patient skin and were tracked by the camera. First trials were with a dummy and animal-cadavers. Afterwards, the system was brought to the OR, and vertebroplasty was performed conventionally but with the additional assistance of AR. Time-saving for entry point identification by 70% was reported. One major limitation was the unreliable registration, in case that the patients' posture changed between CT and surgery.

Abe et al. [66] simulated needle-insertion into vertebral bodies in a phantom-

study. Point and angles of insertion were identified on patients' pre-operative CT scans. During the procedure, the surgeon wore a video see-through HMD (Moverio, Epson) with a webcam. The visual information was observed by the webcam and transmitted to a computer for processing. Registration between the patient and the CT required several manual steps and involved using a few fluoroscopy images. After estimating the spatial relations between the pre-operative planning and the patient, the desired trajectories were streamed to the HMD and overlaid on the patient (Fig. 1.8). Post-interventional CT was used to calculate deviation with respect to the planned trajectory. Significantly higher precision was reported compared to the conventional approach. Following the phantom-trials, they validated their system in five clinical cases and reported successful assistance to the performing surgeon.

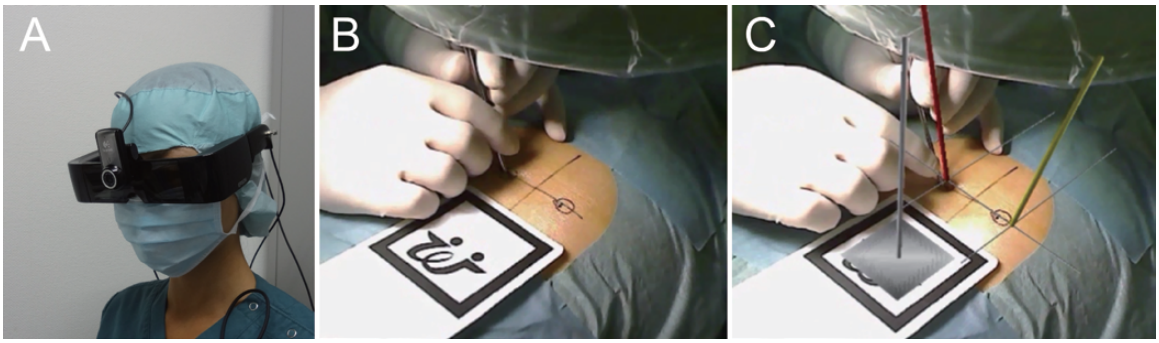

Figure 1.8: **(a)** HMD with camera, **(b)** the raw image captured by the camera, and **(c)** AR view of the surgeon are shown, respectively. The software created an augmented view and indicated the ideal insertion point and needle trajectory. Image courtesy of the Journal of Neurosurgery Spine, *A novel 3D guidance system using augmented reality for percutaneous vertebroplasty: technical note*, Yuichiro Abe, Shigenobu Sato, Koji Kato et al., Copyright 2013.

Navab et al. [67] focused on AR-supported vertebroplasty with a system consist-

ing of a mobile C-arm and a video camera attached near the X-ray source, so-called camera augmented mobile C-arm (CAMC). This system was designed using a double mirror construction, allowing the origin of the optical and X-ray cameras to virtually coincide. A major advantage of this design was that C-arm fluoroscopy images and video camera frames from the surgical site were fused without the need to warp the images. This system was self-contained and did not require any external navigation system. One goal of their solution was to perform vertebroplasty with one initial X-ray image, which was overlaid onto the video camera image. Five simulated vertebroplasties in a spine model were reported. Maximum of three X-rays were required, which is close to the goal of one. Three of these five procedures showed perfectly positioned needles, and two showed a medial perforation. The main reason for these perforations was reported as the undetected motion of the spine. As a consequence, they implemented markers to detect displacement automatically. In the same work, they performed interlocking of intramedullary nails as well as pedicel screw placements in cadavers. The two experiments were conducted successfully, and the procedure required less radiation and time compared to the standard C-arm technique. For the interlocking of intramedullary nails and pedicel screw placement, surgeons required two and three X-rays images, respectively. In an earlier study, Heining S. M. et al. [68] also investigated pedicle screw placement using the CAMC system. In two cadaver studies, in different levels of the lumbar and thoracic spine, all needle insertions were possible.

Another study by Gibby J. T. et al. [69] also investigated pedicle screw placement while using Microsoft HoloLens (Microsoft, Redmond, WA). In a lumbar saw bone model, they placed 36 needles, representing the pedicle screws. Using pre-interventional CT data, needle trajectory was estimated and superimposed into the surgeon's view, Post-interventional CT indicated that 97% of the needles were placed within the pedicle. Calculation with pedicle screws of a diameter up to 7 mm still demonstrated that 86% of screws were placed completely inside the pedicle.

U-Thainual P. et al. [70] suggested an AR-based technique for MRI-guided musculoskeletal interventions. The proposed magnetic resonance image overlay system (MR-IOS) provided an MRI vision for the operator and was used for needle insertions on a spine phantom. Main hardware components included a transverse plane laser, an MRI compatible monitor, and a semi-transparent mirror (Fig. 1.9). Onto this mirror, the MR image and the desired insertion path were jointly projected. This system was mounted in the mouth of an MRI scanner that provided 2D transverse slices. The alignment between the virtual medical images in the mirror and the patient was achieved by manually rotating and translating the virtual image plane until the anatomical landmarks on the patient and image coincide. Another study from Fischer G.S. et al. [71] also investigated the usefulness of MR-IOS in performing arthrography in porcine and humans' shoulder and hip joints. In their trial, every needle insertion was successful in the first attempt. A similar construct was built and evaluated by Fichtinger G. et al. [72], where instead of MRI, CT data was used as

the baseline. They successfully performed spinal nerve blocks, facet joint injections, shoulder and hip arthrographies, and needle insertions for musculoskeletal biopsy in cadaver experiments. Limitations included complex calibration phase, interference of the room light with the overlay, and the parallax effect.

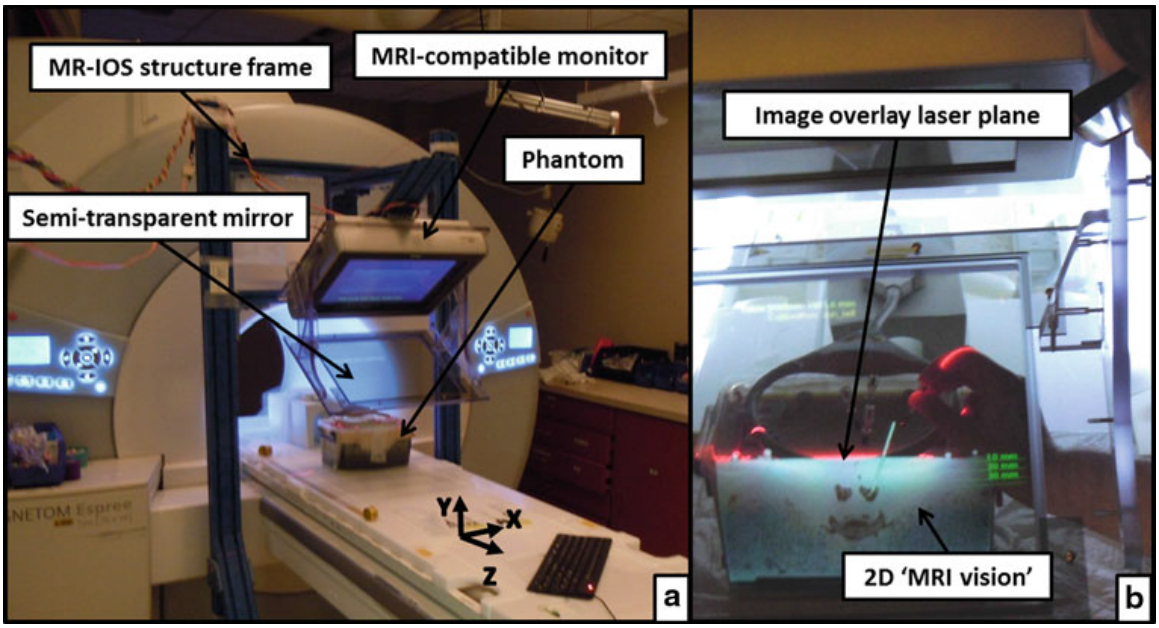

Figure 1.9: The system components of MR-IOS is shown in **(a)**. Using this system, the surgeon can look through the semi-transparent mirror that is augmented with the insertion path **(b)**. Image courtesy of the International Journal of Computer Assisted Radiology and Surgery, *MR image overlay guidance: system evaluation for preclinical use*, Paweena U-Thainual, Jan Fritz, Choladawan Moonjaita et al., Copyright 2012.

Londei R. et al. [73] used the camera augmented C-arm proposed by Navab et al. [67] and performed studies on distal locking of intramedullary (IM) nails, a procedure which requires a large number of C-arm fluoroscopic images. In this work, *down-the-beam* view of the IM nail was achieved by first acquiring an X-ray image of the nail. They registered the information of this image with the 3D CAD model of

the IM nail to estimate the C-arm pose and subsequently predict a second view that produces the "perfect circle" view of the holes on the IM nail (Fig. 1.10). Therefore, the authors were able to conduct IM nailing and distal locking with only two X-ray images, on average. To track the drill, a cannula with chained cross-ratios was placed on the surgical drill. The markers on this cannula were tracked by a video camera on the C-arm, and the position of the drill tip was estimated with respect to the patient.

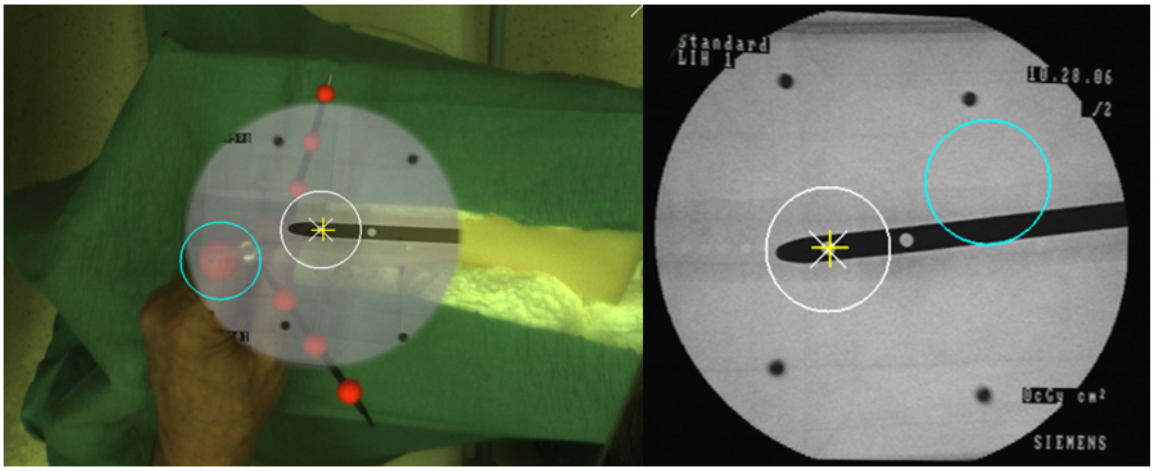

Figure 1.10: Visualization of perfect circles for distal locking

Wang et al. [74] performed AR-based navigation for percutaneous placement of sacroiliac screws in six cadavers. Pre-operatively, the authors acquired CT scans of each pelvis and segmented the bone and vessels from other tissue in the CT data. Given the CT images, ideal entry points and trajectories of the percutaneous screws were calculated. This system was materialized by registering the pre-operative data and the planned trajectories to the cadaver and projecting the surgical plan as a cylinder onto an HMD display. Major limitations of this system were complex setup due

to the use of external navigation system, out-side-in tracking of different components, and the bulky and tethered setup.

Another study by Ogawa H. et al. [75] investigated cup placement in total hip arthroplasty measured by an AR approach. In 56 total hip arthroplasties, the orientation of the acetabular cup (anteversion and inclination) were measured either using an AR smartphone or a goniometer. Three months post-operatively, cup anteversion and inclination were measured in CT.

Liu H. et al. [76] performed a proof of concept study about AR-based navigation in hip resurfacing. Using a robotic system [77] and the Microsoft HoloLens (Microsoft, Albuquerque, New Mexico, USA) they suggested a system to support the drilling of a hole along the axis of a femoral neck. Comparing the post-interventional drill orientation with the pre-interventional plan in a phantom study yielded a mean error of approximately 2 mm and 2°.

1.3.2 Osteotomies

In high tibial osteotomy, knowledge about the mechanical axis is essential. Based on the work of Wang L. et al. [78], Fallavollita P. et al. [79] published a study on intra-operative assessment of the mechanical axis of the lower limb using AR navigation. Using the conventional fluoroscopy-based approach, a large number of X-ray images were required. Using the RGB camera setup on the C-arm, with only three X-ray images - one from the hip, one from the knee and one from the ankle - a parallax-

free panoramic image of the lower limb was created. Their innovative parallax-free solution required a joint rotation of the C-arm scanner around the origin of the X-ray source as well as the translation of the surgical bed. These two motions were applied such that the overall translation around the X-ray origin vanished, and therefore, the parallax effect was eliminated. Using this intra-operatively acquired non-overlapping panoramic image, the mechanical axis and the amount of misalignment that required correction for high tibial osteotomy were measured. Procedures were carried out in 25 human cadaver legs. To evaluate the usefulness of the proposed AR solution, the mechanical axis, and its respective deviation were also measured in the ground-truth CT scan. The AR system proved to be an accurate and low radiation technique.

1.3.3 Tumor Surgery

Information about the 3D expansion of the tumor is crucial for a proper resection. Cho H. S. et al. addressed this problem by incorporating AR support [80]. The first study investigated tumor resection in pig femurs. Multiple visual markers were attached to the subject which was tracked by the integrated camera on a tablet PC. In this overlay, a cylindrical virtual template represented the tumor. The virtual implant was then superimposed on the patient using the tablet PC.

1.3.4 Trauma

Shen F. et al. [81] developed an AR implant design system for pre-operative creation of osteosynthesis plates in unilateral pelvic and acetabular fractures. The proposed solution comprised two sub-systems: *i)* a virtual fracture reduction system in which a repaired model of the fractured pelvis was constructed and an ideal curve indicating the implant model was identified, and *ii)* an AR templating environment to manipulate and bend the implant according to the planned trajectory. This AR system consisted of an external monitor and an HD webcam. The suggested technology enabled the surgeon to visualize the physical implant as well as the augmentation of the virtual model simultaneously and use the AR environment to create the desired implant model. The reduction was performed on six fractured pelves. Pre-operative CT was acquired, and fracture reduction was performed on the computer. For each case, virtual osteosynthesis plates were first drawn in ideal locations. This information was used together with the aid of the AR system to bend the osteosynthesis plates. This allowed pre-bent osteosynthesis plates for ideal fracture reduction. The authors claimed that the intra-operative implant bending could be eliminated using their approach; therefore surgical time and invasiveness could be minimized.

AR fluoroscopy simulation for guide-wire insertion in dynamic hip screws were investigated in [82]. This system included cameras that were orthogonally viewing the operative site and tracking the marked guide-wires. Post-interventionally, the tip-apex distance (TAD) between the guide-wire and the femoral head was measured

with a mean square error of 4.2 mm.

Another study by Hiranaka T. et al. [83] also evaluated guide-wire insertion into five artificial femoral heads by the use of AR navigation. They made use of the PicoLinker wearable glasses (Westunits Co., Ltd., Osaka, Japan) that was connected to the fluoroscopic monitor such that the surgeon was able to observe the fluoroscopic stream through the PicoLinker glasses.

1.3.5 Surgical Training and Education

Yeo C. T. et al. [84] suggested the employment of AR for spine surgery training. The AR simulation display was designed using a semi-reflective glass, where a slice of the CT, as well as the trajectory of the needle, were augmented onto the trainee's view. An electromagnetic tracker was used to estimate needle pose in relation to CT. The trajectory was then projected onto the AR display using a laser-guided system. This study compared two groups: *i)* the first group received AR supported training, and *ii)* the control group received training for conventional freehand facet joint injections. Later, both groups performed injections in a phantom with the conventional freehand technique. The AR-trained group achieved higher rates of successful placement of injections with less tissue trauma compared to the control group.

AR-based surgical training systems were also considered for tele-guided shoulder arthroplasties [85]. This system, so-called virtual interactive presence (VIP), allowed a physically absent surgeon to be virtually present. One video camera in the OR and

one at the remote station were first calibrated. Thereafter, both surgeons were able to observe the surgical site concurrently with a common task field. By using Google Glass (Google Inc., Mountain View, California, USA), the second surgeon that was physically absent was able to join their collaborative virtual experience. The remote surgeon could follow the entire procedure and provide real-time feedback. The authors reported several technical issues they encountered, such as battery life of the HMD, poor video quality, limited field of view, and video mismatch due to delay. In another report, Ponce B. et al. [86] used the VIP technology for rotator cuff and shoulder instability interventions. Six different resident surgeons performed the procedures, and one attending surgeon that was physically located in an adjoining dictation room proctored the procedures using the VIP technology. The attending surgeon was able to see the arthroscopic image at any time and was able to guide the residents.

Condino S. et al. [87] performed a study on how to build a patient-specific hybrid simulator for orthopaedic open surgery. By using a Microsoft HoloLens HMD and patient-specific 3D models, five subjects performed hip arthroplasty. Using their simulator, they reported that the perceived positioning accuracy matched the requirements, and the overall workload was low.

1.4 Problem Definition and Thesis Overview

We investigated several applications that AR, artificial intelligence, and robotics create value in computer-assisted interventions by increasing accuracy and safety, reducing collateral damage, and enhancing the surgeon's and the patient's experience. Fig. 1.11 shows a schematics overview of this thesis with four main components: *i)* enhancing the interventional image acquisition and reconstruction, *ii)* applying medical image computing algorithms using computer vision and machine learning techniques, *iii)* visualizing and interacting with intra-operative data using AR, and *iv)* simplifying the robotic workflows in minimally-invasive procedures.

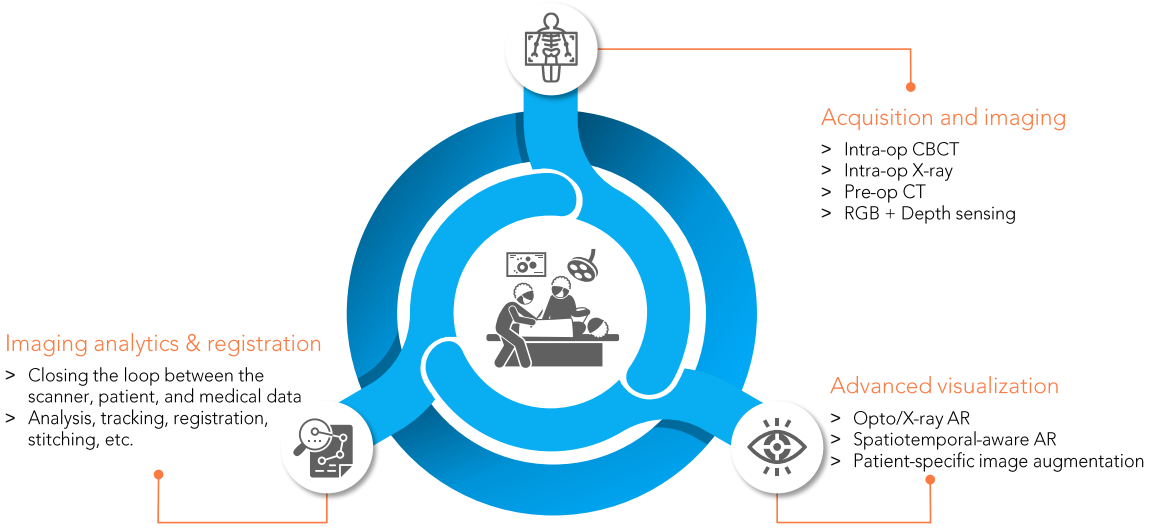

Figure 1.11: The thesis presents the applications of AR and artificial intelligence in a broad spectrum of problems in interventional medicine, namely for imaging and acquisition, image computing and registration, and data visualization.

Chapters 2 and 3 of this dissertation present the methodology for combining optics

and X-ray imaging, and explain its application in improving the quality of 3D intra-operative imaging by combining RGBD and C-arm imaging. Chapter 2 presents the design of the imaging systems that integrate optical sensing into a commonly used mobile C-arm. The chapter presents two distinct calibration mechanisms with vision-based tracking capabilities that impact different steps in the surgical workflow and enable immersive and non-immersive visualization experiences. In chapter 3, different imaging modalities are combined to enhance the intra-operative imaging. We discuss RGBD-supported tomographic reconstruction and present the methodology to correct for undesired rigid motion during CBCT acquisition.

Chapters 4 and 5 present novel approaches to use the optical information from the operative field to address the challenges in 2D/3D and 3D/3D registration for various medical image computing tasks. To bring image-based navigation to surgery, chapter 4 offers automatic initialization of the standard 2D/3D registration between C-arm X-ray images and the pre-operative CT by incorporating real-time inside-out tracking of the C-arm scanner using vision-based tracking. Chapter 5 presents automatic 3D/3D registration and stitching between non-overlapping CBCT acquisitions using visual tracking information from the integrated RGBD sensor. In chapter 6, the methodology for 2D/2D parallax-free stitching and alignment between the X-ray images of the femur anatomy is presented.

In the remainder chapters of this dissertation, various AR systems are introduced and validated for image-guided and minimally-invasive therapies. In chapter 7,

display-based AR solutions are presented for various orthopedic tasks. These systems are realized by fusing clouds of points observed and reconstructed from an integrated RGBD camera and the X-ray/CBCT intra-operative data. Chapter 8 presents the full design of interactive flying frustums for immersive AR visualization and reveals how exemplary workflows are redefined by taking full advantage of HMD when entirely co-registered with the imaging system at all times. The awareness of the system from the geometric and physical characteristics of X-ray imaging allows the exploration of different human-machine interfaces. Chapter 9 introduces a novel methodology to detect imperfect bilateral symmetry in CT of human anatomy for patient-specific augmentations and visualizations. In this chapter, the structurally symmetric nature of the pelvic bone is explored and used to provide interventional image augmentation for unilateral fractures in patients with traumatic injuries. This solution's mathematical basis is on the incorporation of attributes and characteristics that satisfy the properties of intrinsic and extrinsic symmetry and are robust to outliers.

Chapter 10 investigates the intersection of AR and robotics, and presents unique concepts to simplify robotic workflows. The staff equipped with HMD aligns the robot with its planned virtual counterpart. In this user-centric setting, the main challenge is the perspective ambiguities hindering such a collaborative robotic solution. To overcome this challenge, we introduce a novel registration concept for intuitive alignment of AR content to its physical counterpart by providing a multi-view AR experience via reflective-AR displays that simultaneously show the augmentations from multiple

viewpoints.

Chapter 11 provides a summary of contributions, considerations for clinical deployments, outlook, and future works.

1.5 Acknowledgments

Sincere thanks to my clinical partners, Drs. Greg Osgood, Herpal Khanuja, Mazda Farshad, Amol Narang, Alex Johnson, Clayton Alexander, Alex Leob, Nima Befrui, Sebastian Adress, and Marius Fischer for supporting this research throughout all steps, from problem definition to modeling and validation. Special thanks to Dr. Lukas Jud for his support during the systematic review of the state-of-the-art in AR.

Part I

Multi-Modal Imaging and Acquisition

This section of the dissertation presents the material and methods for merging optics and X-ray imaging to achieve a multi-modal imaging platform that benefits the joint use of information from the patient site as well as the knowledge from the anatomy-level X-ray imaging.

We first present multiple approaches to calibrate the C-arm scanner to a rigidly attached camera on the C-arm, or to a moving subject with an augmented reality head-mounted display (chapter 2). Calibration of the RGBD and C-arm imaging devices is performed in two steps: (i) calibration of the RGBD sensor and the X-ray source using a multimodal checkerboard pattern, and (ii) calibration of the RGBD surface reconstruction to the CBCT volume. The co-localization between a moving augmented user and the imaging observer (C-arm scanner) is achieved via a joint localization using the operating room environment.

In chapter 3 of this dissertation, this multi-modal imaging system is exploited to improve interventional tomographic reconstruction. The patient surface is acquired during the CBCT scan and then used as prior information for the reconstruction using Maximum-Likelihood Expectation-Maximization. An RGBD-based simultaneous localization and mapping method is utilized to estimate the rigid patient movement during scanning. We focus primarily on rigid and accidental movements, and not the internal motion caused by respiration.

The contributions in this part have been primarily published in the following manuscripts:

1. Fotouhi, Javad, Mathias Unberath, Tianyu Song, Jonas Hajek, Sing Chun Lee, Bastian Bier, Andreas Maier, Greg Osgood, Mehran Armand, and Nassir Navab. "Co-localized augmented human and X-ray observers in collaborative surgical ecosystem." International journal of computer assisted radiology and surgery 14, no. 9 (2019): 1553-1563.

2. Fotouhi, Javad, Bernhard Fuerst, Wolfgang Wein, and Nassir Navab. "Can real-time RGBD enhance intraoperative cone-beam CT?." International Journal of Computer Assisted Radiology and Surgery 12, no. 7 (2017): 1211-1219.

Chapter 2

Imaging System and Calibration

2.1 Introduction

A significant challenge in C-arm imaging is the lack of direct association between the patient images and the patient's anatomy. To disambiguate this relation, and visualize and process data more optimally, we investigated the benefits of combining RGBD information and C-arm imaging. Before benefiting from a camera-augmented scanner, the first step is to calibrate such a system with appropriate routines that are practical and fit within the pre- and intra-operative workflows. The main technical contributions that are presented in the following chapter are *i)* calibration between the RGBD camera and the C-arm fluoroscopy system and *ii)* real-time co-localization between augmented users and imaging observer systems such as the X-ray imaging device. All the below approaches benefit from offline calibration steps that would not

interfere with the surgical workflow.

2.2 Integration of RGBD Sensing into C-arm Imaging

To include the visual information observed by the RGBD camera into CBCT reconstruction *(i)* the RGBD sensor extrinsic parameters relative to the X-ray source are estimated by tracking the rigid patient movement, and *(ii)* CBCT coordinate system is calibrated with respect to the patient surface coordinate system to allow the masked CBCT reconstruction.

2.2.1 RGBD-to-X-ray Calibration

Co-calibration of the C-arm X-ray source and the RGBD camera is a necessary component for the movement correction during scans. This calibration allows to update the pose of each projection relative to the patient. To this end, we designed a custom-made radiopaque checkerboard pattern to perform stereo calibration of the X-ray and the RGBD imaging devices [88, 89]. This calibration target comprises a standard black-and-white pattern, where every black square is backed with a same size thin metal sheet. Hence, the checkerboard pattern is visible both in the RGB, Infra-Red (IR), and X-ray images. In Fig. 2.1 the checkerboard pattern is shown in

the RGB and X-ray images.

In a C-arm guided intervention, the X-ray source is commonly under the surgical bed, and interventional X-ray images are acquired from bottom to top. In order to simplify the mapping between the displayed X-ray image and the patient laying on the bed, the images are represented in left-hand coordinate system which simulates an image taken from top to bottom (surgeon's view). Hence, the stereo calibration of the X-ray and RGBD camera using checkerboard images requires an additional pre-processing step to transform images to a right-hand coordinate system. The stereo relation of the two imaging devices is then defined as:

$$^{\mathrm{RGBD}}\mathrm{T}_{\mathrm{X}} = ^{\mathrm{RGBD}}\mathrm{T}_{\mathrm{CB}} \cdot \mathrm{T}_{l-r} \cdot ^{\mathrm{CB}}\mathrm{T}_{\mathrm{X}}, \tag{2.1}$$

where X, RGBD, and CB are X-ray source, RGBD camera, and checkerboard coordinate frames, respectively. T_{l-r} is the transformation from left-hand to right-hand coordinate system.

2.2.2 Surface-to-CBCT Calibration

To use the reconstructed RGBD surface information as a prior during the CBCT reconstruction, we directly obtain the relationship of the RGBD sensor and CBCT reconstruction. This reduces the length of the chain of transformations compared to using the RGBD-to-X-ray calibration for this purpose. To obtain the transformation

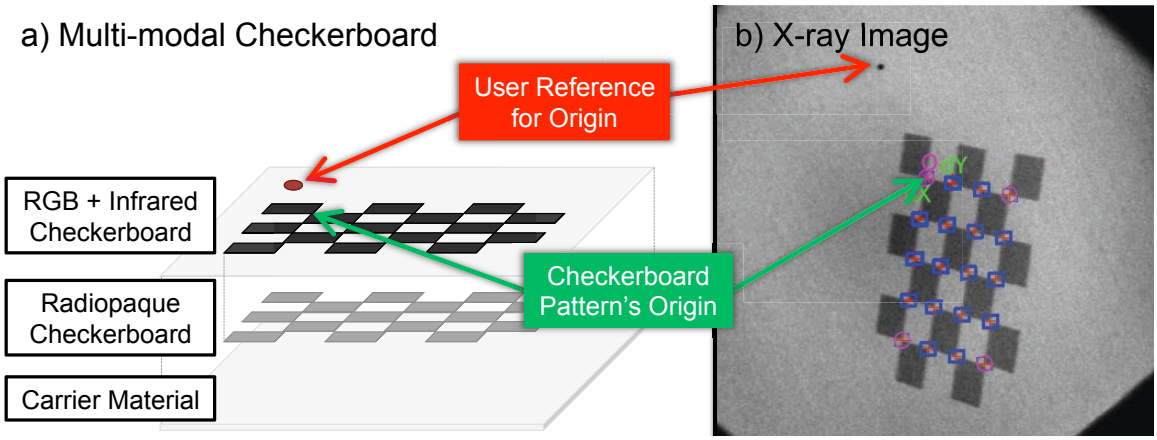

Figure 2.1: Checkerboard is designed to be fully visible in RGB, depth, and X-ray images.

between RGBD and CBCT coordinate systems, we use a calibration phantom which comprises of three hollow radiolucent cylinders of different sizes positioned at different depth and orientation. During the CBCT scan the calibration phantom is placed near the C-arm iso-center, and the surface is simultaneously acquired and reconstructed using depth information from the RGBD camera [90]. Automatic registration of the CBCT coordinate frame with the surface data is performed by first extracting point clouds from each data set, and then initializing the spatial relation of the point clouds using Fast Point Feature Histogram (FPFH) descriptors and SAmple Consensus Initial Alignment (SAC-IA) [91]. The final transformation is obtained after an Iterative Closest Point (ICP) refinement [92–94]. This step allows us to map patient 3D surface to the CBCT reconstructed volume.

In clinical routine, the angular-dependent distortion correction needs to be performed when setting up the C-arm or at predefined service intervals. This distortion

correction may also trigger the need to update the RGBD sensor to C-arm calibration, which could be automatically performed at the same time.

2.2.3 Results

Registration of patient surface and CBCT volume using FPFH and ICP has a mean Target Registration Error (mTRE) of 2.58 mm [92]. The mTRE is measured as the average of absolute Euclidean distances between eight landmarks visible in both modalities. This error can improve with using more advanced RGBD sensors that provide a more reliable cloud of points.

The calibration of the RGBD/X-ray system is achieved using a multi-modal checkerboard (see Fig. 2.1), which is observed at multiple poses using the RGB camera, depth sensor, and the X-ray system. We use a 5×6 checkerboard where each square has a dimension of 12.655 mm. The distance between the black-and-white and the radiopaque metal pattern is negligible. Thus, for the purpose of stereo calibration we assume all three cameras (RGB, infrared, and X-ray cameras) observe the same pattern. 72 image triplets (RGB, infrared, and X-ray images) were recorded for the stereo calibration. Images with high reprojection errors or significant motion blurring artifacts were discarded from this list for a more accurate stereo calibration. The stereo calibration between the X-ray source and the RGB camera was eventually performed using 42 image pairs with the overall mean error of 0.86 pixels. The RGB and infrared cameras were calibrated using 59 image pairs, and an overall reprojection

error of 0.17 pixels was achieved.

2.3 Co-calibrated and Co-localized Augmented User and Imaging Observer

The goal of this offline calibration procedure is to establish a constant pose relation between an inside-out visual tracker attached to and the X-ray source embedded in the gantry of a C-arm machine. The rigid transformation associated with this pose relation is shown as ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ in Fig. 2.2. In Fig. 2.2, ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{W}}$ and ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{W}}$ show the relative poses of the tracker and the augmented surgeon in the world coordinate frame, receptively. Finally, ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ denotes the transformation between the C-arm source and the volume. Other transformations shown with dashed lines are derived.

The visual tracking is intended to localize the C-arm source in the operating theatre, and share the information with an HMD-equipped surgeon (Sec. 2.3.1). Therefore, the tracker was mounted on the gantry such that it observes the static structures in the operating theatre such as walls, corners, ceiling, etc. As a consequence, the fields of view of the tracker and the X-ray imaging system do not exhibit overlap, preventing direct multi-sensor calibration via a common calibration phantom [95–97].

Since both sensors are rigidly connected by the C-arm gantry, the calibration problem can be formulated as a hand-eye system of equations. This system of equations can be solved if simultaneous tracking data are acquired from both the X-ray source

and the tracker.

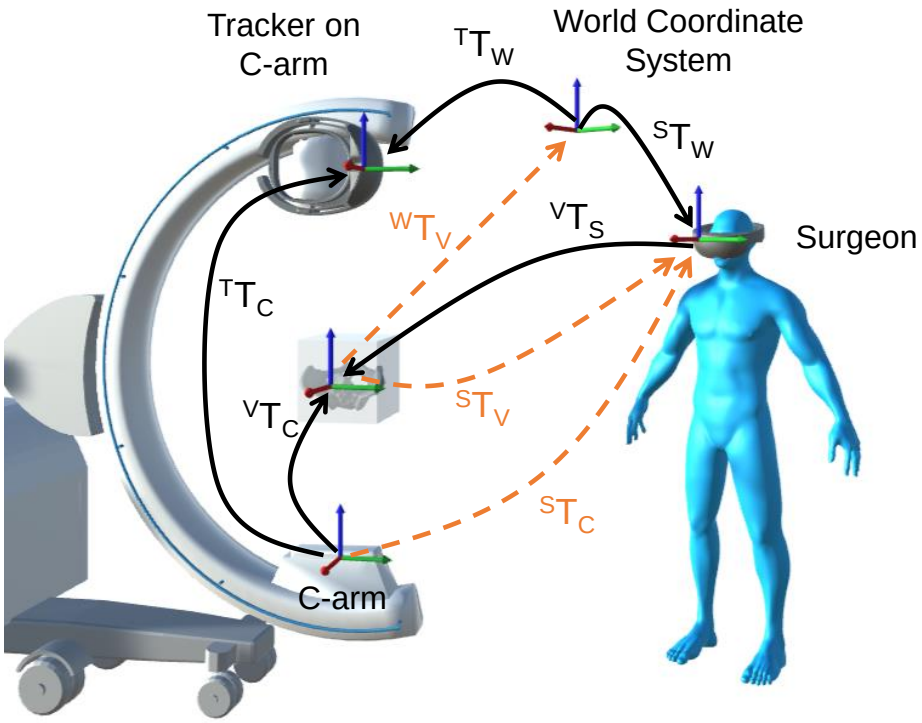

Figure 2.2: The chain of transformations involved in jointly localizing the augmented surgeon and the imaging system in a unified coordinate frame. Note that the transformations shown with solid arrows are acquired, while the transformations with dashed arrows are derived. The transformations ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{W}}$ and ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{W}}$ are estimated using the RGB-based environment tracking sensors integrated in the HMD. ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ is computed by C-arm internal calibration, and ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ is computed using a hand-eye calibration approach.

We estimate the rigid transform ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ using a hand-eye formulation as:

$$\mathbf{A}(t_i)\,{}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}} = {}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}\mathbf{B}(t_i), \tag{2.2}$$

where $\mathbf{A}(t_i)$ and $\mathbf{B}(t_i)$ are the relative poses of the visual-tracker and the X-ray source with respect to their pose at time t_0. By stacking multiple pose data along the rows

of $\mathbf{A}$ and $\mathbf{B}$, we can re-write Eq. 2.2 as:

$$\mathbf{A}\,\mathbf{X} = \mathbf{X}\mathbf{B}. \tag{2.3}$$

where the objective is to estimate $\mathbf{X} := {}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}} = \begin{bmatrix} R_x & \mathbf{t}_x \\ \mathbf{0}^\top & 1 \end{bmatrix}$.

The pose pairs ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}(t_i)$ and ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_i)$ in Fig. 2.2 that are used to construct the rows of $\mathbf{A}$ and $\mathbf{B}$ are estimated via an external infrared-based navigation system and visual simultaneous localization and mapping (SLAM) system, respectively. SLAM successively estimates the transformation ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_i)$ between the tracker T and the world coordinate system W, thereby incrementally constructing a 3D representation of the environment, i. e. the anchor [98]. This is achieved by solving the following equation at every time point t_i:

$${}^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_i) = \underset{{}^{\mathrm{W}}\hat{\mathbf{T}}_{\mathrm{T}}}{\arg\min}\ \mathrm{d}\Big(\mathbf{f}_{\mathrm{W}}\left(\mathbf{P}\,{}^{\mathrm{W}}\hat{\mathbf{T}}_{\mathrm{T}}(t_i)\mathbf{x}_{\mathrm{T}}(t_i)\right), \mathbf{f}_{\mathrm{T}}(t)\Big), \tag{2.4}$$

where $\mathbf{f}_{\mathrm{T}}(t_i)$ are visual features detected at frame t_i, $\mathbf{x}_{\mathrm{T}}(t_i)$ are the 3D locations of these feature estimates, $\mathbf{P}$ is the projection operator, and $d(\cdot,\cdot)$ is the feature similarity distance to be optimized. In our prototype, we rely on a proprietary implementation of the SLAM concept that is provided by the Microsoft HoloLens SDK.

In room scale surgical AR settings that are focused in this chapter, the distance between different sensors are typically large. In such cases, the calibration errors prop-

agate greatly and negatively impact the final pose estimation. Thus, this necessitates an accurate calibration between the involved components. With this knowledge, to compute the poses that construct $\mathbf{B}$, instead of only using the projection transformations from the internal calibration of the C-arm, we acquired pose data from various views that the C-arm underwent along all axis.

Finally, to solve the system of equations in Eq. 2.3, As detailed in [99], we first estimate the rotation component via a unit quaternion representation as follows:

$$Q_a \, Q_x = Q_x \, Q_b. \tag{2.5}$$

Since Q_a and Q_b are known, Eq. 2.5 can be linearized in the form of $\mathbf{M}Q_x = 0$, and solved using least-squares minimization. The rotation parameters Q_x are converted from a unit quaternion form to a rotation matrix form R_x, and are substituted in Eq. 2.3 to estimate the translation parameters $\mathbf{t}_x$.

2.3.1 Inside-out Tracking and Localization

To close the calibration loop between the tracker on the C-arm and the augmented surgeon, we introduce a world coordinate frame $\mathbf{W}$ as shown in Fig. 2.2. The world frame is a 3D representation of the operating theatre (in the Microsoft Hololens community often referred to as anchor) which is conveniently shared among multiple AR-enabled devices as the common point of reference. When the anchor is synchro-

nized, all master and slave devices keep track of its position and the coordinate system attached to it.

It is important to ensure clear line-of-sight during the short time period that reference anchors are exchanged, i. e. uploading anchor by the master system (imaging system) and downloading by the slave system (augmented user). However, after the co-registration to the environment is established, it is no longer essential for the sensors to share the same view. If the view of the master or slave systems are temporarily blocked and the co-localization is interrupted, the shared tracking can be restored by simply re-observing the area in the OR where the anchors were initially constructed from.

The final pose ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{W}}$ of the virtual volume is estimated with respect to the world anchor. This transformation is used to dynamically recover the relative pose of the virtual 3D volume in relation to the augmented surgeon as:

$$ {}^{\mathrm{S}}\mathbf{T}_{\mathrm{V}}(t_i) = {}^{\mathrm{S}}\mathbf{T}_{\mathrm{W}} \underbrace{\left({}^{\mathrm{T}}\mathbf{T}_{\mathrm{W}}^{-1}(t_0)\ {}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}(t_0)\right)\ {}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}^{-1}}_{{}^{\mathrm{W}}\mathbf{T}_{\mathrm{V}}}, \tag{2.6} $$

The final step in establishing a closed loop for an AR experience is to estimate ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$, describing the transformation from 3D virtual pre-operative or intra-operative patient volumes to an intra-operatively acquired X-ray image. This registration is crucial as the offline system calibration only recovers the transformation between the visual tracker and the X-ray source. Hence, the following step is required to find

relation between the X-ray source and the origin of the 3D volume.

If a CBCT capable C-arm is available, ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ is defined as one of the pre-calibrated C-arm poses on the source trajectory, e. g. for convenience the anterior-posterior pose. On the other hand, to incorporate pre-operative 3D patient data, the transformation ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ can be estimated by acquiring X-ray images from patient at the current C-arm angle and registering them to the pre-operative CT data using an image-based 2D/3D registration [95, 96, 100–102]. The registration pipeline typically involves creation of many 2D digitally reconstructed radiographs (DRRs). DRRs are generated iteratively from various view points onto the patient CT ($I_{\mathrm{D}}({}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}})$) and are compared against a C-arm X-ray image (I_{X}). DRR generation continues until an intensity- or gradient-based similarity function $S(.)$ yields maximal agreement between the X-ray and DRR generated from the current view point:

$$\underset{{}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}}{\arg\max}\ S\left(I_{\mathrm{X}}, I_{\mathrm{D}}({}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}})\right) \tag{2.7}$$

Once ${}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ is known, the volumetric images are registered to the operating theatre via ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{V}} = {}^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_0)\,{}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}\,{}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}^{-1}(t_0)$, where t_0 denotes the time of calibration.

2.3.2 System Setup

The CBCT scan is acquired from a motorized ARCADIS Orbic 3D C-arm (Siemens Healthineers, Forchheim, Germany). This C-arm executes orbital rotations of 190°

for acquisition of 100 projection images to enable tomographic reconstruction.

The surgeon's augmented view is enabled via commercially available Microsoft Hololens optical see-through HMD (Microsoft, Redmond, WA). For the proof-of-principle implementation of this system, we affixed another HoloLens HMD to the C-arm as the visual tracker. This enabled us to leverage the SLAM-based inside-out tracking technology of the HMD, and conveniently share joint anchors representing the 3D visual content of the operating theatre between the HMD on the C-arm and the augmented surgeon. The pose data between these HMDs were transmitted via wireless network, that was made possible by HoloToolkit using Unity 3D game engine.

To estimate the C-arm poses during hand-eye calibration, we used a Polaris Spectra (Northern Digital, Waterloo, ON), an external optical navigation system. This external tracker provides large measurement volumes, and allows reliable estimation of relative C-arm rotation and translation at different configurations. It is important to note that the external tracker is only used during the offline calibration phase, and is not used intra-operatively. The system setup is illustrated in Figs. 2.3 and 2.4.

2.3.3 Experiments and Results

In the following we first evaluate the co-calibration between the visual-tracker and the C-arm source and analyze the calibration accuracy as the number of input pose data increases. Next, the overall user-in-the-loop error is evaluated using a multi-level hybrid phantom.

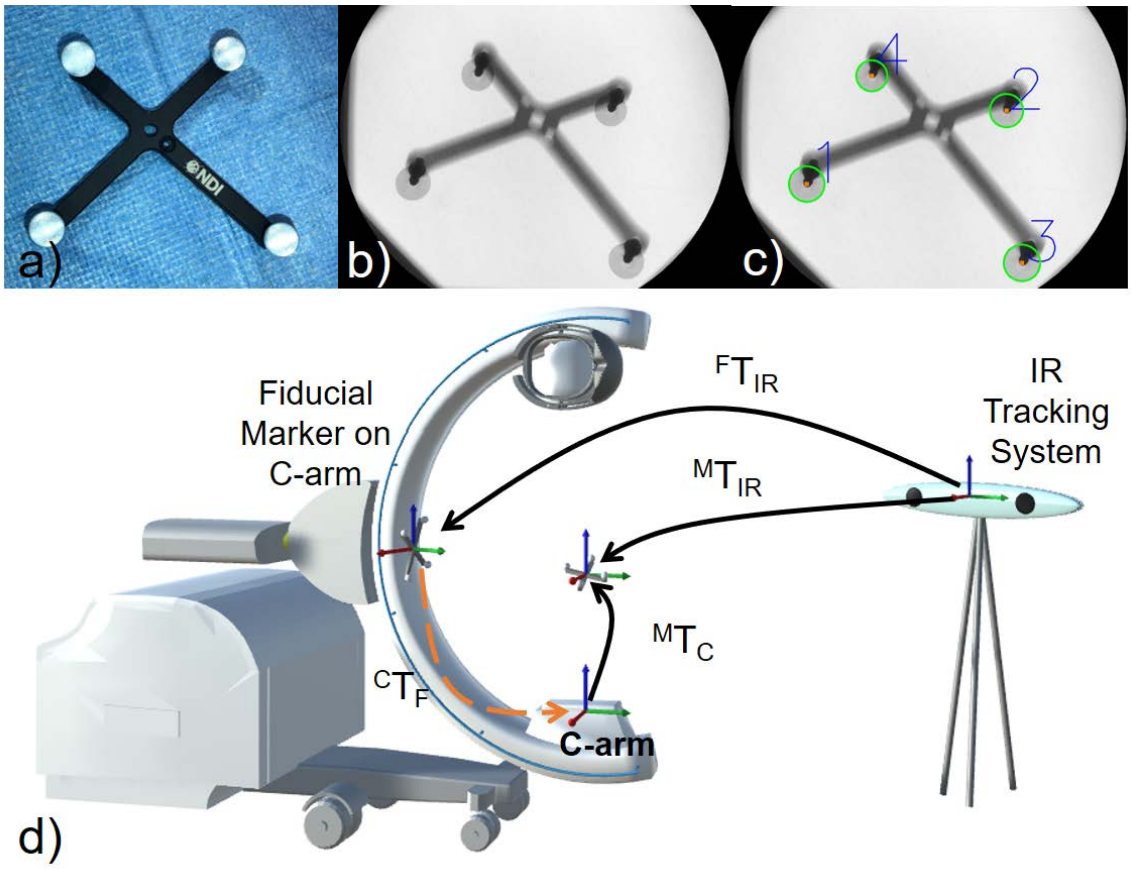

Figure 2.3: An infrared reflective marker is attached to the gantry and calibrated to the X-ray source using a second marker that is imaged by the navigation system and the C-arm simultaneously. This enables calibration of the optical tracker to the C-arm source. a-c) shows the marker used for calibration and the corresponding X-ray image with detected centroids of the spheres.

To optimally estimate and evaluate the hand-eye calibration between the X-ray source and the tracker on the C-arm, we obtained 120 pose pair data from both the external tracking of the C-arm as well as the inside-out SLAM-based tracking of the visual tracker. The calibration accuracy was evaluated for rotation $\overline{e_R}(N)$ and translation $\overline{e_p}(N)$ with respect to the number of pose data N. Alg. 1 demonstrates

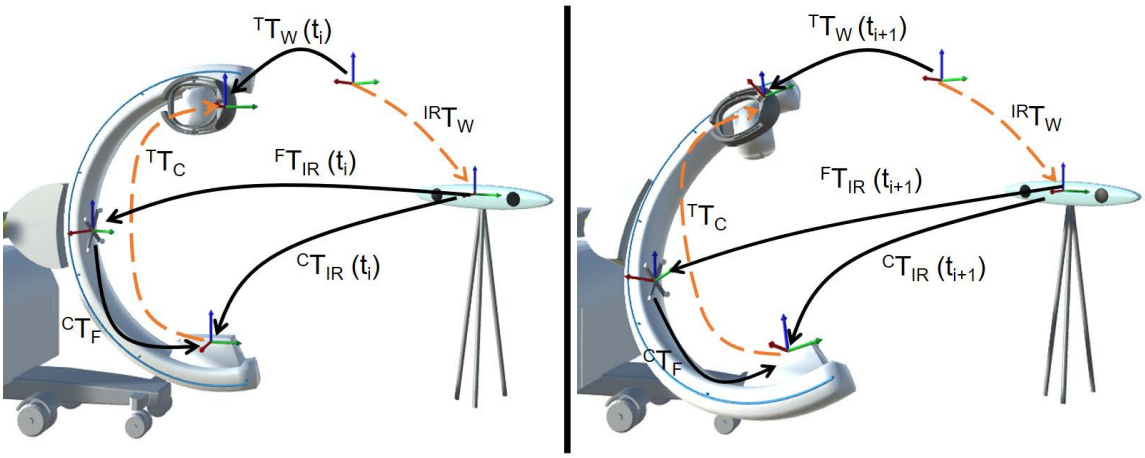

Figure 2.4: The C-arm gantry, and therefore, the tracker and the optical marker are moved and corresponding pose pairs in the respective frames of reference are collected that are then used for hand–eye calibration.

the procedure for calculating the average errors $\overline{e_R}(N)$ and $\overline{e_p}(N)$ [99]. The mean and standard deviation for $\overline{e_p}(N)$ and $\overline{e_R}(N)$ errors are shown in Figs. 2.5 and 2.6. Results in these two plots indicate that the improvement in accuracy when using more than 50 pose pairs are minimal, suggesting convergence.

Here we present the end-to-end error assessment of the proposed system considering augmented users in the loop. In the previous work, we reported a point-to-line averaged error of 11.46 mm for this system [103]. In this work, we evaluate a more geometrically relevant point-to-point distance. Fig. 2.7 illustrates these two error measures.

To calculate the Target Augmentation Error (TAE), we designed a phantom (Fig. 2.8) with $L = 4$ X-ray opaque landmarks at different (x, y, z) positions, such that all landmarks could be localized in a single CBCT scan. In a pre-clinical user study,

Algorithm 1 Assessment of hand-eye calibration

$p(.)$: translation component of a homogeneous rigid transformation
$tr(.)$: matrix trace
$\overline{(.)}$: expected value

procedure HAND-EYE ACCURACY ASSESSMENT (M pairs)
  for `N = 4 to M` **do** ▷ M = 120
    for `i = 1 to T` **do** ▷ T = 800
      `Randomly select N pairs from M`
      `Estimate hand-eye calibration` ${}^{\mathtt{C}}\mathbf{T}_{\mathtt{T}}$ `with the N pairs`
      for `j = 1 to N` **do**
        ${}^{\mathrm{IR}}\mathbf{T}_{\mathrm{W}}(j) = {}^{\mathrm{IR}}\mathbf{T}_{\mathrm{C}}(j)\ {}^{\mathrm{C}}\mathbf{T}_{\mathrm{T}}\ {}^{\mathrm{T}}\mathbf{T}_{\mathrm{W}}(j)$
      end for

      `Estimate mean transformation:` ${}^{\mathtt{IR}}\overline{\mathbf{T}}_{\mathtt{W}}$
      `Randomly select M pairs from M`
      for `k = 1 to M` **do**
        ${}^{\mathrm{C}}\mathbf{T}_{\mathrm{Wa}}(k) = {}^{\mathrm{C}}\mathbf{T}_{\mathrm{IR}}(k)\ {}^{\mathrm{IR}}\overline{\mathbf{T}}_{\mathrm{W}}$
        ${}^{\mathrm{C}}\mathbf{T}_{\mathrm{Wb}}(k) = {}^{\mathrm{C}}\mathbf{T}_{\mathrm{T}}\ {}^{\mathrm{T}}\mathbf{T}_{\mathrm{W}}(k)$
        $R = {}^{\mathrm{C}}\mathbf{R}_{\mathrm{Wa}}(k){}^{\mathrm{C}}\mathbf{R}^{-1}_{\mathrm{Wb}}(k)$
        $e_R(N,i,k) = \cos^{-1}(\frac{tr(R)-1}{2})$
        $e_p(N,i,k) = \left\| p({}^{\mathrm{C}}\mathbf{T}_{\mathrm{Wa}}(k) - {}^{\mathrm{C}}\mathbf{T}_{\mathrm{Wb}}(k)) \right\|_2$
      end for

      $\overline{e_R}(N,i) = \frac{1}{M}\sum_{k=1}^{M} e_R(N,i,k)$
      $\overline{e_p}(N,i) = \frac{1}{M}\sum_{k=1}^{M} e_p(N,i,k)$
    end for

    $\overline{e_R}(N) = \frac{1}{T}\sum_{i=1}^{T} e_R(N,i)$
    $\overline{e_p}(N) = \frac{1}{T}\sum_{i=1}^{T} e_p(N,i)$
  end for

end procedure

Figure 2.5: Translational error in mm units with respect to number of pose pairs. The shaded area represents standard deviation of the error.

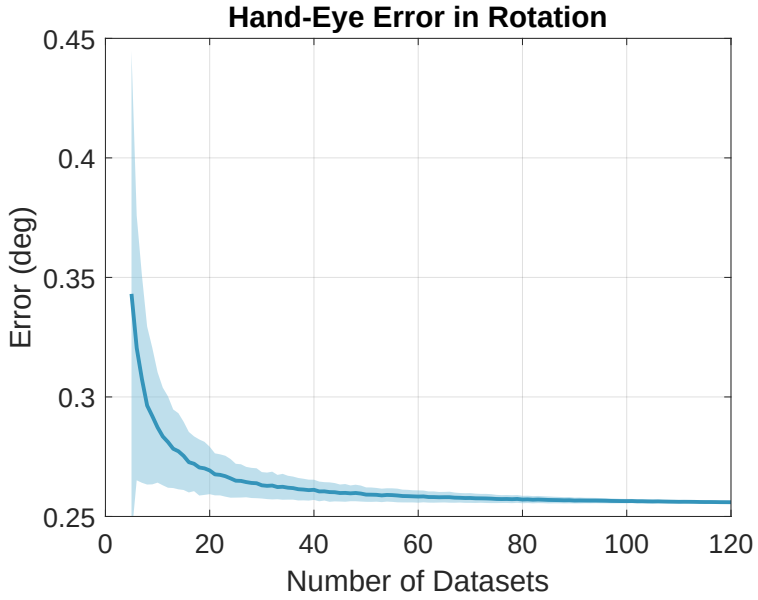

Figure 2.6: Rotational error in degree units with respect to number of pose pairs. The shaded area represents standard deviation of the error.

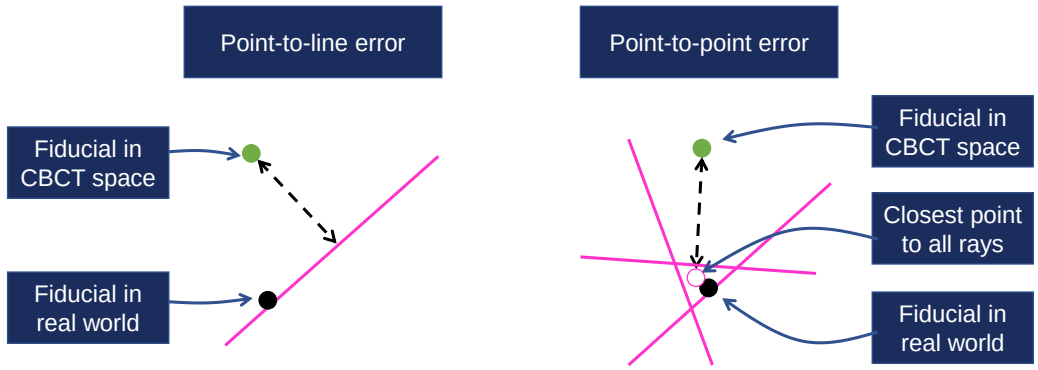

Figure 2.7: Illustration of point-to-line and point-to-point error measures. The rays shown in magenta are estimated by casting rays from the HMD to the 3D points selected by the user on the spatial map of the scene.

$U = 9$ participants were asked to select the landmarks at the surface of the phantom by confirming that the gaze cursor was on the target landmark. Confirmation was communicated either via interactive air-tap hand gesture or using voice commands. Per selection, a single ray was cast connecting the the HMD and the targeted point on the phantom. To localize the 3D position of the landmarks along the ray, users targeted the same point 4 times from various views around the phantom.

Each ray is defined via two elements: *1)* the 3D position of HMD p_i, and *2)* u_i as the unit direction vector normal to HMD. After all rays were identified, we estimated the closest point x_l^* to all the rays corresponding to each landmark l via a least-squares minimization strategy as follows:

$$x_l^* = \arg\min_{x \in \mathbb{R}^3} \sum_{i=1}^{N} \|(I - u_i u_i^\top)x - t_i\|^2 \text{ , where}$$
$$t_i = (I - u_i u_i^\top)p_i \text{ .} \tag{2.8}$$

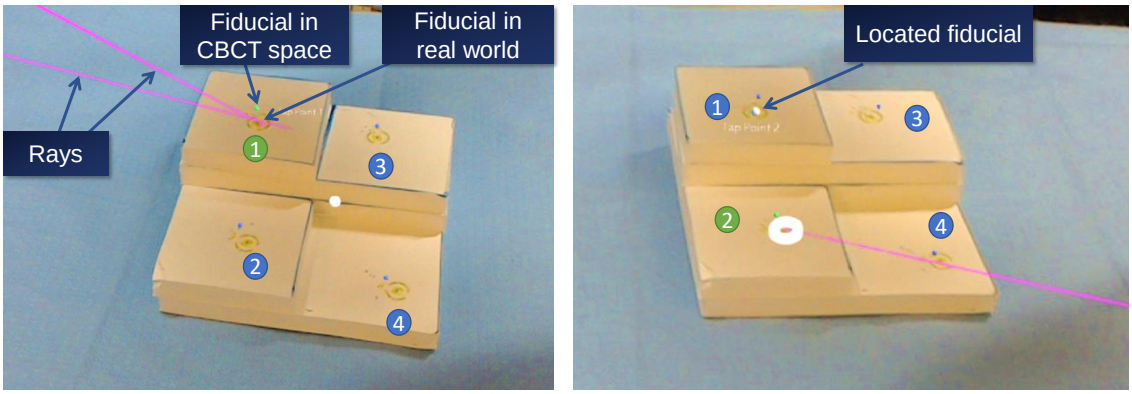

Figure 2.8: The multi-level hybrid phantom with X-ray visible landmarks along $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$ is used to evaluate the augmentation error.

Table 2.1: TAE measurements from 9 user experiments.

Target Augmentation Error (TAE)	Mean	Standard Deviation	Median	RMS
	10.8 mm	3.45 mm	11.2 mm	11.3 mm

Finally, the average TAE error was calculated as the average distance between the corresponding landmarks selected by the user, and the landmarks identified in the CBCT scan as follows:

$$\overline{\mathrm{TAE}} = \frac{1}{L \times U} \sum_{l=1}^{l=L\times U} \left\| x_l^* - {}^{\mathrm{S}}\mathbf{T}_{\mathrm{C}}\ {}^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}^{-1}\ x_l^{\mathrm{V}} \right\|_2, \tag{2.9}$$

where x_l^{V} corresponds to the l-th landmark in the CBCT volume coordinate frame. TAE measurements are presented in Table 2.1 and Fig. 2.9.

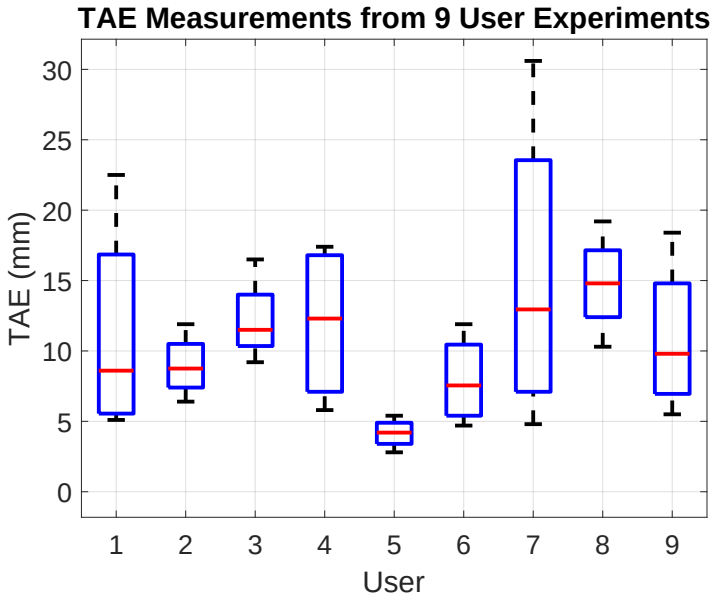

Figure 2.9: TAE plot for $U = 9$ users, with red markers defining the median of the measurements for each user.

2.4 Discussion and Conclusion

In this chapter we introduced several multi-modal systems relevant to different fields of image-guided surgery. These flexible solutions provide means to intra-operatively fuse 3D patient data for surgical AR and beyond. The methodologies suggested in this chapter are all fiducial-free alternatives to complex calibration and registration setups that were traditionally employed for image-to-patient registration. The applications of these systems may extend to different realms of surgery, e.g. fracture management surgeries in orthopedic trauma and neuro-spinal surgery. These systems are not intended to be surgical trackers, instead, are spatially-aware platforms that assist surgeons to better understand the geometric relations by using spatially registered data.

While vision-based tracking does not yet provide accuracy en a par with fiducial-based approaches, we envision that our systems for *in situ* visualization will benefit residents in training that observe surgery to fully understand the actions of the lead surgeon with respect to the deep-seated anatomical targets. These applications in addition to simple tasks such as optimal positioning of C-arm systems, do not require the accuracy needed for surgical navigation and, therefore, could be the first target for OST-HMD visualization in surgery. These proposed systems are low-cost compared to surgical navigation systems both in terms of hardware as well as the cost due to additional OR time that is required for pre-calibration of the navigational setup.

The accuracy for co-calibrating the tracker on the C-arm and the X-ray source were 5.7 mm and 0.26° in translation and rotation, respectively. Considering lower angular errors and high translational errors, this solution may become suited for surgical procedures during which orientational alignment is the key measure of success. One such procedure in orthopedic surgery is acetabular component placement in total hip arthroplasty using the direct anterior approach, where translation alignment of the implant is easily achieved by pivoting the cup inside the hip joints relying on tactile feedback, while achieving proper rotation requires multiple X-ray images.

The localization and tracking between the visual tracker on the C-arm and the surgeon's HMD was performed in a master-slave configuration. To construct the pose-aware system, the inside-out tracker on the C-arm generates an anchor of the operating theatre, which serves as a common spatial reference for all AR devices

within the operating room environment. The anchor is uploaded via a shared service and downloaded by other devices. Thus, all AR devices are co-localized with respect to a shared coordinate system. As a consequence, the number of transformations involved in the transformation chain is fewer compared to an outside-in navigation system with optical markers, and error propagation is minimized.

Marker-free tracking is enabled via visual SLAM that is computed in real-time based on the spatial map of the room. The spatial map is a triangulated mesh that is constructed from the visual structures present in the operating theatre. Therefore, the performance of the SLAM-based inside-out navigation depends on the richness of the structures within the field of view of the tracking systems.

Naturally, this paradigm can extend to any other robotic imaging device entering the operating theatre, that will require each of the involved viewers to have a calibration between the real world and their views of the world. In HMDs, these calibrations are performed by algorithms such as SPAAM [104]. In the case of X-ray viewer, this is done by the proposed hand-eye strategy. Finally, using the same SLAM-based reconstruction and tracking environment, the surgeon, surgical crew, and the X-ray viewer become the natural parts of an augmented surgical team.

2.5 Acknowledgments

I want to thank Mr. Sing Chun Lee for assisting with the surface to CBCT calibration and designing the appropriate calibration phantoms, and Mr. Jonas Hajek and Mr. Tianyu Song for their help in calibrating the HMD with the C-arm scanner and acquiring multiple data sets for the hand-eye calibration. Sincere thanks to Drs. Mathias Unberath and Bernhard Fuerst for their collaboration during the design and implementation of these systems. I also thank Drs. Nassir Navab and Mehran Armand for their supervision towards the realization of each system.

Chapter 3

Multi-Modal Acquisition and Reconstruction

3.1 Motivation and Problem Statement

Modern C-arms are motorized to enable intra-operative Cone-Beam Computed Tomography (CBCT). However, they suffer from poor tomographic reconstruction quality due to object truncation, undesired movement during scanning, incomplete projections, limited view angles, and higher influence of scattering due to the lack of collimation. For many years, Filtered Back Projection (FBP) has been the traditional reconstruction technique, which weights all projections equally, mostly ignores beam hardening and the physical model of the reconstruction volume. Therefore, artifacts are encountered in the tomographic reconstruction.

Alternatively, Algebraic Reconstruction Techniques (ART) have more artifact reduction effects compared to FBP when the projections are irregularly spaced (not uniformly distributed around the object) [105, 106]. More reliable reconstruction can be generated when the field of view is limited, or some views are missing [107]. Hence, ART has become an interesting field of research in the recent years, including attempts to reduce the radiation dose or compensate for known implants [108]. ART is mainly based on iterative optimization algorithms, and is computationally more expensive. However, by using larger computational capacities, mainly modern parallel hardware, computational effort has become less of an issue [109].

Classic ART assumes an empty space as the initial estimate of the reconstruction where the weights are uniformly distributed inside. The shape and dimensions of this space are defined based on the characteristics of the imaging machine. In order to initialize the reconstruction with a more accurate estimate, a FBP of the volume is computed before the iterative reconstruction, and the contours of the surface of the scanned object are extracted and used as the imaging volume for the algebraic reconstruction, which may reduce artifacts near the metal objects [110]. The metal-trace is then located in the projections and compensated in the reconstruction using appropriate confidence parameters. Pre-operative CT data can be used as a priori information under the assumption that the C-arm projections are registered with CT scans [111]. This method is used to avoid the artifacts, and to improve the reconstruction from truncated projections. However, it will not be applicable to

situations with no prior CT scan.

Undesired movement of the object during scanning is a major source of artifacts. A typical application is perioperative imaging when the patient is not fully anesthetized, such as the procedures in interventional radiology [112], catheter placements [113], and CBCT guided biopsies [114]. Unintended movements also occur during dental CBCT for children and elderly, or imaging of extremities (hand, knee, etc.). Avoiding these artifacts is also crucial for veterinary imaging where anesthesia causes severe complications [115], and keeping the animal static is a challenge. These movements result in shifted and displaced projections, which in return will cause the reconstruction to be blurred, and exhibit shadows or duplicates.

A possible solution was presented in [116], where 2D/3D registration in combination with an iterative reconstruction to correct projection data inconsistencies due to motion and positional error is proposed. The rigid transformation of the projections are computed by estimating the 3D pose of the C-arm source, and minimizing the distance of the acquired projections and the digitally reconstructed radiographs. However, when several major rigid movements occur during the scanning process, the reconstruction of the reference volume for 2D/3D registration becomes problematic. Therefore, an alternative method for the correction of the pose of each projection may be desired.

Artifacts caused by respiration have been addressed by using different patient prior models for motion compensation [117]. However, we attempt to solve a different type

of motion artifact caused by accidental and mostly rigid movements, and investigate the effect of these movements on tomographic reconstruction.

An irregular orbit of the C-arm is another source for motion artifacts, duplicates and shadows in a reconstructed volume. Different methods suggest to attach beads to the patients and use the projection information from the beads, or place various calibration patters under the surgical bed and recover the correct C-arm pose using projections [118]. These only consider an irregular orbit, and do not address sudden, rigid patient movements. External tracking systems have also been utilized to track the C-arm and the patient using multiple fiducials [4]. These systems are impractical in an operating room setting due to high complexity, disruption of workflow, introduction of hardware, and line of sight issues.

In this chapter, we propose a novel tomographic reconstruction using real-time patient surface information. A Red-Green-Blue-Depth (RGBD) sensor is mounted on the image intensifier of a CBCT-enabled C-arm. In contrast to time-of-flight cameras, this structured-light RGBD camera is less dependent on temperature, lighting, or materials, and provides color information. During CBCT scan the surface of the patient is acquired using the RGBD sensor, which is used as a priori information to mask the CBCT reconstruction using Maximum-Likelihood Expectation-Maximization (MLEM). Consequently, this method uses the patient's surface as the initial estimate of the reconstruction volume. Previous approaches to jointly use an RGBD camera and C-arm were limited to stitching several CBCT volumes after

individual reconstructions [119]. To the best of our knowledge, this is the first tomographic reconstruction techniques using surface information from a single, integrated RGBD sensor. This method improves the reconstruction quality inside the patient, and has a low sensitivity to missing projections.

Rigid patient movements during scanning can be observed in real-time using the RGBD camera. During the reconstruction process these rigid movements can be compensated by correcting the pose of the projections. The estimated and irregular path of the X-ray machine relative to the patient motivates the use of ART. Our movement correction applies to only accidental rigid movements during CBCT scans, which can take up to one minute using a motorized C-arm.

The improved reconstruction can benefit several interventions, particularly when the patient is not fully anesthetized. In cases the surgical scene is very homogeneous, or there are large gaps between the patient's body and the covering drape the system may not be applicable. However, in typical interventional radiology scenario, the patient is draped around the point of entry, not the actual area of interest (e.g. head or extremities). In an orthopedic or trauma scenario, the surgical site is usually exposed, and the surrounding drapes stuck to the patient.

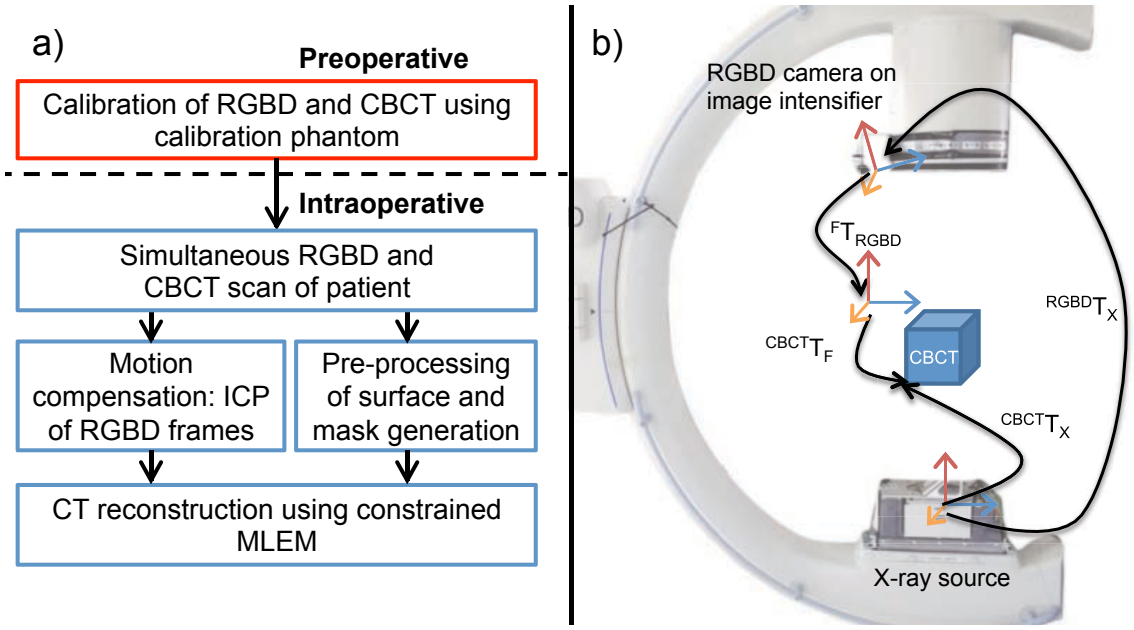

Figure 3.1: The workflow of this technique includes an offline calibration of RGBD surface and CBCT Volume **(a)**. This calibration together with patient RGBD and CBCT scans are used to mask and correct for rigid movement in the reconstruction. System **(b)** comprises a mobile C-arm, and an RGBD sensor near the detector. The transformation chain is illustrated in **(b)**.

3.2 Workflow and System Design

CBCT reconstruction using patient surface requires an offline pre-interventional calibration of RGBD surface and CBCT volumes (Fig. 3.1a). Next, simultaneous RGBD and CBCT scans are acquired from the patient. The RGBD camera mounted near the detector plane (Fig. 3.1b) observes the scene during reconstruction, and provides C-arm poses relative to the patient. This information is used to correct the reconstruction if undesired rigid movement occurs by the patient. Lastly, patient surface is converted to masks and used as the reconstruction space.

3.3 Rigid Movement Correction using RGBD-based Simultaneous Localization and Mapping System

The rigid movement of the object scanned is observed using the RGBD sensor. The transformation is estimated using RGBD-based Simultaneous Localization and Mapping (RGBD SLAM), which allows tracking of objects independent of the background RGBD data. The relative transformations of RGBD sensor origin and X-ray source are equal when no movement occurs. Under this assumption, the pose of the X-ray origin for each projection is corrected [120].

RGBD devices comprise multiple cameras and at least one projector, and allow the fusion of color images and depth information. Consequently, these devices allow us to recover the scale for visual features extracted in a color image. In this work, we aim at utilizing the co-calibrated RGB and depth channels concurrently to track the displacement of patient relative to a C-arm during CBCT scanning.

RGBD SLAM is used to track the C-arm relative to the patient, which performs simultaneous estimation of the pose of perceived landmarks, and updating the position of a sensing device [121]. RGBD-based SLAM uses the similar concept and augments the 2D extracted color features with co-calibrated depth information [98]. These 3D features are used to estimate the transformation between frames. The transformations

are then used to initialize the iterative closest point algorithm to further refine the mapping. Finally, a pose graph solver is used that optimizes the trajectory using a non-linear energy function.

After scanning, the recorded trajectory is matched with the CBCT geometry parameters and trajectory [120], using the previously described RGBD-to-X-ray calibration based on the multi-modal radiopaque checkerboard (see Sec. 2.2). In the presence of dissimilarities in the two trajectories, the RGBD SLAM trajectory is utilized to correct for patient movement during the reconstruction. The RGBD SLAM computes a dense feature space, so-called feature (patient) coordinate frame $\mathbf{F}$, using the initial frame, and performs tracking with respect to this frame. The extrinsics of the C-arm trajectory is then defined as:

$$ {}^{\mathbf{CBCT}}\mathbf{T}_{\mathbf{X}} = \{ {}^{\mathrm{CBCT}}\mathbf{T}_{\mathrm{X}}^{(i)} \mid {}^{\mathrm{CBCT}}\mathbf{T}_{\mathrm{F}} \cdot {}^{\mathrm{F}}\mathbf{T}_{\mathrm{RGBD}}^{(i)} \cdot {}^{\mathrm{RGBD}}\mathbf{T}_{\mathrm{X}} \}_{i=1}^{n}, \tag{3.1} $$

where n is the number of projection images, and CBCT, F, RGBD, and X are CBCT, feature, RGBD, and X-ray source coordinate frames, respectively. Note that the acquisition of color and depth information is subject to the line of sight. However, we do not anticipate any objects or surgeons to be between the detector and the patient during the scan. In most surgical scenarios, drapes are stuck to the patient's surface or - in case of interventional radiology or veterinarian - may not be present at all. In these cases the depth data does provide sufficient information to estimate the surface and relative movement. Furthermore, our system is not limited to a movement

occurring between two consecutive frames, but for each projection the displacement is estimated relative to the feature (patient) coordinate frame.

3.4 RGBD-Enhanced Tomographic Reconstruction

In this section, we will present the basics of iterative reconstruction for CBCT, followed by the introduction of the incorporation of real-time surface data to enhance the reconstruction quality. The underlying problem to be solved can be expressed through a system of linear equations: $\mathbf{p} = A\mathbf{f}$+noise, where $\mathbf{p}$ is the set of j projections p_j, the entries of $\mathbf{f}$ represent the unknown attenuation coefficients of the discretized volume, and A is the system characterization operator (also described as system matrix). A is comprised of individual entries a_{ji} expressing the relation between the observed attenuation along a ray and voxels this ray passes through:

$$a_{rj} = \frac{\text{illuminated area of pixel } r \text{ by ray } j}{\text{total area of pixel } j}. \tag{3.2}$$

The system matrix A is sparse and large, and ART algorithm performs well under this condition by iteratively projecting N-dimensional image estimates onto N-dimensional projection hyperplanes, where N is the number of voxels to be reconstructed. Due to inherent property of the X-ray attenuation, the noise in the projec-

tions could be modeled as a Poisson distribution. Using this model, MLEM has been shown to outperform the classic ART reconstruction [122].

RGBD camera is used to scan and reconstruct the patient surface concurrently with the CBCT scan [123]. To use the patient's surface as prior information, we propose to create 3D closed mask λ from the surface, where each element λ_r in the mask is zero for voxel elements outside of the patient's body, and non-zero for voxel elements inside the body. The masks are generated in real-time using the RGBD data acquired during the scanning procedures. First, the surface data is automatically pre-processed (morphological operations as well as smoothing), and then converted to meshes. In the presence of large holes, manual interaction by the user is needed to create a closed surface. Finally, the mesh is converted to the volumetric mask. The mask is then used as a relaxation factor λ:

$$f_r^{(i+1)} = \lambda_r f_r^{(i)} \frac{1}{\sum_j a_{rj}} \sum_j a_{rj} \left(\frac{p_j}{\sum_k a_{kj} f_k^{(i)}}\right), \tag{3.3}$$

where r is the voxel element, j is the projection element (ray), k is the set of voxels that intersect with projection j, and i is the iteration number. Deriving λ from RGBD data and integrating it into MLEM is an important contribution of this work. λ enforces zero constraints in the reconstruction for air voxels, and using the correct attenuation coefficient of air, it allows a more precise reconstruction inside patient's body. Reconstruction is constrained outside patient surface and voxel intensities will

remain zero, and inside patient surface no constraints are applied, therefore it is equivalent to unmodified MLEM. This is equivalent to uniformly distributing the weights of the system matrix A inside the body rather than the entire imaging volume. Note that at each step, $f_i^{(n+1)}$ is computed using all the voxel intensities from the previous estimation. Therefore, correct values of air voxels contribute to a more precise reconstruction inside the patient's body. The implementation steps are described in Alg. 2.

Algorithm 2 MLEM reconstruction from projections using surface masks

f: reconstruction volume, p: projections, A: system matrix
λ: masked surface, i: iterations
nv: number of voxels, np: number of projections
fp: forward-proj., bp: backward-proj., rt: ratio of estimate to measurement

for $(i = 1 : runs)$ **do**
  for $r = 1 : nv$ **do**
    rt $= 0$
    for $j = 1 : np$ **do**
      $fp = \sum(A_j * f + \epsilon)$
      $measurement = A_{jr} * p_j$
      $rt \mathrel{+}= measurement/fp$
    end for
    $bp_r = f_r * \lambda_r * rt/(\sum(A_r + \epsilon))$
  end for
end for

3.5 Experimental Validation and Results

The RGBD-enhanced CBCT reconstruction is evaluated with simulated and real data, where our method is compared to unmodified MLEM-based CBCT reconstruc-

tions. Our reconstruction supports total variation regularization to achieve higher quality at lower dose (statistical reconstruction) [124]. First we present the reconstruction quality, followed by the results of correction of rigid movement.

3.5.1 Experimental Setup

All projections are acquired using a motorized C-arm. The Arcadis Orbic 3D, Siemens Healthineers, can execute orbital rotations of 190° for acquisition of 100 projection images. The close-range coded-light Intel RealSense SR300 RGBD camera, Intel Corporation, is rigidly attached to the image intensifier. SR300 comprises a small 3D depth camera with 640×480 resolution and an HD color camera. The effective depth range from the camera origin is between 20 to 150 cm. The RGBD-SLAM tracking is performed at 30 fps.

Real-time tracking of the patient to C-arm, generating and pre-processing masks, and the tomographic reconstruction are all performed on the same PC. Reconstruction of a CBCT volume of dimensions $512 \times 512 \times 512$ with spacing of 0.2475 mm from 100 2D projection images, each of dimension 1024×1024, 16 bit data, takes < 3 sec. The reconstruction algorithms are implemented in the ImFusion software[1] utilizing a GeForce GTX TITAN.

[1] `http://imfusion.de/products/imfusion-suite`

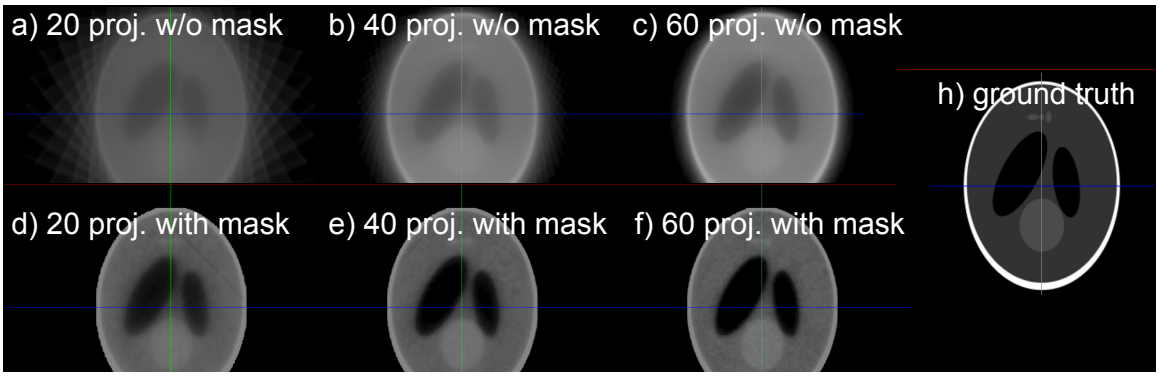

Figure 3.2: Reconstruction quality using an unconstrained MLEM-based reconstruction highly depends on the number of projections **(a-c)**, while the RGBD-enhanced reconstruction performs better with fewer projections **(d-f)**.

3.5.2 Reconstruction Quality and Number of Projections

The reconstruction quality is analyzed using simulated projections of a Shepp-Logan phantom. This enables the direct comparison of MLEM reconstruction without masking (Fig. 3.2a-c) and with masking (Fig. 3.2d-f) with respect to the ground truth (Fig. 3.2h). The impact of a reduction of the number of projections is of great interest. Fig. 3.2 clearly illustrates that a lower number of projections reduces the reconstruction quality when performed using MLEM, while our method is less sensitive to a reduced number of projections. This is quantified by comparing several reconstructions with lower number of projections with the ground truth by computing the Sum of Absolute Differences (SAD) of the voxel intensities in the CBCT scans. The differences are illustrated in Fig. 3.3a.

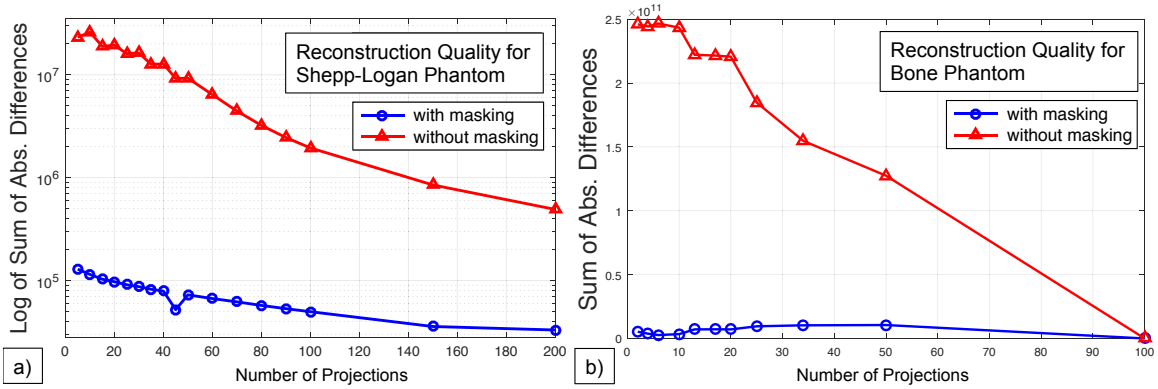

Figure 3.3: **(a)** Using simulated reconstructions of a Shepp-Logan phantom, it can be shown that surface masking during reconstruction yields a better quality when compared to the ground truth (all projections) using SAD. The mask is separates the phantom from the background. The enhanced reconstruction is less sensitive to a reduced number of projections. **(b)** Quality of reconstruction of bone model using MLEM without (red) and with surface masking (blue), where SAD is computed for each method between the best reconstruction using all projections and using fewer projections. It is clear that the classic reconstruction is more sensitive to fewer projections than the RGBD-enhanced method. Note that in **(a)** logarithmic scale is used to compare the data with different orders of magnitudes.

To validate the reconstruction quality in a more realistic setup, RGBD data and projection images are acquired from a bone phantom in which a guide wire has been inserted. The reconstruction using MLEM is shown in Fig. 3.4a. It exhibits artifacts inside and around the phantom, and the guide wire is seen as blurry and wide area. To consider the uncertainties due to poor depth resolution of the RGBD camera the volumetric mask is smoothed using a Gaussian transition function in the neighborhood of the object surface (2 mm). Using the mask extracted from the RGBD data (see in Fig. 3.4b), the reconstruction quality in terms of clarity of the wire tip and metal artifacts (Fig. 3.4c) is improved.

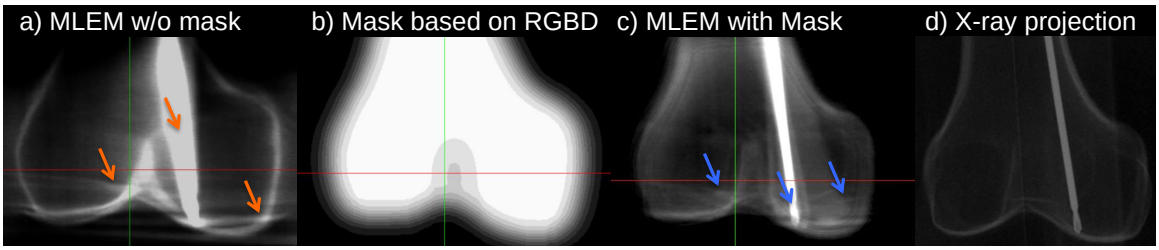

Figure 3.4: Reconstruction of a bone model with guide wire yields strong artifacts in and outside the structure (orange arrows in panel **(a)**) and a blurry, wide guide wire. Incorporating RGBD data (mask in **(b)**), results in an improved reconstruction quality (blue arrows in **(c)**) with clear guide wire tip. The log-normalized X-ray projection **(d)** shows the guide wire inside the bone.

When comparing the reconstruction with all projections against reconstructions with fewer projections, it is clear that the classic approach is highly sensitive to the reduction of projections, while the RGBD-enhanced reconstruction is relatively stable. Fig. 3.3b shows this sensitivity by comparing each method with its best performance using all projections.

3.5.3 Correction of Rigid Movement during CBCT Scan

In this section we use the RGBD SLAM system for C-arm to patient tracking [95], and will evaluate the effect of rigid movement correction in CBCT reconstruction. In this experiment, the bone phantom was moved 1.27 cm and rotated ($< 2°$) planar to the surgical bed (arbitrary rigid movement in the C-arm coordinate frame), between the 50^{th} and 51^{st} projection. The observed point clouds are shown in

Fig. 3.5a. The reconstruction using 100 projections without incorporation of RGBD data yields clear motion artifacts. Duplicates and shadows marked with orange arrows in Fig. 3.5b. This rigid movement is corrected in the RGBD-enhanced reconstruction (see Fig. 3.5c), which significantly reduces motion artifacts. The average thickness of the observable guide wire in Fig. 3.5b is 3.31 mm. The movement artifacts cause duplicates to appear in a neighborhood of 5.53 mm. After correction of movement, duplicates no longer appear, but the thickness of the reconstructed wire is 4.61 mm (Fig. 3.5c). The thickness of the wire from physical measurement is 2.5 mm.

We further evaluate the system performance using a pig specimen. The animal cadaver is covered with self-adhering drape to simulate a realistic surgical scenario, and placed near the ico-center of the C-arm using the guidance laser attached to the C-arm. Next, we acquire four CBCT scans. During each scan an arbitrary rigid transformation is applied to the bone with the norm of the translation component between 2 to 5 cm, and rotational component $< 5°$. Each scan takes 60 sec, and the movements take 2 to 4 sec at arbitrary intervals during the acquisitions. An intensity-based mask is applied to the projection images to crop padding created by the detector. Next, the MLEM ART method is used to reconstruct the volumetric data using all 100 projections. Normalized CBCT volumes are compared to a ground-truth reconstruction with no rigid movement using SAD. Without movement correction the SAD is $(21.1 \pm 1.1) \times 10^6$, and improves to $(2.97 \pm 0.97) \times 10^6$ after

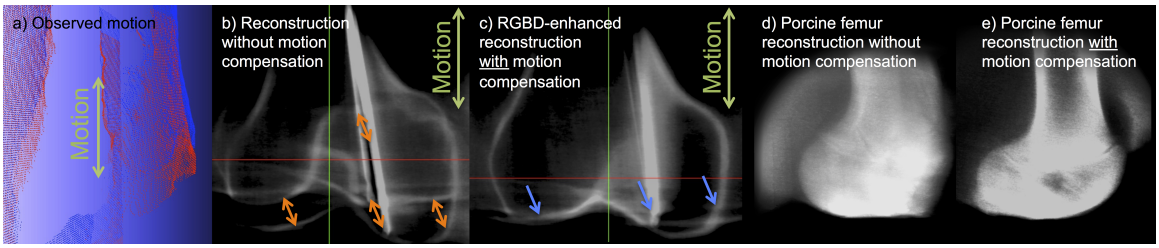

Figure 3.5: Movement during the CBCT scan can be observed using the RGBD sensor **(a)**. Without correction, artifacts are clearly visible (orange arrows in panel **(b)**. The RGBD-enhanced reconstruction does not exhibit such artifacts (blue arrows, panel **(c)**. In **(d)** and **(e)**, CBCT reconstruction before and after motion correction are shown on a pig femur.

correcting for the rigid movement[2]. The average SAD per voxel for normalized CBCT volumes reduces from 0.157 without correction, to 0.022 with correction.

We finally evaluate the influence of reducing the number of projections. Reconstruction a ground truth volume and the four volumes using 50 projections yielded an SAD of $(16.8 \pm 0.7) \times 10^6$ before correction, and $(2.33 \pm 0.84) \times 10^6$ after movement correction. This is equivalent to reduction of average SAD per voxel from 0.125 to 0.017. Results indicate improved reconstruction quality after correcting the bone rigid movement with both 100 and 50 projections.

3.6 Discussion and Conclusion

We propose a CBCT reconstruction algorithm incorporating the patient's surface from an RGBD sensor and use MLEM algorithm. The first major contribution of

[2]values reported as mean ± standard deviation

this work is the computation of the relaxation factors for each voxel from the real-time depth information provided by the RGBD sensor. Due to the rigid system construction, the calibration only needs to be performed once. Our method improves the reconstruction quality, and the streak artifacts near metal are reduced. The second contribution is the automatic correction of rigid patient movements using both RGB and depth information. The image inconsistencies in terms of motion artifacts, caused by accidental rigid movements of the patient during scanning and slow C-arm movement, are observed by the RGBD camera, and are then corrected during the reconstruction. Results indicate that using the surface information results in lower sensitivity to fewer projections.

Due to the rigid movements during scanning, the assumption of an orbital trajectory of C-arm is no longer valid. This newly observed, highly irregular path of the X-ray source relative to the target may cause FBP to fail. In this work we use the iterative reconstruction which is not affected by the dynamically changing scanner motion or the parallax effect caused by the translation component of the movement. To compute the relative pose of a C-arm with respect to a moving patient, we use RGBD SLAM and perform vision-based tracking. When the patient is partially covered with drapes (clinically realistic scenario) fewer color features are identified in the background and the tracking quality increases [95].

Calibration between C-arm and RGBD sensor is performed using a multi-modal checkerboard calibration target. This calibration is used to correct the projection

poses relative to the patient. When incorporating the surface information into the CBCT reconstruction as prior information, we avoid using the information from the checkerboard pattern, as this produces a long calibration chain (intrinsic calibration of the C-arm as well as the transformation from the RGBD origin to the surface reconstruction coordinate frame become necessary), and will increase the error. Alternatively, we perform direct 3D to 3D calibration using ICP.

Mask quality is dependent on the depth resolution of the RGBD camera. This also results in RGBD/CBCT calibration error of nearly 2 mm. Moreover, holes in the meshes currently require manual pre-processing and morphological operations. This pre-processing also includes smoothing to compensate for uncertainties near the boundary. Similar approaches of extracting a mask from reconstructed volume and subsequently performing a reconstruction have been performed in [110, 111]. Our approach of integrating RGBD camera into the CBCT device, and simultaneously recording data has a far larger range of possible applications. One of which is demonstrated by correcting for the unwanted rigid movement during the scan. This correction would not be possible using the two-stage approach as the automatic extraction of a mask from a motion blurred image is difficult. Note that we only compensate for rigid movement during scans, which can be observed using a camera. The assumption of rigid motion is valid for imaging the extremities and dental applications where the movements are mainly rigid with minimal deformations. Respiratory motion is a complex process which involves internal movements, deformations and sliding motion of

lungs along the pleura. None of these motions can be observed directly using a camera, but our system may provide a sufficient surrogate signal to drive biomechanical models [125].

The advantages of using cameras with C-arm for radiation and patient safety, augmented reality, mosaicing, etc. is emphasized repeatedly in the literature. The fully integrated opto-X-ray imaging system represents the future of C-arm imaging. We believe that incorporating real-time RGBD data leads to a dramatic improvement of CBCT reconstruction quality. Considering that radiation dose for high quality imaging is a major clinical concern, we believe this technique can contribute as a radiation dose reduction measure enabling reconstruction with fewer projections, and avoid repeated scans by correcting accidental patient movements.

3.7 Acknowledgments

My sincere thanks to Drs. Nassir Navab and Bernhard Fuerst for their support and insightful comments. Special thanks to Dr. Wolfgang Wein and his team at imFusion for giving us access to use imFusion suit for the algebraic tomographic reconstruction.

Part II

Imaging Analytics and Registration

In part II of this dissertation, using computer vision and artificial intelligence tools, we perform analytics and processing for various medical image registration tasks using the data that were acquired and reconstructed in part I.

Chapter 4 addresses some of the challenges regarding 2D/3D registration between X-ray and CT, where accurate initialization is one of the important aspects that determines a successful registration. Chapter 5 investigates 3D/3D registration for automatic stitching of non-overlapping cone-beam CT acquisitions. This stitching problem is particularly important because the limited field-of-view of CBCT prohibits us from taking important measurements, such as the bone's total length, or verifying malrotations intra-operatively, using a single acquisition. Hence, we explored different vision-based stitching approaches that stitch and combine CBCT volumes independent of their overlap. Finally, chapter 6 presents the novel methodology for 2D/2D stitching between intra-operative X-ray images that, because of their projective nature and parallax constraints, require special considerations.

This chapter's work is based on the following manuscripts:

1. Fotouhi, Javad, Bernhard Fuerst, Alex Johnson, Sing Chun Lee, Russell Taylor, Greg Osgood, Nassir Navab, and Mehran Armand. "Pose-aware C-arm for automatic re-initialization of interventional 2D/3D image registration." International journal of computer-assisted radiology and surgery 12, no. 7 (2017): 1221-1230.

2. Fotouhi, Javad, Bernhard Fuerst, Mathias Unberath, Stefan Reichenstein, Sing

Chun Lee, Alex A. Johnson, Greg M. Osgood, Mehran Armand, and Nassir Navab. "Automatic intraoperative stitching of nonoverlapping cone-beam CT acquisitions." Medical Physics 45, no. 6 (2018): 2463-2475.

3. Fotouhi, Javad, Xingtong Liu, Mehran Armand, Nassir Navab, and Mathias Unberath. "From Perspective X-ray Imaging to Parallax-Robust Orthographic Stitching." arXiv preprint arXiv:2003.02959 (2020).

Chapter 4

2D/3D: Pose-Aware C-Arm for Automatic Re-Initialization of the Image Registration

4.1 Introduction and Background

Intensity-based image registration is one group of image-based registration techniques, and of special interest in this chapter. Registration is typically performed by simulating 2D radiographs, so-called Digitally Reconstructed Radiographs (DRRs), from pre-interventional patient data, and matching them with the intra-interventional C-arm image [126–128]. Intensity-based registration becomes challenging where bony structures in the pre- and intra-interventional data differ due to deformations caused

by the surgery. C-arm image registration is in particular a complex problem due to the limited views of the C-arm images. Atlases of the target anatomy and deformable shape statistics are suggested to support the registration in such cases of image truncation [129]. In spite of promising results from using shape models, in several trauma and orthopedic cases where the bony anatomy is severely damaged and deformed, the registration is yet prone to failure. Last but not least, intensity-based registration is very sensitive to initialization due to the limited capture range of the intensity-based similarity cost. These are the challenges that prohibit the 2D/3D registration to become a part of standard surgical routine.

Reliable registration is often performed using external navigation systems [3]. In navigation-guided fluoroscopy, it is a common practice to drill the fiducials into the bone to maintain the registration in the presence of patient movement [4]. Generally, these fiducials are implanted before the 3D pre-interventional data is collected, and will remain inside the anatomy until after the intervention. Fiducial implantation requires a separate invasive surgery, and increases the risk of fracture in the osteoporotic bones.

Fiducial-based C-arm tracking was used for intensity-based image registration in [4]. This method achieves sub-millimeter accuracy, however requires the implantation of a custom-made in-image fiducial. The mean Target Registration Error (mTRE) is 0.34 mm for the plastic bone phantom with 90° rotation, and 0.99 mm for the cadaveric specimen with images 58.5° apart. In [130], anatomical features or

beads on patient surface are located in the C-arm images and used for C-arm pose estimation. This work tackles the problem of intra-interventional calibration and reconstruction, and does not address registration to pre-interventional data. Despite the high registration accuracy of navigation-guided systems near the fiducial markers, the registration error and uncertainty increase in distant areas. Furthermore, these systems have a complex setup, occupy additional space, and change the surgical workflow. Last but not least, the line of sight issues limit the free space in the surgical site.

In order to initialize the 2D/3D registration for several number of C-arm images, Uneri et al. [131] proposed to use each successful registration of a C-arm image to 3D data as the initialization for the latterly C-arm image. This work avoids the use of any external trackers by solely relying on image-based registration. But on the other hand, in order to robustly initialize the registration, the allowed displacement between consecutive C-arm images was limited to only 10 mm in translation.

The initial alignment and outcome verification are introduced as two main bottlenecks of the registration problem [132]. In order to address the problem associated with initialization of a 2D/3D registration task, [133] suggested an interactive initialization technique where the user performs the alignment by utilizing a gesture-based interface or an augmented reality environment together with a navigation system. The tedious initialization procedure makes the system impractical in a surgical setting.

Generalized Hough Transform is used in [134] to learn large number of 2D tem-

plates over a variety of 3D poses. During intervention this information is used to estimate the 3D pose from the 2D C-arm images and initialize the 2D data with respect to 3D pre-interventional CT. Projection-slice theorem and phase correlation are also used to estimate the initialization of the 2D/3D registration problem [135].

We investigate the use of vision information from an RGB-Depth (RGBD) camera mounted near the detector of a mobile C-arm to estimate the pose of the C-arm relative to the patient. We refer to this C-arm as *pose-aware*. The proposed C-arm tracking will result in estimating the projection geometry of the C-arm relative to the surgical scene at arbitrary poses of the C-arm. The tracking information will be used to transform the 2D C-arm images globally near their correct alignment with respect to the pre-interventional data. Next, intensity-based registration is utilized to align images locally. Our methodology aims at improving the automation during C-arm guided interventions. This technique is more effective when the registration has to be repeated for multiple C-arm poses. The workflow (Fig. 4.1) initiates by performing an initial registration of the C-arm to the pre-interventional data. The registration loop is then closed by using this pose as the reference for the tracking of different C-arm arrangements. This system is self-contained, requires a single factory calibration, and the surgical workflow remains intact.

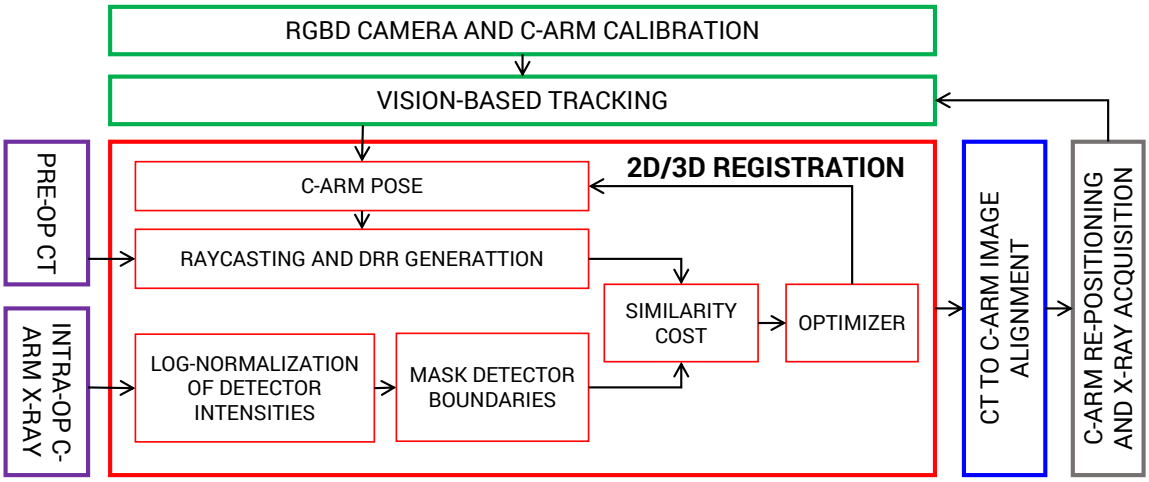

Figure 4.1: Workflow of a pose-aware C-arm system for intensity based 2D/3D registration. Every C-arm image is globally aligned (initialization for 2D/3D registration) with pre-interventional CT based on vision-based tracking. The main contribution of this work is shown in green.

4.2 Estimation of C-arm Trajectory Using Visual Tracking

The tracking of C-arm source during C-arm re-positioning is performed by first observing the surgical site (patient) using the RGBD camera. These relationships, denoted as ${}^{\mathrm{P}}\mathbf{T}_{\mathrm{RGBD}}$ and ${}^{\mathrm{P}}\mathbf{T}_{\mathrm{RGBD'}}$, are shown in Fig. 4.2. The relative transformation between the RGBD camera poses is then derived as:

$$ {}^{\mathrm{RGBD'}}\mathbf{T}_{\mathrm{RGBD}} = {}^{\mathrm{P}}\mathbf{T}^{-1}_{\mathrm{RGBD'}} \cdot {}^{\mathrm{P}}\mathbf{T}_{\mathrm{RGBD}} \tag{4.1} $$

Since the RGBD camera is rigidly mounted on the C-arm, the multi-view relation

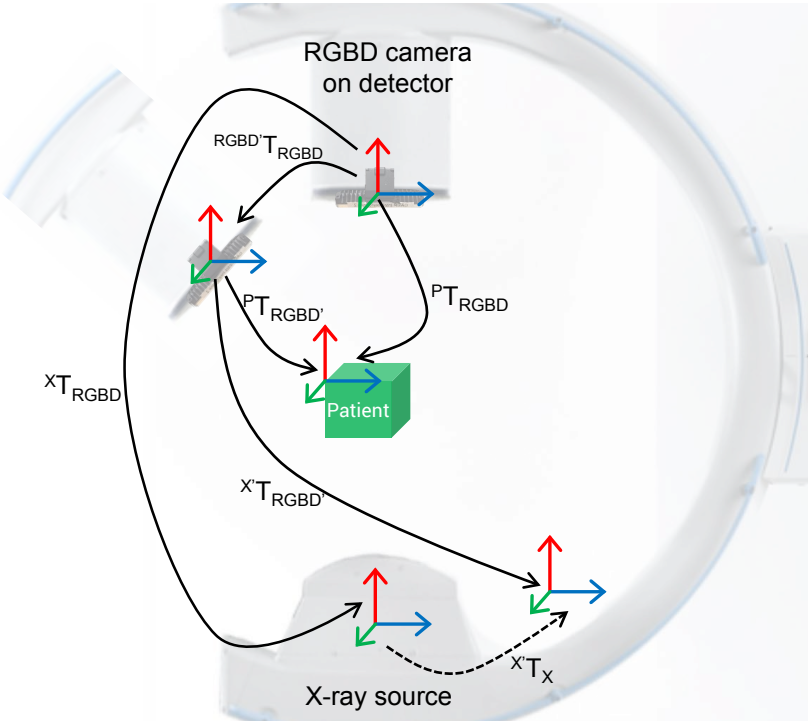

Figure 4.2: The relative displacement between camera poses ${}^{\text{RGBD'}}\mathbf{T}_{\text{RGBD}}$ is used for estimation of the relative displacement of the X-ray source ${}^{\text{X'}}\mathbf{T}_{\text{X}}$.

between the RGBD origin and X-ray source remains fixed (${}^{\text{X}}\mathbf{T}_{\text{RGBD}} = {}^{\text{X'}}\mathbf{T}_{\text{RGBD'}}$). We can then obtain the relative displacement of the X-ray source as:

$$ {}^{\text{X'}}\mathbf{T}_{\text{X}} = {}^{\text{X'}}\mathbf{T}_{\text{RGBD'}} \cdot {}^{\text{RGBD'}}\mathbf{T}_{\text{RGBD}} \cdot {}^{\text{X}}\mathbf{T}^{-1}_{\text{RGBD}} \tag{4.2} $$

In the following we discuss two distinct vision-based methods for tracking the surgical site with depth-only and RGB+depth data.

4.2.1 Tracking and Surface Reconstruction Using Depth Data

We deploy an algorithm to automatically estimate the depth sensor trajectory, and create a smooth dense 3D surface model of the objects in the scene [123]. The relative camera poses are acquired by iteratively registering the current depth frame to a global model. This iterative stage is performed using multi-scale Iterative Closest Point (ICP), where depth features among frames and the global model are matched using projective data association [136].

4.2.2 RGBD-based Simultaneous Localization and Mapping

Modern RGBD cameras are small devices which provide color information as well as depth data. Therefore, greater amount of information is extracted and processed compared to a single RGB or infrared camera. We aim at using depth and color information concurrently, and acquire 3D color information from the surgical site. Ultimately, we use this knowledge to track the C-arm relative to the patient. For this purpose we use an RGBD SLAM system introduced in [98]. The underlying tracking method uses visual features (e.g. SURF [137]) extracted from the color images, and constructs feature descriptors. The feature descriptors are matched among the

consecutive frames to form a feature correspondence list. The initial transformation is then computed using the correspondence list and a RANdom SAmple Consensus (RANSAC) method [138] for transformation estimation. The outcome is later used to initialize ICP to further refine the mapping between consecutive frames.

4.3 Pose-aware C-arm for Intensity-based 2D/3D Image Registration

Intensity-based 2D/3D image registration is an iterative method which 6 DOF parameters are estimated for a rigid body transformation that brings the input 2D and 3D data to a common coordinate frame. The key component of this process is the creation of many 2D artificial radiographs, so-called Digitally Reconstructed Radiographs (DRRs), from the CT data and comparing them against the C-arm X-ray image. This process continues until the similarity cost between the DRR and X-ray images is maximized using an iterative optimization.

CT image is a collection of voxels representing X-ray attenuation. In order to generate a DRR from a given CT data, 6 DOF rotation and translation parameters are used together with the intrinsics of the C-arm to position a virtual source and a DRR image plane near the CT volume [139]. The intrinsic calibration of the C-arm, allows us to estimate the source to detector arrangement, and use this information to generate DRR images [140]. Next, by means of raycasting [141], the accumulation

of intensities along each ray (that originates from the virtual source and intersects with the image plane) is computed. The accumulated values then form a DRR image. In this work, we first perform a single image-based C-arm image to CT registration using manual initialization. Thereafter, we track the C-arm new pose relative to the surgical site. The tracking results of the C-arm are later used as the 6 DOF initialization parameters of the new coming C-arm images.

Next, the C-arm X-ray images are converted to linear accumulation of X-ray attenuation by means of log-normalization. Hence, the DRRs and detector images become linearly proportional. Furthermore, we apply a circular mask on the detector images to discard the saturated areas near the detector boundaries and only include relevant information in the similarity cost.

After estimating the re-initialization transformation $\mathbf{T}_0$ obtained as results of the composition of transformations ${}^{\mathrm{X'}}\mathbf{T}_{\mathrm{X}}$ estimated by the pose aware system, and that of the first pose of the C-arm, the DRR image is generated and compared against the C-arm X-ray image I_{X} using a similarity cost S. In this work we utilize the Normalized Cross-Correlation (NCC) as the similarity cost defined as:

$$NCC := S\left((\mathbf{R},\mathbf{t})|\mathbf{T}_0, I_{\mathrm{X}}, I_{\mathrm{D}}\right) = \sum^{|\Omega_{\mathrm{X,D}}|} \frac{I_{\mathrm{X}} \cdot I_{\mathrm{D}}^{((\mathbf{R},\mathbf{t})|\mathbf{T_0})}}{\sigma_{\mathrm{X}}\sigma_{\mathrm{D}}} \tag{4.3}$$

where I_{X} and I_{D} are the mean-normalized X-ray and DRR images, $(\mathbf{R},\mathbf{t})$ are the rotation and translation parameters used to form the DRR image, $\Omega_{\mathrm{X,D}}$ is the common

spatial domain of the two images, and σ_{X} and σ_{D} are the standard deviations of the X-ray and DRR image intensities in $\Omega_{\mathrm{X,D}}$.

The final parameters $(\hat{\mathbf{R}}, \hat{\mathbf{t}})$ are estimated by an iterative optimization:

$$[\hat{\mathbf{R}}, \hat{\mathbf{t}}] = \underset{\mathbf{R},\mathbf{t}}{\arg\max}\ S\left((\mathbf{R}, \mathbf{t})|\mathbf{T}_0, I_{\mathrm{X}}, I_{\mathrm{D}}\right) \tag{4.4}$$

The optimization is performed using Bound Constrained By Quadratic Approximation (BOBYQA) algorithm [142].

4.4 Experimental Validation and Results

4.4.1 Experimental Setup

X-ray images are acquired using an Arcadis Orbic 3D mobile C-arm from Siemens Healthineers with the detector size of 230 mm × 230 mm. In order to remove low-noise, the X-ray images are captured as Digital Radiographs (DRs) where a weighted average filter is applied to the images. The DR images are acquired using 0.2 to 23 ma and 40 to 110 kV. For safety purposes the maximum power for taking a DR image is set to 1000 W. The CT scan data is acquired using a Toshiba Aquilion One CT scanner, where the slice spacing and thickness are 0.5 mm, and the volume elements are created using 16 bits. An Intel RealSense SR300 camera from Intel Corporation is rigidly mounted on the image intensifier using a custom-made 3D printed mount. The

SR300 has a small form factor (X= 110.0±0.2mm, Y= 12.6±0.1mm, Z= 3.8−4.1mm), and integrates a full-HD RGB camera, an infrared projector, and an infrared camera. To ensure a wide visibility range of the patient surface, the RGBD camera is positioned with a particular tilt angle that the center of the image is nearly aligned with the C-arm iso-center. The C-arm imaging device is connected via Ethernet, and the RGBD camera is connected via powered USB 3.0 to the development PC which runs the tracking and the registration software. The tracking is performed in real-time, and the 2D/3D registration module is fully parallelized using Graphical Processing Unit acceleration.

4.4.2 Results

In the following we evaluate the pose-aware C-arm system by presenting the calibration accuracy, tracking results, and 2D/3D registration performance with and without the automatic re-initialization.

4.4.3 Calibration Outcome

A 5 × 6 checkerboard calibration target is employed to calibrate the X-ray, RGB, and infrared (depth channel of the RGBD camera) imaging devices. The length of each side of a single checkerboard square is 12.66 mm, and the total dimensions of the checkerboard are 63.26 mm and 75.93 mm. The checkerboard size is carefully selected according to the field of view of the C-arm and the RGBD camera images to ensure

full visibility among a variety of checkerboard poses. Next, 72 X-ray/RGB/IR images were simultaneously recorded to perform the calibration. A subset of images that produced high reprojection error were discarded as outliers. Ultimately, we used 42 image pairs and performed stereo calibration between RGB camera and X-ray source where the stereo mean reprojection error was 0.86 pixels. The stereo calibration in this work does not refer to the calibration of a single moving camera, instead it refers to estimation of the extrinsic parameters of two stationary cameras (RGB and X-ray imaging devices). RGB and X-ray images had individual mean reprojection errors of 0.75 and 0.97 pixels, respectively. RGB and IR cameras were calibrated using 59 images and the mean stereo reprojection error was 0.17 pixels, with individual mean reprojection errors of 0.18 and 0.16 pixels for each camera respectively.

4.4.4 Tracking Accuracy

The accuracy of the vision-based tracking methods is evaluated by rotating the C-arm around different axis, computing the pose of the X-ray source, and comparing the outcome to the ground truth provided by an external optical tracking system. an exemplary rotation is shown in Fig. 4.3, where the C-arm orbits around the surgical bed.

The optical position sensor tracks a single-faced passive rigid body fiducial mounted on the C-arm which is pre-calibrated to the X-ray source. The tracker is a Polaris Vicra System from Northern Digital Inc., where the measurement area for the mini-

mum distance (557 mm away from the tracker) is 491 mm $\times$ 392 mm, and 938 mm $\times$ 887 mm for the maximum distance (1336 mm) from the tracker.

The C-arm is rotated along three main axis and the tracking results from the depth-only algorithm and RGBD SLAM are recorded and compared to an optical tracking system. First, C-arm is orbited 45° around the cranial/caudal axis. The tracking errors are shown in Fig. 4.4 (a)-(b) in the form of Euclidean distance, as well as the rotational error. The rotational error are computed as the norm of mal-rotation angle along the three axis with respect to the ground truth. Next, the C-arm is swiveled by $\pm$10°, and the tracking errors are plotted in Fig. 4.4 (c)-(d). Lastly, the C-arm is rotated to anterior-posterior oblique view with $\pm$30°. The results of this experiment are shown in Fig. 4.4 (e)-(f).

Note that the C-arm rotation does not only result in the rotation of the X-ray source, but also by construction of the device a significant translation is applied to the X-ray source. In an exemplary case, considering the C-arm source to iso-center distance (approximately 600 mm), a 45° orbit of the C-arm results in approximately 430 mm displacement of the X-ray source. Therefore, the aforementioned experiments all involve rotation as well as translation of the X-ray source.

The ranges selected for these experiments were solely defined by the limited capture volume or line of sight constraints of the tracking system, or the range of C-arm motion (e.g. swivel is only possible up to 10° for this C-arm).

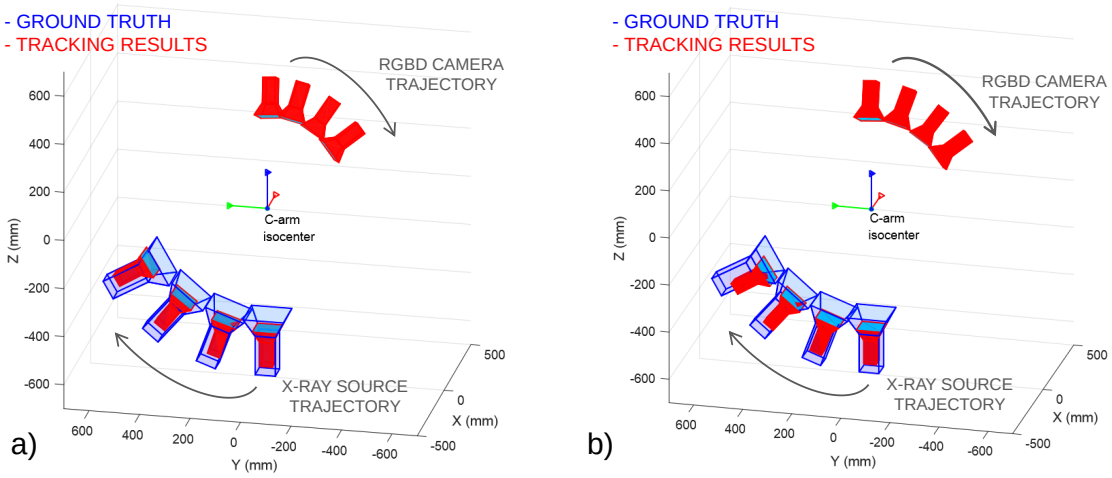

Figure 4.3: The pose of the X-ray source is tracked at different C-arm positions. Tracking based on (a) depth-only, and (b) an RGBD SLAM system is shown in red, and the tracking outcome based on the external tracking system is shown in blue. While drift (mistranslation) is observed in the tracking of the X-ray source, misrotation compared to the ground-truth is minimal.

4.4.5 2D/3D Registration with Automatic Re-initialization

To evaluate the effect of re-initialization on intensity-based 2D/3D image registration, we attempt to register CT and C-arm images of a dry femur-pelvis phantom with the automatic re-initialization. In addition to re-initialization, the optimizer search space is limited according to the aforementioned tracking errors, i.e. the optimizer bounds are 10 mm and 20 mm for $< 15°$ and $> 15°$ rotation of the C-arm, respectively. For speedup, the registration software is fully parallelized using GPU acceleration.

A CT scan is acquired from the femur-pelvis phantom that is encased in gelatin, and partially covered with drapes (to simulate realistic surgical scenario). Next, the

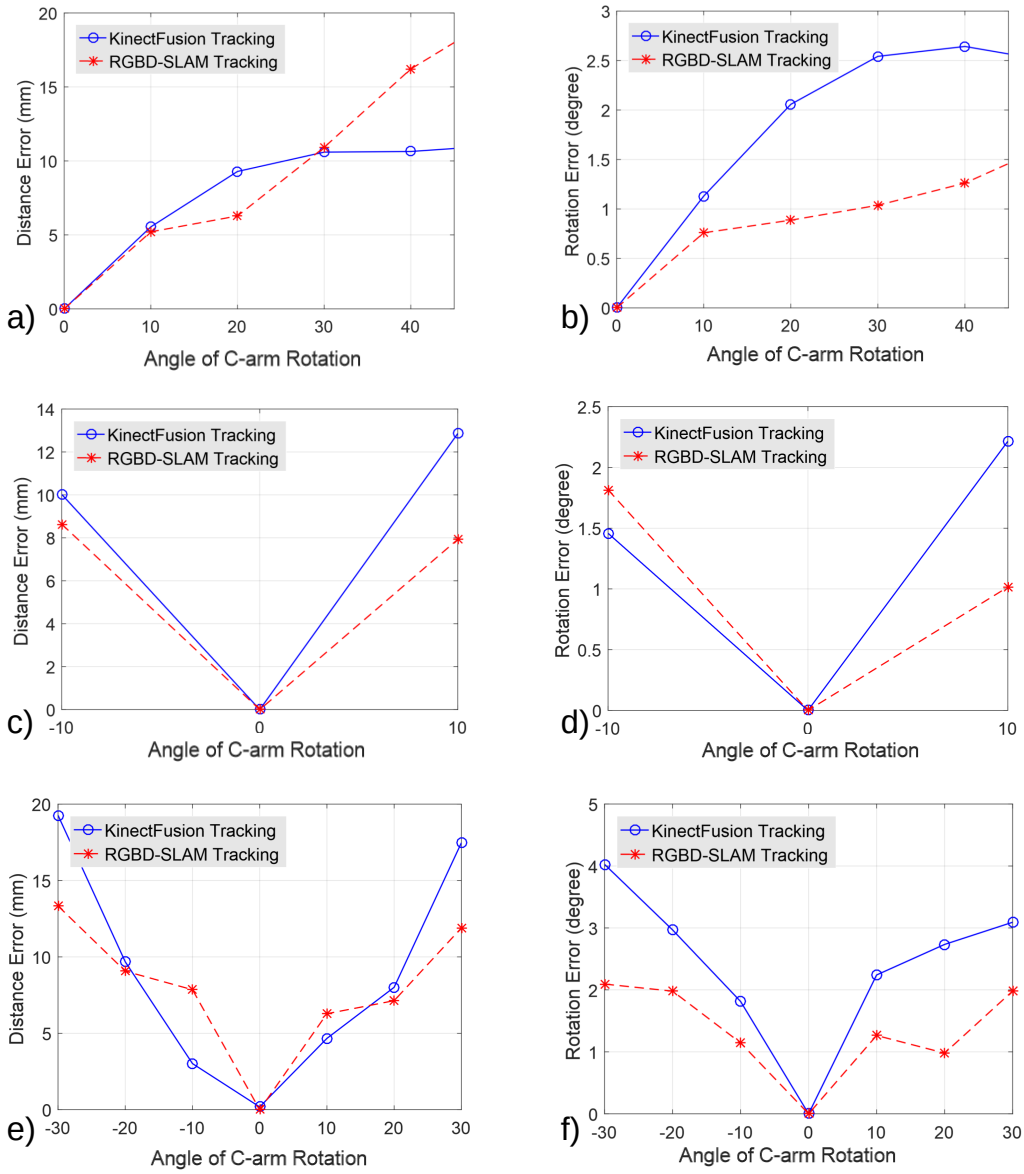

Figure 4.4: Tracking errors of X-ray source pose estimation with respect to ground truth are presented in Euclidean distance as well as rotational misalignment. The plots indicate the results when using depth-only data, and RGBD SLAM with RGB+depth information. (a)-(b) are the errors for 45° C-arm rotation in cranial/caudal direction, (c)-(d) correspond to ±10° C-arm swivel, and (e)-(f) are errors for ±30° of C-arm rotation in oblique direction. Regarding the composition of errors, results show a relatively small rotational and large translational component. Note that here the C-arm extrinsics recovered from vision-based methods are compared to tracking results using an external optical tracking system. Therefore, any changes in the intrinsic parameters (due to mechanical deformation of the C-arm) do not contribute to the errors.

Table 4.1: NCC similarity measures for the intensity-based 2D/3D registration are presented at two stages: 1. after re-initialization 2. after registration. The experiments are conducted using 13 arbitrary C-arm poses, and the mean and standard deviation (SD) of the similarity measure are presented in the table. Moreover, mTRE values are presented in mm after the re-initialization step. We consider a registration with mTRE < 2.5 mm as a successful attempt. The rate of success using each method is presented in the last column.

Tracking Method	After Re-initialization			After Registration		
	NCC	SD(NCC)	mTRE	NCC	SD(NCC)	Success Rate
depth-only	0.243	0.18	13.23	0.721	0.04	69%
RGBD SLAM	0.310	0.24	11.81	0.749	0.06	75%

phantom is placed on the surgical bed using laser guidance where the femur head is near the iso-center of the C-arm orbit. A C-arm X-ray image is then acquired and registered to the CT data using manual initialization. Thereafter, we rotate the C-arm within a range of 30°, and collect multiple C-arm images while tracking the C-arm using vision-based methods in Sec. 4.2. For each C-arm image, the tracking outcome is used for re-initialization of the intensity-based 2D/3D registration. In Table 4.1 the registration outcome is presented for all successful registration attempts. We consider a registration attempt successful only if the mTRE error after registration is < 2.5 mm. mTRE is computed as the mean of the absolute Euclidean distances between radiolucent landmarks on the phantom in the CT coordinate frame. Registration using random initialization yielded a 23% success rate, where vision-based re-initialization yielded a 75% success rate. All registrations were performed using the ImFusion software[1] with NCC similarity measure. An example demonstrating the registration steps is presented in Fig. 4.5.

[1] `http://imfusion.de/products/imfusion-suite`

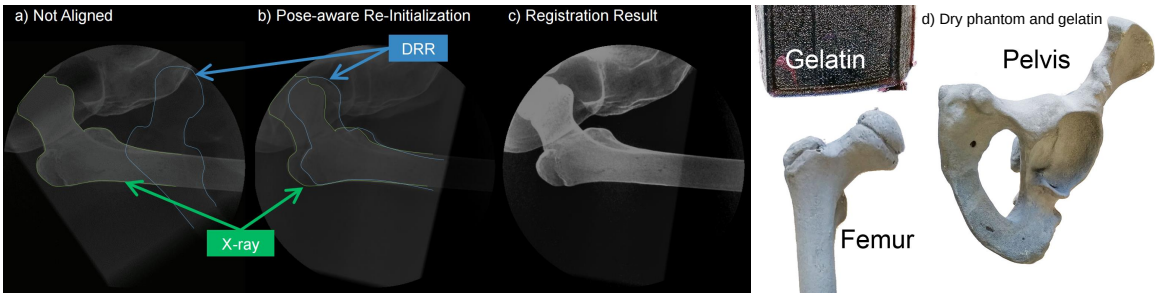

Figure 4.5: DRR is overlaid with C-arm X-ray image (a) before re-initialization, (b) after pose-aware re-initialization, and (c) after the 2D/3D registration. The pelvic-femur phantom and the gelatin are shown in (d). During the experiment the bone phantom is encased in gelatin and covered with drape to simulate a realistic surgical condition.

4.5 Discussion and Conclusion

The marker-free pose-aware C-arm system proposed in this work is an RGBD enhanced C-arm system that uses vision-based techniques to track the relative displacement of the C-arm with respect to the surgical site. Before the vision-based tracking takes place, C-arm is registered to pre-interventional CT data using a single 2D/3D registration with manual interaction. Thereafter, the relative C-arm poses are estimated with respect to the initial C-arm pose and consequently to pre-interventional 3D data. An error in the initial registration of the C-arm to CT data will directly affect the accuracy of the vision-based tracking system. The fusion of 3D pre- and 2D intra-interventional data allows the surgeon to understand the relationship between the current state of surgery, the complex 3D structures, and the pre-operative planning. This increases the surgeon's confidence, reduces the mental task load, and lowers the probability of a revision surgery.

Classically tracking the C-arm/patient using external optical trackers may provide better accuracy, but involves longer and more complex preparation, line of sight issues, and the invasive implantation of pins into bone; hampering surgical navigation from being widely adopted in orthopedics. In contrast to these systems, the pose-aware C-arm is non-invasive, requires no interaction time for manual re-initialization, and the workflow remains intact (manual initialization for authors as expert users takes between 30 to 60 sec). The registration step after a re-initialization takes < 10 sec.

The tracking accuracy is evaluated at various C-arm poses. The results in Fig. 4.4 indicate that an RGB+depth tracking method slightly outperforms a depth-only tracking system. The translation error for a wide range of displacement is 10-18 mm, and the rotation error is below 2°. The reported mTRE values are only for the re-initialization step. The range of acceptable mTRE for clinical application depends on the registration method and its capture range. Surgical navigation can only be performed after the 2D/3D registration takes place.

Registration using the re-initialization has 75% success rate, where random initialization only yielded 23% success rate. Our method increases the likelihood of a successful registration, which therefore improves the practical applicability and usefulness of surgical navigation. A similar fiducial-less initialization method in [135] yielded only 68.8% success rate, where in 95% of the cases the initialization error was below 19.5 mm. Spine imaging is the focus of the 2D/3D registration proposed in [134], where DRRs are generated from each vertebra and used to generate Hough

space parametrization of the imaging data. The success rate of the registration using Generalized Hough Transform is 95.73%. The interactive initialization method in [133] with average interaction time 132.1 ± 66.4 s has an error of 7.4 ± 5.0 mm.

RGBD camera is mounted on the image intensifier that allows the X-ray source to remain under the surgical bed, hence the workflow is not disrupted. Despite the benefits of mounting the camera near the detector, large distance between RGBD camera and the X-ray source (approximately 1000 mm) may result in significant propagation of error. In other words, minute tracking error in the RGBD coordinate frame, may produce large errors in the X-ray source coordinate frame.

The depth-only algorithm performs a global tracking and utilizes points from foreground as well as the background. Therefore, patient movement in a static background is not tracked. Moreover, this method requires a complex scene with dominant structures. Lastly, the tracking is only maintained when the C-arm is re-arranged slowly. On the other hand, RGBD SLAM uses color features and assigns a dense feature area in the image (e.g. surgical site) as the foreground, and discards the background. Therefore, RGBD SLAM accounts for patient rigid movement.

The errors from the vision-based tracking shows that the translational components are subject to larger errors than the rotational components. The errors depend on the choice of the color features and the accuracy of depth data. The error in tracking the X-ray source using the RGBD-SLAM is the result of unreliable color features, drift from the frame-to-frame tracking, large translations of the RGBD camera in

the world coordinate frame, and the propagation of error due to the large distance between the X-ray and the RGBD imaging devices. Fast C-arm movement, as well as the presence of reflective and dark objects in the surgical scene that reflect and absorb IR are factors that result in poor tracking. Tracking accuracy can be further improved by incorporating application specific details such as drape color and using model-based tracking methods based on the known representations of the surgical tools. Additionally, the drift caused by frame-to-frame tracking can be reduced using bundle adjustment. Other improvements can take place by using stereo RGBD cameras on the C-arm, or using a combined tracking algorithm with both RGBD SLAM and depth-only. Alternative feature detectors that provide better representations can be used to extract relevant information from the surgical scene. Redundancy tests may also be necessary to reject color features where the depth information is sparse or unreliable. Lastly, high-end RGBD imaging devices with more reliable depth information can significantly improve the tracking quality.

The tracking accuracy of the vision-based methods presented in this chapter are within the capture range of an intensity-based image registration, and provide reliable registration. We believe that an RGBD enhanced C-arm system can contribute to automated fusion of pre- and intra-interventional data.

4.6 Acknowledgments

I want to thank Drs. Mehran Armand, Nassir Navab, and Russell Taylor for their insightful comments and supervision. Special thanks to Dr. Bernahrd Fuerst for his help during the design of the system.

Chapter 5

3D/3D: Automatic Intra-Operative Stitching of Non-Overlapping Cone-Beam CT Acquisitions

5.1 Motivation and Problem Statement

Intra-operative 3D X-ray Cone-Beam Computed Tomography (CBCT) during orthopedic and trauma surgeries has the potential to reduce the need of revision surgeries [25] and improve patient safety. Several works have emphasized the advantages that C-arm CBCT offers for guidance in orthopedic procedures for head and neck surgery [143, 144], spine surgery [145], and (Kirschner wire) K-wire placement in pelvic fractures [93, 94]. Other medical specialties, such as angiography [146],

dentistry [147] or radiation therapy[148], have reported similar benefits when using CBCT. However, commonly used CBCT devices exhibit a limited field of view of the projection images, and are constrained in their scanning motion. The limited view results in reduced effectiveness of the imaging modality in orthopedic interventions, particularly in imaging long bones, due to the small volume reconstructed.

To produce larger volumes, panoramic CBCT is proposed in [148] by stitching overlapping X-ray images acquired from the anatomy. Reconstruction quality is ensured by requiring sufficient overlap of the projection images, which in return increases the X-ray dose. Moreover, the reconstructed volume is vulnerable to artifacts introduced by image stitching. An automatic 3D image stitching technique is proposed in [149]. Under the assumption that the orientational misalignment is negligible, and sub-volumes are only translated, the stitching is performed using phase correlation as a global similarity measure, and Normalized Cross Correlation (NCC) as the local cost. Since NCC depends only on information in the overlapping area of the 3D volumes, sufficient overlap between 3D volumes is imperative. To reduce the X-ray exposure, Lamecker *et al.*[150] incorporated prior knowledge from statistical shape models to perform 3D reconstruction. To optimally support the surgical intervention, our focus are CBCT alignment techniques that do not require the change of workflow or additional devices in the operating theater. To avoid excessive radiation, we assume that no overlap between CBCT volumes exists (Fig. 5.1). Therefore, this chapter presents the first work that stitches fractured bones, recovers the necessary

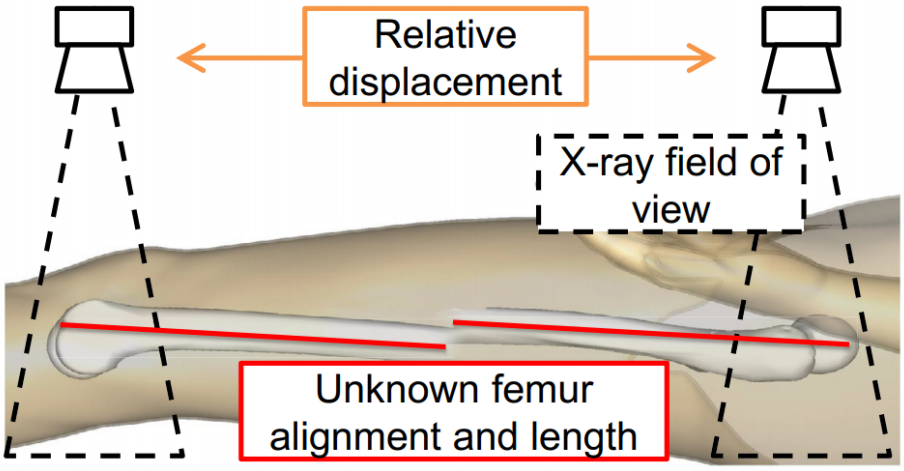

Figure 5.1: The 3D misalignment of bones (red lines) may be difficult to quantify using 2D images. CBCT contributes as a valuable tool for interventions in which the 3D alignment is of importance, for instance in acute fracture treatment or joint replacement. Background image courtesy of BodyParts3D, Center for Life Science, Japan.

information regarding their total length, and verifies mal-rotations independent of the acquisitions overlaps. We also highlight that recovery of such critical information is accomplished without imaging the mid-shaft of the fractured bone.

To avoid the introduction of additional devices, such as computer or camera carts, we co-register the X-ray source to a color and depth camera, and track the C-arm relative to the patient based on the RGBD observations [67, 92, 151, 152]. This allows the mobile C-arm to remain self-contained, and independent of additional devices or the operating theater. Additionally, the image quality of each individual CBCT volume remains intact, and the radiation dose is linearly proportional to the size and number of individual CBCT volumes.

5.2 Vision-Based Stitching Techniques for Non-Overlapping CBCT Volumes

After calibration of the cameras and X-ray source, the intrinsics and extrinsics of each imaging device are known. The calibration allows to track the patient using the RGB camera or depth sensor and apply this transformation to the CBCT volumes. In Sections 5.2.1- 5.2.3 we introduce stitching using visual markers, 3D color features, and surface depth information.

5.2.1 Vision-based Marker Tracking Techniques

This tracking technique relies on flat markers with a high contrast pattern that are easily detected in an image. The pose can be retrieved as the true marker size is known [153]. In the following we investigate two approaches: first, a cubical marker is placed near the patient and the detector is above the bed. Second, an array of markers is attached under the bed and the detector is below the bed while the C-arm is re-positioned.

Visual Marker Tracking of Patient: To enable visual marker tracking, we deploy a multi-marker strategy and arrange markers on all sides of a cube, resulting in an increased robustness and pose-estimation accuracy. The marker cube is then rigidly attached to the anatomy of interest and tracked using the RGB stream of the camera.

After performing the orbital rotation and acquiring the projection images for the

reconstruction of the first CBCT volume, the C-arm is rotated to a pose R for which the projection matrix $\mathbf{P}_{\mathtt{R}}$ is known. Ideally, this pose is chosen to provide an optimal view of relative displacement of the marker cube, as the markers are tracked based on the color camera view. The center of the first CBCT volume is defined to be the world origin, and the marker cube M is represented in this coordinate frame based on camera to X-ray source calibration:

$$^{\mathtt{CBCT}}\mathbf{T}_{\mathtt{M}} = ^{\mathtt{CBCT}}\mathbf{T}_{\mathtt{X}} \cdot ^{\mathtt{X}}\mathbf{T}_{\mathtt{RGB}} \cdot ^{\mathtt{RGB}}\mathbf{T}_{\mathtt{M}}. \tag{5.1}$$

The transformations are depicted in Fig. 5.2. The surgical table or the C-arm is re-positioned to acquire the second CBCT volume. During this movement, the scene and the marker cube are observed using the color camera, allowing for the computation of the new pose of the marker cube $^{\mathtt{RGB}'}\mathbf{T}_{\mathtt{M}'}$. Under the assumption that the relationship between CBCT volume and marker (Eq. 5.1) did not change as the marker remained fixed to the patient for the duration between two CBCT scans, the relative displacement of the CBCT volumes is expressed as:

$$\begin{aligned} ^{\mathtt{CBCT}'}\mathbf{T}_{\mathtt{X}} &= ^{\mathtt{CBCT}}\mathbf{T}_{\mathtt{M}} \cdot ^{\mathtt{RGB}}\mathbf{T}_{\mathtt{M}'}^{-1} \cdot ^{\mathtt{X}}\mathbf{T}_{\mathtt{RGB}}^{-1}, \\ ^{\mathtt{CBCT}'}\mathbf{T}_{\mathtt{CBCT}} &= ^{\mathtt{CBCT}'}\mathbf{T}_{\mathtt{X}} \cdot ^{\mathtt{CBCT}}\mathbf{T}_{\mathtt{X}}^{-1}. \end{aligned} \tag{5.2}$$

Visual Marker Tracking of Surgical Table: In many orthopedic interventions, the C-arm is used to validate the reduction of complex fractures. This is mostly done by moving the C-arm rather than the injured patient. Consequently, we hypothesize

that the patient remains on the surgical table and only the relationship between table and C-arm is of interest, which has also been assumed in previous work [78].

A pre-defined array of markers is mounted on the bottom of the surgical table, which allows the estimation of the pose of the C-arm relative to the table. While re-arranging the C-arm to acquire multiple CBCT scans, the C-arm detector is positioned under the bed where the RGBD camera observes the array of markers. Again, this allows for the estimation of ${}^{\mathtt{RGB}}\mathbf{T}_{\mathtt{M}}$, and, thus, stitching.

5.2.2 RGBD Simultaneous Localization and Mapping for Tracking

RGBD devices allow for fusion of color and depth information and enable scale recovery of visual features. We aim at using RGB and depth channels concurrently to track the displacement of patient relative to a C-arm during multiple CBCT acquisitions.

Simultaneous Localization and Mapping (SLAM) has been used in the past few decades to recover the pose of a sensor in an unknown environment. The underlying method in SLAM is the simultaneous estimation of the pose of perceived landmarks, and updating the position of a sensing device [121]. An RGBD SLAM was introduced in [98] where the visual features are extracted from 2D frames, and later the depth associated with those features are computed from the depth sensor in the RGBD

camera. These 3D features are then used to initialize a RANdom SAmple Consensus (RANSAC) method to estimate the relative poses of the sensor by fitting a 6 DOF rigid transformation [138].

RGBD SLAM enables the recovery of the camera trajectory in an arbitrary environment without prior models; rather, SLAM incrementally creates a global 3D map of the scene in real-time. We assume that the global 3D map is rigidly connected to the CBCT volume, which allows for the computation of the relative volume displacement using Eq. 5.3, where $\mathbf{f}_{\texttt{RGB}}$ and $\mathbf{f}_{\texttt{RGB'}}$ are sets of features in `RGB` and `RGB'` frames, π is the projection operator, $\mathbf{d}$ is the dense depth map, and $\mathbf{x}$ is the set of 2D feature points.

$$\begin{aligned} {}^{\texttt{RGB'}}\hat{\mathbf{T}}_{\texttt{RGB}} &= \underset{{}^{\texttt{RGB'}}\mathbf{T}_{\texttt{RGB}} \in \mathbb{SE}(3)}{\arg\min} \left| \mathbf{f}_{\texttt{RGB}}\left(\mathbf{x}\right) - \mathbf{f}_{\texttt{RGB'}}\left(\pi^{\texttt{RGB'}}\mathbf{T}_{\texttt{RGB}}\left(\mathbf{dx}\right)\right)\right|, \\ {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}} &= {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{RGB}} \cdot {}^{\texttt{RGB'}}\hat{\mathbf{T}}_{\texttt{RGB}} \cdot {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{RGB}}^{-1}. \end{aligned} \tag{5.3}$$

5.2.3 Surface Reconstruction and Tracking Using Depth Information

Surface information obtained from the depth sensor in an RGBD camera can be used to reconstruct the patient's surface, which simultaneously enables the estimation of the sensor trajectory. KinectFusion provides a dense surface reconstruction of a complex environment and estimates the pose of the sensor in real-time [123]. Our goal is to use the depth camera view and observe the displacement, track the scene,

and consequently compute the relative movement between the acquisition of CBCT volumes. This tracking method involves no markers, and the surgical site is used as reference (real-surgery condition).

KinectFusion relies on a multi-scale Iterative Closest Point (ICP) with a point-to-plane distance function and registers the current measurement of the depth sensor to a globally fused model. The ICP incorporates points from both the foreground as well as the background and estimates rigid transformations between frames. Therefore, a moving object with a static background causes unreliable tracking. Thus, multiple non-overlapping CBCT volumes are only acquired by re-positioning the C-arm instead of the surgical table.

Similar transformations shown in Fig. 5.2 are used to compute the relative CBCT displacement ${}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}}$, where D defines the depth coordinate frame, ${}^{\texttt{D'}}\hat{\mathbf{T}}_{\texttt{D}}$ is the relative camera pose computed using KinectFusion, $\mathbf{V}_{\texttt{D}}$ and $\mathbf{V}_{\texttt{D'}}$ are vertex maps at frames D and D', and $\mathbf{N}_{\texttt{D}}$ is the normal map at frame D:

$$
\begin{aligned}
{}^{\texttt{D'}}\hat{\mathbf{T}}_{\texttt{D}} &= \underset{{}^{\texttt{D'}}\mathbf{T}_{\texttt{D}} \in \mathbb{SE}(3)}{\arg\min} \left\| ({}^{\texttt{D'}}\mathbf{T}_{\texttt{D}}^{-1}\mathbf{V}_{\texttt{D'}} - \mathbf{V}_{\texttt{D}})^{\top}\mathbf{N}_{\texttt{D}} \right\|_2, \\
{}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}} &= {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{D}} \cdot {}^{\texttt{D'}}\hat{\mathbf{T}}_{\texttt{D}} \cdot {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{D}}^{-1}.
\end{aligned}
\tag{5.4}
$$

5.3 Reference Techniques

To provide a reasonable reference to our vision-based tracking techniques, we briefly introduce an infrared tracking system to perform CBCT volume stitching

(Sec. 5.3.1). This chapter concludes with a brief overview of our previously published vision-based stitching technique [119] (Sec. 5.3.2) put in context of our chain of transformations.

5.3.1 Infrared Tracking System

In the following we first discuss the **calibration** of the C-arm to the CBCT coordinate frame, and subsequently the C-arm to patient **tracking** using this calibration.

Calibration: This step includes attaching passive markers to the C-arm and calibrating them to the CBCT coordinate frame. This calibration later allows us to close the patient, CBCT, and C-arm transformation loop and estimate relative displacements. The spatial relation of the markers on the C-arm with respect to the CBCT coordinate frame is illustrated in Fig. 5.3 and is defined as:

$$^{\texttt{CBCT}}\mathbf{T}_{\texttt{Carm}} = ^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}} \,.\, ^{\texttt{Carm}}\mathbf{T}_{\texttt{IR}}^{-1}. \tag{5.5}$$

The first step in solving Eq. (5.5) is to compute $^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}}$. This estimation requires at least three marker positions in both CBCT and IR coordinate frames. Thus, a CBCT scan of another set of markers (M in Fig. 5.3) is acquired and the spherical markers are located in the CBCT volume. Here, we attempt to directly localize the spherical markers in the CBCT image instead of X-ray projections [154]. To this end, a bilateral filter is applied to the CBCT image to remove noise while preserving

edges. Next, weak edges are removed by thresholding the gradient of the CBCT, while strong edges corresponding to the surface points on the spheres are preserved. The resulting points are clustered into three partitions (one cluster per sphere), and the centroid of each cluster is computed. Then an exhaustive search is performed in the neighborhood around the centroid with the radius of $\pm(r+\delta)$, where r is the sphere radius (6.00 mm) and δ is the uncertainty range (2.00 mm). The sphere center is localized by a least-square minimization using its parametric model. Since the sphere size is provided by the manufacturer, we avoid using classic RANSAC or Hough-like methods as they also optimize over the sphere radius. We then use the non-iterative least-squares method suggested in [155] and solve for ${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}}$ based on singular value decomposition. Consequently, we can close the calibration loop and solve Eq. 5.5 using ${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}}$, and ${}^{\texttt{Carm}}\mathbf{T}_{\texttt{IR}}$ which is directly measured from the IR tracker.

Tracking: The tracking stream provided for each marker configuration allows for computing the motion of the patient. After the first CBCT volume is acquired, the relative patient displacement is estimated before the next CBCT scan is performed.

Considering the case where the C-arm is re-positioned (from `Carm` to `Carm'` coordinate frame) to acquire CBCT volumes (`CBCT` and `CBCT'` coordinate frames), and the patient is fixed on the surgical table, the relative transformation from `IR` tracker

to CBCT volumes are defined as follows:

$$\begin{aligned} {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}} &= {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{Carm}} \cdot {}^{\texttt{Carm}}\mathbf{T}_{\texttt{IR}}, \\ {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{IR}} &= {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{Carm'}} \cdot {}^{\texttt{Carm'}}\mathbf{T}_{\texttt{IR}}. \end{aligned} \tag{5.6}$$

The relation between the C-arm and the CBCT is fixed, hence ${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{Carm}} \overset{\text{def}}{=} {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{Carm'}}$. We can then define the relative transformation from CBCT to CBCT' as:

$${}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}} = {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{IR}} \cdot {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}}^{-1}. \tag{5.7}$$

To consider patient movement, markers (coordinate frame M in Fig. 5.3) may also be attached to the patient (e. g. screwed into the bone), and tracked in the IR tracker coordinate frame. ${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{M}}$ is then defined as:

$${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{M}} = {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{Carm}} \cdot {}^{\texttt{Carm}}\mathbf{T}_{\texttt{IR}} \cdot {}^{\texttt{M}}\mathbf{T}_{\texttt{IR}}^{-1}. \tag{5.8}$$

Assuming that the transformation between CBCT and marker is fixed during the intervention (${}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{M}} \overset{\text{def}}{=} {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{M}}$) and combining Eq. 5.6 and Eq. 5.8, volume poses in the tracker coordinate frame are defined as:

$$\begin{aligned} {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{IR}} &= {}^{\texttt{CBCT}}\mathbf{T}_{\texttt{M}} \cdot {}^{\texttt{M}}\mathbf{T}_{\texttt{IR}}, \\ {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{IR}} &= {}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{M'}} \cdot {}^{\texttt{M'}}\mathbf{T}_{\texttt{IR}}. \end{aligned} \tag{5.9}$$

solving Eq. 5.9 leads to recovery of CBCT displacement using Eq. 5.7.

5.3.2 Two-dimensional Feature Tracking

We use the camera attached to the mobile C-arm, and the positioning laser in the base of the C-arm to recover the 3D depth scales, and consequently stitch the sub-volumes as presented in [119]. The positioning-laser in the base of the C-arm spans a plane which intersects with the unknown patient surface. The laser line can be observed as a curve in the camera image, and used to approximate the scale of features nearby.

Calibration: To determine the relationship between camera and laser plane, we perform a calibration using multiple checkerboard poses. At each of the n poses the laser intersects the origin of the checkerboard, which allows us to recover points on the laser plane in the camera coordinate frame. By performing RANSAC-based plane fitting, the plane coefficients are computed.

Tracking: Following [119], the tracking algorithm comprises the following steps: *(i)* Automatic detection of Speeded Up Robust Features (SURF) in every frame; *(ii)* Matching features between frames, and rejecting outliers by estimating the Fundamental Matrix; *(iii)* Automatically detecting the laser line and computing the 3D shape based on the known laser plane; *(iv)* Recovering the scale of the features using the scale of the nearby laser line; *(v)* Estimating the 3D transformation for the sets of 3D features; and *(vi)* Validating transformation estimation by applying it to 3D laser line. Fig. 5.4 illustrates the feature relations across multiple frames. Finally, the frame-by-frame transformations are accumulated to estimate ${}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}}$.

5.4 Experiments and Results

In this section, we report the results of our vision-based methods to stitch multiple CBCT volumes as presented in Sec. 5.2. The same experiments are preformed using the methods outlined in Sec. 5.3, namely using a commercially available infrared tracking system, and our previously published technique [119]. Finally, we compare the results of the aforementioned approaches to image-based stitching of overlapping CBCT volumes.

To acquire a CBCT volume, the patient is positioned under guidance of the lasers. Then, the motorized C-arm orbits 190° around the center visualized by the laser lines, and automatically acquires a total of 100 2D X-ray images. Reconstruction is performed using a maximum-likelihood expectation-maximization iterative reconstruction method [110], resulting in a cubic volume with 512 voxels along each axis and an isotropic voxel size of 0.2475 mm. For the purpose of reconstruction, we use the following geometrical parameters provided by the manufacturer: *source-to-detector distance*: 980.00 mm, *source-iso-center distance*: 600.00 mm, *angle range*: 190°, *detector size*: 230.00 mm × 230.00 mm.

5.4.1 Stitching Results

Our vision-based tracking methods are all tested and evaluated on an animal cadaver (pig femur). For these experiments, we performed the stitching of CBCT

volumes with each method individually under realistic surgery conditions. The C-arm was translated for the acquisition of multiple CBCT volumes when the detector was located at AP orientation. The geometric relation of the AP view to the C-arm was estimated using an intensity-based 2D/3D registration with a target registration error of 0.29 mm. Subsequently, we measured the absolute distance between the implanted landmarks inside the animal cadaver and compared the results to a ground-truth acquired from a CT scan. The outcome of these experiments were compared to an infrared-based tracking approach (baseline method), as well as image-based stitching approach. Stitching errors for all proposed methods are reported in Table 5.1.

The lowest tracking error of 0.33 ± 0.30 mm is achieved by tracking the cubical visual marker attached to the patient. Marker-less stitching using RGBD-SLAM exhibits sub-millimeter error (0.91 mm), while tracking only using depth cues results in a higher error of 1.72 mm. The alignment of CBCT volumes using an infrared tracker also has errors larger than a millimeter. The stitching of overlapping CBCT volumes yielded a substantially higher error (9.27 mm) compared to every other method in Sec. 5.2 and 5.3. Fig 5.6 shows the convergence of the registration cost when stitching using image information. In Table 5.1, we also report the angles between the mechanical and the anatomical axes of the femur (Tibio Femoral Angle), as well as the angle between the mechanical axis and the knee joint line (Lateral-distal Femoral Angle) using the vision-based stitching methods. The results indicate minute variations among different methods. In Fig. 5.5, the non-overlapping stitching of the

CBCT volumes of the pig femur are shown.

These methods are also evaluated on a long radiopaque femur phantom. The stitched volumes are shown in Fig. 5.7, and the stitching errors for each method are reported in Table. 5.2.

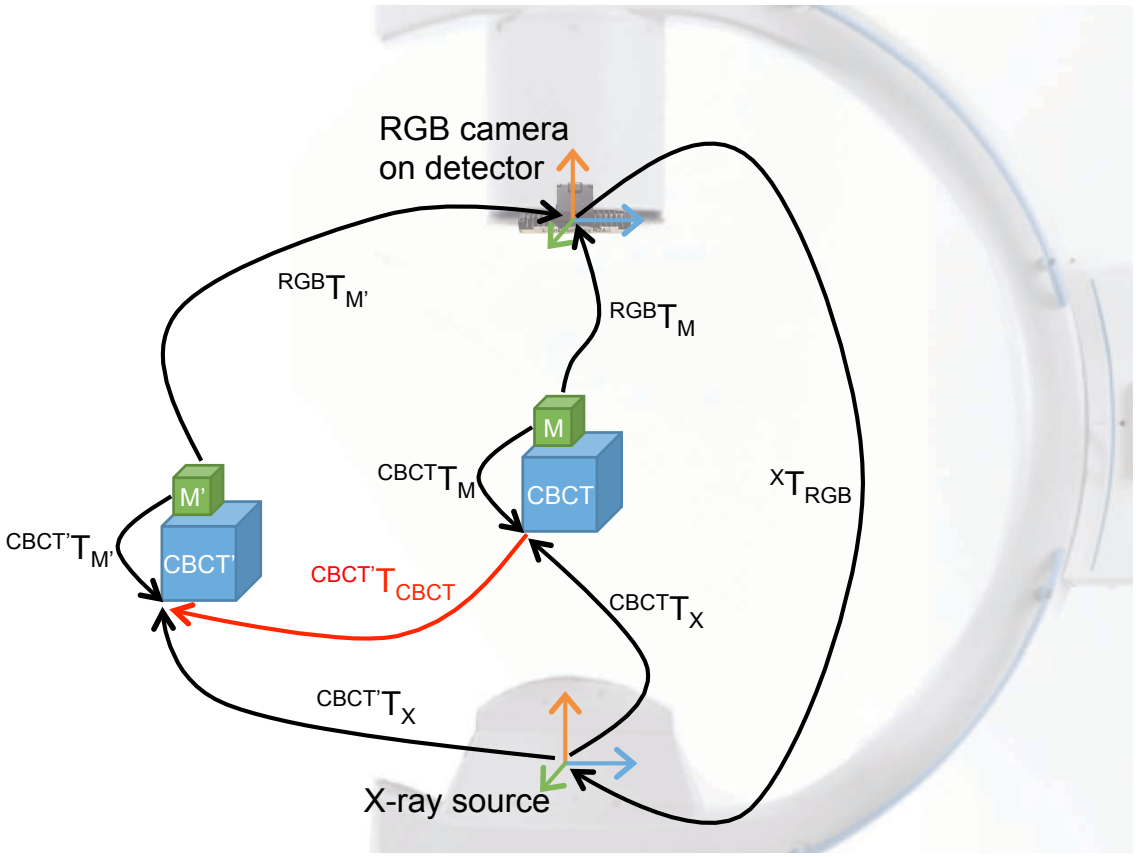

Figure 5.2: The relative displacement of CBCT volumes (${}^{\texttt{CBCT'}}\mathbf{T}_{\texttt{CBCT}}$) is estimated from the tracking data computed using the camera mounted on the C-arm. This requires the calibration of camera and X-ray source (${}^{\texttt{X}}\mathbf{T}_{\texttt{RGB}}$), and the known relationship of X-ray source and CBCT volume (${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{X}}$). The pose of the marker is observed by the camera (${}^{\texttt{RGB}}\mathbf{T}_{\texttt{M}}$), while the transformation from marker pose to CBCT volume (${}^{\texttt{CBCT}}\mathbf{T}_{\texttt{M}}$) is computed once and assumed to remain constant.

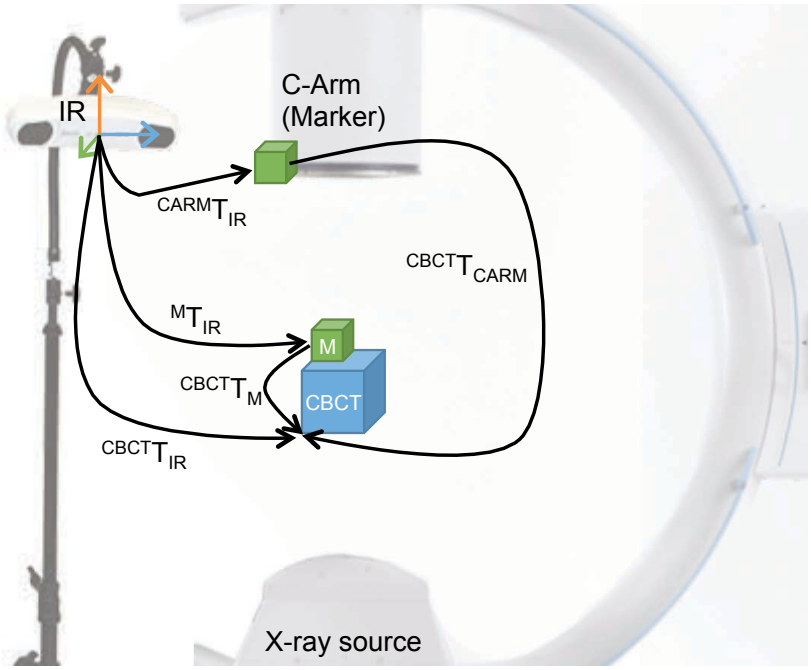

Figure 5.3: An infrared tracking system is used for alignment and stitching of CBCT volumes. This method serves as reference standard for the evaluation of vision-based techniques.

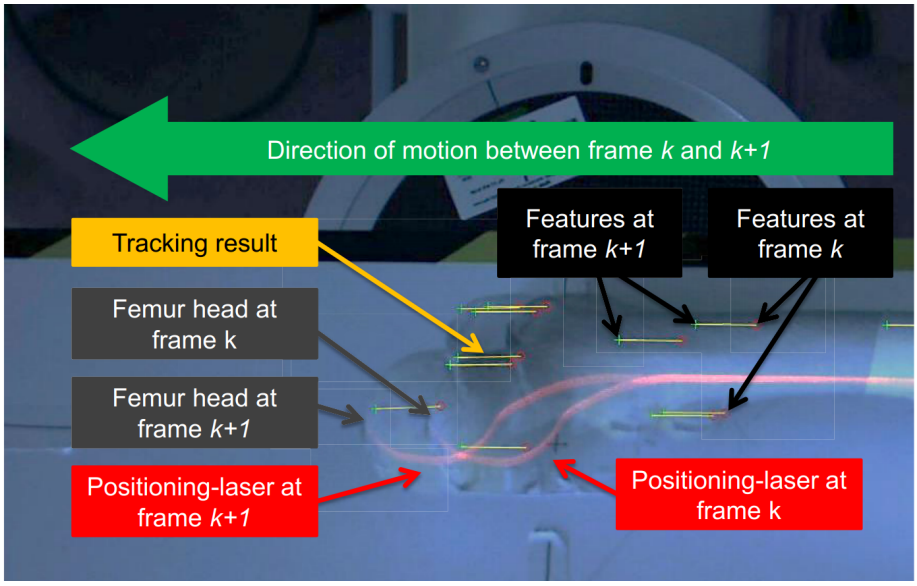

Figure 5.4: The figure shows the overlay of two frames to illustrate the feature correspondences to estimate the movement of the patient. From both frames, the positioning-laser (red) and natural surface features are extracted. The tracking results of the matched features in frame k (+) and frame k +1 (∘) are illustrated as yellow lines.

Table 5.1: Errors are computed by measuring the average of the absolute distances between 8 radiolucent landmarks implanted into the femur head, greater trochanter, patella, and the condyle. The residual distances are measured between the opposite sides of the femur (hip to knee). Errors in angular measurements for tibio femoral (TF) and lateral-distal femoral (LDF) are reported in the last two columns. Each method is tested twice on the animal cadaver. The C-arm translation was nearly 210 mm to acquire each non-overlapping CBCT volume. The first four rows present the results using vision-based methods suggested in this chapter. We then present the errors of registration using external trackers as well as image-based stitching of overlapping CBCT volumes with NCC similarity measure. Note that in this table the results of stitching using 2D features (Sec. 5.3.2) are not presented as measurements on a similar animal specimen were not reported in [119]. All errors are measured by comparing the stitching measurements with the measurements from a complete CT of the porcine specimen as ground-truth.

Tracking Method	**Stitching Error (mm)**	**Standard Deviation (mm)**	**Absolute Distance Error (%)**	**TF Error (°)**	**LDF Error (°)**
Marker Tracking of Patient (Sec. 5.2.1-a)	0.33	0.30	0.14	0.6°	0.5°
Marker Tracking of Surgical Bed (Sec. 5.2.1-b)	0.62	0.21	0.26	0.7°	2.2°
RGBD-SLAM Tracking (Sec. 5.2.2)	0.91	0.59	0.42	0.5°	0.6°
Surface Data Tracking (Sec. 5.2.3)	1.72	0.72	0.79	1.0°	3.1°
Infrared Tracking (Sec. 5.3.1)	1.64	0.87	0.73	0.3°	2.4°
Image-based Registration	9.27	2.11	6.52	1.2°	2.7°

Table 5.2: The errors on a long femur phantom are reported similar to the measurements in Table 5.1. The length from the femur neck to the intercondylar fossa of the dry phantom is approximately 369 mm. To measure the distance errors, a total of 12 landmarks are attached to the femur (6 metal beads on each end). Stitching with each method is repeated three times, and all errors are computed by comparing the measurements to the ground-truth measurements in a CT scan of the phantom.

Tracking Method	**Stitching Error (mm)**	**Standard Deviation (mm)**	**Absolute Distance Error (%)**	**TF Error (°)**	**LDF Error (°)**
Marker Tracking of Patient (Sec. 5.2.1-a)	0.59	0.37	0.20	0.8°	2.9°
Marker Tracking of Surgical Bed (Sec. 5.2.1-b)	0.66	0.18	0.23	0.7°	2.3°
RGBD-SLAM Tracking (Sec. 5.2.2)	1.01	0.41	0.38	0.8°	0.9°
Surface Data Tracking (Sec. 5.2.3)	2.53	1.11	0.87	1.9°	4.1°
Infrared Tracking (Sec. 5.3.1)	1.76	0.99	0.61	1.1°	2.7°
2D Feature Tracking (Sec. 5.3.2) [119]	1.18	0.28	0.62	-	-
Image-based Registration	68.6	22.5	23.4	3.9°	5.2°

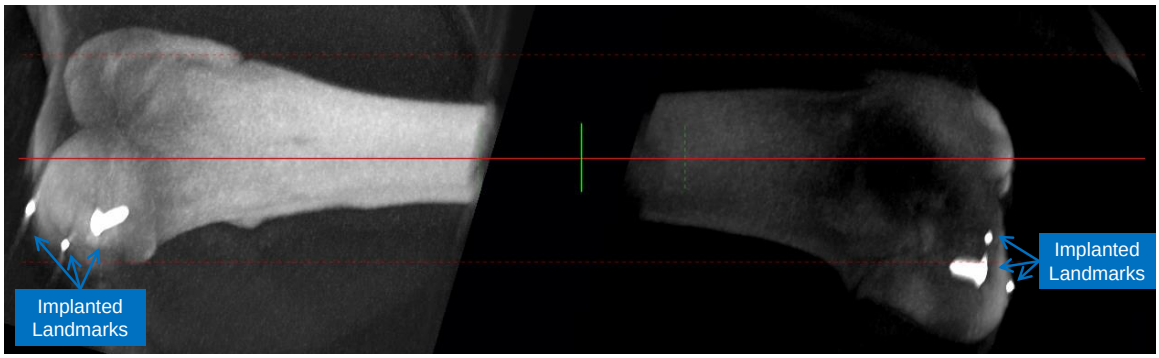

Figure 5.5: Parallel projection through both CBCT volumes to create a DRR-like visualization.

5.5 Discussion and Conclusion

The stitching solutions we suggested allowed for tracking of the patient or C-arm movement with minimal increase of workflow complexity and without introduction of external tracking systems. We attached an RGB and depth camera to a mobile C-arm, and deployed computer vision techniques to track changes in C-arm pose and, consequently, stitch the sub-volumes. The proposed methods employ visual marker tracking, RGBD-based SLAM, and surface tracking by fusing depth data to a single global surface model. These approaches estimate the relative CBCT volume displacement based on only RGB, a combination of RGB and depth, or only depth information. As a result, stitching is performed with lower dose, linearly proportional to the size of non-overlapping sub-volumes. We anticipate our methods to be particularly appropriate for intraoperative planning and validation for long bone fractures or joint replacement interventions, where multi-axis alignment and absolute distances

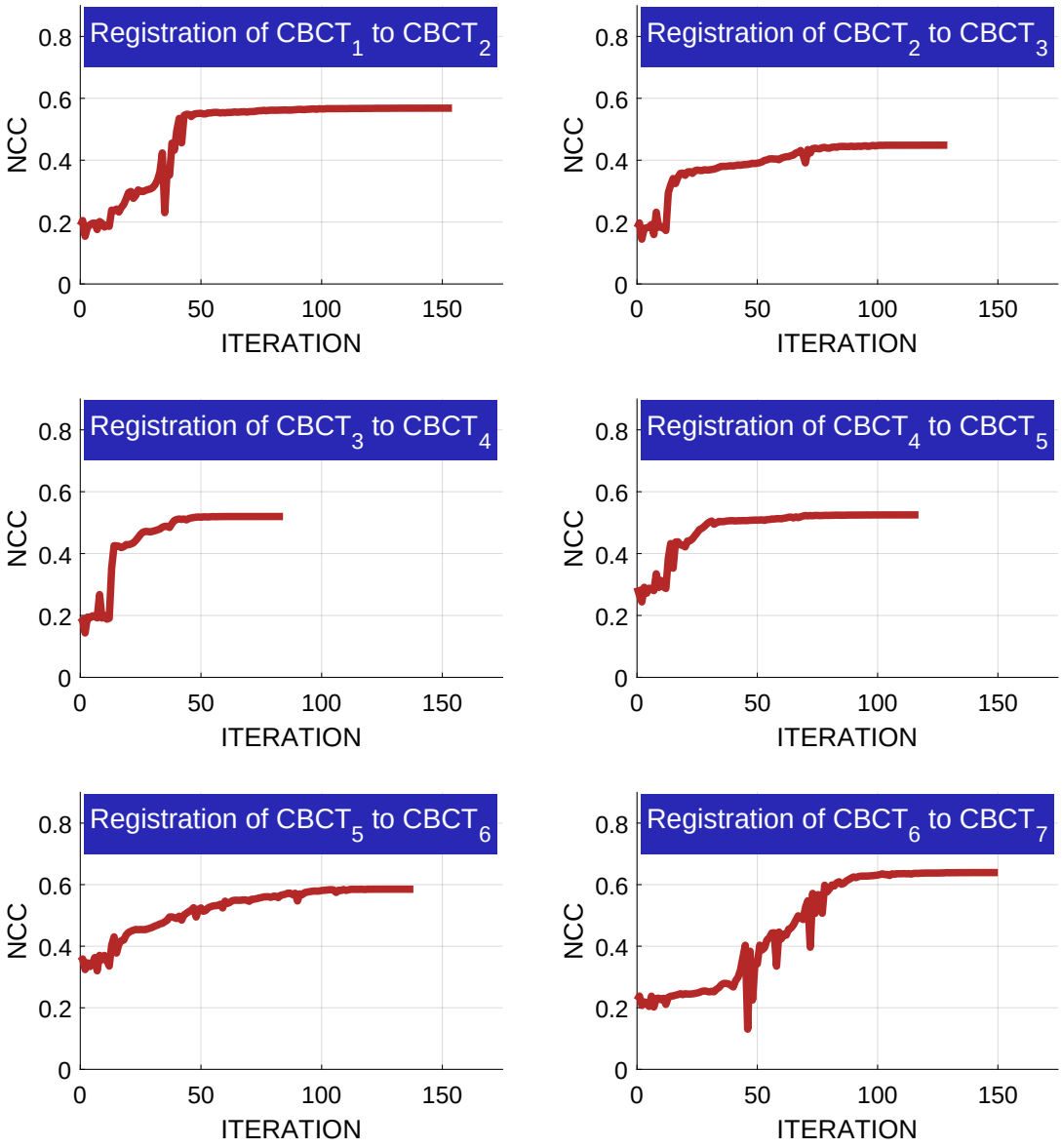

Figure 5.6: Optimization of the NCC similarity cost for registering multiple overlapping CBCT volumes. Seven CBCT scans were acquired to image the entire phantom. Every two consecutive CBCT scans were acquired with 50.0 mm to 60.0 mm in-plane translation of the C-arm in between to ensure nearly half volume overlap (CBCT volume size along each dimension is 127 mm). The optimization never reached the maximum number of iteration threshold that was set to 500. Image-based registration was performed on the original volumes, with no filtering or down-sampling of the images. The NCC similarity measure reached higher values (0.6 ± 0.04) when registering CBCT volumes acquired from the two ends of the bone which had more dominant structures, and yielded lower similarity scores at the shaft of the phantom.

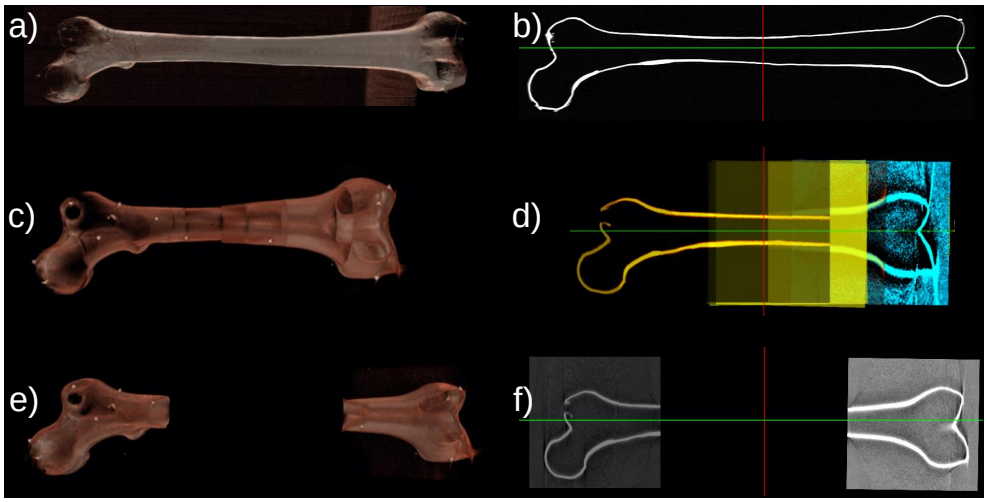

Figure 5.7: **(a-b)**, **(c-d)**, and **(e-f)** are volume rendering and single CT slice of the CT, stitched volume using image-based registration, and the non-overlapping stitched volume, respectively. Image-based registration shown in (c-d) uses seven overlapping CBCT volumes and results in significantly shorter total length of the bone (results in Table 5.2). This incorrect alignment is due to insufficient amount of information in the overlapping region, especially for volumes acquired from the shaft of the bone. Shaft of the bone is a homogeneous region where the registration optimizer converges to local optima.

are difficult to visualize and measure from the 2D X-ray views.

During the re-arrangement of the C-arm and the patient for the next CBCT acquisition, the vision-based tracking results are recorded. For this re-arrangement we consider the clinically realistic scenario of a moving C-arm and a static patient. However, as extensively discussed in Sec. 5.2, for marker-based methods the relative movement of the patient to C-arm is recorded, hence there are no limitations on allowed motions.

We performed the validation experiments on an animal cadaver, and compared the non-overlapping stitching outcome to an infrared tracking system and image-

based registration using overlapping CBCT volumes. In these experiments we used a CT scan of the animal cadaver as the ground truth data. The visual marker-based tracking achieved the lowest tracking error (0.33 mm) among all methods. The high accuracy is due to utilizing a multi-marker strategy which avoids tracking in shallow angles. The RGBD camera has a larger field of view compared to the X-ray imaging device. Therefore, the marker can be placed in the overlapping camera views. For example, in the case of imaging the femoral head and the condyle, the visual marker can be placed near the femoral shaft. The marker only needs to remain fixed with respect to the patient for the duration which the C-arm is re-positioned and need not be present for the CBCT acquisitions. Therefore, certain clinical limitations, such as changes to the scene, draping, patient movement, or presence of surgical tools in the scene are not limiting factors.

Stitching based on the tracking with RGB and depth information together has 0.91 mm error, and tracking solely based on depth information has 1.72 mm error. In a clinically realistic scenario, the surgical site comprises drapes, blood, exposed anatomy, and surgical tools which allows the extraction of large number of useful color features in a color image. The authors believe that a marker-less RGBD-SLAM stitching system can use the aforementioned color information, as well as the depth information from the co-calibrated depth camera, and provide reliable CBCT image stitching for orthopedic interventions.

The use of external infrared tracking systems to observe the displacement of pa-

tients are widely accepted in clinical practice, and are usually not deployed to automatically align and stitch multiple CBCT volumes. A major disadvantage of external tracking systems is the introduction of additional hardware to the operating room, and the accumulation of tracking errors when tracking both the patient and C-arm.

The stitching errors of the vision-based methods are also compared to image-based stitching of overlapping CBCT volumes in Tables 5.1 and 5.2. Image-based approach yielded high errors for both the animal cadaver (9.27 mm) as well as the dry bone phantom (68.6 mm) because of insufficient and homogeneous information in the overlapping region. The errors are reported lower when registering the porcine specimen due to shorter length of the bone and presence of soft tissue in the overlapping region.

We also avoided stitching of projection images due to the potential parallax effect which causes incorrect stitching and the length and angles between the anatomical landmarks will not be preserved in the stitched volume.

The benefits of using cameras with a C-arm for radiation and patients safety, scene observation, and augmented reality has been emphasized in the past. This work presents a 3D/3D intraoperative image stitching technique using a similar opto-X-ray system. Our approach does not limit the working space, nor does it require any additional hardware besides one RGBD camera near the image intensifier. The C-arm remains mobile, self-contained, and independent of the operating room.

5.6 Acknowledgments

Sincere thanks to Dr. Bernahrd Fuerst for his support during the implementation and design of different tracking methods. I also thank Mr. Stefan Reichenstein for his help in imaging the porcine phantom and calibrating the system. Stefan acquired multiple series of checkerboard data from the imaging systems. I also thank Drs. Nassir Navab, Mehran Armand, and Mathias Unberath for their suggestions and constructive remarks.

Chapter 6

2D/2D: From Perspective X-ray Imaging to Parallax-Robust Orthographic Stitching

6.1 Introduction and Problem Definition

Any two images of a planar scene are related by a homography $\mathbf{H} \in \mathbb{R}^{3\times3}$, an invertible mapping. Since the homography expresses the warping of the planar image content, these images can easily be stitched. In the special case where the camera motion is limited to rotation about its origin, homography can also describe the relative mapping as if the scene was at infinity. In all other cases, however, if $\mathbf{H}$ is used in an attempt to stitch images, it creates ghosting effects, also known as

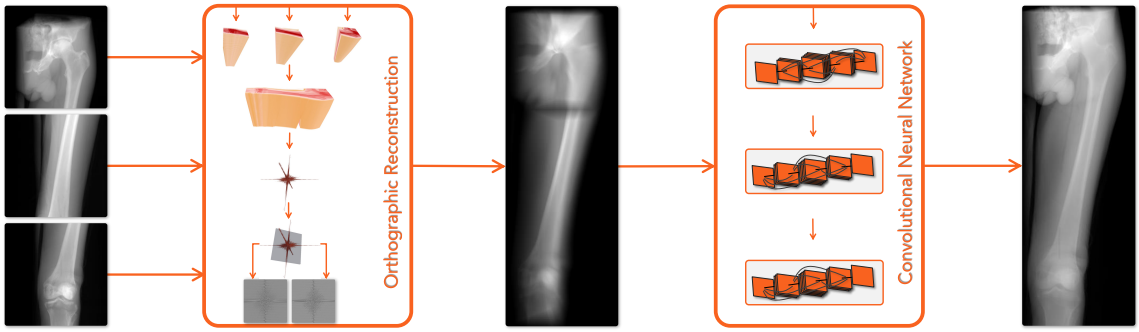

Figure 6.1: The image stitching pipeline includes orthographic 2D reconstruction of multiple 2D acquisitions, followed by restoration of image semantics using ConvNets.

parallax. This has practical implications for every application where multiple images are acquired to capture an extended view of an arbitrary scene. In this case, the motion of the camera would typically comprise of both rotational and translational components. Unfortunately, and as the result of camera translation, parallax occurs for objects at different depths. The parallax effect appears stronger for points closer to the camera, and weaker for points farther from the camera. Due to parallax and inconsistencies between views, no universal mapping exists for the stitching of 2D images.

The problem of stitching has been long investigated in computer vision for various applications, including scene rendering and video frame stitching [156, 157]. In its general form, since the mapping between images cannot be explained by a global homography, multiple early works addressed stitching by using local alignment strategies. In cases of minor translations with weak motion parallax, these works were able to deliver seamless mosaics by employing deghosting models based on local align-

ment within small patches in the overlapping region [158–160]. Blurry areas within the mosaics were reduced by limiting the sampling to the content of a single image per region, selected based on geometric information from prior segmentation [161]. To construct visually appealing renderings, smooth transitions of features were suggested based on different depth cues [162]. Later works introduced hybrid stitching schemes by combining global homography and content-preserving local warping to render natural-looking mosaics [163, 164]. Considering that all the aforementioned works failed to address the stitching problem in the presence of large motion parallax, an iterative stitching method was proposed to identify reliable local alignments and address depth mismatch with larger parallax [165]. Despite improving the warping between images, this work still suffered from poor stitching when parallax took place near the periphery of common regions.

In a clinical context, a common approach to stitching fluoroscopic images is to employ radiopaque planar markers that are placed parallel to the patient, approximately at the same depth from the camera as the anatomy [166–168]. For instance, Yaniv et al. stitched X-ray images using a radiopaque ruler. They considered a simplified motion for the X-ray camera, restricted to fronto-parallel acquisitions, such that the planar mapping between images was parameterized using only two *in-plane* translation and one *in-plane* rotation parameters, hence ignoring all *out-of-plane* variables. Their stitching pipeline was followed by a parallax correction step that estimated parallax errors via approximating reconstruction planes for each individual pixel. In

a different work, Wang et al. proposed a parallax-free X-ray image stitching methodology by restricting the X-ray camera to only undergo pure rotation [78, 169]. To compensate for translation between the C-arm camera and the scene, the patient bed was translated with the same translational parameters as the X-ray source. The movement of patient bed in the surgery room is not practical in most surgical environments. Instead, this solution can be delivered by a robotic C-arm platform that uses kinematics to enforce the rotational motion. Since robotic platforms typically cost an order of magnitude more than non-robotic scanners, from an economic standpoint, their availability for standard operative procedures will remain a concern.

With the availability of 3D intra-operative imaging, instead of 2D X-ray image stitching, recent works suggested stitching of tomographic volumes reconstructed from X-ray images acquired on a circular trajectory [119, 170]. While volumetric data do not suffer from the effects of parallax and can, therefore, be stitched easily, circular shortscans required for 3D tomographic reconstruction expose the patient to much higher radiation doses compared to simple X-ray acquisitions, suggesting that these methods are best employed only once, e.g., for verification.

We suggest an end-to-end solution that combines information from 2D X-ray acquisitions in parallax-free domains, i.e., in 3D spatial and 3D Fourier spaces, and provides stitching of X-ray images with no constraints on the motion of the X-ray camera. Our mapping for stitching is no longer a homography; instead, it directly uses the projection matrices. The global structure of the orthographically stitched

image is recovered by leveraging the Fourier slice theorem and principles governing image formation. The details in the stitched image are restored via a convolutional neural network (ConvNet) with regularization losses on frequency, as well as sparse and dense spatial features.

As the result of employing a data-driven approach for learning the stitching parametrization and performing image-based rendering, the model we present is anatomy-specific, therefore, not directly comparable to the previous work that was invariant to the content of the image but was constrained by the motion of the camera [78]. To this end, we only focused on images of healthy human femurs in this work.

Our object is to deliver a methodology that *i)* is invariant to parallax, *ii)* provides orthographic reconstruction which enables direct metric measurements on the image without using any priors, *iii)* is robust to minor gaps between input images such that missing information is recovered based on structure continuity, and *iv)* does not require explicit blending of content between multiple sources in their overlapping regions.

6.2 Methodology

Our solution to transmission image stitching shown in Fig. 6.1 is designed based on two fundamental steps. First, we leverage the close relationship between the Fourier

slice theorem and the Radon transform to provide a stitched image in an orthographic geometry from back-projected rays (Sec. 6.2.1). Second, to restore the missing and blurred content, we use a series of ConvNets with adversarial losses and regularizes for structural, intensity, contrast, frequency, and sparse feature similarities between input and ground-truth orthographic images (Sec. 6.2.2).

We simultaneously train an anatomical landmark detector network, which serves two purposes: *i)* it automatically detects anatomical landmarks that are critical to metric and angular measurements from the bone on the stitched image, and *ii)* it integrates into the stitching pipeline and enforces the network to predict images closer to the ground-truth domain such that the landmark detection with an identical network performs well on both the ground-truth and prediction domains.

6.2.1 Orthographic Reconstruction

The back-projection of each pixel element (x_m, y_n) in a 2D X-ray image $g(x_m, y_n)$ is defined as:

$$V(\mu(d, i, m, n)) = P_i^{+} g(x_m, y_n) \quad \text{, where} \tag{6.1}$$

$\mu(d, i, m, n)$ is the ray characterized by the projection matrix P_i and pixel coordinates (x_m, y_n) in the i-th image, and $V(.)$ is the volume constructed by smearing out μ into the 3D space. The parameter $d \in [0, 1]$ is used identically for all projections. It defines the depth of the backprojected volume, where 1 refers to backprojection within the

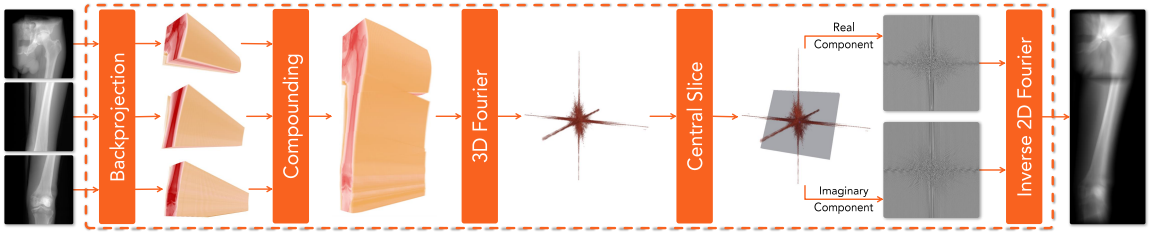

Figure 6.2: The orthographic reconstruction pipeline leverages the Fourier slice theorem to transform multiple 2D X-ray images acquired using cone-beam geometry, to a single extended-view image in parallel-beam geometry. In this example, the backprojection is performed with $d = 0.5$, hence in a volume between the detector plane and 50% of the focal length.

entire imaging cone, and 0.5 denotes backprojection between the detector plane and the depth equivalent of 50% of the focal length. The $(.)^+$ denotes the pseudo-inverse operation. Since all projection images are acquired from an identical static scene, the back-projected rays from all acquisitions can be compounded into a single volume as:

$$\Omega = \sum_{i,m,n} V(\mu(d, i, m, n)). \tag{6.2}$$

In the remainder of this section, we treat the stitching of multiple input images, as a reconstruction problem that aims to reconstruct an orthographic 2D view given the incomplete 3D data Ω. As shown in Fig. 6.2, within the orthographic image, the global structure of the scene is reconstructed with insufficient details due to *i)* loss of information in Ω as it is only constructed from sparse and single-view set of data, and *ii)* different binning and sampling of frequencies, particularly in the overlapping region, which can be explained by the Fourier-slice theorem [171].

Based on the Fourier slice theorem, also known as the projection slice theorem, the Fourier transform of an orthographic projection of a 3D function $f(x, y, z)$ in 2D represented as $G(\Theta)$, is equivalent to the 2D slice in the 3D Fourier of the function $f(x, y, z)$ that passes through the origin and is parallel to the projection plane Θ. Given the Fourier transform $F = \mathcal{F}(\Omega)$, we represent the central slice that passes through the origin and is parallel to the projection plane Θ_0 as $G(\Theta_0) = F(\theta = \Theta_0)$. Based on the Fourier slice theorem, the stitched image I in Fig. 6.2 is then computed as $I = \mathcal{F}^{-1}(G(\Theta_0))$.

6.2.2 Restoration of Stitching Semantics

In CT reconstruction, if X-ray images are acquired on a circular trajectory, oversampling occurs at the center of the Fourier domain. For the stitching problem in hand, since there is no analytical approach to identifying the appropriate filter for resampling and re-binning, as well as the loss of information from orthographic rendering of a perspective image, we suggest a learning strategy to jointly learn corrections in the spatial and frequency domains.

The learning framework, as shown in Fig. 6.3, consists of three modules that are trained in an end-to-end fashion. The first module, also known as the generator, restores image semantics from the 2D reconstructions introduced in Sec. 6.2.1. To better represent the underlying structural information of the anatomy, the semantics are produced as continuous values instead of discrete categories. The second module

is the discriminator that attempts to distinguish between the predicted images and the ground-truth 2D orthographic projections. This encourages the generator module to predict visually similar images compared with the ground-truth domain. The last module detects anatomical landmarks from the orthographic 2D reconstructions. This module not only facilitates automatic landmark detection but also encourages the predicted images to offer more details near the important structures of the anatomy.

Structural similarity (SSIM) loss [172] and an adversarial loss on the spatial domain as well as a cosine similarity loss on the frequency domain are used to optimize the overall visual similarity between the predicted images and the ground-truth orthographic projections. The SSIM index comprises a weighted multiplication between three distance measurements, namely luminance l, contrast c, and structure s, between the prediction X and ground-truth projection Y. The loss between X and Y is defined as:

$$\mathcal{L}_{\text{ssim}} = 1 - \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} l_{i,j}^{\alpha} c_{i,j}^{\beta} s_{i,j}^{\gamma} \quad \text{, where} \tag{6.3}$$

i, j iterate over the entire image domain. The three terms in SSIM are computed as $l_{i,j} = \frac{2\mu_{i,j}^{x}\mu_{i,j}^{y}+b_1}{{\mu_{i,j}^{x}}^2+{\mu_{i,j}^{y}}^2+b_1}$, $c_{i,j} = \frac{2\sigma_{i,j}^{x}\sigma_{i,j}^{y}+b_2}{{\sigma_{i,j}^{x}}^2+{\sigma_{i,j}^{y}}^2+b_2}$, and $s_{i,j} = \frac{\sigma_{i,j}^{xy}+b_3}{\sigma_{i,j}^{x}\sigma_{i,j}^{y}+b_3}$, such that $\mu_{i,j}^{x}$ and $\sigma_{i,j}^{x}$ are the local window average and standard deviation of X centered at location (i, j), respectively. $\sigma_{i,j}^{xy}$ is the local window covariance of X and Y centered at location (i, j).

The adversarial loss follows the basic idea of relativistic GAN [173]. The adver-

sarial losses in the discriminator and generator cycles are:

$$\begin{aligned} \mathcal{L}_{\mathrm{D}} &= \left(1 - C_y + \overline{C_x}\right)^2 + \left(1 - \overline{C_y} + C_x\right)^2 \\ \mathcal{L}_{\mathrm{G}} &= \left(1 - C_x + \overline{C_y}\right)^2 + \left(1 - \overline{C_x} + C_y\right)^2 \quad \text{, where} \end{aligned} \tag{6.4}$$

C_x is the confidence of the discriminator regarding whether the prediction X is real and $\overline{C_x}$ is the average confidence over a mini-batch. The architecture of the discriminator is DenseNet [174] with linear activation as the final layer.

In the frequency domain, the proposed cosine similarity loss is defined as:

$$\mathcal{L}_{\cos} = 1 - \langle \frac{f_x - f_i}{\|f_x - f_i\|_2}, \frac{f_y - f_i}{\|f_y - f_i\|_2} \rangle \quad \text{, where} \tag{6.5}$$

f_x is the frequency representation of the restored semantics X that is flattened to a 1D vector, and f_i is the frequency representation of the input orthographic reconstruction. Since the low-frequency components in images are much larger than the high-frequency ones, we only use the residual frequency for both the prediction and the ground-truth reconstruction w.r.t. the input reconstruction. This could encourage the network to focus more on the high-frequency texture information in the ground-truth.

To adequately convey task-related information to the generator network (semantic restoration network), we propose to jointly optimize the generator network as well as an anatomical landmark detection network with a specialized training scheme. The

landmark detector model takes either ground-truth reconstruction or the predictions of the generator as input and predicts heatmaps corresponding to distinct anatomical landmarks. The ground-truth landmark locations and heatmaps are generated and annotated from the CT volume and used for supervision. Each ground-truth heatmap is a 2D Gaussian distribution with a corresponding landmark location as the mean and a manually selected standard deviation σ. Binary Cross Entropy (BCE) and Relative Response (RR) losses [175] are used for the landmark detection learning. The BCE loss is defined as follows:

$$\mathcal{L}_{\text{bce}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} m_{i,j}^{\text{gt}} \log m_{i,j} + \left(1 - m_{i,j}^{\text{gt}}\right) \log\left(1 - m_{i,j}\right) \quad \text{, where} \tag{6.6}$$

$m_{i,j}$ and $m_{i,j}^{\text{gt}}$ are the values of the predicted and ground-truth heatmaps, respectively, at the location (i, j). RR loss is defined as:

$$\mathcal{L}_{\text{rr}} = -\log\left(\frac{\mathrm{e}^{\sigma m_{u,v}}}{\sum_{(i,j)\in\Omega} \mathrm{e}^{\sigma m_{i,j}}}\right) \quad \text{, where} \tag{6.7}$$

(u, v) is the ground-truth landmark location and σ is a scale factor.

In a single iteration of training, there are several cycles involved, where only one of all modules are updated. In the cycle of discriminator training, only $\mathcal{L}_{\text{D}}$ is involved. In the cycle of landmark detector training, the overall loss is $\mathcal{L}_{\text{landmark}} = \lambda_{\text{rr}}\mathcal{L}_{\text{rr}} + \lambda_{\text{bce}}\mathcal{L}_{\text{bce}}$, where the input to the landmark detector is the ground-truth 2D orthographic projections. In the cycle of semantics restoration training, the overall

loss is $\mathcal{L}_{\text{restore}} = \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{G}}\mathcal{L}_{\text{G}} + \lambda_{\text{cos}}\mathcal{L}_{\text{cos}} + \lambda_{\text{rr}}\mathcal{L}_{\text{rr}} + \lambda_{\text{bce}}\mathcal{L}_{\text{bce}}$, where the input to the detector is the predicted stitching images. The landmark detector is only updated when the ground-truth reconstruction is fed to ensure that it only learns features that appear in the ground-truth reconstruction. When the stitching predictions are used as input to the landmark model, only the generator network gets updated, which implicitly forces the stitching generator network to learn spatial features in the ground-truth domain that contribute to the task of landmark detection.

6.3 Experiments and Results

6.3.1 Data Set

The training data comprises of eight CT volumes from the left and right legs of four cadaveric specimens. We generated the training data by simulating realistic digitally-reconstructed radiographs (DRRs) using the physics-based DeepDRR pipeline [176, 177]. Rather than relying on full Monte Carlo simulation of image formation, DeepDRR analytically generates forward projections from CT volumes by accounting for the physical interactions that occur during image formation and then estimates the contributions of scattering and noise. Compared to naive DRRs, this mechanism has demonstrated improved generalizability [178, 179].

The intrinsic parameters of the X-ray camera are selected based on the nominal parameters of a commercially available flat panel C-arm, Cios Fusion (Siemens

Healthineers, Forchheim, Germany). We generated a total of $45,000$ X-ray images, where each 3 were stitched together, resulting in $15,000$ training instances, as shown in Fig. 6.4-a. The translation components of the X-ray camera were adjusted such that the first image is acquired from the region of the femoral head, the second image from the shaft, and the third from the knee. We also added a random value in the range of $[-20, 20]$ mm to each translation component in the x, y, z axes, which resulted in gaps or overlaps between simulated acquisition.

For each specimen's leg, we generated DeepDRR images from $-21°$ to $+21°$, and $-6°$ to $+6°$ in LAO/RAO and Cranial/Caudal directions, respectively, around the anterior-posterior view of the bone. These directions which correspond to *out-of-plane* rotations, are shown in Fig. 6.4-b. We also added an offset up to $[-6°, 6°]$ to each *out-of-plane* rotation element of the X-ray camera. This rotation offset is intended to prevent all three images from being exactly parallel to each other, and enforce the training to become robust to unintended *out-of-plane* rotations. In the backprojection step, all projection images were smeared out with $d = 0.5$, which realistically assumes the imaged object is at most 500 mm away from the detector.

A validation set is constructed from 1875 DeepDRR images from a separate cadaveric CT. Finally, for testing, we generated a total of 3750 images from the left and right CT scans of two other patients.

In our supervised training scheme, the ground-truth images were obtained by generating forward projections of the original CT volumes in an orthographic model

(parallel-beam geometry). Given the rotation matrix R associated with the ray direction, the orthographic projection is defined as [180]:

$$\begin{bmatrix} x \\ y \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^\top & 0 \\ \mathbf{r}_2^\top & 0 \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{, where} \tag{6.8}$$

$\mathbf{r}_i^\top$ is the i-th row in R, and $(\frac{x}{\gamma}, \frac{y}{\gamma})^\top$ are the 2D projections of the 3D voxels that are denoted by the homogeneous coordinates $(X, Y, Z, 1)^\top$.

6.3.2 Stitching Performance

The stitching results on two test data sets are shown in Figs. 6.5 and 6.6. It should be noted that in all cases that *out-of-plane* rotations were present between the three input data, the orthographic reconstruction plane was selected to be parallel to the first input image. In Figs. 6.5 and 6.6, we also demonstrate the detection of four bony landmarks, namely the femoral head, greater trochanter, patellar groove (knee), and tibia. These landmarks are particularly important as they are used in defining the length and the biomechanical axes of the bone.

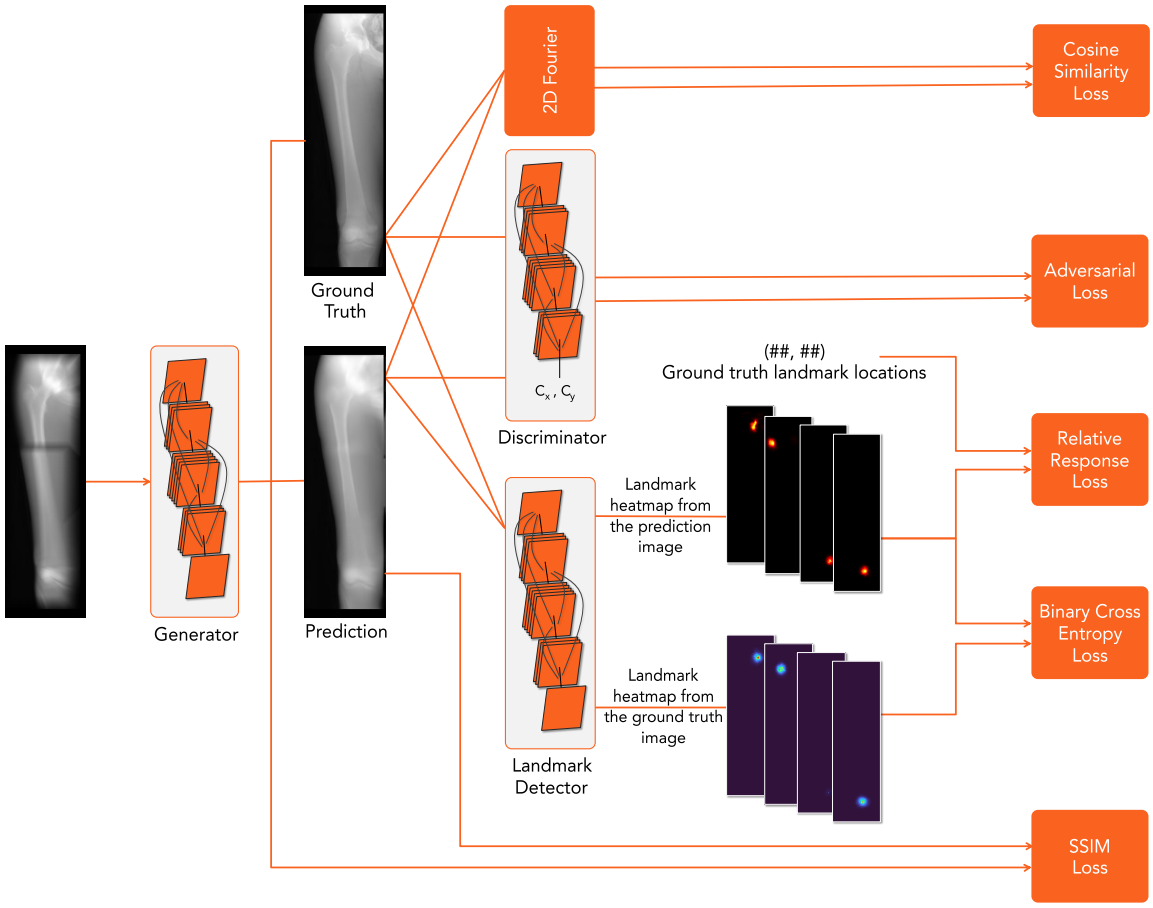

Figure 6.3: The overall network architecture that was used for training is shown. The input orthographic 2D reconstruction in the frequency domain is generated using our proposed method from several X-ray images. SSIM and adversarial losses are used to enforce the network to generate visually similar images compared to the ground-truth. Cosine similarity loss is employed to emphasize more on the high-frequency components of the predictions. RR and BCE loss are used to train a landmark detection network and encourage the semantics restoration to be functionally similar to the ground-truth reconstruction.

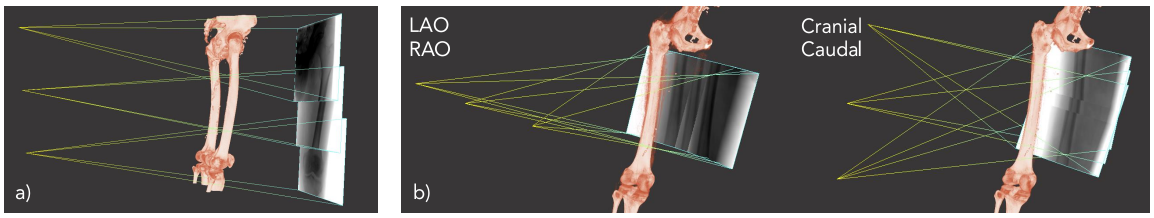

Figure 6.4: **a)** Each training instance comprises three X-ray images. **b)** The training X-ray images are generated in the LAO/RAO and Cranial/Caudal directions.

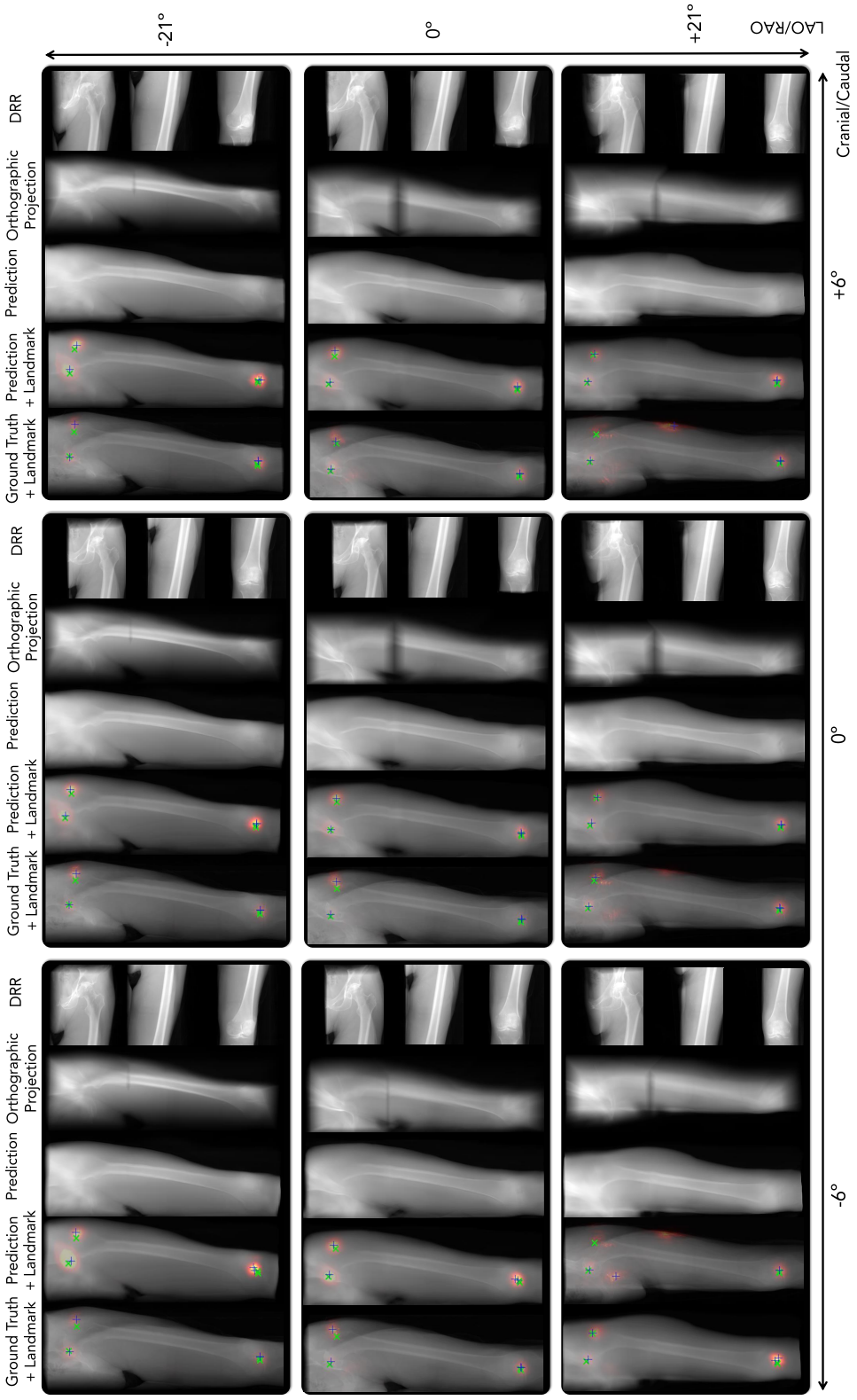

Figure 6.5: Stitching outcomes are shown on test data from the left leg. The heatmaps corresponding to different landmarks are shown as red overlays. The ground-truth and predicted landmark locations are shown as green and blue crosses, respectively.

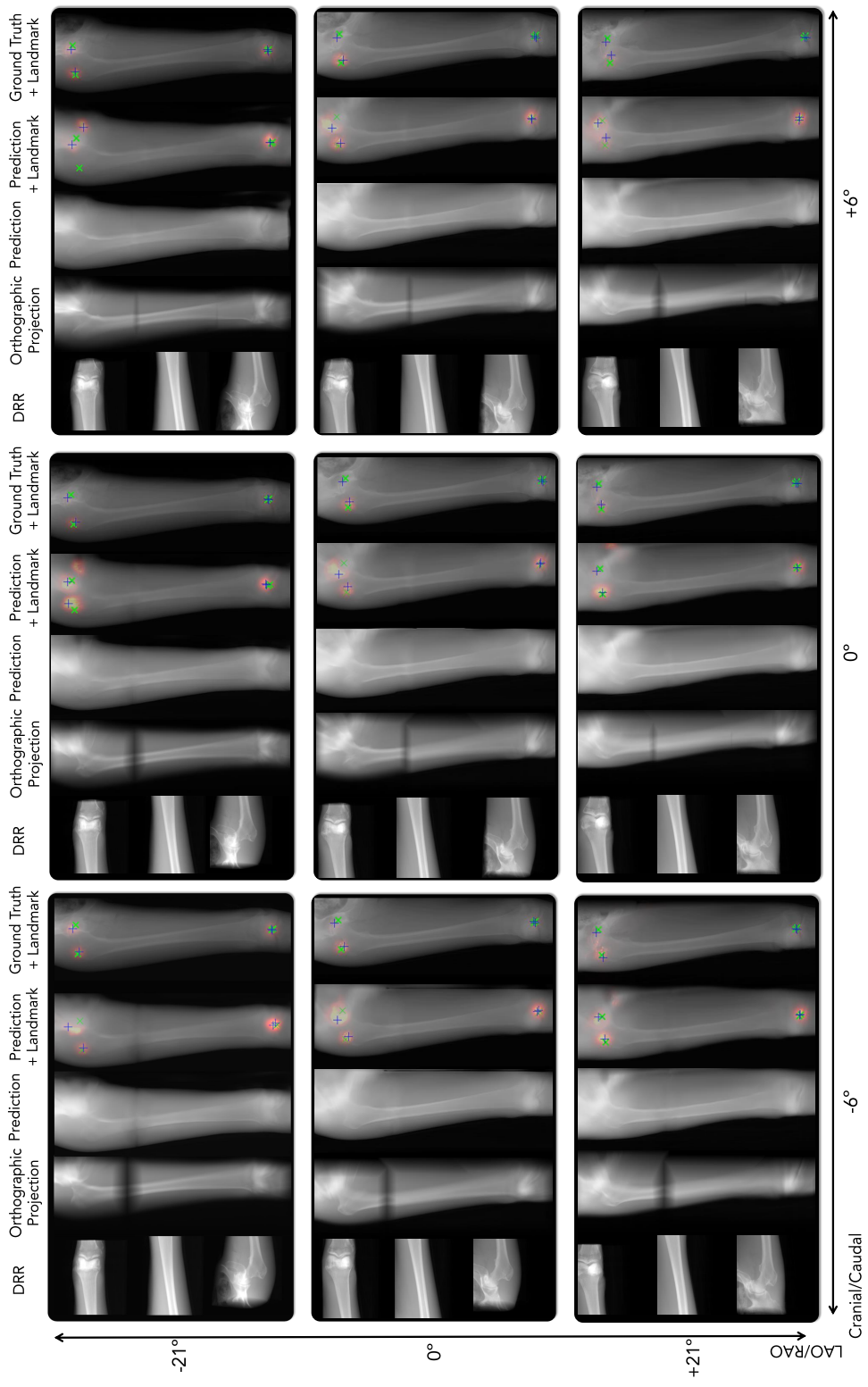

Figure 6.6: Stitching outcomes are shown on test data from the right leg. The heatmaps corresponding to different landmarks are shown as red overlays. The ground-truth and predicted landmark locations are shown as green and blue crosses, respectively.

Table 6.1: Comparing the performance of the network given different regularization factors $\lambda_{\cos}$ for the cosine frequency loss in an ablation study

	$\lambda_{\cos} = 0.0$	$\lambda_{\cos} = 1.0$	$\lambda_{\cos} = 2.0$	$\lambda_{\cos} = 3.0$	$\lambda_{\cos} = 4.0$	$\lambda_{\cos} = 5.0$
SSIM (%)	92.56	94.66	95.81	**95.87**	95.27	94.23
PSNR (db)	24.85	26.80	**26.83**	25.58	24.94	24.83
BCE	**0.0045**	0.0047	0.0047	0.0046	0.0046	0.0049
RR	11.14	10.75	**10.53**	10.60	10.79	11.27

Table 6.2: Comparing the performance of the network given different regularization factors λ_{rr} for landmark detection

	$\lambda_{\text{rr}} = 0.0$	$\lambda_{\text{rr}} = 0.03$	$\lambda_{\text{rr}} = 0.06$	$\lambda_{\text{rr}} = 0.09$	$\lambda_{\text{rr}} = 0.12$	$\lambda_{\text{rr}} = 0.15$
SSIM (%)	95.31	**95.93**	95.59	94.72	94.77	94.47
PSNR (db)	24.81	**26.16**	25.97	25.13	25.06	24.87
BCE	0.0082	**0.0040**	0.0042	0.0047	0.0060	0.0071
RR	26.23	11.00	10.57	10.56	**10.41**	10.74

We trained for 58 epochs on the data with 1 mm pixel spacing, and with the input and output image sizes of 640×640 pixels. Using this model, we achieved an SSIM similarity score of 95.7% and a PSNR of 25.70 (db). The BCE and RR landmark losses were 0.0044 and 13.53, respectively.

6.3.3 Ablation Study

To identify the optimal learning parameters and understand the effects of each component, we performed ablation studies. To keep all experiments tractable, we completed the ablation tests on $15,000$ downsampled training data with 2 mm pixel spacing, with 10 epochs, and a batch size of 2. We kept the regularization factor for the GAN losses fixed and evaluated against all other regularization parameters.

In Tables 6.1, 6.2, and 6.3 we report the best losses by selecting the lowest validation loss. The tables report SSIM and peak signal-to-noise ratio (PSNR) similarity

Table 6.3: Comparing the performance of the network given different regularization factors λ_{bce} for landmark detection. N/A indicates that the training diverged with the corresponding parameters.

	$\lambda_{\text{bce}} = 0$	$\lambda_{\text{bce}} = 100$	$\lambda_{\text{bce}} = 200$	$\lambda_{\text{bce}} = 300$	$\lambda_{\text{bce}} = 400$	$\lambda_{\text{bce}} = 500$
SSIM (%)	N/A	95.17	95.38	95.54	**95.84**	95.42
PSNR (db)	N/A	25.19	26.65	23.63	**27.05**	26.02
BCE	N/A	0.0045	0.0038	0.0038	**0.0036**	0.0039
RR	N/A	11.01	10.86	10.93	**10.69**	11.01

scores, as well as BCE and RR losses. The best results in each experiment are represented with bold digits.

6.3.4 Landmark Detection

We evaluated the accuracy of landmark detection on 3500 test ground-truth and prediction images. The corresponding results are presented in Figs. 6.7 and 6.8. The rightmost columns in these figures contain the automatic measurements of total bone length directly from 2D orthographic ground-truth or prediction images.

6.4 Discussion

The outcome of orthographic stitching presented in Figs. 6.5 and 6.6 suggest that the ConvNet can effectively close small gaps between input images, and substantially reduce more significant gaps. The automatic landmark detection in Figs. 6.7 and 6.8 indicate that the mean landmark detection errors are 5.06 and 6.07 pixels on the ground-truth projections and the prediction images, respectively. It is important to note that, despite the errors in automatic measurement of the bone length based on

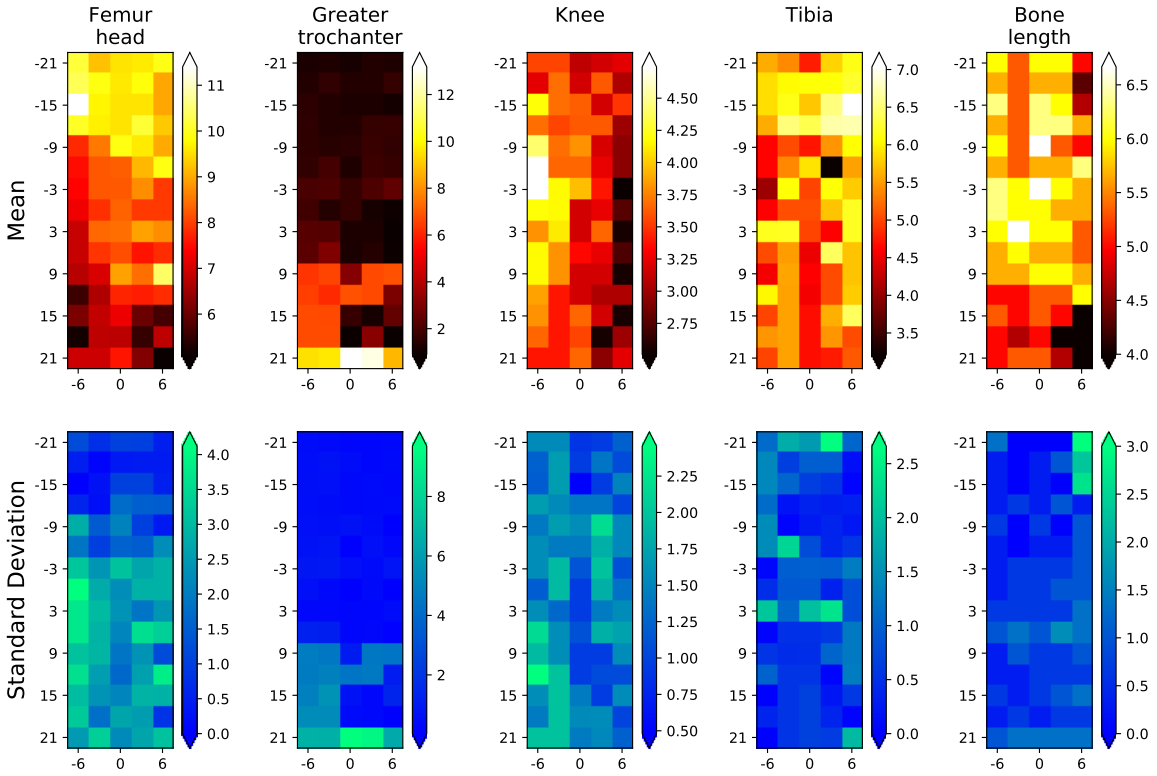

Figure 6.7: Landmark detection performances on the ground-truth orthographic projections for two test patient data sets are shown as error heatmaps. The vertical axes correspond to the LAO/RAO rotations, and the horizontal axes correspond to cranial/caudal rotations around the anterior-posterior view, respectively.

anatomical landmarks, the manual measurements from the prediction images were highly consistent with the measurements from the ground-truth orthographic projections. This suggests that the strict modeling of the geometry of stitching based on the projection matrices and the central slice theorem, and using the ConvNet to recover details, yielded true orthographic predictions.

We observed that object magnification that happens when the imaged object is closer to the origin of the X-ray camera has an adverse effect on orthographic

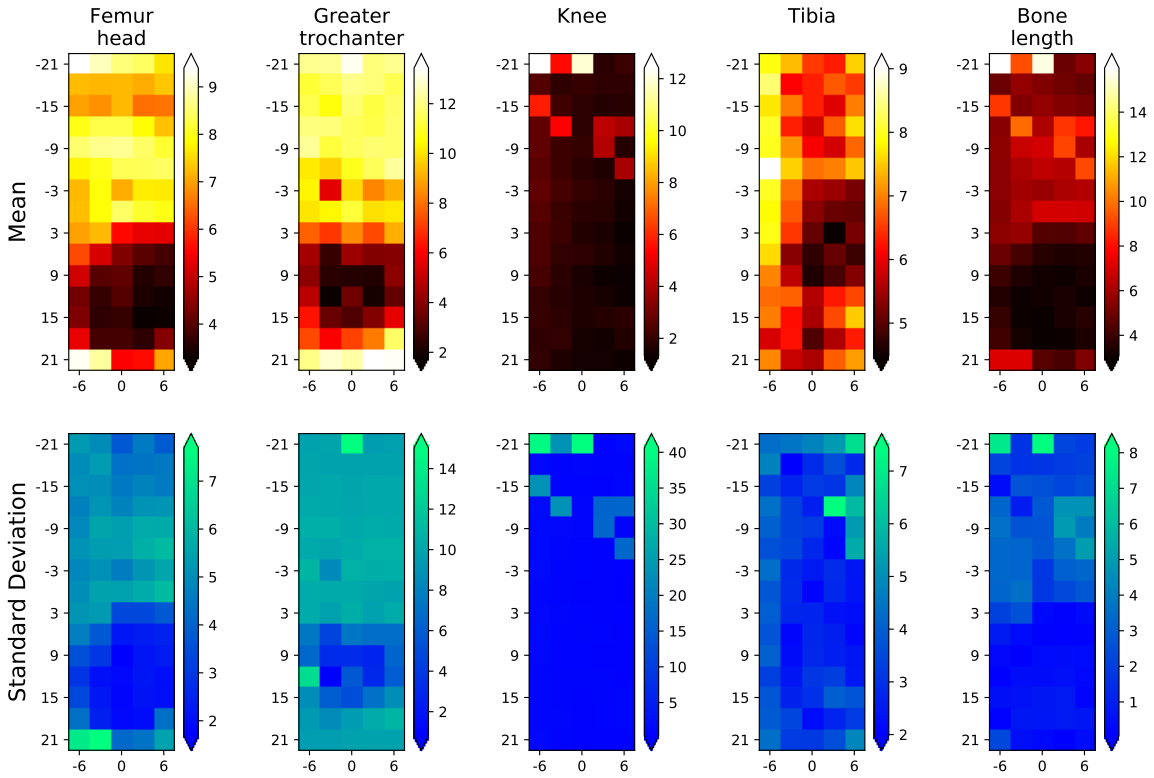

Figure 6.8: Landmark detection performances are shown on the prediction images of the test sets. The vertical axes correspond to the LAO/RAO rotations, and the horizontal axes correspond to cranial/caudal rotations around the anterior-posterior view, respectively.

reconstruction. When the object is closer to the origin, the rays from the source to the image plane that hit the anatomy are more diverged compared to the parallel rays that are used for reconstruction. For the reconstruction, we assumed the patient is within a 500 mm distance from the image detector, which is a clinically realistic assumption.

Lastly, in our ablation studies, we demonstrated that all regularization losses contribute to stronger image similarities for image reconstruction.

6.5 Conclusion

We presented the first work that provides orthographic image stitching by leveraging the principles of image formation and geometric models from computer vision and combining them with ConvNets to recover semantics. In contrast to the state-of-the-art systems, our solution naturally allows for both translation and rotations of the X-ray camera and does not impose any constraints regarding the motion. A direction for future research is to extend the training to complex scenarios where stitching becomes robust to fractures and arbitrary tools in the scene using robust estimators [181–183].

The orthographic representations of images *i)* do not carry perspective properties, hence enable metric measurements directly on 2D images, and *ii)* are described with 5 *degree-of-freedom* (DOF) as opposed to 6 DOF rigid body parameters. As a consequence of this drop of DOF, registration of orthographic images will not require scale disambiguation along their depth.

In reflective images used in standard computer vision applications, the Fourier slice theorem does not hold. Nevertheless, other contributions of our work, such as back-projection, data compounding in parallax-free domains, and the recovery of details using ConvNets, can be directly employed to tackle image stitching or image-based rendering problems in other domains of computer vision applications.

6.6 Acknowledgments

Sincere thanks to Mr. Xingtong Liu for his help during the neural network design, particularly the implementation of the relativistic GAN. I thank Dr. Mathias Unberath for his guidance and support to combine the Fourier slice theorem with data-driven modeling to achieve image stitching. Finally, my appreciation to Drs. Nassir Navab and Mehran Armand for their encouraging feedback.

Part III

Advanced Visualization in Augmented Surgical Environments

The following section of this dissertation suggests various techniques and systems to bring augmented reality into the workflow of image-guided therapies. These solutions provide spatial and/or temporal awareness and allow efficient visualization of surgical information. The main problems that are aimed to address in this section are *i)* lack of geometric correspondence between data and patient, which makes image-based navigation a mentally challenging task, *ii)* inconvenient and off-axis display of data, and *iii)* poor interaction between the surgical crew with each other and with data.

Chapter 7, suggests a display-based, non-immersive augmented reality system using cone-beam CT data. Chapter 8 presents flying frustums that fully exploit all imaging geometries and provide immersive augmented reality experiences for all the surgical crew to visualize data with all their spatial and temporal information. Lastly, chapter 9 introduces patient-specific image augmentations for patients with unilateral pelvic fractures by exploring the partial symmetry that is present in the anatomy.

The contributions in the following three chapters are based on the these publications:

1. Fotouhi, Javad, Bernhard Fuerst, Sing Chun Lee, Matthias Keicher, Marius Fischer, S. Weidert, E. Euler, N. Navab, and G. Osgood. "Interventional 3d augmented reality for orthopedic and trauma surgery." In 16th annual meeting of the international society for computer-assisted orthopedic surgery (CAOS). 2016.

2. Fotouhi, Javad, Clayton P. Alexander, Mathias Unberath, Giacomo Taylor, Sing Chun Lee, Bernhard Fuerst, Alex Johnson et al. "Plan in 2-D, execute in 3-D: an augmented reality solution for cup placement in total hip arthroplasty." Journal of Medical Imaging 5, no. 2 (2018): 021205.

3. Fotouhi, Javad, Mathias Unberath, Tianyu Song, Wenhao Gu, Alex Johnson, Greg Osgood, Mehran Armand, and Nassir Navab. "Interactive flying frustums (IFFs): spatially aware surgical data visualization." International journal of computer-assisted radiology and surgery 14, no. 6 (2019): 913-922.

4. Fotouhi, Javad, Arian Mehrfard, Tianyu Song, Alex Johnson, Greg Osgood, Mathias Unberath, Mehran Armand, and Nassir Navab. "Spatiotemporal-Aware Augmented Reality: Redefining HCI in Image-Guided Therapy." arXiv preprint arXiv:2003.02260 (2020).

5. Fotouhi, Javad, Mathias Unberath, Giacomo Taylor, Arash Ghaani Farashahi, Bastian Bier, Russell H. Taylor, Greg M. Osgood, Mehran Armand, and Nassir Navab. "Exploiting Partial Structural Symmetry for Patient-Specific Image Augmentation in Trauma Interventions." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 107-115. Springer, Cham, 2018.

6. Fotouhi, Javad, Giacomo Taylor, Mathias Unberath, Alex Johnson, Sing Chun Lee, Greg Osgood, Mehran Armand, and Nassir Navab. "Exploring Partial In-

trinsic and Extrinsic Symmetry in 3D Medical Imaging." arXiv preprint arXiv:2003.02294 (2020).

Chapter 7

Display-based Augmented Reality for Orthopedic Surgeries

7.1 Clinical Motivation and Background

Complex and frequent K-wire placements occur after ilio-sacaral joint and superior pubic ramus fractures. The common treatment for both injuries is the placement of screws to stabilize the fracture. During the minimally invasive approach, the surgeon places the K-wire through the muscles and all bone fragments under image-guidance, requiring a good mental alignment of the intra-operative X-ray imaging, the medical instrument, and the patient [7, 8]. Due to the 2D nature of the X-ray images, this results in frequent re–positioning of the C–arm [9]. For instance, in pelvic acetabulum fractures, the surgeon needs to find the correct trajectory of the K-wire through the

superior pubic ramus. The misplacement of the K-wire could cause severe damage to the external iliac artery and vein, obturator nerve, or to structures such as the inguinal canal and intra-articular hip joint [10]. It is not unusual that a single K-wire placement for one screw takes up to ten minutes [11].

In total hip replacement (THA), Poor placement leads to increased impingement and dislocation that promotes accelerated wear. Conversely, proper implant placement that restores the hip anatomy and biomechanics decreases the risk for dislocation, impingement, loosening, and limb length discrepancy, and thus implant wear and revision rate [184–188]. Steps to ensure accuracy and repeatability of acetabular component positioning are therefore essential. Due to the large volume of THA procedures, small but favorable changes to the risk-benefit profile of this procedure enabled by improved implant positioning will have a significant impact on a large scale.

Unfortunately, optimal placement of the acetabular component is challenging due to two main reasons. First, the ideal position of the implant with respect to the anatomy is unknown; yet, a general guideline exists [189] and is widely accepted in clinical practice. This guideline suggests abduction and anteversion angles of the hip joint measured with respect to bony landmarks defining the so-called safe zone, that is indicative for an acceptable outcome. The safe zones for the anteversion and abduction angles are $15° \pm 10°$ and $40° \pm 10°$, respectively. Recent studies suggest that an even narrower safe zone may be necessary to minimize the risk of hip disloca-

tion [190, 191]. Defining the ideal implant position is not as straight-forward as the definition of a range of abduction and anteversion angles when considering a large population [192]. A static definition of the safe zone seems even more prone to error when considering that the position of the pelvis varies dramatically from supine to sitting to standing posture among individuals [193, 194].

Second, even if a clinically acceptable safe zone is known it is questionable whether surgeons are, in fact, able to accomplish acetabular component placement within the suggested margin [191]. In light of previous studies that report mal-positioning of up to 30% to 75% [195–197] when free-hand techniques are used, addressing this challenge seems to be imperative. Most computer-assisted methods consider the direct anterior approach (DAA) to the hip for THA, as it allows for convenient integration of intra-operative fluoroscopy to guide the placement of the acetabular component [198]. The guidance methods reviewed below proved effective in reducing outliers and variability in component placement, which equates to more accurate implant positioning [199–202].

The state of the art approaches that provide guidance using image-less or image-based methods have certain drawbacks. Image-less methods require complex navigation and may provide unreliable registration [47]. Image-based solutions rely on pre-operative CT scans or intra-operative fluoroscopy and often use external navigations systems for tracking [203, 204]. Systems based on external navigation are expensive and increase the operative time due to the added complexity. Use of pre-operative CT

scans increases the radiation exposure and cost to the patient. Moreover, many of the methods used for registering CT to patient seek to solve ill-posed problems that require manual interaction either for initialization or landmark identification and, thus, disrupt the surgical workflow. Manual annotations can take between 3 to 5 minutes during the intervention for *each* image registration [38]. Although proven beneficial for the surgical outcome, neither of these costly and labor-intensive navigation techniques were widely adopted in clinical practice.

Partly due to above drawbacks, surgeons who use the DAA often rely solely on fluoroscopic image guidance [198, 205]. These images, however, are a 2D representation of 3D reality and have inherent flaws that complicate the assessment. The challenges include finding the true anterior pelvic plane as well as eyeballing acetabular component position by eye on the image. Therefore, a technique that provides a quantitative and reliable representation of the pelvis and acetabular component intraoperatively without increasing neither radiation dose or cost while largely preserving the procedural workflow is highly desirable.

7.2 Proposed Solution

We suggest a display-based augmented reality (AR) system for two different orthopedic procedures: *i)* fracture reduction and stabilization using K-wires, and *ii)* implant placement in THA. Our system extends the camera-augmented C-arm sys-

tem by Navab et al. [67] and exploits 3D imaging capabilities.

For K-wire placement, we present an interventional 3D AR system, which utilizes a CBCT-enabled C-arm, and a color and depth (RGBD) camera to enable an intuitive 3D visualization that overlays both physical and anatomical information from arbitrary views. We present a pre-clinical usability study that evaluates the potential benefit based on simulated K-wire placements in the superior pubic ramus. Similarly, we propose a similar AR solution for intra-operative guidance of THA using DAA where the C-arm is kept in place until the correct alignment of the acetabular cup is confirmed [206, 207]. With the proposed solution, the surgeon first plans the position of the acetabular cup intra-operatively based on two fluoroscopy images that are acquired after the dislocation of the femoral head and the reaming of the acetabulum are completed. The orientation of the cup in the X-ray images could be either automatically preset based on desired angles relative to the APP plane (or other known pelvic coordinate frames), or be adjusted by the surgeon. Once the desired pose of the acetabular cup is estimated relative to the C-arm, we use optical information from the co-calibrated RGBD camera that is mounted on the C-arm to provide an AR overlay [92–94] that enables placement of the cup according to the planning. As the cup is not visible in RGBD, we exploit the fact that the acetabular cup is placed using an impactor that is rigidly attached to the cup and is well perceived by the RGBD camera. For accurate cup placement, the surgeon aligns the optical information of the impactor (a cloud of points provided by the RGBD camera) with the planned

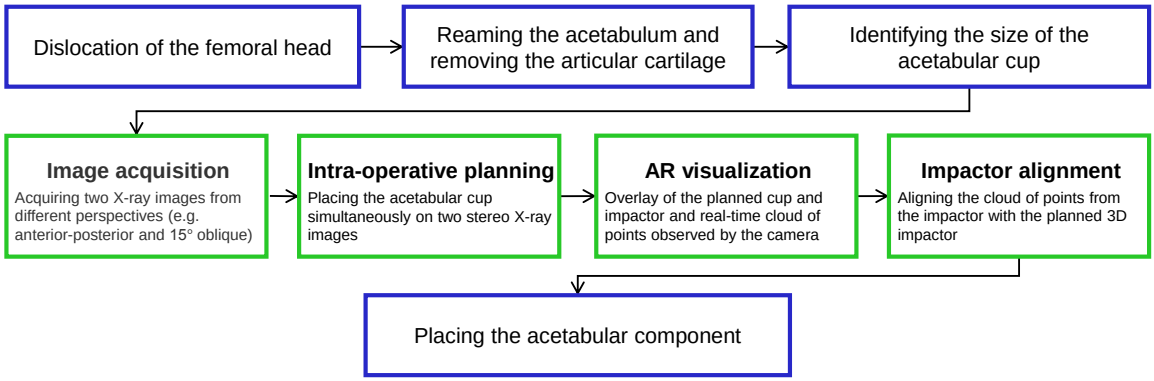

Figure 7.1: After the femoral head is dislocated, the size of the acetabular implant is identified based on the size of the reamer. Next, two C-arm X-ray images are acquired from two different perspectives. While the C-arm is repositioned to acquire a new image, the relative poses of the C-arm are estimated using the RGBD camera on the C-arm and a visual marker on the surgical bed. The surgeon then plans the cup position intra-operatively based on these two stereo X-ray images simultaneously. Next, the pose of the planned cup and impactor are estimated relative to the RGBD camera. This pose is used to place the cup in a correct geometric relation with respect to the RGBD camera and visualize it in an AR environment. Lastly, the surgeon observes real-time optical information from the impactor, and aligns it with the planned impactor using the AR visualization. The green boxes in this figure highlight the contributions of this work.

virtual impactor-cup, that are visualized simultaneously in our AR environment. A schematic of the proposed clinical workflow is shown in Fig. 7.1.

7.3 Augmented Reality for K-wire placement

The 3D interventional Augmented Reality visualization requires the fusion of 3D optical views and 3D CBCT data. The acquisition of the 3D optical views requires

tracking of the position and orientation of the camera, and allows the 3D reconstruction of the surgical site. The conventional technique to perform this is using stereo vision with multiple cameras, or moving a single camera. Stereo vision requires accurate feature detection and tracking in multiple frames, and the scale can only be recovered by using markers or objects with a known size in the scene.

Alternatively, an RGBD camera provides depth information by emitting and automatically detecting an infrared pattern on the surface of the object. This results in the simultaneous acquisition of color and depth information, and can be processed as colored 3D surface point clouds. To achieve an Augmented Reality image-guidance system, we need to mount the RGBD camera rigidly to the C-arm, obtain the CBCT-to-RGBD calibration, and perform a joint visualization by blending the multi-modal data.

The camera is mounted rigidly near the detector plane of the CBCT-enabled C-arm. The calibration can be performed using any arbitrary phantom with an unique shape (rotationally variant), homogeneous radiopacity, and a matte surface. Other properties of the phantom are optimal size (within the CBCT volume) and round surfaces to avoid sharp edges and corners. After introducing a calibration phantom, the surface information and CBCT are acquired simultaneously, followed by an offline one-time calibration of the C-arm and the RGBD camera.

The multi-modal registration of RGBD surface data and CBCT data requires additional pre-processing to limit the registration to common structures of interest,

and discard the noisy background. Additionally, RGBD camera and C-arm have different fields of view, thus additional structures observed by one of the devices may cause the registration to converge to a local minimum. The pre-processing step results in two data sets (RGBD and CBCT) representing the same calibration phantom.

Next, the calibration algorithm performs an automatic registration, for which Fast Point Feature Histograms (FPFH) are used for initialization [208]. The final calibration transformation is obtained after the refinement by using an Iterative-Closest-Point (ICP) registration [92]. The calibration remains valid as long as the RGBD camera is fixed on the C-arm. If the camera is moved or displaced on the C-arm, a re-calibration becomes necessary.

After the offline calibration, the patient's CBCT and body surface scans are acquired. The transformation obtained from the calibration algorithm is used to overlay both modalities in a common coordinate system. An example is shown in Fig. 7.2. The intra-operative 3D Augmented Reality visualization is then provided in form of overlaid an Digitally Reconstructed Radiograph (DRR) computed from the CBCT augmented onto the scanned patient's surface. Additionally, the surgical tools and surgeon's hands are dynamically visualized. This system allows the visualization of the scene from arbitrary views, which are not possible to acquire using a standard C-arm in a conventional OR setup.

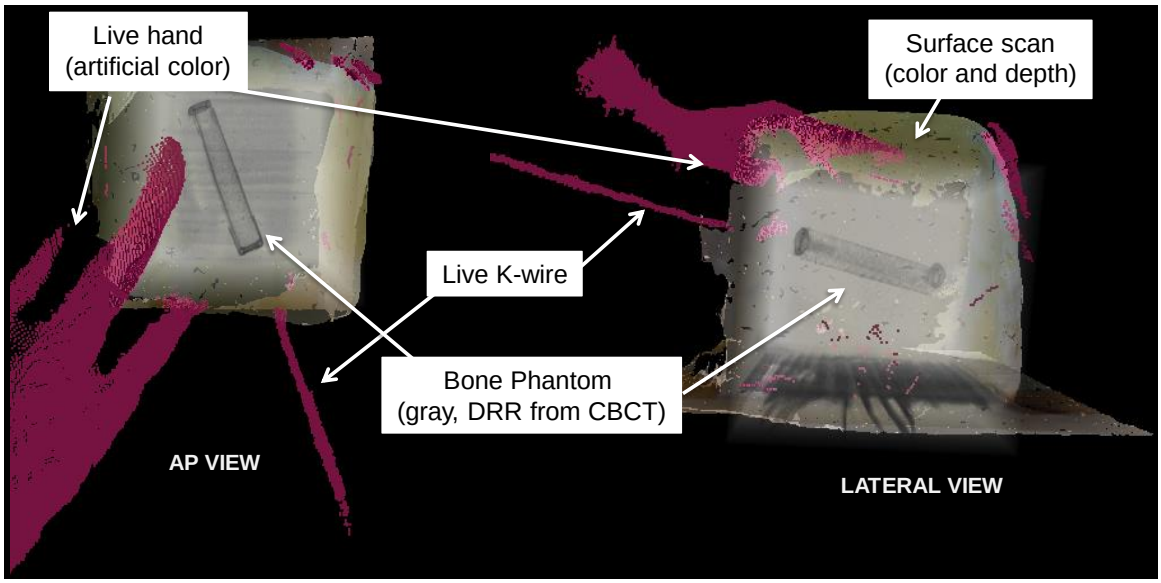

Figure 7.2: The Augmented Reality visualization provides surgeons with multiple views of entire surgical site. The scene includes live representations of the hand and surgical tools (e.g. K-wire) as artificially colored point clouds, the colored surface scan of the patient, and dynamically adjusted DRR from the CBCT scan. This allows the surgeon to dynamically change the view during the procedures.

7.4 Augmented Reality for Total Hip Arthroplasty

The stereo relation between C-arm X-ray images acquired at different poses are estimated by first tracking visual markers in the RGBD camera coordinate frame, and then transforming the tracking outcome to the X-ray coordinate frame:

$$ {}^{\mathrm{X}'}\mathbf{T}_{\mathrm{X}} = {}^{\mathrm{RGBD}'}\mathbf{T}_{\mathrm{X}'}^{-1}\ {}^{\mathrm{M}}\mathbf{T}_{\mathrm{RGBD}'}^{-1}\ {}^{\mathrm{M}}\mathbf{T}_{\mathrm{RGBD}}\ {}^{\mathrm{RGBD}}\mathbf{T}_{\mathrm{X}}\,, \tag{7.1} $$

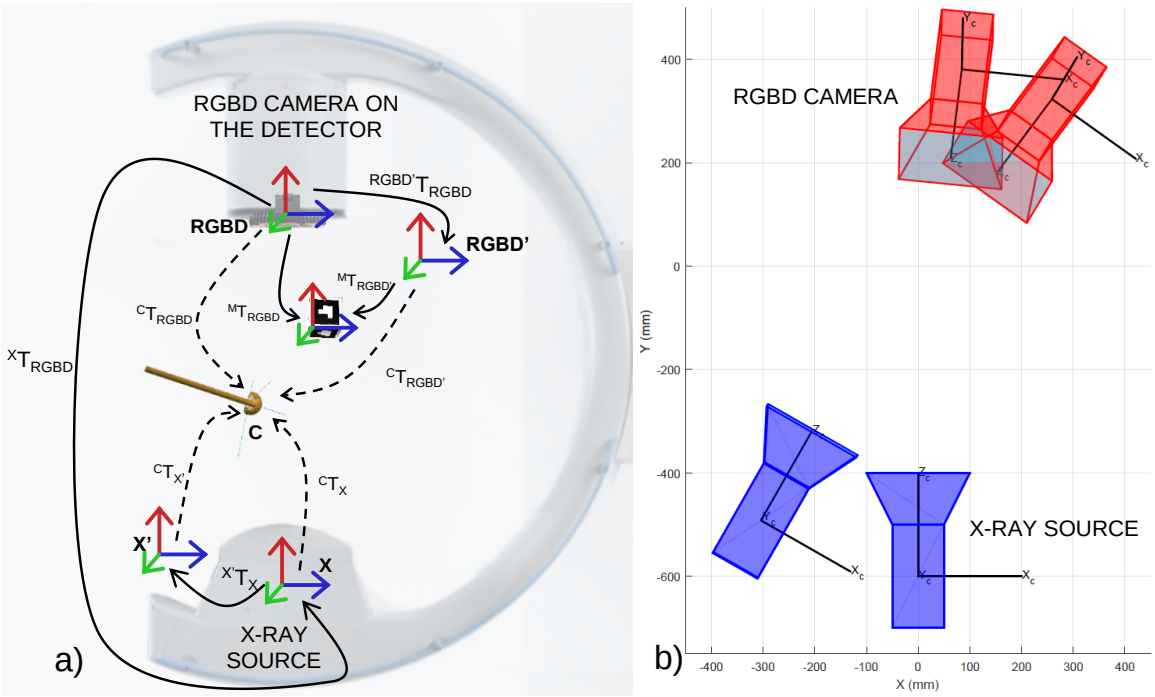

Figure 7.3: In the transformation chain of the RGBD-C-arm system for THA **(a)**, the RGBD, X-ray, visual marker, and acetabular cup coordinate frames are denoted as **RGBD**, **X**, **M**, and **C**, respectively. In an offline calibration step, the extrinsic relation between the RGBD and X-ray (${}^{\mathrm{X}}\mathbf{T}_{\mathrm{RGBD}}$) is estimated. Once this constant relation is known, the pose of the X-ray source can be estimated for every C-arm re-positioning **(b)** by identifying displacements in the RGBD camera coordinate frame.

where ${}^{\mathrm{RGBD'}}\mathbf{T}_{\mathrm{X'}} = {}^{\mathrm{RGBD}}\ \mathbf{T}_{\mathrm{X}}$ due to the rigid construction of the RGBD camera on the C-arm gantry. In Fig. 7.3-b the rigid movement of X-ray source with the RGBD camera origin are shown for an arbitrary C-arm orbit. Fig. 7.3-a illustrates the spatial relation between the RGBD camera and the X-ray source.

Planning of the acetabular component is performed in a user interface where the cup could be rotated and translated by the surgeon in 3D with six degrees of freedom (DOF) rigid parameters, and is forward projected (p_{c_v} and p'_{c_v}) onto the planes of

the two X-ray images acquired from different perspectives:

$$
\begin{aligned}
p_{c_v} &= \mathrm{K}\ \mathrm{P}\ {}^{\mathrm{C}}\mathbf{T}_{\mathrm{X}}^{-1}\ {}^{\mathrm{C}_v}\mathbf{T}_{\mathrm{W}}\,, \\
p'_{c_v} &= \mathrm{K}'\ \mathrm{P}\ {}^{\mathrm{C}}\mathbf{T}_{\mathrm{X}'}^{-1}\ {}^{\mathrm{C}_v}\mathbf{T}_{\mathrm{W}}\,,
\end{aligned}
\tag{7.2}
$$

where K and K′ are the intrinsic perspective projection parameters for each C-arm image, P is a projection operator, and ${}^{\mathrm{C}_v}\mathbf{T}_{\mathrm{W}}$ is the position of vertex v of the cup in the world coordinate frame. Relying on two X-ray views not only provides the ability to plan the orientation of the acetabular component such that it is aligned in two images but, more importantly, also allows adjusting the depth of the cup correctly, which is not possible when a single X-ray image is used. It is worth mentioning, that the size of the acetabular cup does not require adjustment but is known at this stage of the procedure as it is selected to match the size of the reamer.

In addition, if the desired orientation of the cup is known relative to an anatomical coordinate frame (e.g. APP plane), and an X-ray image is acquired from a known perspective in relation to that anatomical frame (e.g. AP view), then the orientation of the cup could be automatically adjusted for the user (equivalent to presetting the orientation in ${}^{\mathrm{C}}\mathbf{T}_{\mathrm{X}}$). It is worth emphasizing that in several image-guided orthopedic procedures, X-ray images are frequently acquired from the AP view.

The transparency of the cup is adjusted by the surgeon in the user interface such

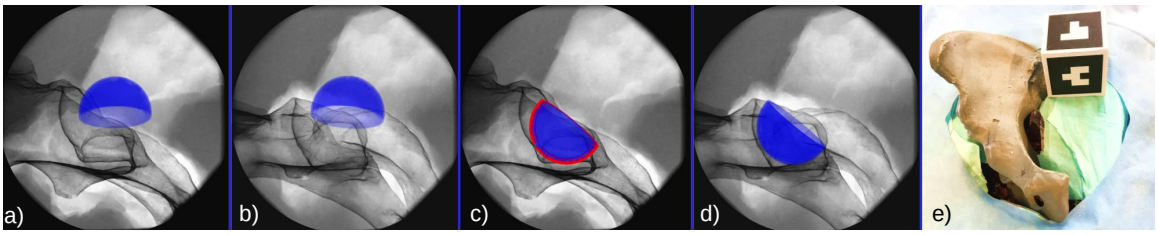

Figure 7.4: The acetabular component is forward projected from an initial 3D pose onto the respective X-ray image plane **(a-b)**. The surgeon moves the cup until satisfied with the alignment in both views **(c-d)**. The X-ray images shown here are acquired from a dry pelvis phantom encased in gelatin. A cubic visual marker is placed near the phantom but outside the X-ray field of view to track the C-arm **(e)**.

that the ambiguity between the front and the back of the cup is optimally resolved. Lastly, the contours around the edge of the cup are estimated and visualized by thresholding the dot product of the unit surface normal $\mathbf{n}_v$ and the intersecting ray $\mathbf{r}_v$:

$$| \mathbf{r}_v \, . \, \mathbf{n}_v | < \tau \, . \tag{7.3}$$

The planning of an acetabular cup based on two X-ray images is shown in Fig. 7.4.

Once the desired cup position is known, guidance of the cup placement using an impactor with an AR visualization is needed to ensure a positioning in agreement with the planning. To construct the AR environment, we first estimate the pose of the RGBD sensor relative to the planned cup as

$${}^{\mathrm{C}}\mathbf{T}_{\mathrm{RGBD}} = {}^{\mathrm{C}}\mathbf{T}_{\mathrm{X}} \, {}^{\mathrm{X}}\mathbf{T}_{\mathrm{RGBD}} \, . \tag{7.4}$$

Within the AR environment, we then render a 3D mesh of the cup and impactor superimposed with the real-time cloud of points observed by the camera, all in the **RGBD** coordinate frame. In the interventional scenario, the acetabular cup is hidden under the skin, and only the impactor is visible. Therefore, the surgeon will only align the cloud of points from the impactor, a cylindrical object, with the 3D virtual representation of the planned impactor.

Ambiguities in the AR environment, among others occlusions or the rendering of a 3D scene in a 2D display, are eliminated by showing different perspectives of the scene simultaneously. Thus, it is ensured that the surgeon's execution fully matches the planning once alignment of the current cloud of points of the impactor and the planned model is achieved in all perspectives. We provide an intuitive illustration of these relations in Fig. 7.5.

7.5 Results

7.5.1 Pre-clinical Analysis for K-wire Placement

We observed 7 trained surgeons performing minimally invasive K-wire placements in superior pubic ramus phantoms. Each surgeon performed the same intervention using the standard C-arm and the 3D Augmented Reality visualization in random order. The time in seconds, number of acquired X-ray images, cumulative area dose product (dose) in cGycm2, surgical task load index, error relative to the ideal path in

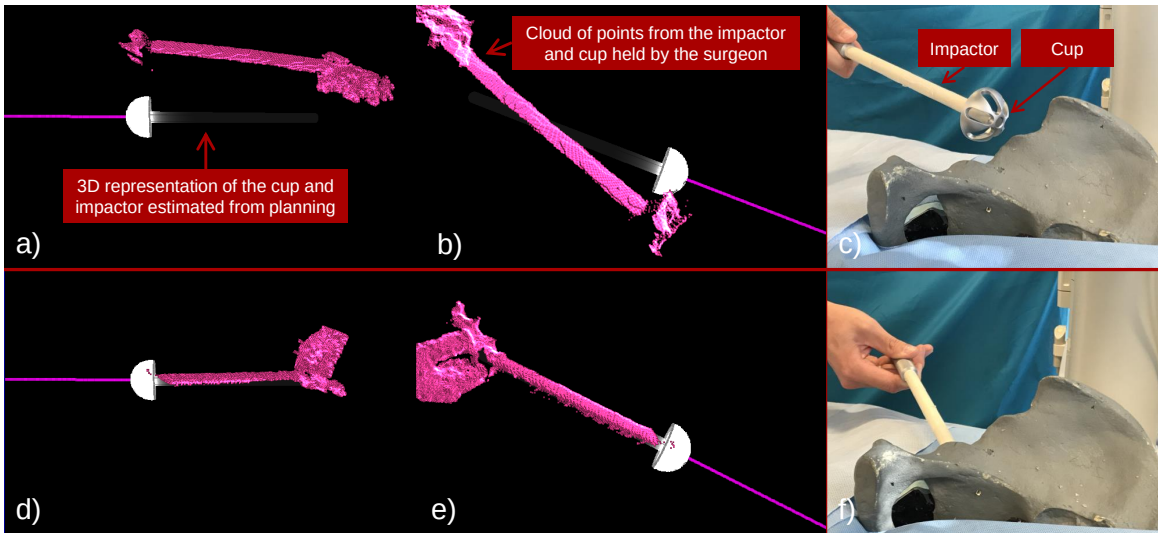

Figure 7.5: Multiple virtual perspectives of the surgical site are shown to the surgeon **(a-b)** before the cup is aligned **(c)**. The impactor is then moved by the user until it completely overlaps with the virtual planned impactor **(d-f)**.

mm (accuracy), and System Usability Scale (SUS) were recorded. Note that the task load is a accumulative scale, for which the score of 5 and 100 represents the lowest and highest possible load, respectively. A lower value is desirable. On the other hand, the SUS is a Likert scale, which gives insight in the subjective assessment of usability. It scales from 0 to 100, where a higher value indicates a better usability.

On average, a procedure lasted 9.9 min and 4.1 min using the standard C-arm and the 3D AR visualization, respectively. This represents a statistically significant improvement ($p < 0.05$). Furthermore, the reduction of the number of acquired X-ray images (40.9 to 2.1), dose (4.43 to 1.60 cGycm2), and surgical task load (43.5 to 17.6) are all significant ($p < 0.05$). The change accuracy (4.6 to 5.1 mm) is not significant, as it is already in an acceptable range. The relative changes of the observed measures

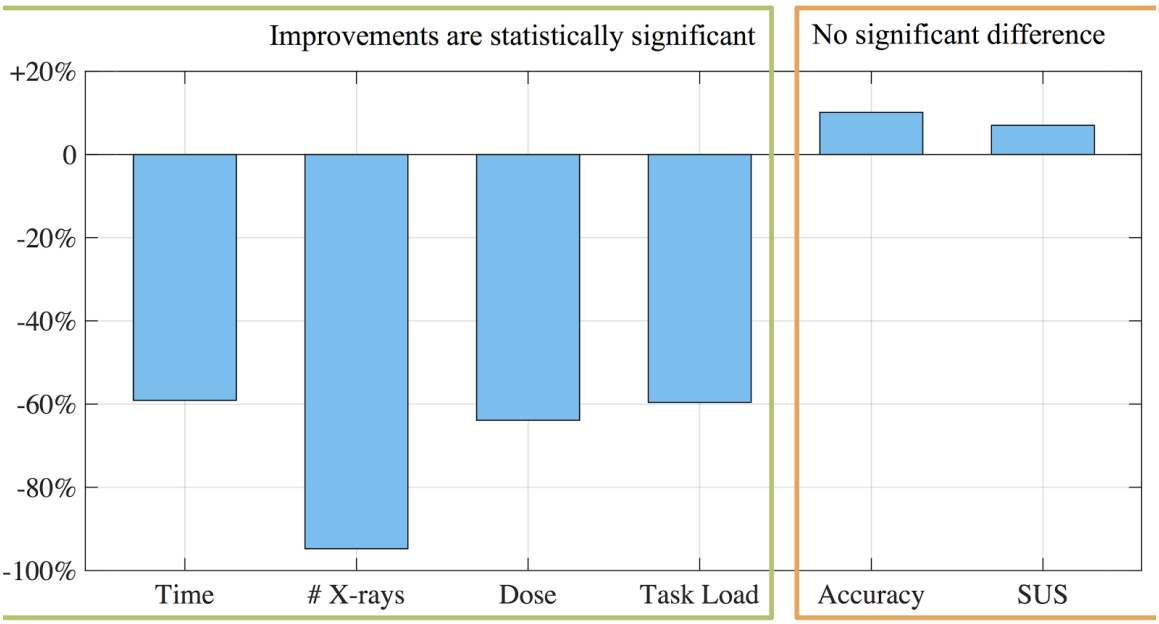

Figure 7.6: During the observed simulated K-wire placements, the improvements in terms of time (−59.1%), number of X-ray images taken (−94.8%), radiation dose (−63.9%) and reduction of task load (−59.6%) are statistically significant ($p < 0.05$). No significant change can be observed for change of accuracy (+10.1%) or System Usability Scale (SUS, +7%).

are illustrated in Fig. 7.6.

The subjective assessment of the usability was evaluated using the SUS. In general, all experts liked the system and the scale indicates a overall good usability (71.43), which is considered to be above average. In comparison, the usability of the standard C-arm was rated with an average score of 66.41.

7.5.2 Error Analysis for Total Hip Arthroplasty

Tracking accuracy is computed by acquiring X-ray images from a phantom with several radiopaque landmarks and measuring the stereo error between the corresponding landmark points in different images.

The phantom is constructed by attaching 9 radiopaque mammography skin markers (bbs) with diameters of 1.5 mm, inside and near the acetabulum on a pelvis model. Next, we acquired 11 X-ray images from $-50°$ to $+50°$ along C-arm oblique rotation, and 9 X-ray images from $-40°$ to $+40°$ on the cranial/caudal direction, with intervals of $10°$. In the planning software, we placed a virtual sphere with the same diameter as the bbs on each of the bb landmarks, and measured the distance of the bb in the second image to the epipolar line from the center of the corresponding virtual sphere in the first image. The error distance is measured as 7.58 ± 3.02 pixels[1] in an X-ray image with pixel size of 1024×1024 and pixel spacing of $0.22\,\frac{\text{mm}}{\text{pixel}}$. In addition, we acquired a cone beam CT (CBCT) scan of the phantom and measured a root mean square error of 1.37 mm between the bbs in the CT and those reconstructed using two X-ray images.

In order to measure 3D errors and ensure precise placement of the cup in two X-ray images during planning, we construct a dry phantom where an implant cup is screwed into the acetabulum. Therefore, the desired implant cup placement is well visible in the X-ray images and serves as a reference. We perform experiments where

[1] values reported as mean ± standard deviation

a virtual cup with the same size of the implant must be aligned precisely with the cup implanted a priori that is visible in the X-ray images. To evaluate the 3D error, we acquire a CBCT scan of the phantom and measure the error between the planning outcome and the ground-truth pose. This yields a mean translation error of 1.71 mm, and anteversion and abduction errors of 0.21° and 0.88°, respectively.

In image-guided DAA hip arthroplasty, the proper alignment of the acetabular component is frequently inferred from AP X-ray images [209]. Thus, the accuracy in estimating the 3D pose based on a single 2D image heavily depends on the surgeon's experience. We conduct another experiment and seek to demonstrate the clinical feasibility of our solution that is based on stereo X-ray imaging, and compare the outcome to image-guided DAA solutions that only use AP X-ray images for guidance. We refer to the latter as "classic DAA". Although the use of a single AP radiograph and the anterior pelvic plane coordinate system have certain drawbacks, it is the frame of reference that is most commonly used in computer-assisted THA solutions [210]. While there may be alternatives (e.g. coronal plane), the use of anterior pelvic plane as the frame of reference will enable direct comparison with the current literature.

We conduct a pre-clinical user study where medical experts use the planning software to place acetabular cups on simulated stereo X-ray images. These results are then compared to conventional AP-based method considering orientational error in abduction and anteversion. For the purpose of the user study, simulated X-ray images or so-called digitally reconstructed radiographs (DRR) are produced from a

cadaver CT data. We generate 21 DRRs from the hip area, starting at $-45°$ and ending at $+45°$ with increments of $+4.5°$ on the orbital oblique axis of the C-arm, where $0°$ refers to an AP image. Each time the users are given a randomly selected DRR together with the DRR corresponding to the AP plane, and are expected to place the acetabular cup such that it is properly aligned in both views.

As the spatial configuration of the DRRs are known relative to the APP plane, we are able to compute the correct rotation of the acetabular component, and preset this orientation for the cup in the planning software. This can occur when an AP image is acquired during the intervention and the desired orientation of the component is known relative to the anterior pelvic plane which allows locking the DOF for rotational parameters. When the orientation is preset, the user only has to adjust a translational component, substantially reducing the task load. Presetting the orientation of the cup is evidently only possible if the X-ray pose is known relative to the APP or the AP image.

Four orthopedic surgery residents from the Johns Hopkins Hospital participate in the user study. The translation error in placing the cup are shown in Fig. 7.7. The abduction and anteversion errors are measured as zero as a result of presetting the desired angles. The abduction and anteversion adjusted by the user solely using AP image (classic DAA) are $6.52° \pm 5.97°$ and $1.82° \pm 1.89°$, respectively. Ground-truth for these statistics includes the 5 DOF pose of the cup in CT data (as the cup is a symmetric hemisphere, 1 DOF, i. e. rotation around the symmetry axis, is redundant),

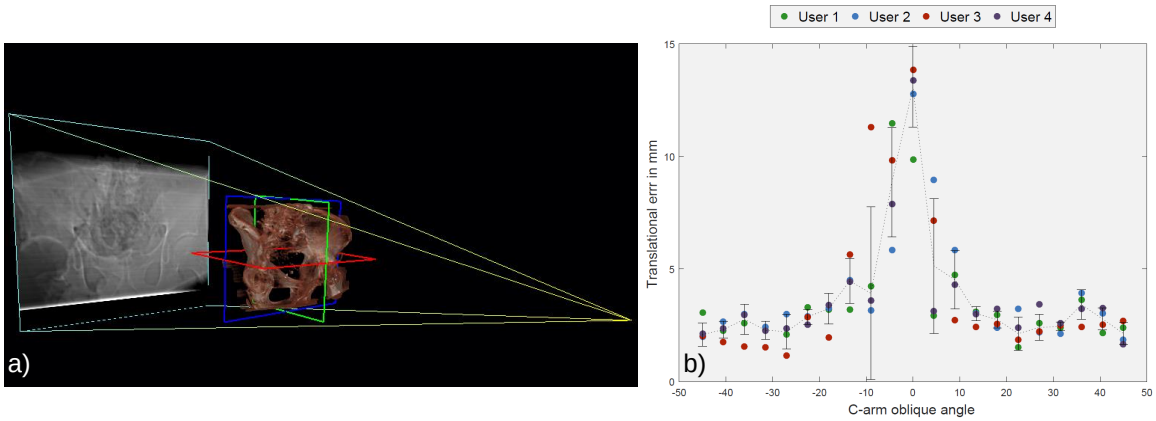

Figure 7.7: DRRs were generated from $-45°$ to $+45°$ around the AP view **(a)**. Participants were each time given two images, where one was always AP, and the other one generated from a different view. The translational errors are shown for all four participants in **(b)**. Note that $0°$ in the horizontal axis refers to where the user performed planning on only the AP X-ray image.

where abduction and anteversion angles are $40°$ and $25°$, respectively.

To evaluate the agreement between surgeons' actions in the AR environment with their intra-operative planning, we measure the orientational error of the impactor after placement with respect to its planning. The axis-angle error between the principal axis of the true and planned impactor in the AR environment are measured as shown in Fig. 7.8. We repeat this experiment for 10 different poses, and each time use four virtual perspectives of the surgical site. The orientational error is $0.74° \pm 0.41°$.

After the cup is placed in the acetabulum using AR guidance, we acquire a CBCT scan of the cup and measure the translation, abduction, and anteversion errors compared to a ground-truth CBCT as 1.98 mm, $1.10°$, and $0.53°$, respectively.

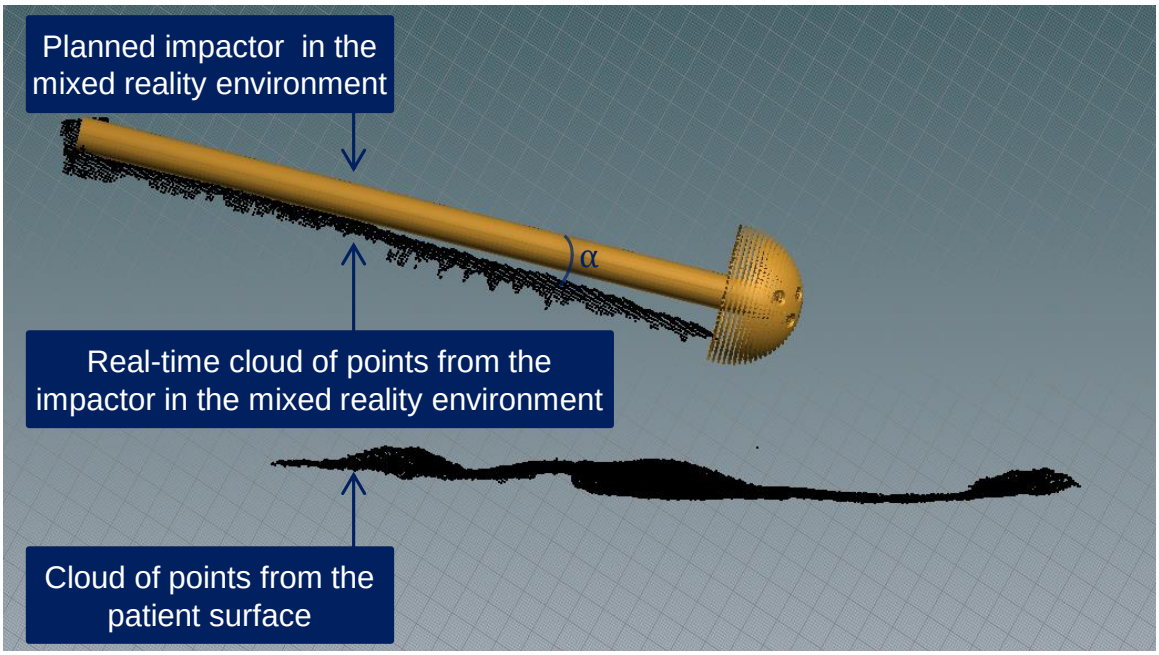

Figure 7.8: The angle between the principal axis of the virtual impactor and the cloud of points represent the orientation error in the AR environment.

7.6 Discussion and Conclusion

The K-wire usability study shows clear indications that an intuitive AR system has the potential to contribute as cost saving measure by significantly reducing OR time (-59.1%), while simultaneously reducing number of acquired X-ray images and accumulated area dose. Additionally, the study shows a reduced surgical task load (-59.6%), which may lead to an overall improved performance by reducing the mental demand of K-wire placements during surgeries.

This technique provides a perceptual fusion of optical and simulated X-ray images without the use of fluoroscopic navigation. In future, this system can complement

fluoroscopic navigation systems by offering a direct view into patient anatomy and allowing real-time visual feedback. This provides the surgeons with a unique and intuitive view of their hands and tools, the patient's anatomy and their relative position to each other in real-time, which allows for an easy, intuitive and accurate alignment of surgical material.

Regarding the method we proposed for THA, the experimental results indicate that the anteversion and abduction errors are minimized substantially compared to the classic DAA approach. The translational error is below 3 mm provided that the lateral opening between two images is larger than 18°. All surgeons participating in the user study believed that presetting the cup orientation is useful and valid, as having access to AP images in the OR is a well-founded assumption. Nonetheless, the authors believe that a pose-aware RGBD augmented C-arm [151] can, in future, assist the surgeon in acquiring and confirming true AP images considering pelvis supine tilts in different planes.

The translational and orientational error of the proposed AR solution is 1.98 mm and 1.22° respectively which shows reduced error compared to navigation-based system proposed by Sato et.al. with translation error of 2.98 mm and orientation error of 4.25° [44]. These results show the clear necessity to continue research and perform user studies on cadaveric specimens and quantify the changes in operating time, number of required X-ray images, dose, accuracy, and surgical task load compared to classic image-guided approaches.

In classic DAA hip arthroplasty, correct translation of the cup is achieved by naturally placing the acetabular component inside the acetabulum, and then moving the impactor around the pivot point of the acetabulum until the cup is at proper orientation. However, in order for our proposed solution to provide reliable guidance, both the translational and orientational alignments need to be planned.

The visual marker is only introduced into the surgical scene for a short interval between acquiring two X-ray images. These external visual markers could be avoided if incorporating RGBD-based simultaneous localization and mapping to track the surgical site [151]. Alternatively, the impactor which is a cylindric object could be used as a fiducial for vision-based inside-out tracking. It is important to note that surgical tools with shiny surfaces reflect IR beam. Tracking the surgical impactor is only done reliably if the surface has a matte finish, or it is covered with a non-reflective adhesive material.

Projection of the 3D hemispheric virtual cup onto the plane of X-ray images are done by utilizing the intrinsics parameters of the X-ray camera. These parameters are estimated while performing the checkerboard calibration. However, at different C-arm arrangements the focal length and principal point could slightly change due to gravity and flex in the C-arm machine. We quantified the drift in the principal point for $\pm 10^{\circ}$, $\pm 20^{\circ}$, and $\pm 30^{\circ}$ of C-arm lateral opening and the average shift was 5.17, 7.3, and 17 pixels on a 1024×1024 X-ray image. Considering the pixel spacing of the detector, these values are equivalent to 1.16 mm, 1.64 mm, and 3.82 mm drift on the

detector plane coordinate frame. To overcome the limitations of change of intrinsics in future, a look-up table could be constructed from pre-calibration of the C-arm at different angulations. The correct intrinsic parameters could then be retrieved from the table by matching the corresponding extrinsics from the inside-out tracking of the C-arm. To avoid small inaccuracies due to image distortion of the image intensifier, we placed the acetabulum near the image center where image distortion is minimal.

Further research is needed to confirm the applicability of the technology to the real surgical scenario, compensate for potential movements during interventions, and validate the observed benefits during larger pre-clinical and clinical studies.

7.7 Acknowledgments

Special thanks to Drs. Nassir Navab, Mehran Armand, Bernhard Fuerst, Greg Osgood, Herpal Khanuja, Alex Johnson, Russell Taylor, and Marius Fischer for their clinical and scientific support during the system's design and study's validation. I thank Dr. Clayton Alexander for his clinical advice for the design of the THA solution using AR. I also want to acknowledge Mr. Sing Chun Lee for setting up the visualization platform to display the joint point cloud and CBCT data, Mr. Giacomo Taylor for designing the environment for planning the acetabular component, and Mr. Emerson Tucker for transferring this technology to the Bayview Hospital and registering the RGBD camera to the pre-operative CT data.

Chapter 8

Spatiotemporal-Aware Augmented Reality Using Head-Mounted Displays

8.1 Introduction

Interventional image guidance is widely adopted across multiple disciplines of minimally-invasive and percutaneous therapies [211–214]. Despite its importance in providing anatomy-level updates, visualization of images and interaction with the intra-operative data are inefficient, thus requiring extensive experience to properly associate the content of the image with the patient anatomy. These challenges become evident in interventions that require the surgeon to navigate wires and catheters

through critical structures under excessive radiation, such as in fracture or endovascular repairs.

As the surgical expectancy increases, the communication between the surgeon, crew, and the information becomes an important concern. Ineffective communication leads to increase of surgery time, radiation, and frustration to a point where in fluoroscopy-guided procedures, instead of the X-ray technician, the surgeons may reposition the scanners to ensure the task-defined views are optimal [8, 215].

To bridge the inefficiency gaps in surgical workflows, researchers have investigated the importance of human factor considerations in improving the usability of surgical data [216, 217]. Recent works focused on facilitating the unmet interaction needs by introducing touch-less mechanisms such as gaze, foot, or voice commands [218–220]. We believe the high stakes of surgery necessitates efficient interaction between all actors in the operating room *i.e.* surgeon, anesthesiologist and staff to communicate and access information. This demands user-centric designs that can also accommodate fluid movement of information and surgical inference across the entire team.

Augmented reality (AR) solutions have gained popularity in computer-integrated surgeries, as they can provide intuitive visualizations of medical data directly at patients' site. Early works on surgical AR focused largely on multi-modal fusion of information and provided display-based overlays [221–223]. Subsequently, AR enabled the utilization of pre- and intra-interventional 3D data during therapies [93, 94, 224, 225].

CHAPTER 8. SPATIOTEMPORAL-AWARE AUGMENTED REALITY USING HEAD-MOUNTED DISPLAYS

The emergence of commercially available optical-see through head-mounted display (HMD) systems has led to development of AR solutions for various image-guided surgical disciplines, including percutaneous vertebroplasty, kyphoplasty, lumbar facet joint injection, orthopedic fracture management, bone cancer treatment, total hip arthroplasty (THA), interlocking nailing, cardiovascular surgeries, and surgical education [69, 75, 80, 82, 217, 226–229].

AR has served as image viewer that directly displays the data at the operative site using virtual fluoroscopy monitors, hence eliminating the conventional off-axis visualization through static monitors [230, 231]. Moreover, AR is used to provide navigational information during interventions [232–234]. These systems often rely on tracking of external markers, which require line-of-sight and invasive implantation into patients tissue that can hinder their usability. Andress et al. suggested a flexible marker-based surgical AR methodology which only required the marker to appear in the X-ray beam during the image acquisition, and was removed immediately after [235]. Recent inside-out localization strategies in AR have greatly favored the fluid workflow over explicit navigation, and have proved effective in eliminating the need for external markers [103, 236, 237].

This chapter introduces the methodology and usability of a novel spatially-aware concept that enables immediate interaction with the medical data and promotes team approach where all stake-holders share a unified AR experience and communicate effectively. Our methodology exploits the viewing frustum of the imaging devices and

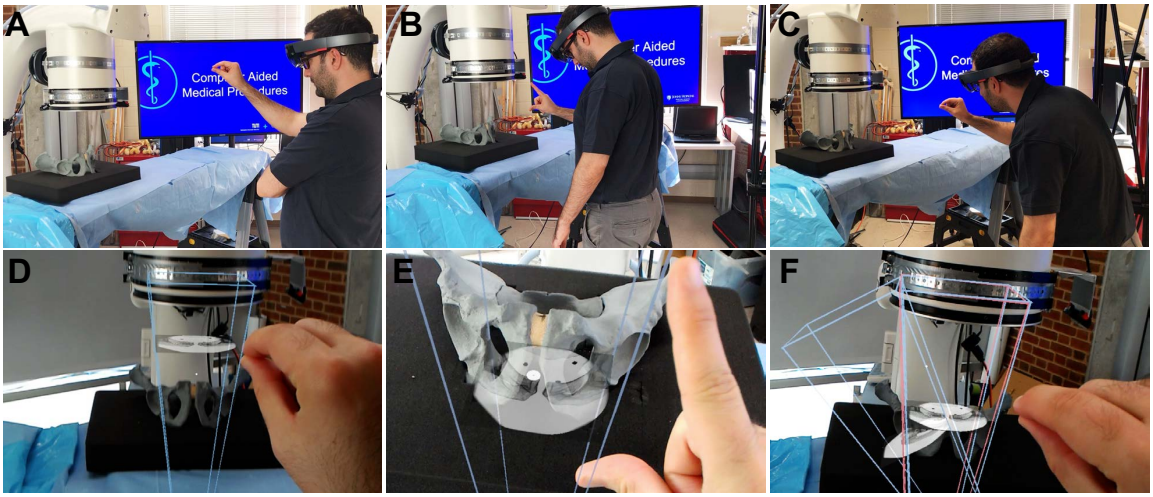

Figure 8.1: Spatiotemporal-aware AR exploits the full imaging geometry. The augmented user interacts with the X-ray images within their viewing frustums (**A-C**). Corresponding AR views are shown in **D-F**.

human observers in the operating room (Fig. 8.1), and provides an engaging and immersive experience for the surgical team. We built upon the concept of image frustums [237], and designed the complete methodology that integrates spatiotemporal-aware AR in the entire workflow. Our contributions are particularly centered around intra-operative planning, integration of the planning into the surgical workflow, and pre-clinical analysis. We showcase this solution in two high volume orthopedic procedures, *i.e.* K-wire placement in fracture care surgery, and acetabular cup placement in THA.

8.2 Methodology

Our main contributions are the collaborative AR concepts using spatiotemporal-aware flying frustums [237] that enable intra-operative planning, define new workflows, support surgical crew, enhance the communication between surgeon and data, and enable intuitive documentation of the surgery for training purposes. We present the methodology for the realization of these concepts in the remainder of this section.

8.2.1 Spatial-Awareness for AR

Visual data from cameras contain a wealth of information that can be used for simultaneous localization and mapping (SLAM). Visual SLAM is an important ingredient in our AR-based interaction recipe that enables co-localization of augmented users and the imaging device, which we will refer to as imaging observer, in a shared operating room environment. Marker-free co-localization will enable the 3D information to easily propagate through different bodies and be spatially-registered for all HMD users.

In the first step, each AR user is localized within the environment. The relative pose between two frames α and β using the environment map $\mathbf{M}$ can be estimated

by minimizing the following reprojection function:

$$
\begin{aligned}
{}^{\alpha}\mathbf{T}_{\beta} = & \underset{{}^{\alpha}\hat{\mathbf{T}}_{\beta}}{\arg\min}\ D({}^{\alpha}\hat{\mathbf{T}}_{\beta}, \mathbf{M}) \\
= & \underset{{}^{\alpha}\hat{\mathbf{T}}_{\beta}}{\arg\min} \sum_{f_i^{(\alpha)} \in I\alpha} |f_i^{(\alpha)} - P(\mathbf{M}(f_i^{(\alpha)}))|^2 \\
+ & \sum_{f_i^{(\beta)} \in I\beta} |f_i^{(\beta)} - P\,{}^{\alpha}\hat{\mathbf{T}}_{\beta}(\mathbf{M}(f_i^{(\beta)}))|^2,
\end{aligned}
\tag{8.1}
$$

where $f_i^{(\alpha)}$ and $f_i^{(\beta)}$ are corresponding features in images I_α and I_β, and P is the projection operator. In direct SLAM, the features include all pixels, and in indirect SLAM, the features are a sparse set of keypoints in the image. In Eq. 8.1, we optimize for the 6 parameters of a rigid transformation that best explains the pose between α and β using only the features present in the images. This step, also known as environment tracking, has become an standard part of most AR applications.

In a similar fashion, all users can be localized with respect to the first user, or with respect to a common spatial anchor in the operating room. The first member joining the shared experience will establish the anchor, *i.e.* OR coordinate system, and every other member of the AR session will share their pose in a master-slave configuration with respect to this OR frame [103]. This relation is shown as ${}^{\mathrm{OR}}\mathbf{T}_{\mathrm{S}}$ in Fig. 8.2.

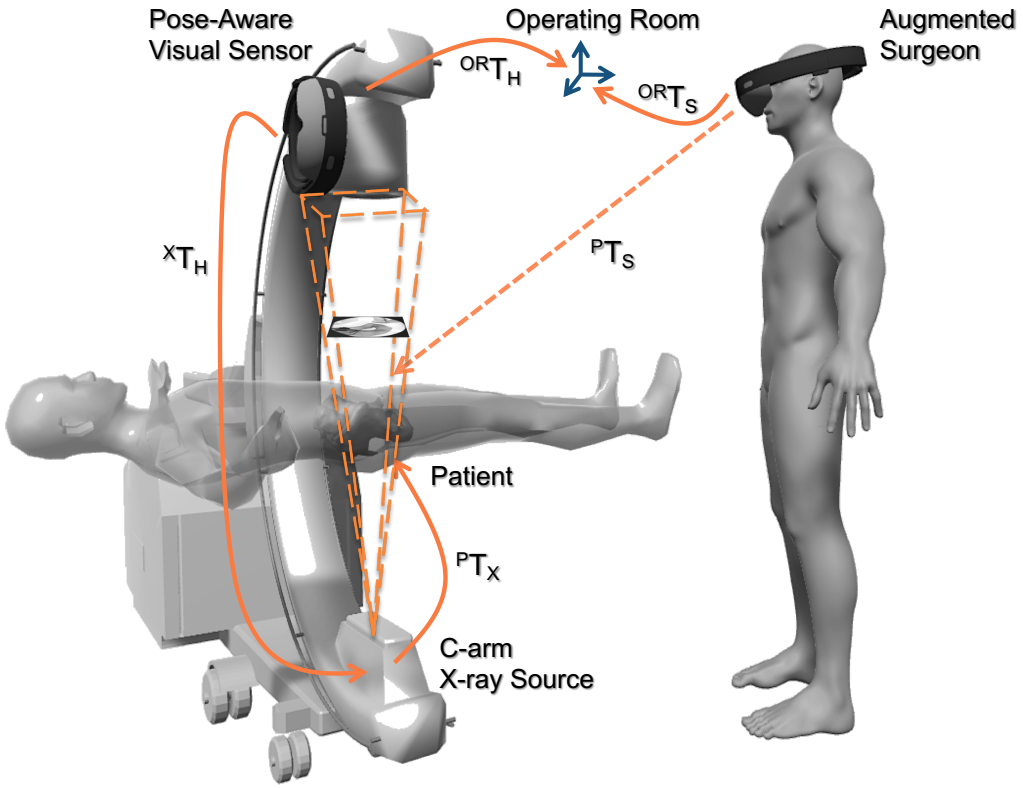

Figure 8.2: The transformation chain of the spatially-aware AR system is shown for a C-arm fluoroscopy system. The transformations layout show the closed-loop between the imaging device and the users at all time. The same relations can be generalized to include multiple users.

8.2.2 Imaging Observer

C-arm scanners offer fluorscopic imaging capabilities for a wide range of less-invasive therapeutic areas. To seamlessly integrate this imaging device into our interactive AR paradigm, we augment the scanner with a rigidly attached visual tracker, that observes the structures in the OR environment, and communicates spatial information to all users. The materialization of this imaging observer system requires a co-calibration between the visual tracker on the scanner, and the X-ray source [236]. The

constant transformation that explains the calibration is denoted as ${}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}$ in Fig. 8.2.

To estimate ${}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}$, we formulate an over-determined system of equations as follows:

$$\begin{aligned} {}^{\mathrm{IR}}\mathbf{T}_{\mathrm{OR}} &= {}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}(t_i)\,{}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}\,{}^{\mathrm{OR}}\mathbf{T}_{\mathrm{H}}^{-1}(t_i) \\ &= {}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}(t_{i+1})\,{}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}\,{}^{\mathrm{OR}}\mathbf{T}_{\mathrm{H}}^{-1}(t_{i+1}). \end{aligned} \tag{8.2}$$

IR denotes the frame of an external tracker that is used to track the motion of the C-arm source as it undergoes different motion at times t_i and t_{i+1}. To construct Eq. 8.2, the scanner is oriented in different poses, at each of which ${}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}$ and ${}^{\mathrm{OR}}\mathbf{T}_{\mathrm{H}}$, which are the poses of the X-ray source in the external tracker frame and the SLAM-based localization of the HMD in the OR coordinate system, respectively, are recorded. It is important to note that the IR tracker is only used for this one-time and offline co-calibration, and it is not used intra-operatively. By re-arranging Eq. 8.2, we formulate the problem in the form of $\mathbf{AX} = \mathbf{XB}$ as presented in Eq. 8.3, such that $\mathbf{X} := {}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}$, and $\mathbf{A}$ and $\mathbf{B}$ represent the relative motion of the X-ray source and the SLAM capable visual sensor on the gantry, respectively.

$${}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}^{-1}(t_{i+1})\,{}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}(t_i)\,{}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}} = {}^{\mathrm{X}}\mathbf{T}_{\mathrm{H}}\,{}^{\mathrm{OR}}\mathbf{T}_{\mathrm{H}}^{-1}(t_{i-1})\,{}^{\mathrm{IR}}\mathbf{T}_{\mathrm{X}}^{-1}(t_i). \tag{8.3}$$

Rotation and translation components of the hand-eye problem are disentangled and

computed separately as:

$$\begin{aligned} R_{\mathbf{A}} R_{\mathbf{X}} &= R_{\mathbf{X}} R_{\mathbf{B}} \\ R_{\mathbf{A}} \mathbf{t}_{\mathbf{X}} + \mathbf{t}_{\mathbf{A}} &= R_{\mathbf{X}} \mathbf{t}_{\mathbf{B}} + \mathbf{t}_{\mathbf{X}}. \end{aligned} \tag{8.4}$$

We estimate the rotation parameters using unit quaternion representation as $q_{\mathbf{A}}\, q_{\mathbf{X}} = q_{\mathbf{X}}\, q_{\mathbf{B}}$. Given that a unit quaternion $q_{\mathbf{i}}$ is formed by a vector $\mathbf{v_i}$ and a scalar $s_{\mathbf{i}}$ such that $q_{\mathbf{X}} = \mathbf{v_X} + s_{\mathbf{X}}$, we re-write the rotation component in Eq. 8.4 using the quaternion product rule as:

$$\begin{aligned} \vec{(.)} &: s_{\mathbf{A}} \mathbf{v}_{\mathbf{X}} + s_{\mathbf{X}} \mathbf{v}_{\mathbf{A}} + \mathbf{v}_{\mathbf{A}} \times \mathbf{v}_{\mathbf{X}} = s_{\mathbf{X}} \mathbf{v}_{\mathbf{B}} + s_{\mathbf{B}} \mathbf{v}_{\mathbf{X}} + \mathbf{v}_{\mathbf{X}} \times \mathbf{v}_{\mathbf{B}} \\ (.) &: s_{\mathbf{A}} s_{\mathbf{X}} - \mathbf{v}_{\mathbf{A}}.\mathbf{v}_{\mathbf{X}} = s_{\mathbf{X}} s_{\mathbf{B}} - \mathbf{v}_{\mathbf{X}}.\mathbf{v}_{\mathbf{B}}. \end{aligned} \tag{8.5}$$

Re-arranging the above formulation yields:

$$\begin{bmatrix} s_{\mathbf{A}} - s_{\mathbf{B}} & (\mathbf{v}_{\mathbf{A}} - \mathbf{v}_{\mathbf{B}})^{\intercal} \\ (\mathbf{v}_{\mathbf{A}} - \mathbf{v}_{\mathbf{B}}) & (s_{\mathbf{A}} - s_{\mathbf{B}}) I_3 + [\mathbf{v}_{\mathbf{A}} + \mathbf{v}_{\mathbf{B}}]_{\times} \end{bmatrix} \begin{bmatrix} s_{\mathbf{X}} \\ \mathbf{v}_{\mathbf{X}} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{0}_3 \end{bmatrix}, \tag{8.6}$$

which is then solved in a constrained optimization fashion as:

$$\min ||M\mathbf{q}||_2^2 \quad s.t. \;\; ||\mathbf{q}||_2^2 = 1, \tag{8.7}$$

where $\mathbf{q} = [s_X, \vec{\mathbf{v}_x}]^{\intercal}$. After the rotation parameters are computed, the translation vector is estimated in a least-squares setting: $(R_{\mathbf{A}} - I_3)\mathbf{t}_{\mathbf{X}} = R_{\mathbf{X}}\mathbf{t}_{\mathbf{B}} - \mathbf{t}_{\mathbf{A}}$.

At the end of this phase, we have achieved R_{X} and $\mathbf{t}_{\mathrm{X}}$ that express the calibration between the X-ray source and the integrated visual tracker; hence, we directly access the pose of the X-ray camera from the pose information acquired from the visual tracker on the scanner.

8.2.3 Geometry-Awareness for AR

In this section, we describe the underlying geometry that allows us to combine the content of 2D X-ray images, directly with the 3D spatial information we computed in Sec. 8.2.1 and 8.2.2. To this end, we explicitly model the viewable region of the X-ray camera, known as the flying frustum [237], and allow interaction with images within their geometries. It is important to note that, the flying frustum refers to the full pyramid of vision (Fig. 8.2), and is different than the truncated pyramids used in the computer graphics community. Despite the similarities in formulation, the conventional frustum model in graphics only applies to reflective images, and cannot accommodate the transmission model used in fluoroscopy. Therefore, we extend the perspective pinhole camera model that is commonly used in the computer vision community [180].

In our paradigm, users can move the images within their frustums on a virtual plane known as the *near* plane, between X-ray source and detector (referred to as the *far* plane), while they remain a valid image of the same anatomy. This interaction enables the users to intersect the images with corresponding anatomies, and

intuitively observe *2D-image-to-3D-anatomy* associations. Additionally, the imaging technologists which operate the scanner, can align the scanner with a desired frustum that is decided by the surgeon.

A flying frustum is defined using the following model:

$$P_f = \begin{bmatrix} \frac{n}{f} & 0 & 0 \\ 0 & \frac{n}{f} & 0 \\ 0 & 0 & 1 \end{bmatrix} K\,P \begin{bmatrix} {}^{\mathrm{OR}}R_{\mathrm{X}} & {}^{\mathrm{OR}}\mathbf{t}_{\mathrm{X}} \\ \mathbf{0}^{\top} & 1 \end{bmatrix}, \tag{8.8}$$

where n refers to the distance to the *near* plane, f is the focal length, K is the matrix of intrinsic parameters, and $0 \leq n \leq f$. The parameter n is controlled by the user, such that when $n = f$, the X-ray image is directly displayed at the detector scale. It is worth mentioning that, with conventional frustum models, the *near* plane can only take values smaller than the *far* plane, which is not the case in our representation.

For each 2D point $x_i \in \mathcal{I}$, where $\mathcal{I}$ is the domain of all acquired images, the corresponding point x_f in the frustum domain $\mathcal{F}$ is scaled by a factor s as $\mathbf{x}_f = s\,\mathbf{x}_i = (\frac{n}{f})\mathbf{x}_i$, such that $0 \leq n \leq f$. Finally, the 3D pose of the interactive image in

the frustum is defined as:

$$
\begin{aligned}
{}^{\mathrm{OR}}\mathbf{T}_{\mathrm{I}} &= \begin{bmatrix} {}^{\mathrm{OR}}R_{\mathrm{X}} & {}^{\mathrm{OR}}\mathbf{t}_{\mathrm{X}} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \begin{bmatrix} & & 0 \\ I_3 & & 0 \\ & & n \\ \mathbf{0}^{\top} & & 1 \end{bmatrix} \\
&= \begin{bmatrix} & r_{13}\, n & \\ {}^{\mathrm{OR}}R_{\mathrm{X}} & r_{23}\, n & +{}^{\mathrm{OR}}\,\mathbf{t}_{\mathrm{X}} \\ & r_{33}\, n & \\ \mathbf{0}^{\top} & 1 & \end{bmatrix},
\end{aligned}
\tag{8.9}
$$

where $R = \{r_{i,j}\}_{i,j:1,2,3}$.

8.2.4 Planning Using Flying Frustums

Flying frustums discussed in Sec. 8.2.1-8.2.3 embed sufficient 3D and 2D information that enable interventional planning for the placement of surgical tools (Fig. 8.3). In this section we introduce two distinct approaches for intra-operative planning. In the first planning approach, we forward-project the 3D virtual surgical implants onto the X-ray images within each frustum using the respective X-ray projective geometry. Hence, the user can observe the implant's projection in the same X-ray image and verify the implant's appearance before placing the real implant. This approach is generic and allows the planning of implants with any arbitrary shape. In the second

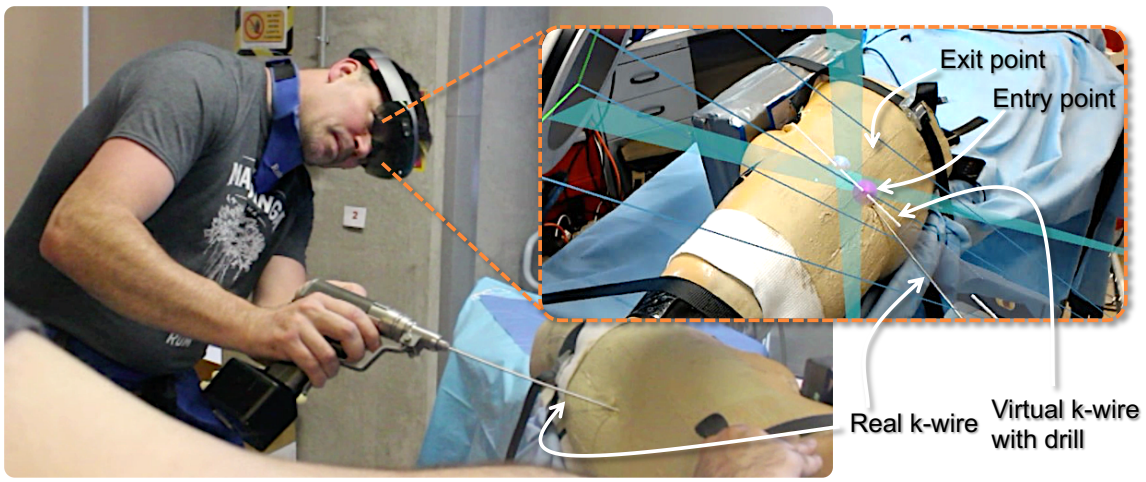

Figure 8.3: In K-wire placement for fracture reduction procedures, the surgeon can plan the entry and exit points of the wire on two X-ray images. After the planning, two triangular planes are constructed by connecting the drilling trajectory defined on the detector plane (X-ray image) and the C-arm source (X-ray origin). The intersection of these two planes is a line that corresponds to the desired drilling trajectory in 3D. By exploiting the imaging frustum, this line is augmented directly on the patient anatomy. The surgeon can then align the physical drill with its virtual counterpart, and advance the wire through the anatomy.

approach, we use multi-view geometry to reconstruct landmark targets or trajectories in 3D. In both methods, after the respective planning on the flying frustums, the resulting 3D information, already registered to the anatomy, is visualized on the patient. In the following, we describe each of the two approaches.

In the first method, virtual tools are manipulated in 3D by the user, and simultaneously projected onto the X-ray images of all valid frustums. A point $X_t \in \mathcal{T}$, where $\mathcal{T}$ is the domain of all 3D points on a virtual tool, is projected onto the i^{th} frustum as $x_{t_i} = P_{f_i} X_t$. The virtual tool is manipulated with complete 6 *degrees-of-freedom* (DOF) by the user to plan on all the frustums simultaneously. However, it may

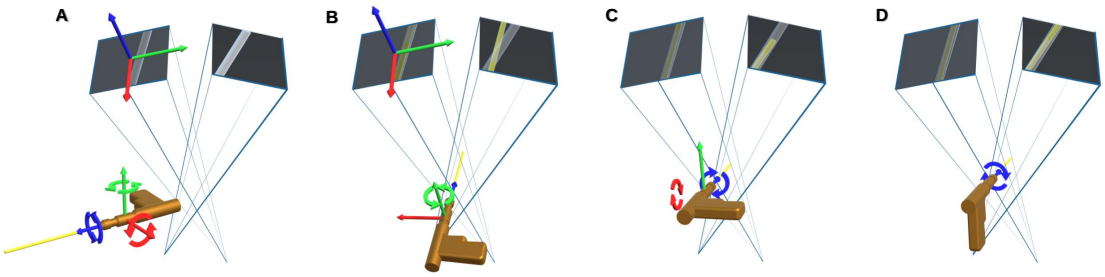

Figure 8.4: The coordinate frame of the virtual drill is defined as shown in **A**, where the Z-axis that is shown in blue points out of the drill along with the K-wire. **B** shows the manipulation of the virtual drill with 4 DOF until the projection of the K-wire is aligned in the first frustum (the yellow projection of the wire is contained within the tube in the left frustum). These 4 DOFs are sufficient to align the tool appropriately with the target anatomy in the first image. In the next step, we change the transformation constraints, as shown in **C**; after the alignment of the drill with the anatomy is verified in the first frustum, the drill maintains rotational freedom around its local X- and Z-axes, and translational freedom along its Y- and Z-axes. These DOF constraints allow the implant's alignment in the second frustum while maintaining the alignment between tool and anatomy in the first frustum. Finally, **D** shows the virtual drill being restricted to only 2 DOF. Moving and rotating along these two DOFs will not influence the alignment in either of the two frustums.

appear challenging to align concurrently in both views. Alternatively, we can apply rotational and translational constraints to the virtual tool, such that the tool can be aligned first in one frustum, and then in the second frustum, while the alignment in the first one is preserved. As shown in Fig. 8.4, in the first stage, the Y-axis of the virtual tool is rotated to hold the same direction as the Z-axis of the first frustum. The 4 DOF transformation model of the virtual tool is defined as:

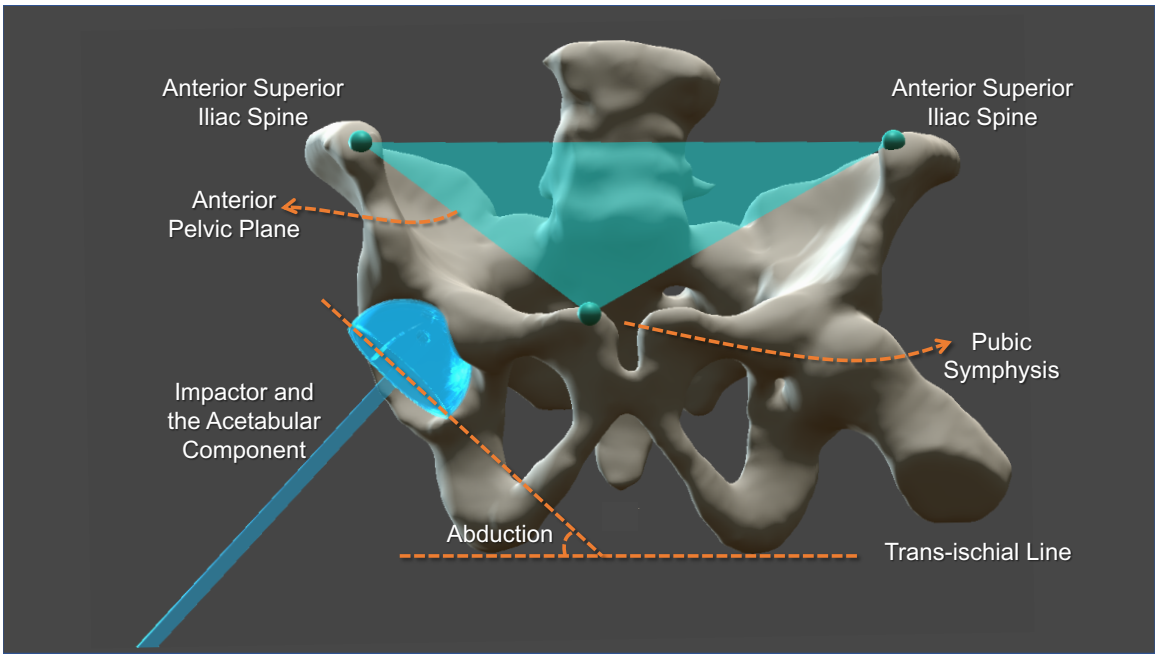

Figure 8.5: In THA, abduction and anteversion angles of the acetabular implant are defined with respect to the anterior pelvic plane (APP). The anterior pelvic plane is defined based on three points: the left and right anterior superior iliac spine landmarks, and the pubic symphysis. Once the surgeon annotates these landmarks, we identify the APP, and subsequently, render the acetabular components at appropriate angles. We also let the user annotate the center of the acetabulum in two or more views, which is used to calculate the 3D position of this landmark on the patient, hence allowing the center of the hemispheric component to render inside the hip socket. In the execution phase, the user aligns the real impactor and cup with their virtual counterparts.

$$\mathbf{T}^{f_1} = \begin{bmatrix} \cos\theta & 0 & \sin\theta & t_x \\ 0 & 1 & 0 & t_y \\ -\sin\theta & 0 & \cos\theta & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{8.10}$$

where θ is the rotation angle around the local Y-axis and t_x, t_y and t_z are trans-

lations along X, Y and Z-axes, respectively. In the second stage, the virtual tool is locked to only allow rotation around the X and Z-axes, and translation in Y and Z-axes using the transformation model as:

$$\mathbf{T}^{f_2} = \begin{bmatrix} \cos\psi & -\sin\psi\cos\phi & -\sin\psi\sin\phi & 0 \\ \sin\psi & \cos\psi\cos\phi & -\cos\psi\sin\phi & t_y \\ 0 & \sin\phi & \cos\phi & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{8.11}$$

where ϕ and ψ are the rotation angles around their X and Z-axes, respectively. As shown in Fig. 8.4, $\mathbf{T}^{f_2}$ does not influence the alignment of the tool and target in the image of the first frustum. Finally, after alignment in both frustums is achieved, the virtual tool then will only be constrained to have 2 DOF which is defined as:

$$T^{tool} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 & 0 \\ \sin\psi & \cos\psi & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8.12}$$

In an exemplary case shown in Fig. 8.7, the virtual drill is rotated and translated until it passes through a desired structure (e.g. through a bone canal) in all frustums. An alignment consensus in all frustums is the equivalent of the alignment of the virtual 3D tool with the imaged anatomy in 3D.

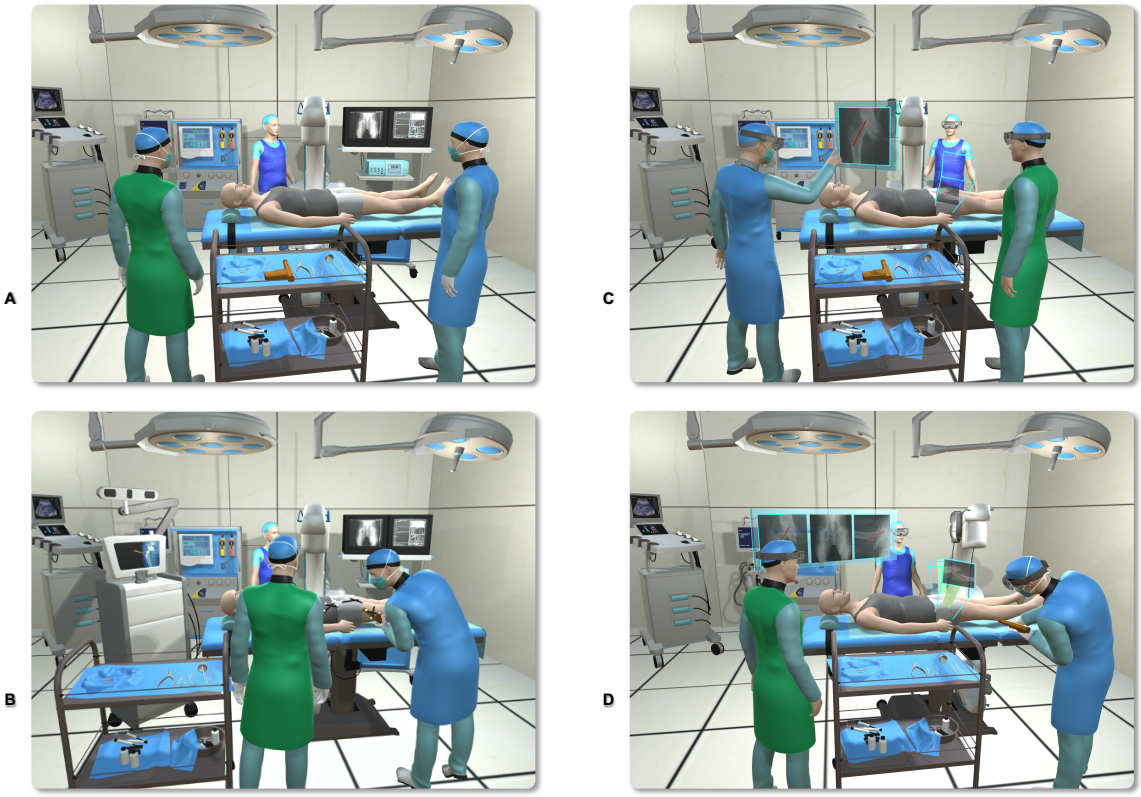

Figure 8.6: The standard operative procedure in percutaneous orthopedic interventions makes extensive use of interventional imaging (**A**). Classic navigation-based solutions use sophisticated tracking hardware and external markers to provide geometric registration between the content in the image and the patient (**B**). On the other hand, in the AR-enhanced OR that we suggest (**C**), the surgeon and crew interactively use the data and pass the information around without explicit navigation. Based on the concepts introduced in Sec. 8.2.4, the planning on X-ray images is directly visualized on the patient. The surgeon takes action based on the information from planning, as well as the X-ray images that are positioned within their respective frustums, both of which are seen through the HMD (**D**).

The second planning approach requires 2D interaction on the frustum X-ray images. In this setting, for each selected landmark on a frustum image, a 3D ray connecting the C-arm source and the target landmark will be rendered into the AR scene. As illustrated in Fig. 8.8, the intersection of two rays from a corresponding

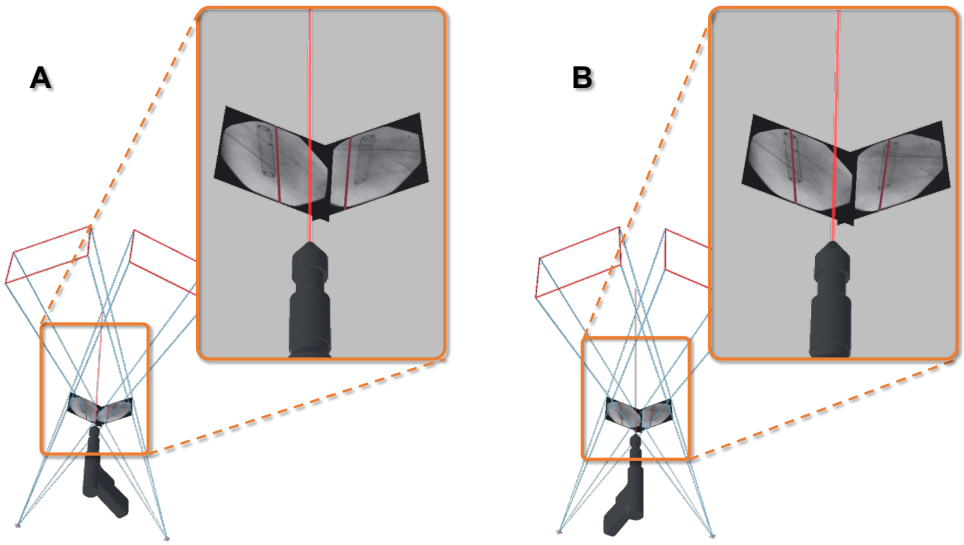

Figure 8.7: The augmented projections allow us to exploit the geometry in AR and plan surgical tools in relation to patient anatomy. The misaligned virtual drill in **A** is repositioned until it appears inside the desired structure in all the frustums (**B**).

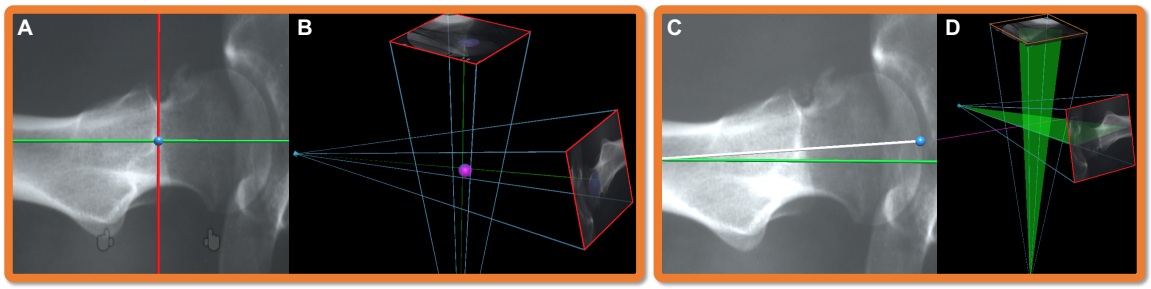

Figure 8.8: Each point in a frustum image corresponds to a ray passing through the landmark in 3D, and connecting the source and detector of the C-arm. Intersection of two rays recovers the 3D point and renders it directly on the patient (**A-B**). Similarly, annotation of lines in each frustum, corresponds to a plane in 3D. The intersection of these planes restores the 3D planning trajectory, and renders it in AR such that it travels through the corresponding anatomical structure (**C-D**).

landmark in two images reconstruct the 3D landmark. Each ray is defined via two elements: *i)* the position of the C-arm X-ray source $\mathbf{c}_i$, and *ii)* the unit direction vector $\mathbf{u}_i$ from the source to the annotated landmark in the frustum. We estimate the clos-

est point $\mathbf{x}_l^*$ to the $N = 2$ rays corresponding to each landmark l via a least-squares minimization strategy as follows:

$$\begin{aligned} \mathbf{x}_l^* &= \arg\min_{\mathbf{x}\in\mathbb{R}^3} \sum_{i=1}^{N} \|(I_3 - \mathbf{u}_i\mathbf{u}_i^\top)\mathbf{x} - \mathbf{t}_i\|^2, \\ &\text{where } \mathbf{t}_i = (I_3 - \mathbf{u}_i\mathbf{u}_i^\top)\mathbf{c}_i\,. \end{aligned} \tag{8.13}$$

Similarly, two points on a frustum i defining the entry and the exit points of a drilling trajectory, associate to two rays $\mathbf{u}_{1i}$ and $\mathbf{u}_{2i}$ in 3D. These two rays span a plane in 3D as shown in Fig. 8.8. The intersection of the planes corresponding to the same entry and exit points on frustums i and j form a 3D line $\mathbf{d}_{12} = (\mathbf{u}_{1i} \times \mathbf{u}_{2i}) \times (\mathbf{u}_{1j} \times \mathbf{u}_{2j})$ that passes through the desired entry and exit points on the patient anatomy.

Our first approach requires a more complex interaction with the augmented surgical implant using the 6 *degrees-of-freedom*, however generalizes to arbitrary structures beyond linear annotations, such as the curved plates used for internal fixations.

8.2.5 Surgical Workflow Integration

Intra-operative planning and execution with the flying frustums support can be used in various fluoroscopy-guided procedures. In THA, the critical points defining the anterior pelvic plane (APP) can be each identified on X-ray images. These anatomical landmarks include the left and right anterior superior iliac spine points on the pelvic wing and the pubic symphysis. Given APP, a virtual acetabular implant

and a rigidly attached impactor are rendered in AR with their desired orientation that is calculated with respect to APP. Likewise, the translational component of the cup implant is identified by defining the center of the patient acetabulum on corresponding fluoroscopic images. These relations are shown in Fig. 8.5. Once these intra-operative planning steps are completed by the surgeon, virtual representations of the cup and impactor are augmented over the patient's acetabulum with the appropriate abduction and anteversion angles. The surgeon could then align the impactor with its virtual counterpart reducing the amount of required X-rays. It is important to note that achieving the desired angles for the hip implant is a crucial step that is mentally challenging to verify from single view X-ray images, as it is commonly practiced in the direct anterior approach.

Another exemplary image-guided procedure is the placement of screws and K-wires during fracture management. As shown in Fig. 8.6, AR provides support for placement of K-wires using the trajectory planning on the corresponding frustums. Fig. 8.6 also depicts the use of our AR solution in the OR, and compares the proposed environment with the current OR.

Our proposed AR landscape is enabled by exploiting all involved frustums to move spatial information between different bodies, hence allowing multiple users to connect simultaneously. In Fig. 8.6, we show hypothetical procedures and further demonstrate that the scanner, crew and the technician can all share this common AR experience through HMDs, thus jointly benefiting from the augmented procedure. As highlighted

Figure 8.9: All acquisitions can be documented and later reviewed with all their corresponding spatial and temporal information. Spatiotemporal-aware AR allows the trainees to watch the surgery's progress and revisit the actions taken based upon each image.

in the figure, our system relies on 2D C-arm fluoroscopy, thus the standard workflow is only minimally altered. The surgeon can always alternate between fluoroscopy-based guidance and the AR view to ensure safe drilling. The figure also signifies the advantage that the surgeon does not need to take his/her gaze away from the patient site during implant placement. As shown in Fig. 8.9, all the spatial and temporal information can be documented for post-operative review and training.

8.3 Experimental results

8.3.1 System Setup

Our system comprises an ARCADIS Orbic 3D C-arm (Siemens Healthineers, Forchheim, Germany) as an intra-operative X-ray device that automatically computes the cumulative area dose for each session. The immersive AR solution was built using the Unity cross-platform game engine (Unity Technologies, San Francisco, CA, US) and was deployed to an optical-see-through HMD, the Microsoft HoloLens (Microsoft, Redmond, WA). To jointly co-localize the augmented surgeon and the C-arm scanner, a second HoloLens device with inside-out SLAM capabilities was attached near the X-ray detector. The two HMDs shared their spatial anchor, a rich feature reference region in the common environment, over a wireless local network, allowing them to remain synchronized and establish spatial awareness. This connection was enabled through a TCP-based sharing service running on an Alienware (Dell, Round Rock, TX, US) laptop server with an Intel i7-7700HQ CPU, NVIDIA GTX 1070 graphics card, 16 GB RAM, and Windows 10 operating system. The 16 bit 1024×1024 single channel X-ray images from C-arm were transmitted to the server computer over a direct Ethernet connection, and then converted to 8 bit grayscale while keeping the original resolution in order to display them properly on the HMD as well as to reduce the data size before uploading them to the HMD. The frame per second update for the HMD device was 60, and the display resolution was 720 p.

To minimize the chance of tracking failure, we mounted the tracking HMD on the C-arm such that the integrated vision cameras on the HMD observe the static structures such as walls and ceiling. Consequently, the chances of blocking the tracking cameras on the HMD during the procedure are diminished.

8.3.2 Experiments

Eight orthopedic surgeons and residents from Johns Hopkins Hospital participated in pre-clinical user studies and performed two surgically relevant tasks while utilizing interactive flying frustums in an immersive AR environment. The surgeons are referenced as **Pi**. **P6** is an attending surgeon. At the time of the study, **P7** and **P4** were in their final year of residency, **P5** in their fourth year, **P3** in their third year, **P2** in their second year, and **P1** and **P8** were first-year residents. All participants were familiar with the surgical tasks involved in this study, however, the more senior participants had naturally more experience in performing such tasks clinically.

In the first procedure, we focused on the correct placement of a K-wire to repair complex fractures. To emulate the K-wire placement through the superior pubic ramus (acetabulum arc), we used radiopaque cubic phantoms, as seen in Fig. 8.10-**C**. For direct comparison, we used the same setup that was used by Fischer et al. [94]. Each cube consisted of a stiff, lightweight, and non radiopaque methylene bisphenyl diisocyanate (MDI) foam. Since the superior pubic ramus is a tubular bone with a diameter of approximately 10 mm, we used a thin aluminium mesh filled with MDI

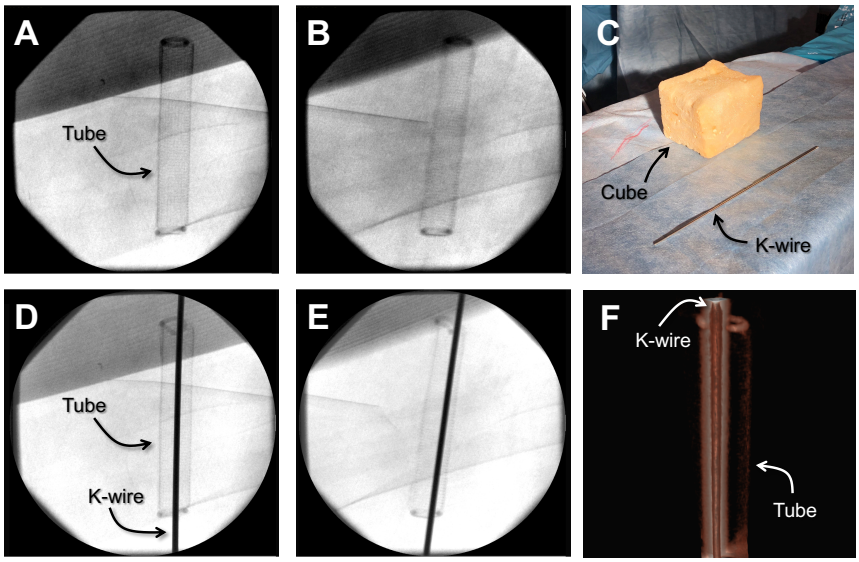

Figure 8.10: **A-B** are the X-ray images of the cubic phantom shown in **C**. In **D-E**, the X-ray images of the same phantom are shown after a K-wire was successfully inserted inside the tube. **F** is the CBCT scan of the phantom which was acquired for verification. Due to metal artifacts, the tube does not exhibit strong contrast.

that was placed inside each cube and served as the bone phantom. The two ends of the tubular structures were complemented with a rubber radiopaque ring. Each subject was asked to place a K-wire with a diameter of 2.8 mm through the tubular phantom using a surgical drill (Stryker Corporation, Kalamazoo, MI, US).

For the second procedure, we constructed a total hip arthroplasty mock setup by using a radiopaque pelvis phantom with a magnetic acetabulum to fixate the acetabular cup (Fig. 8.11). For direct comparison, we adopted the same experimental setup that was suggested by Alexander et al. [238]. The cup was attached to a straight cylindrical acetabular trialing impactor (Smith & Nephew, London, UK) allowing the operator to guide the cup. Since the ideal orientation of the implant is unknown, we use abduction and anteversion angles that lie in a safe zone defined by landmarks on

the pelvis as described in [189].

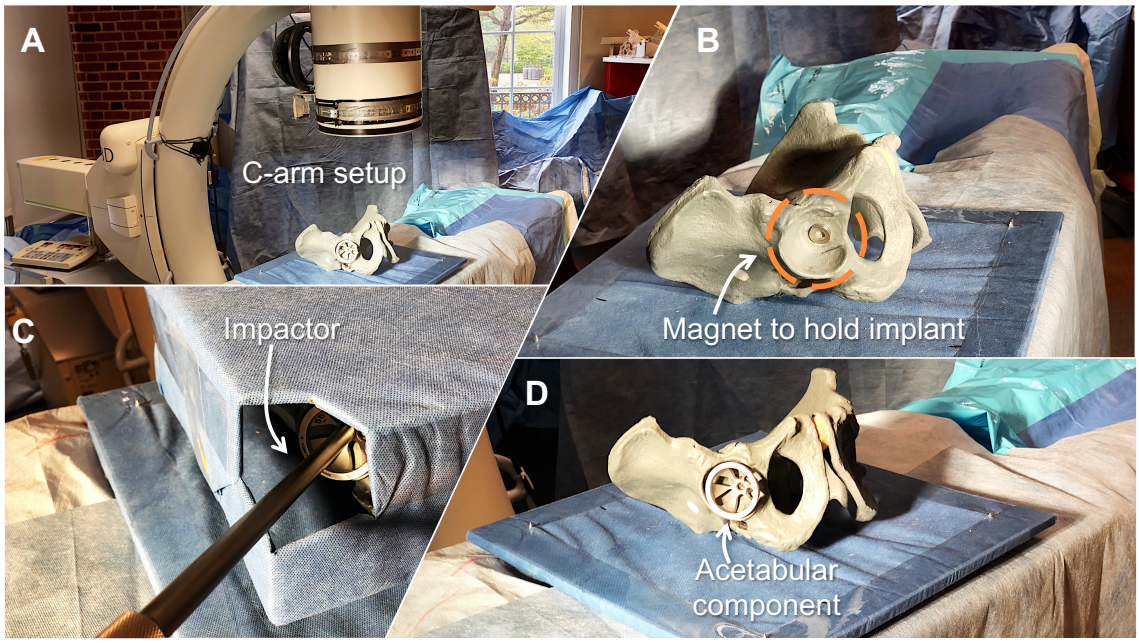

Figure 8.11: In **A** the setup of the C-arm, pelvic phantom, and the acetabular cup are shown. **B** is a close-up view of the phantom with an empty acetabular socket and a magnet for holding the implant in position. Image **C** shows the impactor while it is placed by a surgeon during the experiment, and **D** shows the successfully placed cup in the acetabulum.

Initially, each surgeon received a brief introduction to the Microsoft Hololens, preparing them to properly mount and use the HMD. To further instruct them on our AR application, pre-recorded training X-ray images were loaded onto their HMD, allowing them to become familiar with the task, the desired outcome, interface, planning procedure, and the interaction mechanism using hand gestures.

After the required planning images were acquired by the proctors, each surgeon planned their respective procedure in AR and performed the drilling task into the

Table 8.1: Outcome from the K-wire insertion using our immersive AR system. Individual performances are listed in columns **Pi**. Corresponding mean and SD values can be found in Table 8.4. The last row reports the error that was measured between the inserted K-wire and the center of the tube.

K-wire	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
Planning Time (sec)	125	40	39	58	79	**22**	67	44
Execution Time (sec)	83	74	58	**46**	66	67	63	79
# X-ray images	2	2	2	2	2	2	2	2
Dose (cGY(cm^2))	0.28	0.21	**0.19**	0.27	0.26	0.28	0.27	0.28
Error (mm)	8.23	5.71	9.02	3.26	6.94	**1.13**	1.59	2.23

cube or placed the acetabular component into the pelvis. During the procedure, they were explicitly allowed to order as many X-ray shots from any perspective that they considered necessary.

We recorded the planing time, the time it took them to execute the procedure, number of fluoroscopic acquisitions, and the cumulative radiation dose as it was measured by the scanner. Finally, for the verification and accuracy measurement, we acquired a 3D cone-beam CT (CBCT) scan of the phantoms with their respective implants. It is important to note that distance is not defined for two non-parallel lines. The distance we computed from the drilled path to the desired path, i.e., the average distance from the wire to the center-line of the target structure, is consistent with the past literature [94], and provides an intuition regarding the range of error.

8.3.3 Results

Tables 8.1 and 8.2 comprise the performance of every participant in the experiments. Table 8.1 contains the measurements for the K-wire insertion, and Table 8.2

Table 8.2: Outcome from the placement of the acetabular implant using our immersive AR system. Individual performances are listed in columns **Pi**. Corresponding mean and SD values can be found in Table 8.5.

THA	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
Planning Time (sec)	162	70	117	88	64	**37**	71	110
Execution Time (sec)	87	39	**13**	19	17	35	26	24
# X-ray images	8	8	8	8	8	8	8	8
Dose (cGY(cm^2))	1.27	1.3	1.23	1.26	**1.18**	1.25	**1.18**	1.29
Abduction error (°)	2.1	1	1.1	1.3	2.9	**0.1**	0.4	3.7
Anteversion error (°)	1.1	0.6	2.7	1.4	2.1	0.4	**0.3**	3.1

Table 8.3: Results of the respective SOP presented in [94] and [238]. Columns correspond to individual participants performance. Corresponding mean and SD values can be found in Table 8.4 and Table 8.5.

K-wire SOP

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**
Time (sec)	937	686	617	464	636	388	432
# X-ray images	80	47	44	33	32	21	29
Dose (cGY(cm^2))	7.68	1.73	3.54	4.38	5.62	2.69	5.38
Error (mm)	3.08	7.88	11.43	3.01	1.87	2.27	2.72

THA SOP

	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**
Time (sec)	210	350	195	375	225	150	95	280
# X-ray images	13	15	15	16	13	13	6	19
Dose (cGY(cm^2))	1.93	2	2.09	2.11	1.83	1.88	1.09	2.73
Abduction error (°)	7.9	5.9	5.1	6.5	5.6	2.3	1.5	3.3
Anteversion error (°)	4.9	1.3	9.9	5.1	1.4	6	1.8	7.8

presents the procedural outcome for the acetabular cup placement. We separate the interventional time measurements into *i) planning time*, the time it took each surgeon to plan their procedure in AR, and *ii) execution time*, determining the duration of the insertion/placement of the instruments. Furthermore, we recorded the number of X-ray acquisitions and the respective dose for each user. Finally, to assess the overall performances, we computed the average distance of the K-wire from the center of the

tube at the entry and the exit surface of the tubular structure, and the abduction and anteversion angles of the acetabular implant, based on standard guidelines.

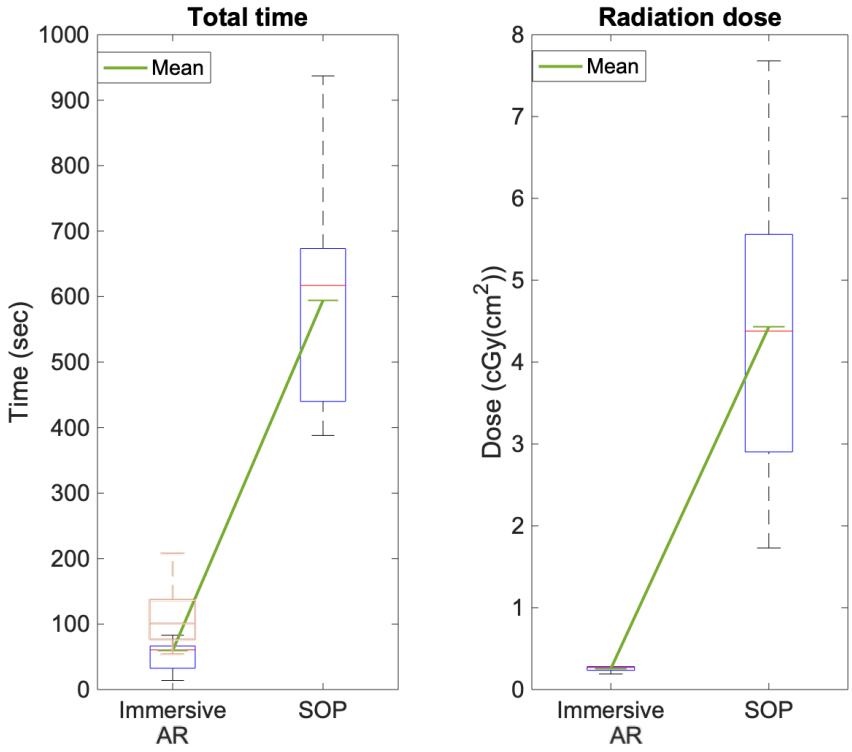

Figure 8.12: The plots present the execution time and total radiation dose during K-wire insertion using the AR supported approach and SOP. On the leftmost plot, the blue boxplot is the execution time with AR, whereas the orange boxplot is the total time including the planning phase. The green lines show the mean values for each of the groups.

Table 8.4 provides a comparison of the mean and standard deviation (SD) values of the K-wire insertion errors using our immersive AR system with a previous non-immersive AR system [94] as well as the standard operating procedure (SOP) using conventional fluoroscopic guidance. Combining the planning and execution times, the AR procedure took on average 111.25 sec versus the 594.3 sec during SOP. Fig. 8.12 depicts this comparison. On average, each surgeon made one trail and used 2 fluo-

Table 8.4: Mean and SD values for K-wire insertion with the immersive AR, NI-AR, and SOP. For each method two rows show the mean and SD values, respectively. For immersive AR, the time is separated into first planning and then execution.

	Method	**Time (sec)**	**# X-ray images**	**Dose (cGY(cm²))**	**Error (mm)**
$\overline{X}$	**AR**	$\mathbf{59.25 + 52}$	**2**	**0.255**	4.76
σ		$(\mathbf{32.02}, \mathbf{24.23})$	(0)	(**0.04**)	(3.11)
$\overline{X}$	**NI-AR**[94]	243.7	2.14	1.6	5.13
σ		(84.00)	(0.69)	(0.17)	(**2.72**)
$\overline{X}$	**SOP**	594.3	40.86	4.43	**4.61**
σ		(188.0)	(19.38)	(2.00)	(3.62)

roscopic shots with a combined dose of $0.255\,\mathrm{cGY(cm^2)}$ per user and committed an insertion error of $4.76\,\mathrm{mm}$. Given the eight samples, the population mean for the drilling error, based on the 95% confidence intervals, is between $[2.60 - 6.92]\,\mathrm{mm}$.

Table 8.5: Mean and SD values for acetabular cup placement with the immersive AR, NI-AR, and SOP. For each method two rows show the mean and SD values, respectively. For immersive AR, the time is separated into first planning and then execution time. In the #X-ray column of NI-AR, only one X-ray is denoted, this references the CBCT that was acquired before the experiment which is reconstructed of 100 digital radiographs.

	Method	**Time (sec)**	**# X-ray images**	**Dose (cGY(cm²))**	**Abd. (°)**	**Ant. (°)**
$\overline{X}$	**AR**	$\mathbf{89.88 + 32.5}$	8	**1.25**	**1.57**	1.46
σ		$(38.85, 23.71)$	(0)	(1.25)	(**1.24**)	(1.07)
$\overline{X}$	**NI-AR** [238]	110.6	1	1.83	1.78	**1.43**
σ		(**15**)	(0)	(**0.06**)	(1.37)	(**0.66**)
$\overline{X}$	**SOP**	235	13.75	1.96	4.76	4.78
σ		(96)	(3.73)	(0.45)	(2.2)	(3.15)

Table 8.6: Results from two sample t-tests of our AR method compared to the respective NI-AR method and the SOP. The upper half of the table shows results of the statistical evaluation of K-wire insertion with the AR and NI-AR version as well as AR and SOP. The lower half contains the corresponding values for the acetabular cup placement.

	P-value	**Time (sec)**	**Dose (cGY(cm²))**	**Error (mm)**	**Abd. (°)**	**Ant. (°)**
K-wire	**AR / NI-AR** [94]	$3.27 \cdot 10^{-5}$	$1.25 \cdot 10^{-11}$	0.81	-	-
	AR / SOP	$0.95 \cdot 10^{-5}$	$4.91 \cdot 10^{-5}$	0.93	-	-
THA	**AR / NI-AR** [238]	$1.64 \cdot 10^{-6}$	0.34	- -	0.49	0.22
	AR / SOP	$0.12 \cdot 10^{-1}$	5.39^{-4}	- -	$0.31 \cdot 10^{-2}$	$0.14 \cdot 10^{-1}$

Correspondingly, we present the outcome for the acetabular cup placement procedure with the immersive AR system in Table 8.5, comparing it to a previous non-immersive AR application [238] and SOP. Using SOP, it took surgeons on average 235 sec to place the cup and under AR a combined time of 122.38 sec was achieved. For the AR setup, we acquired 8 X-ray images with an average dose of 1.25 cGY(cm^2) per surgeon, whereas 14 fluoroscopic images with a dose of 1.96 cGY(cm^2) were acquired during SOP. With the AR system, the average errors were 1.57° and 1.46° for the abduction and anteversion angles, respectively. Based on the 95% confidence interval, the mean error for abduction and anteversion are between $[0.72° - 2.44°]$ and $[0.72° - 2.20°]$, respectively. Under SOP the respective angles were 4.76° and 4.78°. Figs. 8.15 and 8.13 present the outcome with respect to time, radiation dose, and individual rotational measures for the acetabular cup placement experiments using AR and SOP. The immersive AR results show an SD of respectively 89.88 sec and 32.5 sec for planning and execution time, 0 for the number of X-ray images, 1.25 for the dose, 1.24 for the abduction error and 1.07 in the anteversion error. Comparable to Tables 8.1 and 8.2, Table 8.3 displays the individual participants performance during each of the two SOP. To statistically evaluate our findings, a two sample t-test was performed and the results are reported in Table 8.6. We tested the the results of our immersive AR system against the results from the NI-AR system and SOP for both the K-wire insertion procedure and the acetabular cup placement. We did not test the number of acquired X-rays, since in these experiments they turned out to be

a constant value for each procedure.

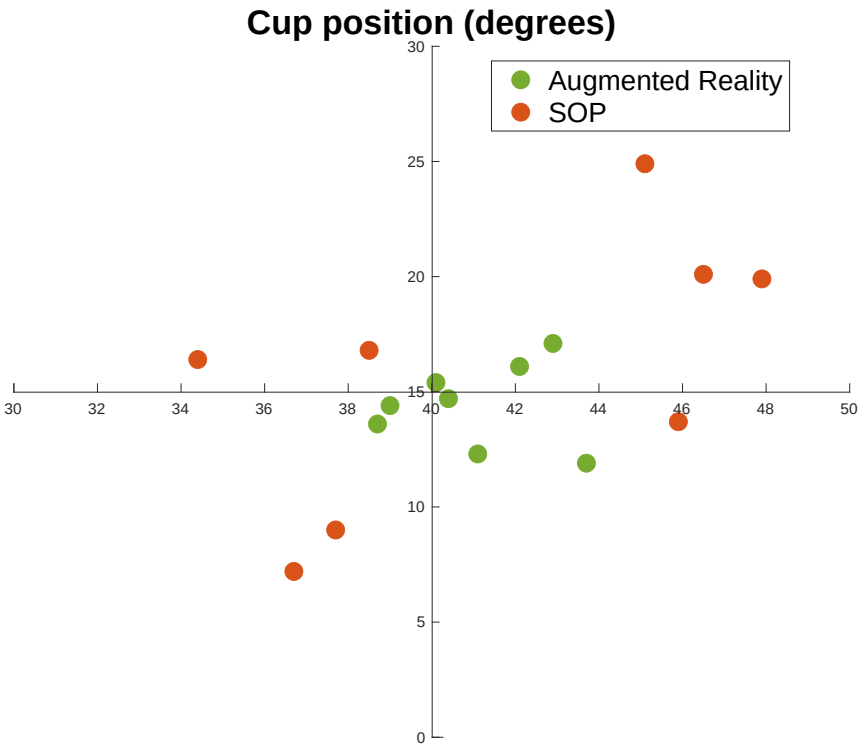

Figure 8.13: Anteversion and abduction angles are shown after acetabular cup placement using AR support and SOP. The horizontal axis represents the abduction angle, and the vertical axis represents the anteversion—the center of the plot corresponds to the desired angles of 40° and 15°. The farther data points from the center signify higher errors committed by the user. The AR method resulted in a stronger cluster near the center, while SOP yielded higher errors and more outliers.

8.4 Discussion

We evaluated our spatially-aware AR system in two clinically relevant procedures, *i)* the placement of K-wires through tubular structures for fracture repair tasks, and *ii)* placement of acetabular components into the hip socket for total hip arthroplasty.

We selected these two high volume procedures among the many other applications which can be enabled by our interactive AR system, as they each represent a class of common orientational alignment and localization tasks that are prevalent across different fields of image-guided surgery.

For the K-wire insertion procedure, the immersive AR system performed significantly faster than the conventional SOP, yielding less than a fifth of the time (Fig. 8.12). Table 8.4 demonstrates a detailed comparison of our system, not only with the SOP as an established baseline, but also with a previously presented non-immersive mixed reality method based on RGBD sensing and intra-operative CBCT imaging [94].

With the AR system every surgeon used exactly 2 X-ray images, which were the 2 images required for procedure planning. Despite explicitly allowing them to take as many radiographs as they desire, no one of the surgeons requested additional X-ray images. As mentioned above, during SOP, surgeons inserted the K-wires with an average number of 40.86 fluoroscopic images and with an average dose of $4.43\,\mathrm{cGY(cm^2)}$, compared to the (statistically) significantly lower dose of $0.255\,\mathrm{cGY(cm^2)}$, which was emitted during the AR procedures. The RGBD-CBCT system in [93, 94] yielded on average 2.14 X-rays, although it required a pre-procedural CBCT scan of the phantom, inducing the higher radiation dose of $1.6\,\mathrm{cGY(cm^2)}$.

Finally, evaluating the outcome of the procedure with regard to the drilling error, AR (4.76 mm) outperforms RGBD-CBCT (5.13 mm), both being marginally worse

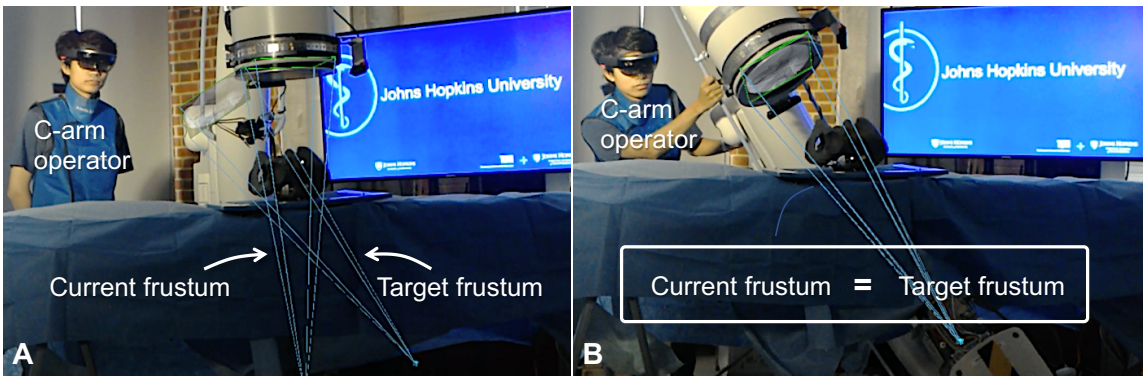

Figure 8.14: Visualization of a target frustum (**A**) allows the C-arm operator to align the current C-arm frustum with the surgeon's desired perspective (**B**) and eliminate the waste of time and radiation during fluoro hunting. This concept is an example of the capabilities of interactive frustums on moving information between different stake holders in the OR, *i.e.* surgeon, patient, X-ray technician, staff, etc.

than SOP (4.61 mm). Considering that we only instructed the surgeons to drill through the tube and not precisely through the center of the tube, we regard these difference as negligible. It is important to note that, our AR system performed similar to the conventional X-ray method in terms of accuracy, while reducing time by a factor of 5, number of fluoroscopic acquisitions by a factor of 20, and the radiation dose by a factor of 17.

We observed that, in addition to the planning information, the surgeons took multiple other considerations into account while deciding on the insertion path. One of which was direct visualization of the X-ray images that they acquired for planning. Observing the C-arm pose with respect to the patient, and the visualization of the images within their viewing frustum, assisted them in better localizing the target structure. In our setup, tactile feedback did not play a significant role; however,

in reducing real fractures, haptic feedback further assists the surgeon in identifying whether the wire is inserted in specific anatomy or not.

The SLAM-based error for HMD is dominantly along the principal axis of the viewer, i.e., direction pointing away from the user to the scene. When the users are presented with a trajectory to follow, we observed that they naturally place their heads in the direction of the target trajectory for optimal visualization and ease of alignment (Fig. 8.3). This configuration is advantageous since the maximum tracking uncertainty is in the direction of the drilling trajectory, which is the direction that is not relevant to the drilling task. This is because the amount of penetration of the drill is easily identified from X-ray images. This uncertainty behavior in AR has been previously investigated by Hoff et al. [239] and Mischke et al. [240].

A similar trend to the K-wire experiment is observed with the measurements for the placement of the acetabular cup, demonstrating the effectiveness of our AR system, we compare it against SOP and a NI-AR system as presented in [238]. As shown in Fig. 8.15, the execution time is considerably lower using AR; even when combining planning and execution time, it took the surgeons 122.38 sec, which is nearly half of the 235 sec that they needed under SOP and comparable to the 110.6 sec with NI-AR. Both differences are statistically significant. Furthermore, the number of fluoroscopic images were reduced; every surgeon used exactly 8 images, which are again merely the images required for planning. This resulted in an average dose of 1.25 cGY(cm^2), which is significantly lower than with SOP, where the surgeons used an average of

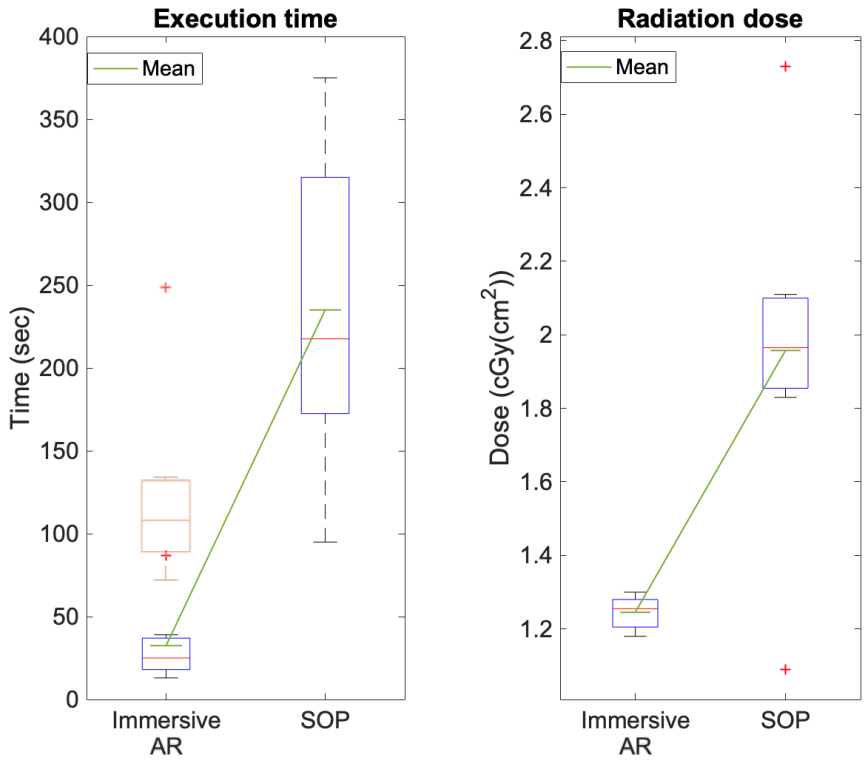

Figure 8.15: Comparison of time and total radiation dose during cup placement with AR and SOP approaches. The orange boxplot represents the total time including the planning time. The red (+) denote outliers, where in the leftmost plot the top sign belongs to the orange boxplot, and the bottom (+) to the blue plot.

13.75 radiographs with an average dose of 1.96 cGY(cm^2), and lower than with NI-AR where one pre-procedure CBCT lead to a dose of 1.83 cGY(cm^2). The objective of this procedure was to achieve abduction and anteversion angles of 40° and 15°, respectively, which lie in the clinical safe-zone [189]. The respective errors are shown in Table 8.5 and Fig. 8.13. The outcome distinctly displays a more accurate cup placement using the spatially-aware immersive AR system (1.57° & 1.46°) compared to the SOP (4.76° & 4.78°), compared to the NI-AR system (1.78° & 1.43°) the abduction error is slightly less, whereas the anteversion error is marginally higher (0.03°). The

differences in the abduction and anteversion errors between the immersive AR system and SOP are both statistically significant.

Concerning the planning and execution time, some participants such as **P2**, **P6**, and **P8**, who had previous experience with AR HMDs and were familiar with our system performed faster than other users, especially compared to **P1** and **P7**, who had no prior exposure to HMDs. In order to reduce the disparity in acquaintance with the technology, each participant completed a short training in which they were familiarized with the headset and interaction techniques used in our software. Despite the same training session, the differences mentioned above still seem related to experience with AR. We expect the performances to level after users gain more experience with the system. These outcomes are not surprising as any new technology requires time and experience to exploit its full potential.

For both procedures the deployment of our AR system lead to a comparable or higher accuracy, fewer X-ray images with a consequently lower radiation dose. For the total time, it has to be noted that our planning time does not include the recording of the X-ray images that were necessary to plan the procedure. This step however, as shown in [241], can be fully automated, resulting in an immediate availability of the fluoroscopic images.

8.5 Conclusion

This chapter presented the embodiment of a novel interaction concept based on spatiotemporal-aware AR. In our work, we aimed to provide meaningful registration and visualization without the need for tracking patients or tools with outside-in navigation hardware. Instead, we brought intuition to visualization by connecting the viewing frustums of the scanner with the surgical team. For the two orthopedic use cases presented in this chapter, our immersive AR system demonstrated improvements in time, number of X-ray acquisitions, radiation dose, and outcome during cup placement.

The most significant source of error is from the localization of the AR HMD using SLAM. This error is present both during calibration and application. If additional sensors such as IMUs and depth cameras would constraint the tracking algorithm in the future, we can expect improved tracking and smaller drifts.

The spatiotemporal awareness inherent in AR overhauls the ill-posed communication between the surgeon, staff, and information; *e.g.* Fig. 8.14 shows the potential role of flying frustums and AR in effectively communicating desired X-ray views to the technician, eliminating unfavorable views and reducing the staff burnout. Though assessment of this concept requires an additional approved study, we believe its introduction to the community paves the way in opening new paths for research in this area and expedites the translation of AR-based solutions into future ORs.

8.6 Acknowledgments

I want to thank Mr. Tianyu Song and Mr. Arian Mehrfard for their help in implementing the flying frustums, and their efforts during the user studies. Tianyu and Arian trained the participants on the new AR platform, described the workflows, and collected the appropriate data for the future analysis. Before the user studies, Tianyu re-calibrated the scanner with more reliable data, that resulted in an improved overall system error. My sincere thanks to Drs. Greg Osgood and Alex Johnson for supporting our user studies and connecting us to the residents at the Johns Hopkins Hospital. Finally, I would like to express my gratitude to Drs. Nassir Navab, Mehran Armand, and Mathias Unberath for their invaluable feedback during the implementation and validation of this work.

Chapter 9

Patient-Specific Image Augmentation using Partial Structural Symmetry

9.1 Introduction

Symmetry is an integral property of nature and is ubiquitous in human anatomy and living organisms. For instance, there is considerable amount of structural correlation across the sagittal plane of the human pelvis. Quantitative analysis of healthy pelvis data indicate that 78.9% of the distinguishable anatomical landmarks on the pelvis are symmetric [242], and the asymmetry in the remaining landmarks are still tolerated for orthopedic surgeries [81].

In the remainder of this section we highlight previous works that aimed at detecting symmetry in shapes and images (Sec. 9.1.1), present the importance of symmetry in surgical practice (Sec. 9.1.2), and propose a novel methodology for exploiting partial symmetry in human pelvis with an end-to-end solution to incorporate the knowledge from symmetry and augment surgeon's information during fracture care procedures (Sec. 9.1.3).

9.1.1 Related Work

There is a great body of work in the computer vision community that investigated symmetries in 2D images. The knowledge from symmetry has found several applications, namely in depth estimation [243], detecting camera projections [244], single-view scene reconstruction [245], and image segmentation [246]. In this chapter, we present the state of relevant art in the computer vision and the computational geometry community. It is important to note that most of these works are either applied to 2D images or 3D meshes. Hence, they can find applications in exploring symmetries in 2D medical images or 3D segmentation of the anatomies. The symmetry on 3D voxelized medical data, which is the center of focus in this chapter, has not been extensively investigated in the past.

Symmetry recognition has leveraged the success of feature detection methods in computer vision, and used image-based key-points to identify local and global symme-

tries. The approach by Loy and Eklundh [247] detects symmetric pair of points and forms local constellations of symmetries that collectively describe a global symmetry. This method searches for symmetries on multiple scales, and all orientations and locations. Bilateral symmetry can also be identified in 2D images under affine and perspective transformations by investigating vanishing points that link quadruplets of feature points in an image [248, 249]. To increase invariance to local illumination and improve robustness, affine invariant edge-based features were suggested as replacement of intensity-based features to locate planar symmetry, where each edge correspondence casts a vote to find the dominant reflection axis [250]. The work by Lee and Liu [251] studied glide-reflection, a combination of translation and reflection. Recent works have primarily focused on using convolutional neural networks to predict symmetries [252, 253].

Despite the advancements in symmetry detection in 2D images, the immediate translation of such techniques to medical imagery data is yet unclear due to different properties and use-cases. For instance, X-ray transmission imaging, in contrast to reflective imaging, is based on different X-ray attenuation from different tissue. A single pixel in an X-ray image relates to all 3D points along the ray. As a result of attenuation-based physics governing the image formation, depending on the viewing direction, anatomical landmarks may vanish or change appearance. Hence, classic feature detection methods fail to identify key-points [178], and symmetry detection based on feature points can become unreliable. Additionally, symmetry recognition

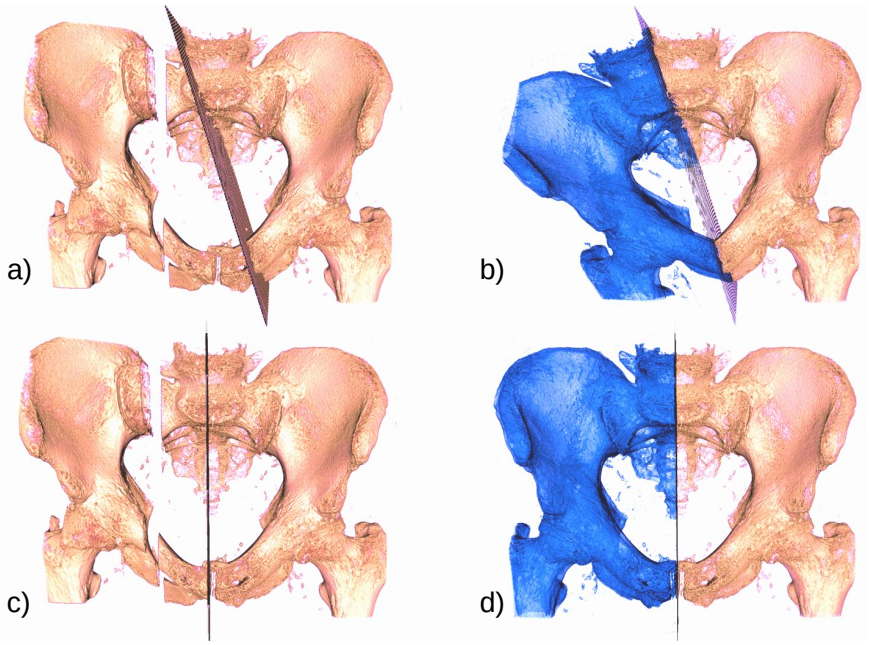

Figure 9.1: During an iterative strategy, the parameters associated to volumetric symmetry are estimated. In **(a-b)** the plane is visualized given the initial estimate, and in **(c-d)** it is visualized given the parameters at the convergence. The color blue represents the reconstructed bone model on the operative side of the patient.

is particularly useful for medical and interventional imaging where the knowledge of symmetry can transfer to 3D, and enable an understanding of the geometry of the contralateral side. This requires a more complex parameterization beyond detecting an in-plane axis of symmetry [254].

Determining symmetry in 3D shapes has been a topic of interest in the field of computational geometry [255, 256]. Kazhdan et. al. proposed a Fourier-based descriptor to score reflective symmetries associated with planes passing through the center of the mass of the objects [257, 258]. Podolak et. al. introduced an approach to recognize all symmetry planes, not limited to the ones passing through the center [259]. Symmetries were also detected in the form of intrinsic symmetries which

includes all self-isometric deformations [260, 261]. Lastly, bilateral reflective symmetries were estimated within point clouds from the real environments using methods inspired from the iterative closes point strategy [262–264].

9.1.2 Clinical Motivation

Ensuring quality of fracture reduction in pelvis and acetabulum surgery is paramount. Studies have consistently shown that anatomic fracture reduction and stable fixation leads to improved outcomes in patients with these complex injuries [265–270]. For example, in a study of 31 patients who underwent open reduction internal fixation of an isolated pelvic fracture, Pastor et. al. demonstrated that clinical outcomes correlate with pelvis symmetry at a six-months time point [265]. Furthermore, studies demonstrate that anatomic articular surface reduction is critical in preventing post-traumatic arthritis in the acetabulum [271].

Surgeons have sought to take advantage of osseous symmetry for performance and assessment of fracture reduction. For example, Zhang et. al. 3D printed a mirrored model of the healthy side for comparison in lower limb long bone fractures [272]. Symmetry has also been shown to be useful in the distal radius and facial fractures [273, 274]. However, neither viewing a 3D printed mirrored version of the healthy contralateral side, nor a mirrored radiographic image of the healthy side allow the surgeon to visualize the live comminuted operative side and the mirrored healthy side simultaneously and in the same position and orientation on the surgeon's display. The ability

to do so, could be substantially beneficial for intra-operative fracture reduction and assessment in pelvis and acetabulum fracture surgery. Similar concept can apply to other fields of surgery including brain and crainiofacial procedures [275–279], or breast reconstruction procedures [280–282].

9.1.3 Proposed Solution

In this work, we outline an end-to-end methodology to detect partial symmetry in human anatomy and exploit this knowledge intra-operatively as a reference to restore structural symmetry of fractured pelvic bone. Our fully automatic solution uses a structure-based cost, an intensity-based robust estimator, and a probabilistic-based loss to identify the plane of symmetry. After identifying the partial symmetry, healthy side of the patient anatomy is mirrored across the bilateral symmetry plane, which then allows simulating the ideal bone fragment configurations. This information is provided by overlaying the fluoroscopic image with a forward-projection of the mirrored anatomy obtained from the CT data. Our contributions enable the surgeon to use patient CT scans intra-operatively, without explicitly viewing the 3D data, but instead using 2D patient-specific image augmentation on commonly used X-ray images. In contrast to surgical navigation systems that provide update in relation to pre-operative data, our solution provides interventional feedback with respect to the desired outcome.

9.2 Materials and Method

9.2.1 Theory

Bilateral reflective symmetry is defined as the group of involutive isometric maps $\mathbf{M}_g \in \bar{E}(3)$, where $\bar{E}(3)$ consists of self-isometries such that $\bar{E}(3) = \{h \in E(3), \mathbf{o} \subseteq \mathbb{P}^3 \mid h(\mathbf{o}) = \mathbf{o}\}$, where $\mathbf{o}$ is defined in the 3D projective space $\mathbb{P}^3$. The group $E(3)$ denotes all isometries of $\mathbb{R}^3$. Transformation $\mathbf{M}_g$ mirrors the object $\mathbf{o}$ across a symmetry plane such that $\mathbf{o}^- = \mathbf{M}_g(\mathbf{o}^+)$, where $\mathbf{o}^-, \mathbf{o}^+ \subseteq \mathbb{P}^3$ are sub-volumes of object $\mathbf{o}$ on the opposite sides of the symmetry plane.

Assuming the plane of symmetry is the Y-Z plane, the extrinsic symmetry is expressed via $\mathbf{M}_g := g\, m_x\, g^{-1}$, where g is a member of the Special Euclidean group $SE(3)$, and m_x reflects the space about the X-axis:

$$m_x = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9.1}$$

Transformation $\mathbf{M}_g$ maps the points $\mathbf{p}, \mathbf{q} \in \mathbb{P}^3$ to $\bar{\mathbf{p}} = g\, m_x\, g^{-1}\, \mathbf{p}$ and $\bar{\mathbf{q}} = g\, m_x\, g^{-1}\, \mathbf{q}$, respectively. The distance between $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$ is then computed as:

$$\|\bar{\mathbf{p}} - \bar{\mathbf{q}}\|_2 = \left\| g\, m_x\, g^{-1}\, (\mathbf{p} - \mathbf{q}) \right\|_2. \tag{9.2}$$

Since, $\det(g\, m_x\, g^{-1}) = \det(g)(-1)(\det(g))^{-1} = -1$, then *i)* $\|\bar{\mathbf{p}} - \bar{\mathbf{q}}\|_2 = \|\mathbf{p} - \mathbf{q}\|_2$, hence $\mathbf{M}_g$ is an isometry, and *ii)* due to the negative determinant, $\mathbf{M}_g$ is orientation reversing, therefore is an anti-conformal map.

9.2.2 Problem Formulation

We propose to estimate extrinsic symmetry parametrization by minimizing the following cost:

$$\arg\min_{g} D(\mathbf{M}_g|g_0) := d_I(\mathbf{o}, \mathbf{M}_g(\mathbf{o})) + \lambda\, d_D(\mathbf{o}, \mathbf{M}_g(\mathbf{o})). \tag{9.3}$$

In Sec. 9.2.3 the prior parametrization g_0 is automatically estimated based on surface correspondences detected from the anatomy. In Sec. 9.2.4 we present a robust estimator term that will be used to minimize an intensity-based distance $d_I(.)$, followed by a distribution-based regularization term $d_D(.)$ which will be discussed in Sec. 9.2.5. Fig. 9.1 demonstrates the iterative step that yields the optimal plane of bilateral symmetry by minimizing the total loss $D(.)$. Finally, surgical image augmentation using interventional image registration is discussed in Sec. 9.2.6.

9.2.3 Automatic Initialization of Extrinsic Symmetry via Global Intrinsic Symmetry

To compute an initial estimate g_0 for the plane of partial symmetry, we first estimate a set of point correspondences on the contralateral sides of the anatomy which satisfy the properties of intrinsic symmetry. Intrinsic symmetry $\mathbf{M}_i(.)$ is associated with all volumetric deformations that preserve pairwise geodesic distances on a symmetric surface. All correspondences are detected automatically and globally [261], and are used to compute an initial estimate of the partial symmetry plane. An overview of this step is demonstrated in Fig 9.2.

9.2.3.1 Symmetry Invariant Candidates

The candidate symmetry invariant correspondences p are selected such that they share a common intrinsic symmetry, *i.e.* $\mathbf{M}_i(p) = p$. Selecting symmetry invariant point candidates are particularly crucial in the pelvic trauma application due to the imperfect and incomplete symmetry which are caused by fractures and dislocations.

The critical points p of a symmetry invariant function $\Phi = \{\Phi(.) : \mathbf{o} \rightarrow \mathbb{R}, \Phi(\mathbf{M}_i(p)) = \Phi(p)\}$, are invariant to symmetry on the surface of the object $\mathbf{o}$. To verify this, from chain rule, we drive:

$$\nabla\Phi(\mathbf{M}_i(p))\ \mathbf{M}_i'(p) = \nabla\Phi(p). \tag{9.4}$$

For the critical point p_c, it can be shown that $\nabla\Phi(\mathbf{M}_i(p_c)) = 0$ *if-and-only-if* $\nabla\Phi(p_c) = 0$, implying that $\mathbf{M}_i(p_c) = p_c$, hence p_c satisfying symmetry invariance condition.

The Average Geodesic Distance function is used as the symmetry invariant function to generate candidate points and is defined as below [261]:

$$\Phi(p) = \int_{q \in \mathbf{o}} d(p, q)\, dq. \tag{9.5}$$

9.2.3.2 Optimal Intrinsic Symmetry via Möbius Transform

Every genus zero surface can be mapped to a unit sphere $\mathbb{S}$, also known as the Riemann sphere [283]. The group of Möbius transformations models all the angle preserving isometries between the Riemann sphere to itself. Via stereographic projection, Riemann sphere can then be mapped to the extended complex plane. Therefore, Möbius transformation is represented as the mapping between extended complex planes. Möbius transformation is formulated on the extended plane via the fractional linear function, also known as homographies:

$$h(z) = \frac{az + b}{cz + d}\,, \qquad a, b, c, d \in \mathbb{C}. \tag{9.6}$$

Möbius group that models all isometries is characterized using only 6 real parameters, hence 3 point correspondences on the complex plane are sufficient to uniquely compute a Möbius transformation in a closed form. This property allows the intrinsic symmetry

on a mesh to appear as extrinsic symmetry on the extended complex plane. Using only 6 parameters has the advantage of computing the parametrization for intrinsic symmetry both globally and efficiently.

The search for the parametrization of symmetry takes place on the anti-Möbius group, which includes the Möbius group augmented with reflections [261]. The anti-Möbius transforms cover the set of orientation reversing isometries which are instances of the anti-conformal maps.

To use the Möbius parameterization, we first construct a binarized volume by segmenting bone from tissue and air using histogram thresholding. Next, a genus zero surface is constructed from the object using Reeb graphs [284]. The pseudo-code for this step is presented in Alg. 3.

Algorithm 3 Construction of the genus zero mesh

initialize empty graph $G = (V = \{\}, E = \{\})$
initialize empty vertex set $V_{\text{previous}} = \{\}$

for each transverse slice $s = 0, 1, 2, \ldots$ **do**
    initialize empty vertex set $V_{\text{current}} = \{\}$
    for each connected component $c = 0, 1, 2, \ldots$ **do**
        add vertex with tag (s, c) into V and V_{current}
        for each element (s_p, c_p) in V_{previous} **do**
            if (s, c) is connected to (s_p, c_p) **then**
                add edge $(s, c) - (s_p, c_p)$ to E
            end if
        end for
        $V_{\text{previous}} = V_{\text{current}}$
    end for
end for

close all cycles in the Reeb graph G by adding the convex hull of the connected components

On the genus zero surface, a set of feature points (triplets and quadruplets of point correspondences) are selected iteratively, and each time the parameters of a unique Möbius transformation is estimated. Followed by each step, all surface elements on the complex plane are warped given the current Möbius estimate, and pair-wise geodesics are measured. The mapping that yields the most number of mutually closest points (most inliers) is selected to parametrize the intrinsic symmetry on the surface [285].

9.2.3.3 Refinement of Point Correspondences

To incorporate the most reliable candidates for estimating the parameters of extrinsic symmetry from the set of point correspondences, we employ a two-stage RANdom SAmple Consensus (RANSAC) strategy [138]. After the completion of this step, the outlier correspondences that are not with agreement with the global symmetry properties are eliminated from the list.

The vector connecting a point p on the surface to its reflection $p' = gm_xg^{-1}p$ is given by $(gm_xg^{-1} - \mathbb{I})p$. For any arbitrary point p, it can be shown that this vector is parallel to $\mathbf{g}_1$, where $g = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \mathbf{g}_3 & \mathbf{g}_4 \end{bmatrix}$. This equality can be intuitively explained given the convention that the plane normal is parallel to the X-axis of the local frame. Therefore, bilateral reflections across the symmetry plane occur in the direction of the local X-axis. The local X-axis of the plane is parallel to $\mathbf{g}_1$ which expresses the image of the world X-axis: $\mathbf{g}_1 = g \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\top$. Hence, the vectors of all correspondences that satisfy extrinsic symmetry are jointly parallel to $\mathbf{g}_1$. We

leverage these relations in a RANSAC setting and eliminate all correspondences with non-agreeing connecting vectors.

During a second RANSAC, we seek consensus among the midpoints of the correspondences. The midpoint $\bar{p}$ lies on the symmetry plane and is invariant to reflection:

$$\begin{aligned} \mathbf{M}_g\,\bar{p} &= g m_x g^{-1}(\frac{p+p'}{2}) = g m_x g^{-1}(\frac{p+g m_x g^{-1} p}{2}) \\ &= g m_x g^{-1}(\frac{\mathbb{I}+g m_x g^{-1}}{2} p) = \frac{g m_x g^{-1}+\mathbb{I}}{2} p = \bar{p}. \end{aligned} \tag{9.7}$$

We employ the second RANSAC to remove outliers from the set of correspondences where there are no strong agreements on the midpoints. At the end of this stage, we identify the key points that only satisfy extrinsic symmetry, selected from a subset of the correspondences that satisfied intrinsic symmetry on the surface.

9.2.3.4 Extrinsic Symmetry from Point Correspondences

The parametrization g_0 of the partial symmetry plane in 3D is defined via a point $\mathbf{n}_0$ and a plane normal $\vec{\mathbf{n}}$. To obtain these parameters which describe extrinsic symmetry, we compute the mean-normalized matrix $\bar{P}$ of all N midpoints. Singular value decomposition of this matrix yields:

$$\bar{P}_{n\times 3} = U_{n\times n}\Sigma_{n\times 3}V^T_{3\times 3}. \tag{9.8}$$

In this formulation, U contains the left singular vectors of $\bar{P}_{n\times 3}$, Σ is a diagonal matrix with singular values of $\bar{P}_{n\times 3}$ as the diagonal elements, and V contains the right singular vectors of $\bar{P}_{n\times 3}$. The last column vector $\mathbf{v}_3$ of $V = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix}$ defines the least principle component of the data, *i.e.* axis with minimum variance. This vector defines the normal to a plane that has the closest distance to all midpoints. Lastly, the point $\mathbf{n}_0$ is computed as the mean of all midpoints: $E[\bar{\mathbf{p}}]$.

9.2.4 Robust Estimators for Detecting Imperfect Symmetry

After an initial parametrization g_0 of the plane is obtained, the intensities are compared iteratively between the voxel elements across the plane of partial symmetry, until the parameters $\mathbf{M}_g$ that minimize the intensity loss $d_I(\mathbf{o}, \mathbf{M}_g(\mathbf{o}))$ are estimated.

A major challenge associated with using an intensity-based approach is the presence of severe outlier regions which may result in incorrect symmetry parametrization given an intensity-based criteria. The symmetry outliers are the results of *i)* unilateral dislocations and fractures, and *ii)* imperfect symmetry in the original anatomy. Normalized Cross-Correlation (NCC) is a commonly used intensity-based measure which is greatly sensitive to noise and outliers [286]. Therefore, it is crucial to employ a symmetry detection mechanism that is robust to the outlier regions. We suggest using Tukey biweight robust estimator which automatically downweights or supresses

the regions that exhibit consistently high errors, and prohibits those elements from contributing to the total loss [287]. Inspired by Tukey robust regression we suggest the following loss:

$$d_I(\mathbf{o}, \mathbf{M}_g(\mathbf{o})) = \sum_{i=1}^{|\Omega_s|} \frac{|\rho(e_i(\mathbf{M}_g))|}{|\Omega_s|}, \tag{9.9}$$

where Ω_s is the spatial domain of the volumetric data. The element-wise error $\rho(e_i(\mathbf{M}_g))$ is computed as follows:

$$\rho(e_i(\mathbf{M}_g)) = \begin{cases} e_i(\mathbf{M}_g)\left[1 - \left(\frac{e_i(\mathbf{M}_g)}{c}\right)^2\right]^2 & ; |e_i(\mathbf{M}_g)| \leqslant c, \\ 0 & ; \text{otherwise.} \end{cases} \tag{9.10}$$

Parameter c sets a threshold that is used for classifying the voxel elements as outliers. The element-wise weighted residuals are computed as:

$$e_i(\mathbf{M}_g) = \frac{r_i(\mathbf{M}_g)}{S}\,, \text{such that } \; r_i(\mathbf{M}_g) = \mathcal{I}(o_i) - \mathcal{I}(\mathbf{M}_g(o_i)). \tag{9.11}$$

In Eq. 9.11, the term $\mathcal{I}(.)$ denotes voxel intensity.

It is suggested in the literature that $c = 4.685$ provides around 95% asymptotic efficiency of linear regression for normal distributions [287]. This value is computed assuming the residuals e_i are drawn from a unit variance distribution. To relax this constraint, the factor S is used as a scaling parameter and is computed based on

median absolute deviation:

$$S = \frac{MD}{0.6745}, \quad MD = \underset{j \in \Omega_s}{\text{median}}(r_j). \tag{9.12}$$

As demonstrated in Fig. 9.3, in contrast to L2 and L1 norms, Tukey biweight disparity term completely suppresses the residuals beyond a threshold, regarding the elements yielding excessive errors as outliers. This behaviour is desired for traumatic cases, since the outlier regions, *i.e.* dislocated bone, should be completely excluded during symmetry identification, allowing the plane parameter estimation to solely rely on the partial symmetry present in the anatomy.

9.2.5 Regularization based on Bone Distribution

To further support the identification of partial symmetry, we exploit a biological fact that, even if injured, dislocated bone fragments remain within the body. On this basis, a regularization term is introduced to maximizes the similarity between the distribution of bone densities across the symmetry plane. This concept is visualized in Fig. 9.4.

To materialize this notion, probability distribution functions are computed in the form of image histograms from voxel intensities. A regularizer based on Normalized Mutual Information (NMI) is used to acquire a similarity score between the density

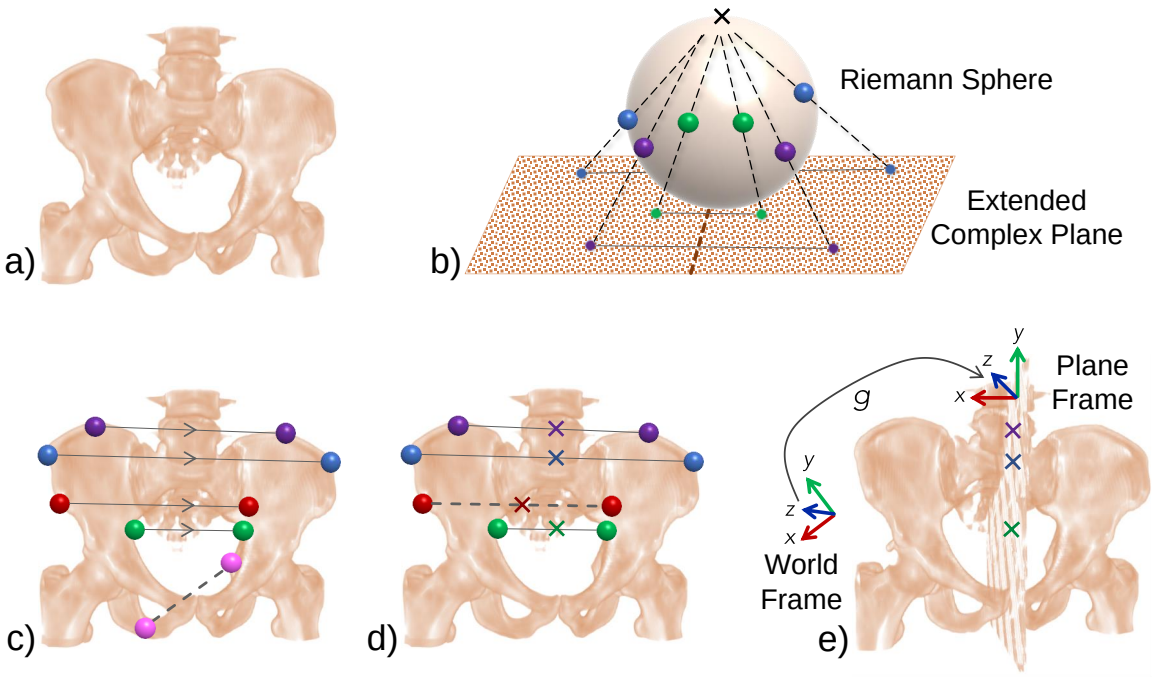

Figure 9.2: First step in computing the initial parametrization of extrinsic symmetry on an object that exhibits imperfect symmetry involves transforming the surface of the object to the Reimann sphere (uniformization) and consequently to the extended complex plane (stereographic projection). These transformations are shown in parts **(a-b)**. Anti-Möbius group is then used to model the intrinsic symmetry on the complex plane and assign self-symmetry correspondences to vertices on the surface of the pelvis. During a two-step RANSAC scheme, the correspondences are pruned and narrowed down to only inliers which satisfy the extrinsic symmetry properties **(c-d)**. As the result of the first and second RANSAC, correspondeces shown in dashed lines are removed due to dissimilar directions and midpoints, respectively. Extrinsic symmetry is subsequently computed using least squares minimization **(e)**.

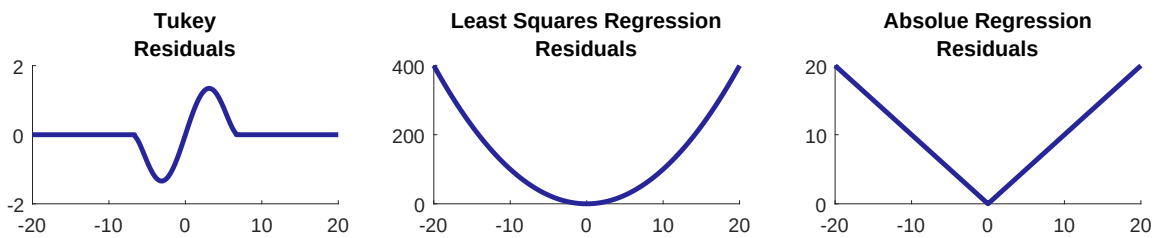

Figure 9.3: Comparison of the Tukey bi-weight with L2 and L1 norms. The horizontal axis represents the residual error, and the vertical axis shows the corresponding loss.

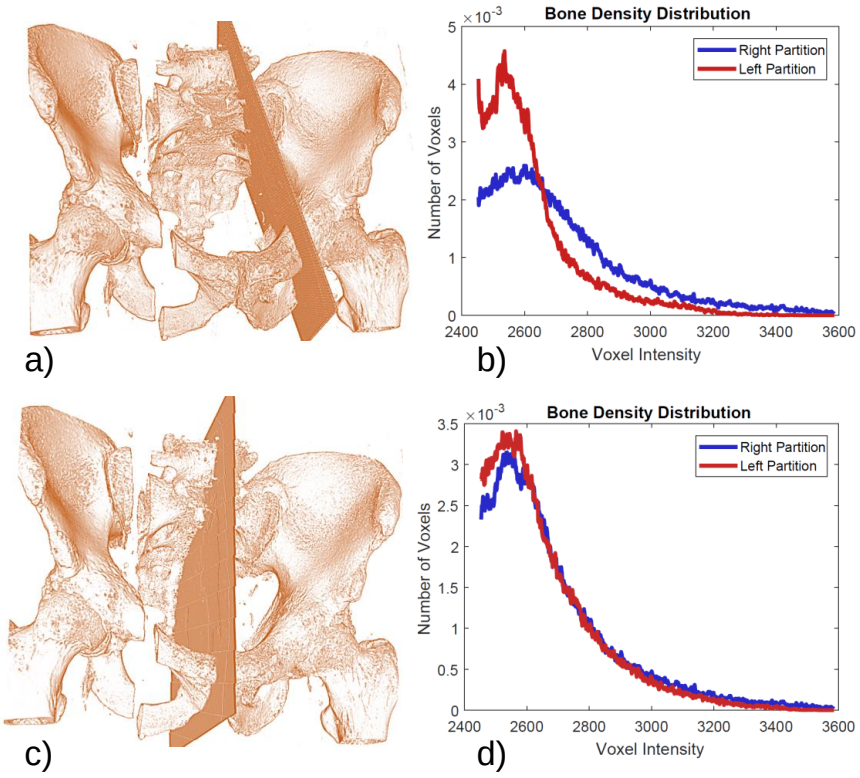

Figure 9.4: Distribution of the bone HU values across the symmetry plane before and after the estimation of symmetry plane. Comparing the histograms in (b) and (d) suggests high similarity when the plane dissect the volume bilaterally in the center.

distributions on the contralateral sides:

$$d_D(\mathbf{o}, \mathbf{M}_g(\mathbf{o})) = -\frac{H\Big(\mathcal{I}(\mathbf{o})\Big) + H\Big(\mathcal{I}(\mathbf{M}_g(\mathbf{o}))\Big)}{H\Big(\mathcal{I}(\mathbf{o}), \mathcal{I}(\mathbf{M}_g(\mathbf{o}))\Big)}. \tag{9.13}$$

In the formulation presented in Eq. 9.13, $H(.)$ is the entropy of voxels' intensities.

The regularizer term $d_D(.)$ is globally non-convex and yields local minimas for various plane parametrizations. An example case is shown in Fig. 9.5, where the NMI-based score is nearly equal for two cases, while the robust estimator cost measured based on Tukey biweight disparity is substantially (99.1%) lower at the configuration

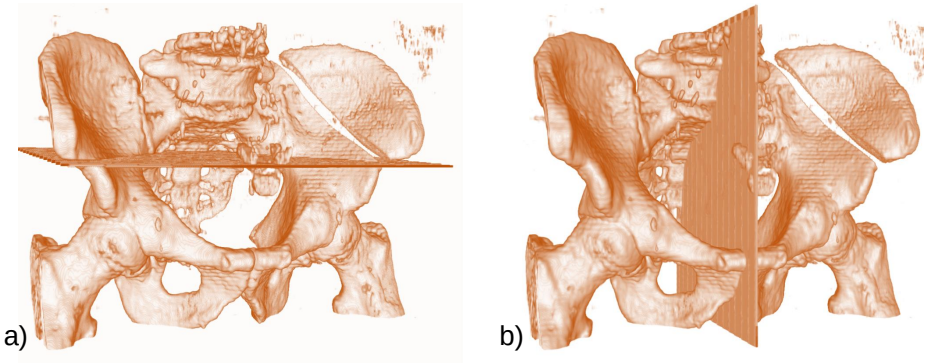

Figure 9.5: Two different plane estimates with near identical distribution score, and substantially different Tukey-based score.

in Fig. 9.5-b. Therefore, the density-based cost cannot replace the Tukey-based term, and is merely used as a regularizer to ensure similar bone distributions contralaterally.

9.2.6 Interventional Image Registration and Augmentation

The plane of partial symmetry with the parametrization $\mathbf{M}_g$ dissects the volume bilaterally, yielding sub-volumes $\mathbf{o}_i$ and $\mathbf{o}_h$, denoting the injured and healthy regions, respectively ($\mathbf{o} = [\mathbf{o}_i, \mathbf{o}_h]$). Using this parametrization, the non-fractured portion of the data can be mirrored across the plane as $\bar{\mathbf{o}}_h = \mathbf{M}_g(\mathbf{o}_h)$, resulting in a non-fractured model of the patient: $\bar{\mathbf{o}} = [\bar{\mathbf{o}}_h, \mathbf{o}_h]$. This patient-specific reconstructed model is then used as a template of patient anatomy, representing the anatomical structures *"as if they were repaired"*. It is important to stress that, although the human pelvic skeleton is not entirely symmetric, it is common for an orthopedic trauma surgeon to consider

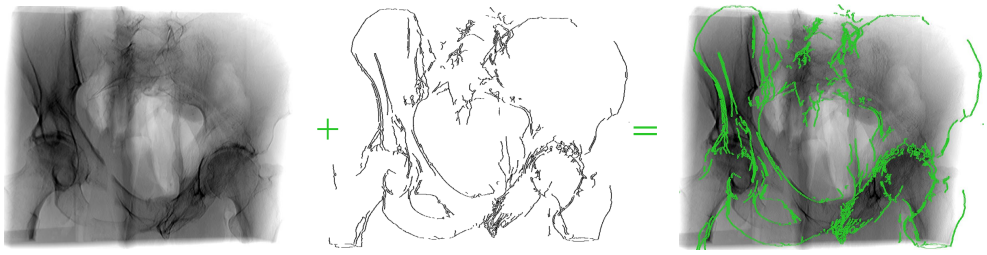

Figure 9.6: Interventional X-ray images are augmented with the contours of the bone extracted from the mirrored CT volume. The green contours serve as road-map, demonstrating desired configurations for bone fragments in the perspective of each X-ray image such that bilateral symmetry is restored.

it symmetric, and use the contralateral side as reference.

To exploit the patient-specific template $\bar{\mathbf{o}}$ intra-operatively, each interventional C-arm fluoroscopy image is augmented with the contours of the reconstructed bone in the mirrored CT volume as demonstrated in Fig. 9.6. Enabling such augmentation requires two steps. In the first step, the transformation that describes the projective relation between the pre-operative CT image and the intra-operative X-ray image is computed via 2D/3D image registration by maximizing NCC score defined below:

$$\underset{R,\mathbf{t},k}{\arg\max}\ NCC\left(R,\mathbf{t},k|\mathcal{I}_X,\mathcal{I}_D\right) = \sum^{|\Omega_{X,D}|} \frac{\mathcal{I}_X \cdot \mathcal{I}_D(R,\mathbf{t},k\big|\mathbf{o})}{\sigma_X \sigma_D}. \tag{9.14}$$

This formulation optimizes over the parameters of the rotation R, translation $\mathbf{t}$, and intrinsic geometry k. The parameters $\mathcal{I}_X$ and $\mathcal{I}_D$ are the mean-normalized X-ray and Digitally Reconstructed Radiographs (DRRs) generated from the fractured patient data given the parameters $(R,\mathbf{t},k)$. Finally, $\Omega_{X,D}$ is the common spatial domain of the two images, and σ_X and σ_D are the standard deviations of the X-ray and

DRR images within $\Omega_{X,D}$. In the second step, augmented image $\mathcal{I}_A$ is constructed by overlaying the X-ray image and the 2D DRR that is computed from the mirrored volume, *i.e.* $\mathcal{I}_A := \mathcal{I}_X \cup \mathcal{I}_D(R, \mathbf{t}, k, \bar{\mathbf{o}})$.

9.3 Experimental Results

In this section we evaluate the proposed methodology for detecting and exploiting symmetry under a variety of different conditions. We present both quantitative and qualitative outcome on data with synthetic fractures, as well as patient data with complex unilateral fractures. To generate ground-truth for the evaluations in sections 9.3.1-9.3.4, we synthetically construct symmetric pelvis data using patient cases from the NIH Cancer Imaging Archive. In Sec. 9.3.5, the distance between anatomical landmarks on pelvic cases with simulated unilateral fractures were compared before and after applying the symmetry transform. Finally in Sec. 9.3.6, the symmetry detection and surgical image augmentation is demonstrated on three patient data with unilateral pelvic fractures.

Bound constrained by quadratic approximation method was used for the optimization of the non-linear cost in Eq. 9.3. For all experiments presented in this section, the maximum number of iterations was set to 100. The rotation errors are measured as the angle between the normal $\hat{\mathbf{r}}$ of the estimated symmetry plane and the ground-truth normal $\mathbf{r}_1$. Since the plane normal is in the direction of the X-axis of the local

frame, hence the normal vectors are defined via the first column vector of the rotation matrix associated to the plane pose. The rotation error is formulated as:

$$\theta_e = \cos^{-1}(\hat{\mathbf{r}}_1^\top \mathbf{r}_1). \tag{9.15}$$

Since Euclidean distance is not explicitly defined for non-parallel planes, we define a translation measure as the projection of the translation difference $(\mathbf{t} - \hat{\mathbf{t}})$ onto the plane normal:

$$\mathbf{t}_e = \left|(\mathbf{t} - \hat{\mathbf{t}})^\top \mathbf{r}_1\right|. \tag{9.16}$$

9.3.1 On the Effect of Regularization

We evaluated the performance of the suggested symmetry detection cost in Eq. 9.3 at different regularization levels, where λ was varied between 0.0 and 1.0. In Fig. 9.7 results are presented for data without fractures as well as data with different fracture patterns. Each case is evaluated 4 times, each with a random initialization in a neighborhood around the ground-truth within the ranges of $\pm 15\,\text{mm}$ and $\pm 15^\circ$ for translation and rotation along each axis, respectively. As shown in the results, all regularizers improve the convergence compared to the case where no regularization is used, however no significant differences is observed between different regularization parameters.

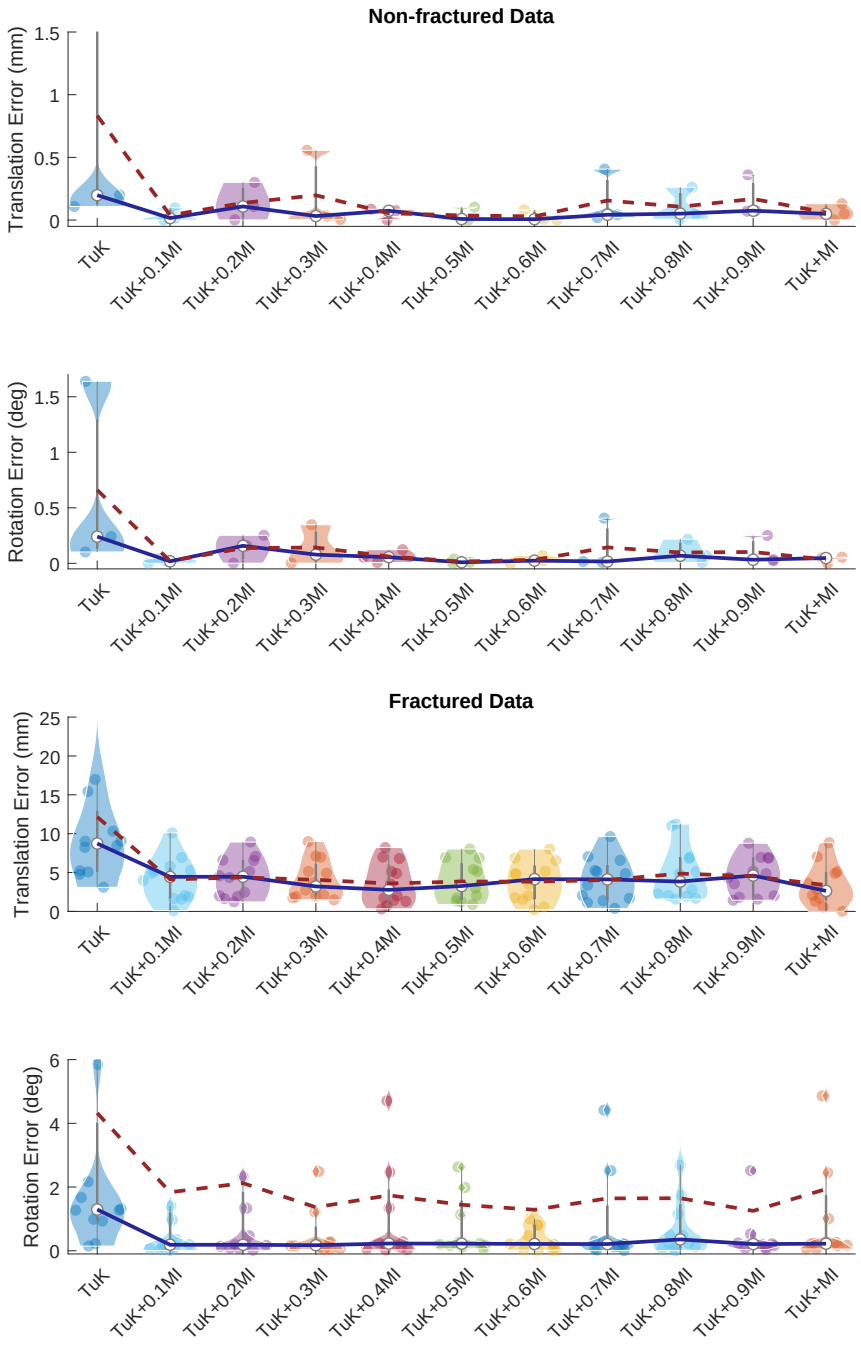

Figure 9.7: Translation and rotation errors given different regularization factor $\lambda \in [0, 1]$

9.3.2 Sensitivity to Imperfect Symmetry

The performance of the symmetry cost $D(.)$ presented in Eq. 9.3 was tested against different levels of noise and bone dislocations. In all experiments, the regularization factor λ was set to 0.5, allowing the $d_I(.)$ to be the dominant term driving the total cost, and $d_d(.)$ serving as a fidelity term. The amount of dislocation (outlier) was varied between 0% to 30% of the entire volume, and the Gaussian noise between 0% and 40% of the highest intensity in the volume. For each given outlier and noise level, the symmetry detection was repeated 20 times, each time randomly sampling an initialization parameter within the maximum range of $\pm 15\,\text{mm}$ and $\pm 15^\circ$ around the ground-truth. Results are presented in Fig. 9.8. The left column are heatmaps from the average error in translation and rotation, and the right column illustrates the corresponding heatmap for their standard deviations. Results indicate smaller errors in the top left quadrant in each sub-plot, which correspond to dislocations between 0% and 15% of the entire volume, and the Gaussian noise between 0% and 20%.

9.3.3 Capture Range

We characterize the dependence of the regularized Tukey cost on the initialization parameters. In Fig. 9.9, the mean rotation and translation errors are presented for varying initialization parameters. The elements on the horizontal axis represent the level of misalignment at the initial configuration, where at each step the ranges for the

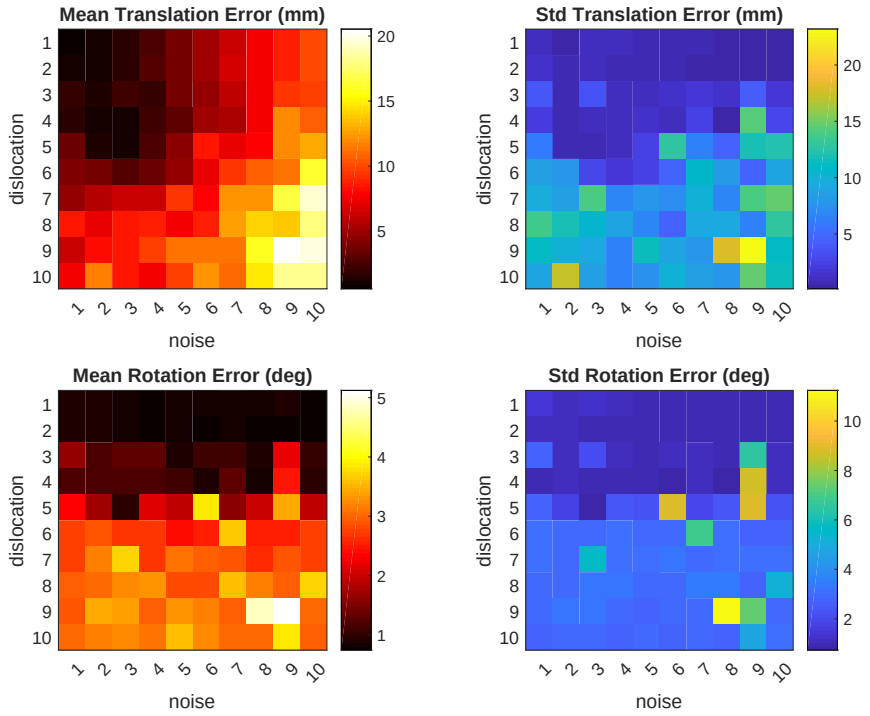

Figure 9.8: Error in detection of partial symmetry plane evaluated against different amounts of noise and bone dislocation

initialization misalignment are increased by 5 mm and 5° along each axis of translation and rotation, respectively. For instance, the first and second elements on the horizontal axis in Fig. 9.9 which have yielded lower errors show initialization samples for each axis between the ranges of $[(0\,\text{mm}, 5\,\text{mm}) - (0°, 5°)]$ and $[(5\,\text{mm}, 10\,\text{mm}) - (5°, 10°)]$, respectively. Given each range, the sampling is repeated 10 times.

9.3.4 Accuracy of Global Initialization

The initialization strategy that leverages the combined properties of intrinsic and extrinsic symmetry (Sec. 9.2.3) was investigated to assess whether global initialization

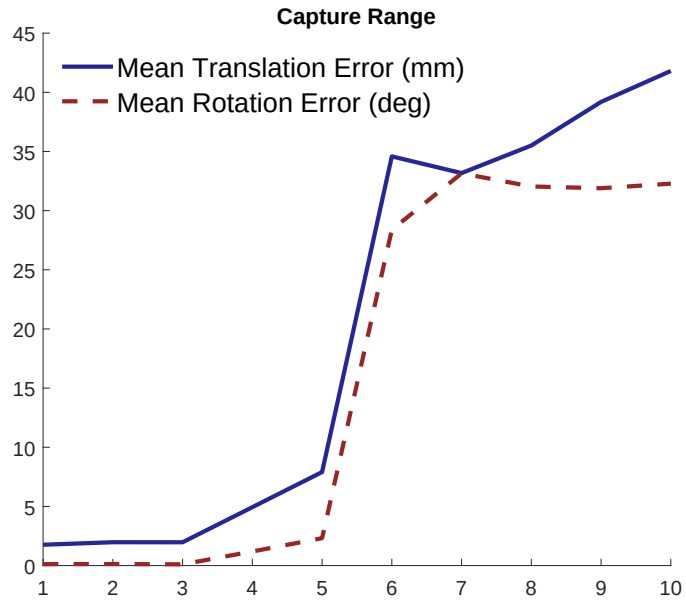

Figure 9.9: Dependence of the regularized Tukey cost on initialization. The horizontal axis corresponds to the extent of misalignment at the intialization step, and the vertical axis represents the translation and rotational errors after convergence.

parameters yield outcome within the capture range. To this end, fractures on the iliac wing bone were applied to three pelvic data. The errors in plane detection, as well as the details on the two RANSAC steps are presented in Table 9.1.

Table 9.1: Errors in detecting the bilateral symmetry plane are estimated given the initialization parametrization. The last three columns represent the total number of landmarks, inliers with agreement on the direction of the vectors connecting the correspondence, and inliers with consensus on the mid-points.

Volume	**Rotation Error**	**Translation Error**	**# Landmarks**		
			initial	RANSAC 1 inliers	RANSAC 2 inliers
1	5.04	3.44	101	31	24
2	8.17	7.57	99	15	6
3	10.23	1.27	92	22	14

Table 9.2: Distances between the four anatomical landmarks on the surface of the pelvis with their counterparts on the mirrored template are presented as mean ± SD.

Fracture	**L1**	**L2**	**L3**	**L4**
Iliac wing	3.49 + 2.41	2.83 + 2.55	3.44 + 1.52	2.58 + 1.28
Pelvic ring	2.24 + 1.13	3.96 + 2.62	1.78 + 0.79	1.84 + 0.86
Vertical shear	4.96 + 1.94	3.35 + 2.18	4.07 + 2.17	4.52 + 2.81

9.3.5 Estimation of Partial Symmetry on Data with Synthetic Fractures

Three prevalent unilateral pelvic fractures, namely, iliac wing fracture, pelvic ring fracture, and vertical shear fracture were applied to four patient data. Given each case, the plane of partial symmetry were detected, and symmetric patient templates were constructed. In Table 9.2 we report the distance from the anatomical landmark on the original volume before applying the fractures to their reconstructed correspondence on the mirrored model. For the measurements, we considered four separate landmarks that were distributed on the surface of the bone, including, **L1**: anterior superior iliac spine, **L2**: posterior superior iliac spine, **L3**: ischial spine, and **L4**: ischial ramus. Fig. 9.10 shows three simulated radiographs from each fracture model that are augmented with the edge-map extracted from gradient-weighted DRRs of the mirrored template. These contours represent the bone at desired configuration *"if the bilateral symmetry was completely restored"*. This figure also highlights the abnormal area on the bone that was automatically detected using the Tukey-based cost as the symmetry violator region.

9.3.6 Estimation of Partial Symmetry on Patient Data with Trauma Injuries

Qualitative outcomes are visualized in Fig. 9.11 for three patient cases with severe unilateral traumatic injuries. For each case, two fluoroscopic images were separately registered to the patient CT. As the result of 2D/3D registration, the relative projective transformations describing the spatial relation between the X-ray and CT were computed. Next, for all three models, symmetry was detected and patient-specific templates were reconstructed by mirroring the healthy side of the bone across the extrinsic symmetry plane. Finally, DRRs were generated from the patient templates using the projective transformations associate with each X-ray image, and were augmented onto their corresponding fluorscopic images.

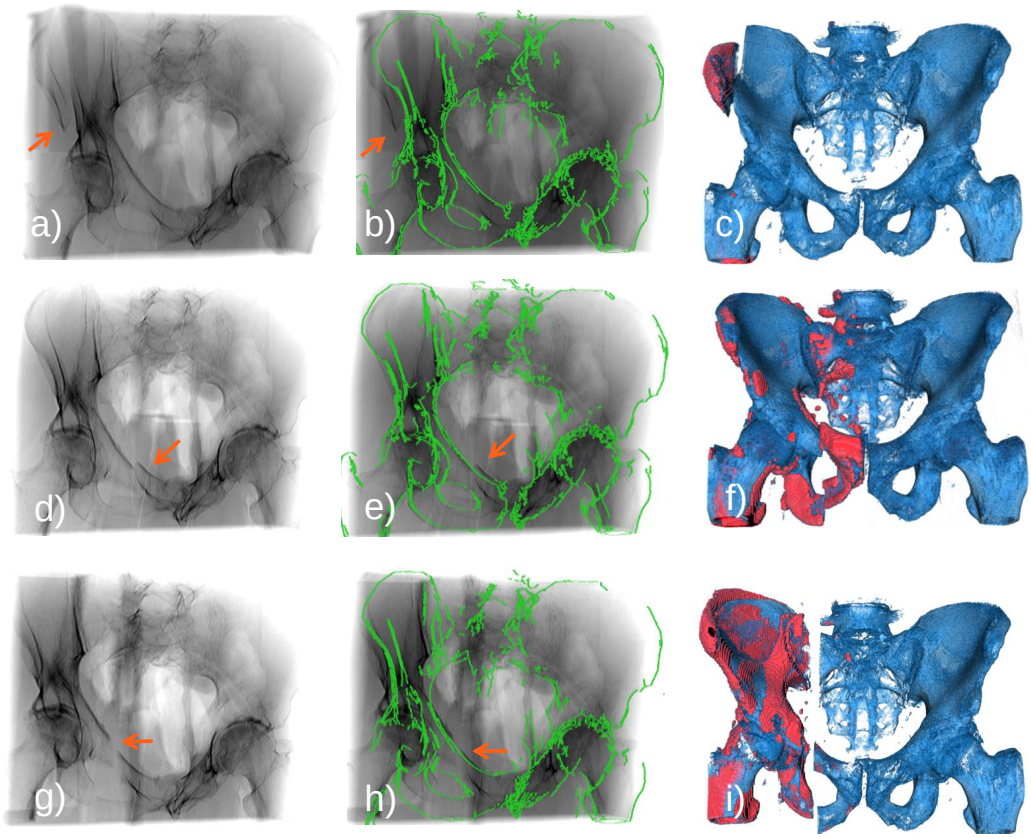

Figure 9.10: Ilic wing fracture **(a-c)**, pelvic ring fracture **(d-f)**, and vertical shear fracture **(g-i)** are shown on a pelvis data. The orange arrows in the first two columns represent the area with the fracture. The green contours that are computed from the symmetrically reconstructed model suggest road-maps in each image perspective that can result in fracture reduction and symmetry restoration. The region colored in red in the last column represents the area that was considered as symmetry violator (outlier) by the Tukey-based term $d_I(.)$.

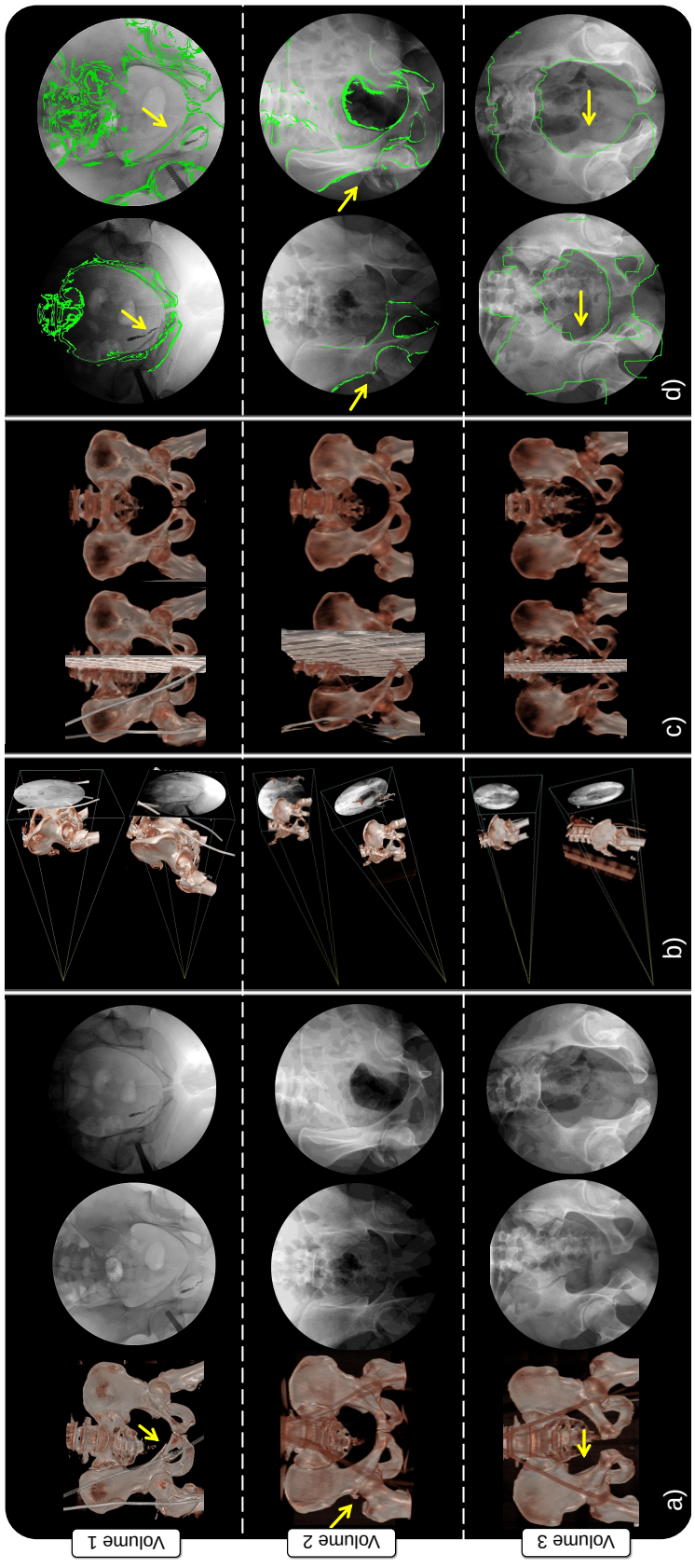

Figure 9.11: Partial symmetry was exploited to augmented X-ray images with desired configurations of bone fragments. Three cases with severe trauma injuries were investigated here. The first case had displaced pelvic ring fracture, the second case had displaced fracture of the right acetabular roof and the posterior wall, and the third case exhibited asymmetric widening of the left sacroiliac joint. The yellow arrows indicate the fracture location in the figure. We present results given the pre-operative CT as well as two X-ray images **(a)** for each patient. The relation of each X-ray image was estimated with respect to patient CT **(b)**. As shown in **(c)**, the plane of symmetry was detected and symmetric patient-specific template was generated for each case. Finally, DRRs were generated from the symmetric template and were augmented onto patient X-ray as bone outlines.

9.4 Discussion and Conclusion

In this work we present an end-to-end methodology for automatic identification of global symmetry in pelvis data with severe unilateral fractures, and exploit the knowledge from symmetry to provide interventional image augmentation. Three measures are combined to identify partial symmetry: *i)* the structural geometry is used in the Möbius space to determine intrinsic and extrinsic symmetry on the surface, *ii)* voxels are used with Tukey robust estimator to score the similarities between the intensities, and *iii)* normalized mutual information is used to match the distribution of bone across the sagittal plane of the patient. Regularization is important when the amount of bone dislocation is large, and Tukey's cost cannot solely drive the symmetry plane to the optimal pose. Each of these three novel steps are designed with the consideration of being insensitive to outlier regions that are caused by the injury.

The clinical relevance of this solution is manifested by considering common practices in surgical routine where orthopedic traumatologists aim at bringing displaced bone fragments into alignment with their natural biological configurations. This is achieved by replicating the contralateral side, hence restoring symmetry in the internal structures. It should be noted that this solution is merely admissible for unilateral fractures, that according to pelvis fracture classification [288], involves a considerable number of cases. Consequently, direct comparison of bony structures across the sagittal plane becomes possible for such cases.

A preeminent criteria in determining symmetry is the Tukey-based robust esti-

mation which automatically suppresses voxel elements that consistently produce high errors. To improve the estimation of the symmetry plane, a novel regularization term based on bone density distribution is added to the overall loss function. In Sec. 9.3.1 we evaluated the accuracy of symmetry estimation with respect to different regulation factors. Results in Fig. 9.3 indicate substantial improvement when using regularization. However, the results do not vary significantly when different λ factors are used. For consistency, we used $\lambda = 0.5$ for all other experiments in Sec. 9.3.

From the results in Fig. 9.9 we conclude that the first three initialization ranges yielded average translation error of $< 2\,\text{mm}$ and rotation error of $< 0.2°$. These are associated with initialization parameters within the ranges of $(0\,\text{mm}, \pm 15\,\text{mm})$ translation and $(0°, \pm 15°)$ rotation near the ground-truth. A comparison between these results and the errors of the proposed automatic initialization approach presented in Table 9.1, proves that our suggested initialization yields outcome within the capture range of the cost in Eq. 9.3. As also appears in Table 9.1, the correspondence-based initialization demonstrates higher performance in predicting the translation parameters compared to rotation.

We simulated severe and unstable unilateral dislocations, and reconstructed a fully symmetric patient template. Comparing the relevant anatomical landmarks on the original and reconstructed pelvis yielded a mean discrepancy of 3.26 mm between different bony features. Finally, we also presented view-specific road-maps to guide towards an optimal repair of pelvic fractures on patient data with trauma injuries.

An essential characteristic of our solution is the automatic outlier detection that is highlighted in the last column of Fig. 9.10. Our approach towards outlier identification can enable several other applications in different disciplines of radiology and surgery where regions that violate symmetry can be classified, and consequently be used to improve pre-operative planning as well as provide real-time feedback to surgeons on whether biological symmetry is properly restored. Learning-based solutions can also substantially benefit from such outlier detection mechanisms by automatically shifting the focus of the artificial agent to relevant regions with structural anomaly.

In conclusion, we presented a solution that exploits partial symmetry in human anatomy and provides intuitive image augmentation for fracture care procedures. It should be noted that, this solution enables patient-specific data augmentation and guidance, that is unattainable by using statistical shape models [289]. Constructing atlases for such procedures require a large population of patient pelvis data for different age, sex, race, disease, fractures, etc. which are not available. We hope that our theoretical findings and methodology can lead to safer and more reliable surgical care.

9.5 Acknowledgments

I want to thank Mr. Giacomo Taylor for his help during the design of the symmetry detection pipeline. Giacomo also contributed significantly by running ablation

experiments and creating demonstration videos. I also appreciate the support of Drs. Nassir Navab, Mehran Armand, Mathias Unberath, Russell Taylor, Greg Osgood, and Alex Johnson at every step of the way during the design and validation of this work.

Part IV

Augmented Robotics: Transcending human limitations in medical interventions

Robot-assisted minimally invasive surgery has shown to improve patient outcomes, as well as reduce complications and recovery time for several clinical applications. While increasingly configurable robotic arms can maximize reach and avoid collisions in cluttered environments, positioning them appropriately during surgery is complicated because safety regulations prevent automatic driving. This final chapter of the dissertation proposes a head-mounted display (HMD) based augmented reality (AR) system designed to guide optimal surgical arm set up. The staff equipped with HMD aligns the robot with its planned virtual counterpart. In this user-centric setting, the main challenge is the perspective ambiguities hindering such collaborative robotic solution. To overcome this challenge, a novel registration concept is introduced for intuitive alignment of AR content to its physical counterpart by providing a multi-view AR experience via reflective-AR displays that simultaneously show the augmentations from multiple viewpoints. Using this system, users can visualize different perspectives while actively adjusting the pose to determine the registration transformation that most closely superimposes the virtual onto the real. The work in this chapter is based on the below manuscript:

1. Fotouhi, Javad, Tianyu Song, Arian Mehrfard, Giacomo Taylor, Qiaochu Wang, Fengfan Xian, Alejandro Martin-Gomez et al. "Reflective-AR display: An interaction methodology for Virtual-to-Real alignment in medical robotics." IEEE Robotics and Automation Letters 5, no. 2 (2020): 2722-2729.

Chapter 10

Reflective-AR Display: An Interaction Methodology for Virtual-to-Real Alignment in Medical Robotics

10.1 Introduction

Robotic-assisted minimally invasive surgery is becoming increasingly common due to its associated benefits that include higher accuracy, and tremor and fatigue reduction. Robotic systems can augment the surgeon's abilities with stereo endoscopic imaging and intuitive control which help the surgeon's hand-eye coordination and

Figure 10.1: AR-assisted robot arm positioning

reduce physical workload during surgery [290]. Furthermore, robotic surgery has benefits over traditional laparoscopic techniques with patients experiencing reduced blood loss and shorter post-operative hospital stays [291].

Quick and accurate set up of robotic systems leading up to surgery remains a major challenge in the endeavor of making robotic surgery the standard of care. After a patient has been positioned, anesthetized, and trocars inserted, the robotic arms must be positioned and docked before operation can begin. This procedure is a crucial step of workflow and grows more complex the more joints the robotic arms have. While many different configurations of the robot's joints may allow the

robotic arm to dock, sub-optimal positioning increases the likelihood of collisions and inadequate reach during teleoperation. Any repositioning or undocking necessary to adjust the robot significantly decreases operating room efficiency [292]. For many procedures including minimally invasive gastrectomy, "junk time", the time taken to set up or reposition robotic arms, is often the sole reason for increased procedure time in robotic procedures over purely laparoscopic approach [293].

Optimal set up of robotic arms, consequently, is critical to increasing the efficiency of robotic surgery and foster acceptance. Due to safety and regulatory concerns, having a robot automatically drive itself to a pre-operative position is infeasible. In modern surgical robotic systems, the set up of the arms can be supported by lasers, as shown in Fig. 10.2. Though lasers assist the staff in aligning the robot, it can still exhibit challenges in a complex system with joint redundancies, as it does not directly show the desired configuration of all joints. As most manual methods are error prone and induce a steep learning curve to operating room staff unfamiliar with a system, we investigate an augmented reality (AR) solution for guidance during robotic set up. Using an optical see-through head-mounted display (OST-HMD), setup staff can be interactively guided through joint-by-joint steps to optimally position the robot in an efficient manner.

The works by Qian et al. are similar to our solution in the spirit of using AR for robotic surgery [294–296]; however, the focus of their works were on optimal instrument insertion and manipulation by showing the extension of the arms inside

the abdomen using AR. Our methodology addresses the alignment of robot arms for optimal reach and minimum collision. Other early applications of AR in robot-assisted surgery focused on multi-modal registration of medical imaging data with the endoscopic view [297, 298].

Several studies have discussed the challenges of aligning virtual and real objects, and have emphasized the importance of this step for various room-scale and spatially-aware AR solutions [299–302]. Nuernberger et al. suggested a semi-automatic alignment strategy to register virtual and real spaces [303]. Their work relied on scene content and environment constraints such as edges and surfaces for snapping the virtual content to real. In a different study, various rendering and visualization techniques were compared for alignment of different virtual models in fully immersive environments [304]. Results indicated that static visualization techniques which exhibited lower occlusion in a single view yielded better alignment.

In order to properly augment virtual assistance on a physical robot, we require an intuitive and fast approach to align the AR environment provided by a OST-HMD to the robot. This registration must be robust to perceptual ambiguities that arise during AR alignment [305, 306]. To this end, we propose virtual-real active alignment (ViRAAl) to register a virtual model of the robot to its real counterpart. Our method enables the user to create and view multiple AR mirrors which show the current 3D scene (including real and virtual robot) from different viewpoints. By providing this overlay from multiple perspectives simultaneously, users can actively adjust the 6

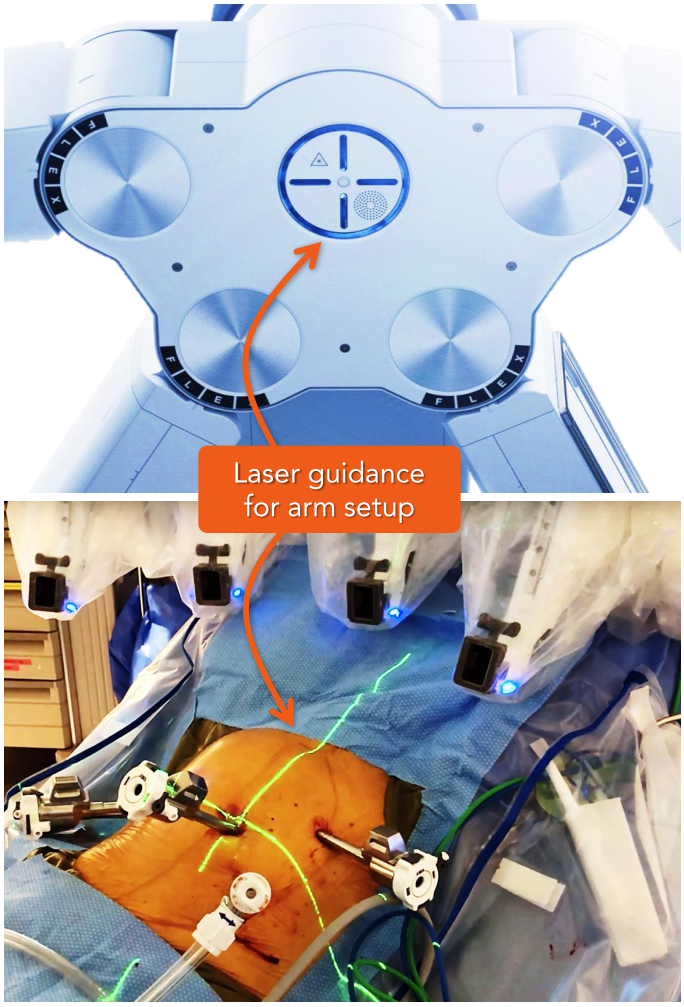

Figure 10.2: Da Vinci Xi surgical robot uses multiple lasers to assist the positioning and docking of the robotic arms.

degree-of-freedom (DOF) transformation parameters that best align the virtual and real objects in all views.

We summarize the contributions of this chapter as *1)* reflective-AR displays as the multi-view and marker-free paradigm for co-registration between the virtual and real spaces, hence enabling spatially-aware AR, and *2)* using AR for assistance during robot set up (Fig. 10.1).

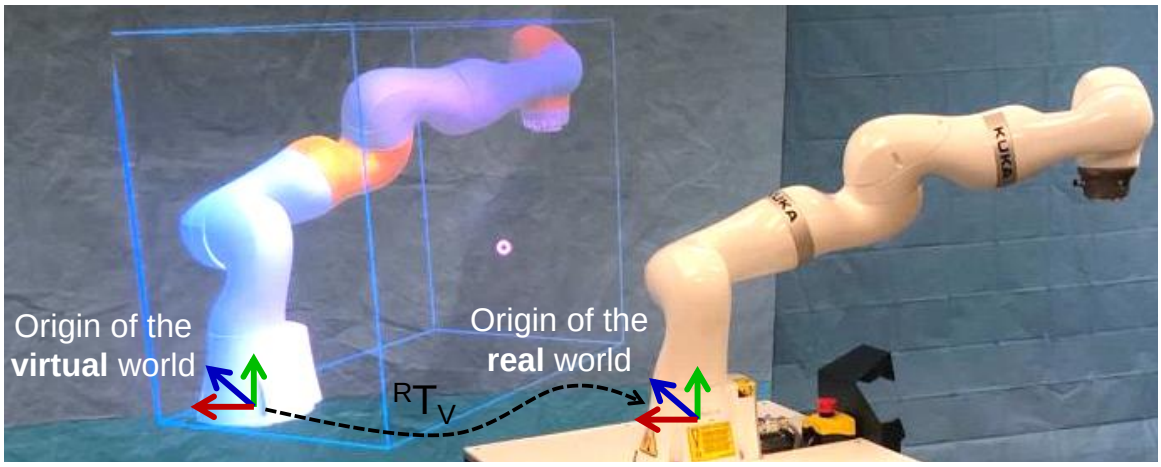

Figure 10.3: ViRAAl strategy estimates the virtuality to reality transformation ${}^{\mathrm{R}}\mathbf{T}_{\mathrm{V}}$.

10.2 Methodology

An important step in many AR scenarios is to bring the virtual content that lives in a controlled environment into alignment with the physical reality that is present in the unmodeled environment [305]. In this work, to enable seamless interaction of a surgical robot manipulator and its virtual representation during an AR experience, we introduce reflective-AR displays that enable multi-view visualization and interactive alignment of virtual and real objects. In Sec. 10.2.1, we present the problem formulation for registering virtual-to-real. A key contribution of this work, which is the AR reflectors, is presented in Sec. 10.2.2. Finally, in Sec. 10.2.3, we suggest AR guidance to facilitate robot set up during surgical interventions. It is important to note that in Sec. 10.2.1 and 10.2.2 we discuss the problem of *"virtual-to-real"* alignment to enable spatially-aware AR, and in Sec. 10.2.3 we discuss *"real-to-virtual"*

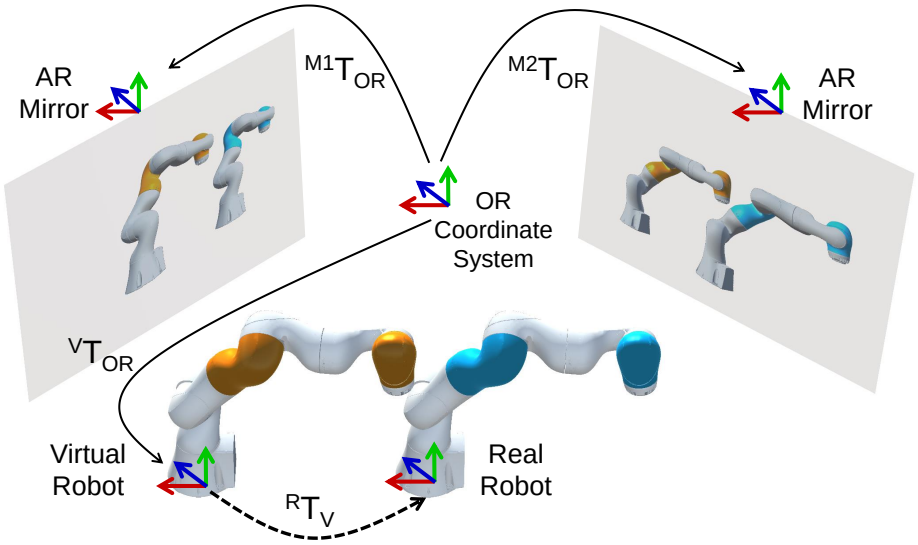

Figure 10.4: Reflective-AR displays enable simultaneous alignment from multiple views.

alignment to provide spatially-aware AR guidance.

10.2.1 Virtual-Real Active Alignment (ViRAAl)

To estimate the virtual-to-real 6 DOF alignment shown in Fig. 10.3, we estimate the transformation ${}^{\mathrm{R}}\mathbf{T}_{\mathrm{V}} = (\bar{R}, \bar{t})$ via interactively registering a robot with its virtual model at N pre-defined joint configurations. Each time a rigid-body transformation $\{(R_i, t_i)\}_{i=1}^{N}$ is obtained, where $(R_i, t_i) \in SE(3)$, and SE is the Special Euclidean group.

We hypothesize that the average transformation computed from these N estimates, can yield a closer approximation of the true virtual-to-real alignment compared to each individual N transformations. Hence, from these N estimates, we

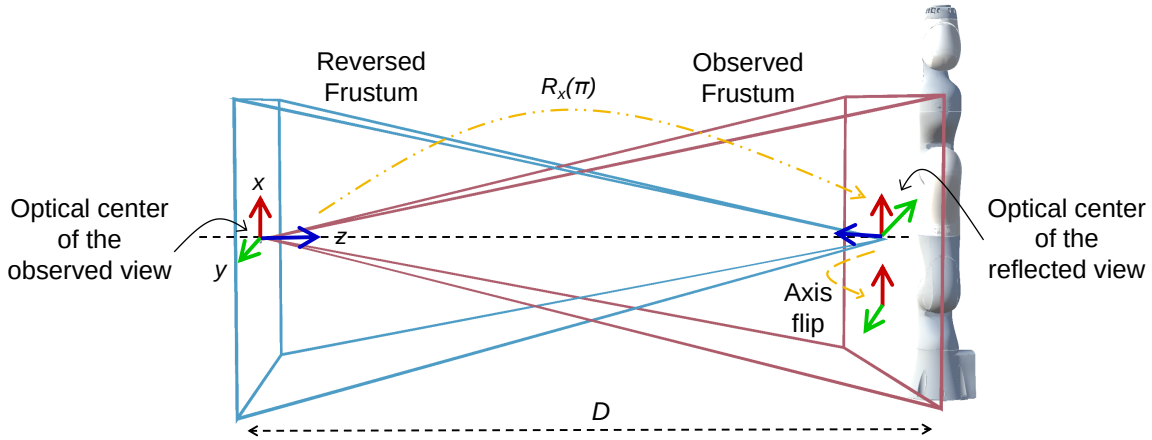

Figure 10.5: Imaging geometries of the observed and reversed frustums in relation to the robotic manipulator

seek to compute the mean rotation and translation. The mean rotation matrix $\bar{R}$ is computed on the Special Orthogonal group $SO(3)$ by minimizing:

$$\underset{\bar{R}\in SO(3)}{\arg\min} \sum_{i=1}^{N} d(R_i, \bar{R})^2, \tag{10.1}$$

where $d(.)$ denotes a distance function on the Riemannian manifold. To establish $d(.)$, the rotation matrix is expressed in the Lie algebra (tangent space) of the Lie group as $R = e^{\hat{\mathbf{w}}}$. The tangent space $\mathbf{w}$ is then obtained as $\log(R) = \hat{\mathbf{w}}$, such that $\hat{\mathbf{w}}$ is the skew-symmetric matrix constructed from the vector $\mathbf{w}$. Consequently, the mean rotation is estimated as [307]:

$$\underset{\bar{R}\in SO(3)}{\arg\min} \sum_{i=1}^{N} \left\| \log(R_i^\top \bar{R}) \right\|_F^2, \tag{10.2}$$

where $\|.\|_F^2$ is the Frobenius norm. The mean translation $\bar{t}$ is computed in Euclidean space as:

$$\bar{\mathbf{t}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{t}_i. \tag{10.3}$$

10.2.2 Reflective-AR Display

Due to the projective property of human visual system and the differences in perceptual cues in virtuality and reality, the scale and depth between real and virtual objects are easily misjudged [306]. To overcome depth ambiguities and enhance 3D perception during an AR experience, we introduce reflective-AR displays that allow simultaneous visualization of the scene from various viewpoints. The reflective-AR displays shown in Fig. 10.4 are constructed by displaying images from the integrated camera sensor of the OST-HMD as if the user observed the real scene from different viewpoints simultaneously, and are augmented with the projections of the 3D virtual objects. To compute a geometrically relevant pose for displaying these images, we compute the associated observer poses to the coordinate frame of the AR scene in the operating room ($^{\mathrm{O}}R_{\mathrm{OR}}$,$^{\mathrm{O}}\mathbf{t}_{\mathrm{OR}}$) via simultaneous-localization and mapping (SLAM).

The observer imaging geometry in Fig. 10.5 is formulated as:

$$P_o = K_o\, P \begin{bmatrix} {}^{\mathrm{O}}R_{\mathrm{OR}} & {}^{\mathrm{O}}\mathbf{t}_{\mathrm{OR}} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \tag{10.4}$$

where K_o is the matrix of intrinsic parameters and P is the projection operator. Next,

to simulate a mirror-like view, we construct a reversed frustum as (Fig. 10.5):

$$P_m = K_m\, P \begin{bmatrix} {}^{\mathrm{O}}R_{\mathrm{OR}} & {}^{\mathrm{O}}\mathbf{t}_{\mathrm{OR}} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} R_x(\pi) & \begin{bmatrix} 0 \\ 0 \\ D \end{bmatrix} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \tag{10.5}$$

$$K_m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} K_o.$$

In Eq. 10.5, the optical center of the observer frustum is rotated by the amount π around the x axis, and translated by the amount D along the principle ray of the frustum. Distance D is approximated as the Euclidean distance between the camera center, and an arbitrary point on the surface of the robot that is acquired by colliding the gaze cursor with the spatial map of the AR scene. The distance is merely used as a reference to position the optical center of the reversed frustum, and does not affect the rendering content in the reflective-AR display. To compute the reversed frustum's intrinsic matrix K_m, the y-axis of the image plane is flipped according to Eq. 10.5. Lastly, to give rise to a mirror-like AR display, the 3D virtual structures are projected into the image plane with the imaging geometry P_m, thus, enabling joint visualization of real and virtual in the reflective display.

Since the reflective AR displays are constructed based on the imaging geometry

of the observer frustum, we adopted the same axis convention used in the computer vision community. We set the z-axis in the direction pointing away from the camera, along the principal ray, connecting the origin to the principal point on the image plane.

10.2.3 Augmented Reality Assistance for Robot Set Up

After the registration transformation is established between real and virtual worlds, a collision-free and safe virtual robot configuration can be presented to the medical assistant. The desired configuration can either be estimated via the inverse kinematics of the robot, or can be adjusted interactively using the virtual robot and the patient position on the surgical bed. The robot set up is then performed in joint-by-joint steps, following the virtual planning.

10.3 Experimental results

We first evaluate the ViRAAl strategy using a virtual and real robotic manipulator. Next, in a simulated surgical setup, we assess the errors in moving the robot joints to achieve a desired joint configuration using AR guidance where a trocar must be inserted at a mannequin's umbilicus (Fig. 10.6). The umbilicus is commonly cho-

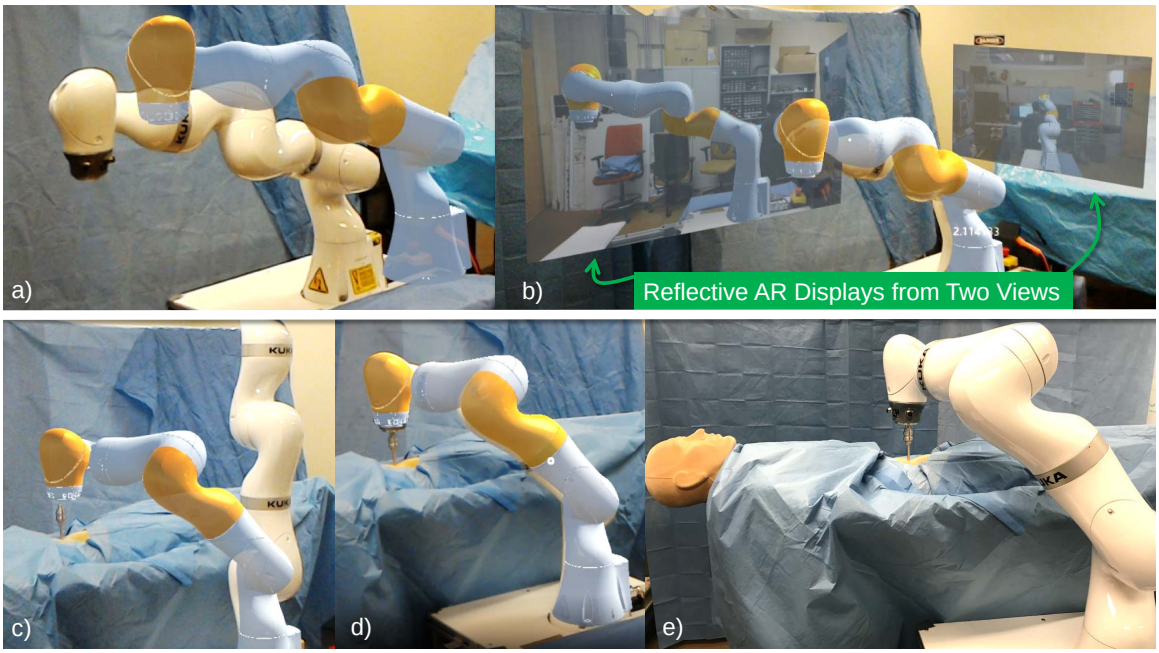

Figure 10.6: During the surgical AR experience, the virtual model of the robot is first visualized at a known configuration (**a**). The alignment between the real and virtual is established in multiple views via reflective-AR displays (**b**). Once the 6 DOF rigid-body transformation is identified between the real and virtual content, a virtual robot is rendered into the scene at a safe surgical configuration (**c**). The robot assistant can then align the robot with the virtual counterpart (**d**), and dock it to the trocar (**e**).

sen as a robotic port and remote center of motion (RCM) for abdominal surgery. Training for port and trocar placement in umbilicus and optimal docking of the robot has a steep learning curve [308, 309].

10.3.1 System and Design

For the experiments we used a 7 DOF KUKA LBR Intelligent Industrial Work Assistant (iiwa) 7 R800 redundant robot manipulator (KUKA AG, Augsburg, Ger-

many). The joint configuration and end-effector pose of the robot was obtained through a ROS interface [310]. It is important to note that our solution is designed to address challenges in surgical settings, and the KUKA arm was merely used as an exemplary robot that was available for this research. The AR environment was delivered by a first-generation Microsoft HoloLens OST-HMD (Microsoft, Redmond, WA).

10.3.2 Alignment of Virtual-to-Real

The ViRAAl strategy is evaluated by aligning the virtual and real robots with and without a reflective-AR display. Each experiment is repeated 10 times for 4 users. The error measurements are presented in Table 10.1. We did not incorporate an external marker-based tracking approach as the base-line since marker tracking exhibits high errors due to propagation, and does not include the user in the loop, i.e. it only determines the registration error and not the augmentation error in the user's view. Instead, to quantify the amount of misalignment, for each iteration we located three pairs of distinct 3D landmarks on the surfaces of both real and virtual robots. These points were identified interactively by intersecting rays from the OST-HMD to the landmark using a gaze cursor.

We define each ray i using the position of the user's head $\mathbf{h}_i$, and the unit direction vector $\mathbf{u}_i$ from the head to the annotated landmark on the spatial map of the environment. The landmark $\mathbf{x}_i^*$ is estimated in a least-squares fashion from the

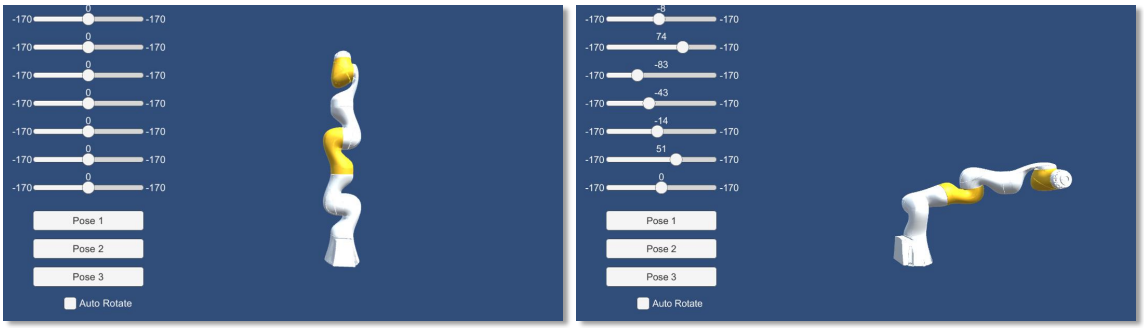

Figure 10.7: Interactive fixed display demonstrating two different target joint configurations

intersection of two rays as:

$$\mathbf{x}_i^* = \arg\min_{\mathbf{x}\in\mathbb{R}^3} \sum_{i=1}^{2} \|(I_3 - \mathbf{u}_i\mathbf{u}_i^\top)(\mathbf{x} - \mathbf{h}_i)\|^2. \tag{10.6}$$

To quantify the error of our ground-truth measurement mechanism using 3D landmarks, we computed the Euclidean distance between different sets of targets on an optical table for a total of 12 times. We selected four combinations of landmarks which were 5 cm, 10 cm, 15 cm, and 20 cm apart. The average error for measuring distances using AR annotations was 3.6 mm.

For virtual-to-real object alignment, the results in Table 10.1 indicate a total error of 16.5 ± 11.0 mm when using the reflective-AR display, and 30.2 ± 23.9 mm when using AR without the additional mirror view. To demonstrate the change in alignment error when averaging multiple alignment transformations on the $SE(3)$ manifold as presented in Sec. 10.2.1, we computed the average transformation given

Eq. 10.2 and Eq. 10.3 when using the AR reflective display. This experiment yielded a total error of 11.3 ± 1.01 mm, which is lower than each individual alignment trial.

10.3.3 Augmented Reality for Robot Set Up: Accuracy Analysis

During a simulated robot-assisted trocar placement, $U = 8$ users moved the robot joints to achieve different target joint configurations. Each user performed this task with the guidance from *1)* AR, *2)* AR with reflective display, and *3)* interactive fixed display, all in randomized orders. The fixed external monitor was used as a non-immersive baseline, displaying the desired robot configurations to the user at all time during the execution, and allowing the users to interact with the visualization by rotating or scaling (Fig. 10.7). The users performed each test three times, and each time with a different target joint. To set up AR we used the average Euclidean transformation over $N = 3$ trials as described in Eq. 10.2 and Eq. 10.3. Errors in joint angles are demonstrated in Fig. 10.8. The total errors are shown in Fig. 10.9. The violin plots show the distribution of error within each guidance method.

The statistics for the overall error in joint positioning using each guidance techniques is compared in Table 10.2. Statistical significance measures are shown in Table 10.3. In Fig. 10.10 we present a comparison between the errors from each joint, and highlight the errors for twisting and revolving joints. A joint is characterized as

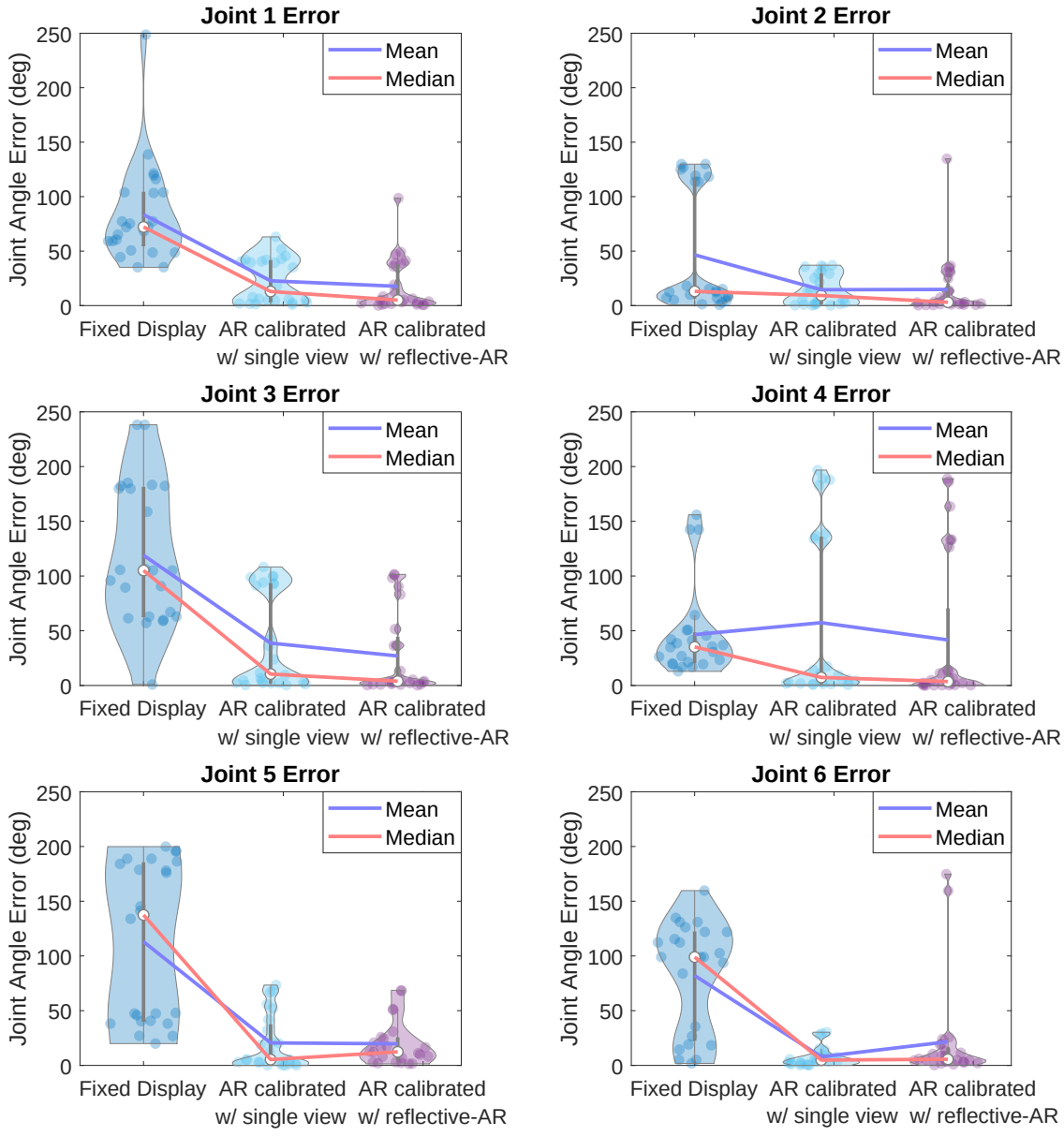

Figure 10.8: Distribution of errors evaluated for each joint separately when guided by *1)* non-immersive fixed display, *2)* AR calibrated without reflective-AR display, and *3)* AR calibrated with reflective-AR display.

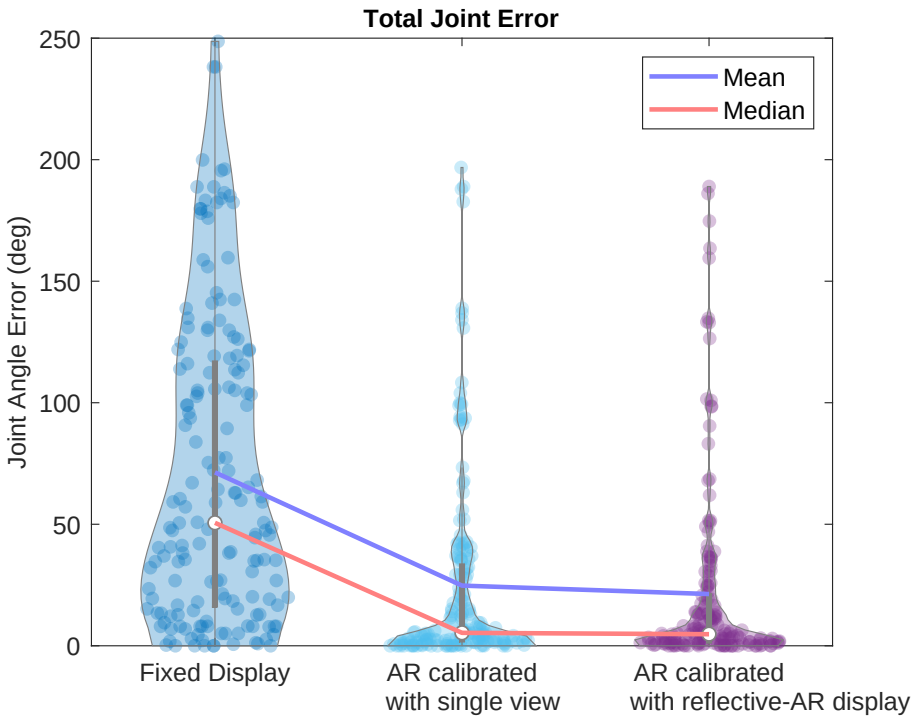

Figure 10.9: Total error distribution for all joints using guidance by *1)* non-immersive fixed display, *2)* AR calibrated without reflective-AR display, and *3)* AR calibrated with reflective-AR display.

twisting if the axis of rotation is parallel to the robot link, and revolving if the axis of rotation is orthogonal to the robot link.

Table 10.1: Mean and standard deviation of misalignment errors in mm.

Alignment Method	$(\overline{t_x}, \sigma_{t_x})$	$(\overline{t_y}, \sigma_{t_y})$	$(\overline{t_z}, \sigma_{t_z})$	$(\|\overline{t}\|_2, \|\sigma_t\|_2)$
ViRAAl	(17.4, 16.1)	(11.9, 6.24)	(21.6, 16.5)	(30.2, 23.9)
ViRAAl + Reflective-AR	(9.00, 5.64)	(10.3, 7.45)	(9.18, 5.77)	(16.5, 11.0)

Table 10.2: Comparison of the error for re-positioning the robot joints in degree units

Joint Error	Mean	Median	Min	Max	Std
Reflective AR	23.7	4.93	0.02	188	41.1
AR	26.8	5.88	0.00	197	42.4
Fixed Display	71.4	50.6	0.64	249	61.8

Table 10.3: P-values for each individual joint, as well as for all joints combined.

P-value	**Joint 1**	**Joint 2**	**Joint 3**	**Joint 4**	**Joint 5**	**Joint 6**	**Total**
Reflective AR / AR	0.61	0.50	0.23	0.71	0.59	0.27	0.54
Reflective AR / Fixed Display	0.12e−6	0.15e−1	0.24e−6	0.77	0.22e−6	0.76e−4	0.12e−15
AR / Fixed Display	0.16e−6	0.39e−1	0.62e−4	0.87	0.12e−6	0.58e−7	0.20e−17

Table 10.4: Time required for ViRAAl and re-positioning the robot joints in minute:second units

	Execution Time				Mean	**Alignment Time**			
	Mean	Median	Min	Max	Std		Median	Min	Max	Std	
Reflective AR	2 : 00	1 : 50	0 : 39	3 : 47	0 : 50		4 : 32	4 : 42	2 : 06	6 : 59	1 : 41
AR	1 : 31	1 : 23	0 : 22	5 : 06	0 : 59		2 : 29	1 : 52	1 : 11	4 : 26	1 : 12
Fixed Display	1 : 34	1 : 35	0 : 27	3 : 20	0 : 48		-	-	-	-	-

10.3.4 Augmented Reality for Robot Set Up: Time Analysis

Table 10.4 presents the observed time for all eight users. Guidance using AR and AR reflector both required registration between the virtual and real content. Therefore, each time the users were given a unique joint configuration target, prior to moving the real robot, the user aligned the virtual model of the robot with its real counterpart three times $\left\{{}^{\mathrm{R}}\mathbf{T}_{\mathrm{V}}(i)\right\}_{i=1}^{3}$. Using the ViRAAl approach presented in Sec. 10.2.1, the transformation ${}^{\overline{\mathrm{R}}}\overline{\mathbf{T}_{\mathrm{V}}}$ which expressed the geometric mean in $SE(3)$ was computed. The average transformation was applied to the AR scene to register the origins of the real and virtual environments, and enable AR guidance. Table 10.4 also presents the time required for ViRAAl with and without AR reflectors for all users.

10.4 Discussion

The experimental results indicated an improved alignment when using a reflective-AR display. The L2-norm average misalignment error in this case was 16.5 ± 11.0 mm, and showed improvement compared to 30.2 ± 23.9 mm error when no reflective displays were used. Averaging the transformations on $SE(3)$ manifold yielded an even lower error of 11.3 ± 1.01 mm. This alignment error does not seem sufficient for tasks

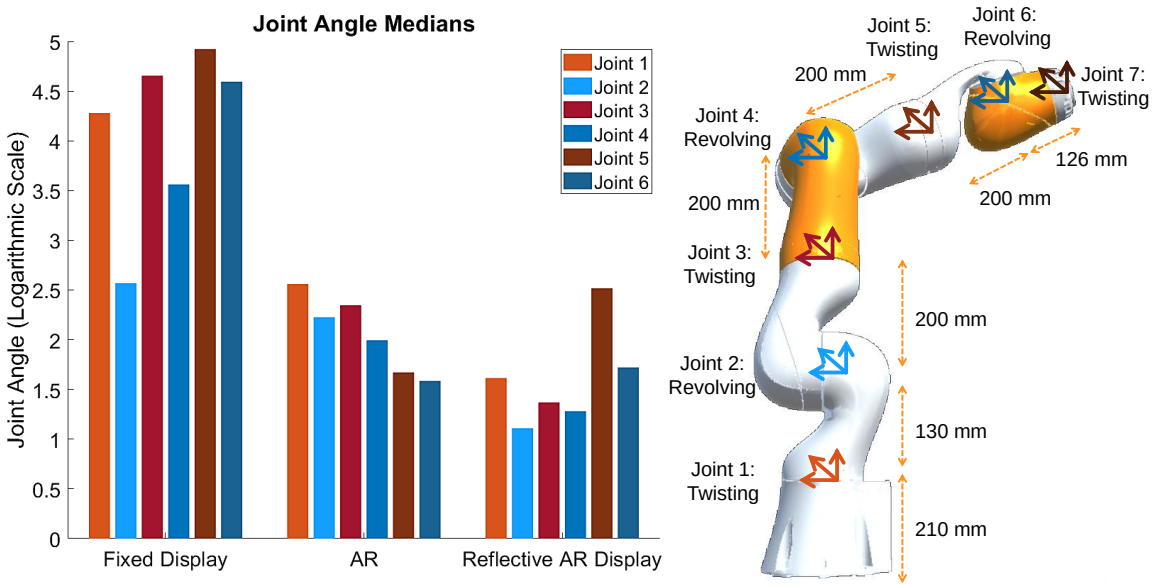

Figure 10.10: The plot demonstrates an abstract comparison between the errors contributed by each joint. Since AR-based approaches yielded substantially smaller errors, we used Logarithmic scale for optimal visualization and comparison of errors with different orders of magnitude. Revolving joints with even indexes are shown in blue, and twisting joints with odd indexes are shown in red.

that require high accuracy such as defining biopsy targets in AR, however seems acceptable for providing intuition during robotic arm set up. Using more than one reflective-AR display did not improve the alignment due to two main reasons. First, the limited field of view of Microsoft Hololens prohibited optimal view of multiple mirrors in their frustums simultaneously when standing in close proximity to the robot. Second, the poor quality of SLAM-based tracking and the unreliable spatial map of the HMD resulted in drifts, hence achieving alignment consensus in all views became challenging. Larger field of view, reliable head tracking, enhanced form factor, enhanced gesture input, and eye tracking capabilities can greatly improve the current

limitations.

We evaluated the results in Table 10.1 based on a novel user-in-the-loop concept. Alternatively, these errors can be evaluated by using an external navigation system and fiducials. However, the latter will exclude the user and will only evaluate the registration error, instead of the augmentation error. Considering the large improvement ($> 45\%$) reported across multiple trials in Table 10.1, we expect the conclusion will not change when using optical navigation.

For modern surgical platforms, the first assistants are trained to set up, dock, tear down, and re-configure the robot using extensive pre-operative E-learning or instructor-led training. In addition to the training that they receive by the manufacturer, general guidelines, and demonstrations in the form of text or visualizations are available in the operating room. In our experiments, to exclude the bias of training, we substituted the conventional training-based approach by recruiting users with no background in setting up surgical robots and focused on demonstrating the effectiveness of intra-operative AR guidance. Comparison with base-line training is a subject of future work, which requires randomized studies in clinical settings with larger populations.

The alignment between the real and virtual, which establishes the registration, is an entirely user-dependent step as the registration chain implicitly takes into account the internal relations between the user's eyes, AR camera on the headset, and the AR displays. Since human visual systems are different, these relations differ, and

consequently cannot be done in a user-out-the-loop setting. Therefore, automatic approaches will not generalize.

Fig. 10.9 presented the misalignment errors for bringing the real robot to a desired configuration. The expected accuracy for the robotic set up depends on the design, number of arms, and number of joints of the particular surgical system. These parameters all vary depending on the surgical use case. The higher the number of joints and arms, the higher the chance for collision; therefore, higher accuracy is demanded for a more complex system. This step during the AR-assisted workflow, which involves the alignment of a real object to virtual, can in future leverage from multi-view strategies by using collaborative AR devices or external cameras.

Several AR publications have shown that time-saving of AR cannot be quantified immediately with dedicated user studies [311], partly because of the unfamiliar interface and exposure to additional information. Time-saving only manifests after the user is proficient with the system. We hypothesize that while there is overhead in setting up AR, the rate of failure/collision would drop leading to a net reduction of overall junk-time. The Riemannian averaging in Eq. 10.2 and Eq. 10.3 can increase this set up time, but is a one-time process which takes place before intervention and can result in a more accurate fusion of information and improved AR experience. We hope AR assistance minimizes the training time and allows operators to verify and inspect the proper alignment of robot arms, both quantitatively and visually.

HMD-based AR is challenged by recurrent estimation of transformations and cor-

rections such as user-to-display and drift, respectively. There is a great wealth of literature directed towards addressing these issues [312], which our solution benefits from as new hardware becomes available. It should be noted that drift will be quite limited because the working volume near the patient bed and the robot is restricted.

Our proposed approach enables the co-registration between the real and virtual spaces and delivers spatially-aware AR. We also demonstrated the application of ViRAAl for AR guidance during minimally-invasive robotic surgery. The estimation of the overall registration greatly benefits from the averaging strategy presented in Eqs. 10.1-10.3, which are suggested to compute the mean transformation that satisfies the properties of $SE(3)$ manifold. This average estimate is a rigid transformation that has the shortest distance to all other estimates around the true pose.

The range of errors exhibited by all intra-operative guidance methods, particularly by the non-immersive fixed display, prove the complexity and importance of this problem for robot manipulation. We computed p-values and compared all pairs of methods in Table 10.3 to identify the most effective assistive approach. Statistical significance was considered if $p < 0.05$. Results suggested that guidance using AR with and without reflective display yielded significantly lower errors compared to non-immersive fixed display. The AR guidance approach using reflective displays outperformed the AR system with no mirrors, however in this comparison statistical significance was not achieved.

In Fig. 10.10, we compared the error contributed by revolving and twisting joints

separately. Results indicated that re-positioning of revolving joints in all three guidance methods are consistently more accurate than the twisting joints. We hypothesize that the higher error is the result of the inherent symmetry in twisting motion that may lead to ambiguities.

10.5 Conclusion

In this chapter we presented a novel multi-view strategy to align virtual and real content, and demonstrated an application of it for improving surgical robotic workflows. The reflective-AR displays were introduced to eliminate the 3D scale ambiguities and improve the AR scene realism. We have demonstrated an AR interface that accommodates multiple reflective displays, and allows the users to scale the images within their viewing frustum [237].

The virtual-to-real registration approach, ViRAAl, is an interactive and user-specific method that calibrates the real and virtual worlds directly to the user's display. No external camera or tracking system, other than the HMD itself, is used in order to keep maximum flexibility and transferability of the system into different surgical environments.

Seamless overlays of virtual content onto the reflective AR displays are achieved by placing virtual cameras at the optical centers where the images were acquired, hence allowing to render virtual and real from an identical imaging geometry. The reflective

displays require a static scene, therefore, are suited for aligning virtual-to-real, and not vice-versa.

The focus of this work is beyond the KUKA robot and its redundant design; it is instead on complex surgical platforms with multiple arms and various joints. The task of alignment is expected to be more difficult for redundant manipulators with more joints. Nonetheless, we expect AR to provide an effective guidance mechanism to reconfigure complex redundant arms at the bedside.

Surgical robots are only certified to be controlled by the surgeon in its active mode, and due to safety reasons, their set up by the surgical staff are performed entirely manually. Our solution is designed around this concept of full manual interaction. It should be noted that our contribution is not on computing joint configurations that minimize collision, but instead we show that if such configuration exists, then with the support of AR it can be manually achieved by the first assistant.

A user interface such as the AR reflectors, can accelerate interaction during surgery. By measuring the exact time of the staff this could be validated in future work. We believe that the proposed alignment strategy can extend to other realms of computer-assisted surgery, namely for surgical training and AR guidance during image-guided therapies.

10.6 Acknowledgments

I express my sincere gratitude to Mr. Arian Mehrfard and Mr. Tianyu Song, for their help with the implementation of the reflective-AR displays. I also thank Mr. Qiaochu Wang for his efforts during the user study and setting up the system for testing in Malone Hall and Mr. Fengfan Xian for running statistical analysis. Lastly, I thank Drs. Nassir Navab, Mehran Armand, Mathias Unberath, and Bernhard Fuerst for their insightful feedback.

Chapter 11

Conclusion

In this dissertation, we proposed and evaluated a broad spectrum of solutions for image-guided and robot-assisted interventions. These span image computing and vision-based approaches for reconstruction, registration, tracking, planning, and stitching of the interventional imaging data. We also investigated the benefits of augmented reality (AR) using immersive and non-immersive technologies for image guidance and staff support during surgical robotic interventions.

It is crucial to consider the tradeoff between workflow flexibility and system accuracy for the design and integration of computer-integrated surgical systems. External tracking hardware delivers high accuracy; however, it compromises flexibility and results in slow adoption of the technology. On the other end of the spectrum, marker-less methods, such as the inside-out tracking accompanied by most AR systems, do not limit the user's movement and do not need other external devices other

than the head-mounted displays, but deliver poor tracking errors. The technologies designed for the surgery rooms of the future should carefully investigate the surgery requirements, and design appropriate solutions by accounting for the ergonomics and needs for each particular intervention.

AR has been traditionally considered as an alternative to fixed displays that enable in-line visuailzation. We extended this concept, and suggested solutions with temporal and spatial awareness gained from the environment and the characteristics of imaging. Based on this concept, we suggested new surgical workflows, and used AR for use-cases beyond visualization, namely for improving interaction, communication, and promoting team approach in the surgical theater.

Our display-based AR environment is built upon an RGBD enhanced C-arm, that enables visualization of 3D optical information from the surgical site superimposed with the planning target. This solution uses a self-contained C-arm which only needs a one-time offline calibration, requires no external trackers, and does not depend on out-dated pre-operative patient data. We believe that this system, by enabling quick planning and visualization, can contribute to reduction of radiation, time, and frustration and increase the efficiency and accuracy for placing surgical implants. Ultimately, this approach may aid in reducing the risk of revision surgery.

A significant advantage of our immersive AR system deliver by head-mounted display is that it is based on 2D C-arm fluoroscopy — and therefore does not diverge from the standard workflow. The standard C-arm is always present in the OR to take

confirmatory images. Consequently, the surgeons are not required to solely rely on the AR system during a procedure. Rather, its value lies in guiding the surgeon to a narrow area of interest and more importantly, to align them with the right trajectory. The surgeon can then seamlessly switch between standard fluoroscopic images and the AR view to guarantee accurate drilling. We believe this is useful, translatable, safe, and novel.

In a training environment, such as in most academic centers with residents and fellows -— this ability for all users to wear the HMD and view the same annotations and AR guidance would be helpful. The more experienced user could ensure that the anatomy had been annotated appropriately and that the AR guidance trajectories are at the correct location.

Direct visualization of X-ray images within their corresponding viewing frustums delivers intuition that effectively unites the content of the 2D image with the 3D imaged anatomy. In this setting, images from various perspectives can be grouped within their frustums to form multi- or extended-view representations of the anatomy. The interlocked frustums shown in Fig. 11.1 are examples for such visualization concept, that can particularly benefit interventions where leg-length discrepancies or malrotations in tibio- and lateral/distal-femoral angles are major concerns. Though assessment of each of these concepts requires an additional approved study, we believe their introduction to the community paves the way in opening new paths for research in this area and expedites the translation of AR-based solutions into future ORs.

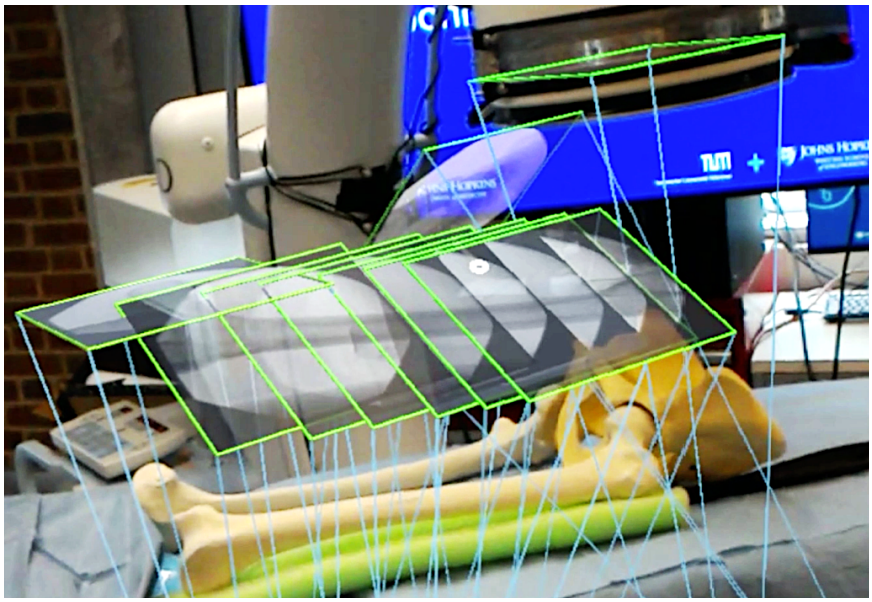

Figure 11.1: The interlocking of multiple X-ray frustums enables visualization of large anatomical structures. In this figure, multiple images are acquired on a co-linear trajectory and are locked to each other to form a quasi-panoramic view of the bone.

Despite that the above solutions deliver spatial awareness, they should not be regarded as replacements for surgical navigation systems. This is because marker-less tracking, currently, cannot deliver the level of accuracy achieved by marker-based surgical navigation or robotic systems. Our solutions are merely advanced visualization platforms that enhance the interaction across the surgical ecosystem and promote effective collaboration.

In designing data-driven models, we exploited the well-established principles and characteristics governing the formation of visual data. This especially manifested in the direct use of Fourier-slice theorem and projection geometry when learning the proper parameterization for the orthographic reconstruction of perspective X-ray

images.

All of the proposed solutions demand the entire crew to get trained appropriately and be able to interact with the systems comfortably. Therefore, time, money, and effort need to be spent to prepare the surgical team for such a digital transformation.

11.1 Considerations for Clinical Deployment:

The success of translation for each of the proposed registration, tracking, and visualization approaches depend on the requirements of the surgery. For instance, for stitching of Cone-Beam CT volumes, while the visual marker-based approach yielded very low stitching errors, it increased the setup complexity by requiring external markers to be fixed to the patient during C-arm re-arrangement. Conversely, RGBD-SLAM tracking allows for increased flexibility as no external markers are required. However, this flexibility came at the cost of slightly higher stitching errors. Since the stitching errors for both marker-based and marker-less methods were well below 1.00 cm, which is considered *"well tolerated"* for leg length discrepancy in the orthopedic literature [313], the deployment of marker-less approaches with higher flexibility may be of greater benefit despite their lower positional accuracy.

11.2 Outlook for Future Work:

The future's surgery rooms are expected to take full advantage of the enhanced and immersive visualization delivered by AR, the autonomy and dexterity realized by the robotic systems, and the data-driven nature and pattern understanding capabilities achieved by machine learning.

To use machine learning in interventional medicine, a significant challenge is access to annotated patient data. Patient privacy and lack of documentation are among the concerns that prevent access to large-scale data. As a result of this scarcity, simulation data have been extensively used as substitutes for real patient data. However, the difference in distributions between simulation and real domains often results in domain gaps that hinder the generalization and transferability of data-driven solutions. Moreover, in real patient data, symptoms and injuries exhibit with diverse patterns, which may be unfeasible to comprehensively include during simulation. For instance, in trauma interventions, fracture patterns are very diverse, and interventional data often contain different surgical tools that could be absent in the simulation data. These challenges highlight the need to understand domain shifts and facilitate data access and documentation in clinics.

Another important area for future research is reinforcement learning to accomplish highly ill-posed and under-constrained tasks for medical image computing. For instance, 2D/3D registration has been traditionally solved using optimization-based strategies by minimizing a loss function. However, these loss terms are highly non-

convex, and as a consequence, the optimizer often converges to local minima. Alternatively, reinforcement learning optimizes the discounted future reward at the cost of taking immediate steps that may deteriorate the short-term similarity scores.

It is also worthwhile to incorporate rich knowledge from imaging characteristics and computer vision principals when using imagery data. Examples include the direct use of epipolar constraints when using multiple views, epipolar consistency conditions when using multiple X-ray images, or forward- and backward-projection principals when dealing with 2D and 3D domains concurrently. The integration of such knowledge can simplify the training and enhance the generalization.

To bring AR to operating rooms, there is a need for further research to identify interactions that promote a human-centric design, such that the interface and the interaction with the interface both appear intuitive for the surgeon.

Alignment between real and virtual content has emerged as a foundational step in many AR applications. For instance, in a clinical setting, the planning data are shown at their corresponding spatial pose, and the surgeon is expected to align the real tools with their virtual counterparts. In another example, to register the AR environment with the patient, specific manual steps may be required to close the registration loop. Unfortunately, due to different perceptions that humans have between real and virtual and the vergence-accommodation conflict, the alignment could be highly inaccurate. Therefore, we ought to better understand the perceptual limitations involving alignment and address those concerns in the future.

AR applications may also extend to outside-OR settings. One of such applications is *"surgical replay"*, where the residents can review the surgery, accompanied with its temporal and spatial information including all the X-ray acquisitions and optical point-clouds from the patient site. This enables the medical trainees to identify distinct actions that were taken by the experienced surgeon based upon each image. Access to such 3D post-operative analysis has the potential to dramatically improve the quality of surgical education.

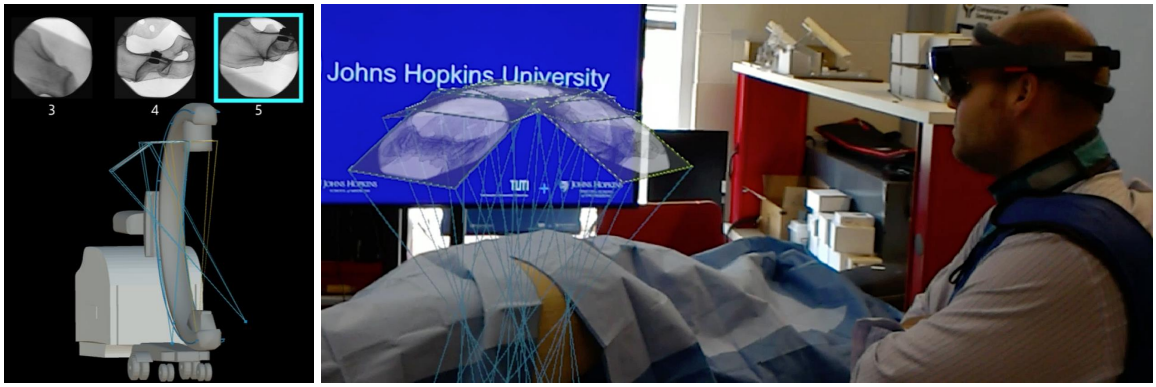

Figure 11.2: Spatial and temporal information from the surgery can be recorded and reviewed after surgery. On the left side an interface is shown allowing the surgeon to select images, which he can then observe geometrically accurate in space, as shown on the right side.

Finally, development of AI strategies can *i)* create semantic understanding from the surgical environment and augment surgeon's intelligence, and *ii)* enhance the spatial mapping and co-localization, thus improving the stability of marker-less AR systems.

Bibliography

1. Meara, J. G. *et al.* Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *The Lancet* **386,** 569–624 (2015).

2. Kurtz, S., Ong, K., Lau, E., Mowat, F. & Halpern, M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *JBJS* **89,** 780–785 (2007).

3. Euler, E., Heining, S., Riquarts, C. & Mutschler, W. C-arm-based three-dimensional navigation: a preliminary feasibility study. *Computer Aided Surgery* **8,** 35–41 (2003).

4. Otake, Y. *et al.* Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration. *IEEE transactions on medical imaging* **31,** 948–962 (2012).

5. Ricci, W. M. *et al.* Angular malalignment after intramedullary nailing of femoral shaft fractures. *Journal of orthopaedic trauma* **15,** 90–95 (2001).

BIBLIOGRAPHY

6. Jaarsma, R., Pakvis, D., Verdonschot, N., Biert, J. & Van Kampen, A. Rotational malalignment after intramedullary nailing of femoral fractures. *Journal of orthopaedic trauma* **18,** 403–409 (2004).

7. Routt Jr, M. C., Simonian, P. T. & Mills, W. J. Iliosacral screw fixation: early complications of the percutaneous technique. *Journal of orthopaedic trauma* **11,** 584–589 (1997).

8. Starr, A., Jones, A., Reinert, C. & Borer, D. Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the acetabulum. *Injury* **32,** SA45–50 (2001).

9. Stöckle, U., Schaser, K. & König, B. Image guidance in pelvic and acetabular surgery—expectations, success and limitations. *Injury* **38,** 450–462 (2007).

10. Guy, P., Al-Otaibi, M., Harvey, E. J. & Helmy, N. The 'safe zone'for extra-articular screw placement during intra-pelvic acetabular surgery. *Journal of orthopaedic trauma* **24,** 279–283 (2010).

11. Starr, A. J., Reinert, C. M. & Jones, A. L. Percutaneous fixation of the columns of the acetabulum: a new technique. *Journal of orthopaedic trauma* **12,** 51–58 (1998).

12. Hankemeier, S., Hufner, T., Wang, G., *et al.* Navigated open-wedge high tibial osteotomy: advantages and disadvantages compared to the conventional tech-

nique in a cadaver study. *Knee Surgery, Sports Traumatology, Arthroscopy* **14,** 917–921 (2006).

13. Hofmann, A. A., Bolognesi, M., Lahav, A. & Kurtin, S. Minimizing leg-length inequality in total hip arthroplasty: use of preoperative templating and an intraoperative x-ray. *American Journal of Orthopedics-Belle Mead-* **37,** 18 (2008).

14. Plaass, C., Clauss, M., Ochsner, P. E. & Ilchmann, T. Influence of leg length discrepancy on clinical results after total hip arthroplasty-a prospective clinical trial. *Hip international* **21,** 441–449 (2011).

15. Blum, L. *et al.* Vertical shear pelvic injury: evaluation, management, and fixation strategies. *International orthopaedics* **42,** 2663–2674 (2018).

16. Mauffrey, C., Stacey, S., York, P. J., Ziran, B. H. & Archdeacon, M. T. Radiographic evaluation of acetabular fractures: review and update on methodology. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* **26,** 83–93 (2018).

17. Lefaivre, K. A., Blachut, P. A., Starr, A. J., Slobogean, G. P. & O'Brien, P. J. Radiographic displacement in pelvic ring disruption: reliability of 3 previously described measurement techniques. *Journal of orthopaedic trauma* **28,** 160–166 (2014).

18. Hamer, O. W. *et al.* Amorphous silicon, flat-panel, x-ray detector: reliability of digital image fusion regarding angle and distance measurements in long-leg radiography. *Investigative radiology* **39,** 271–276 (2004).

19. Boewer, M., Arndt, H., Ostermann, P., Petersein, J. & Mutze, S. Length and angle measurements of the lower extremity in digital composite overview images. *European radiology* **15,** 158–164 (2005).

20. Guenoun, B., Zadegan, F., Aim, F., Hannouche, D. & Nizard, R. Reliability of a new method for lower-extremity measurements based on stereoradiographic three-dimensional reconstruction. *Orthopaedics & Traumatology: Surgery & Research* **98,** 506–513 (2012).

21. Gendrin, C. *et al.* Monitoring tumor motion by real time 2D/3D registration during radiotherapy. *Radiotherapy and oncology* **102,** 274–280 (2012).

22. Mitrović, U., Pernuš, F., Likar, B. & Špiclin, Ž. Simultaneous 3D–2D image registration and c-arm calibration: application to endovascular image-guided interventions. *Medical physics* **42,** 6433–6447 (2015).

23. Otake, Y. *et al.* Automatic localization of vertebral levels in x-ray fluoroscopy using 3D-2D registration: a tool to reduce wrong-site surgery. *Physics in medicine and biology* **57,** 5485 (2012).

24. Otake, Y. *et al. An image-guided femoroplasty system: development and initial cadaver studies* in *SPIE Medical Imaging* (2010), 76250P–76250P.

BIBLIOGRAPHY

25. Carelsen, B., Haverlag, R., Ubbink, D., Luitse, J. & Goslings, J. Does intraoperative fluoroscopic 3D imaging provide extra information for fracture surgery? *Archives of Orthopaedic and Trauma Surgery* **128,** 1419–1424 (2008).

26. Kim, C. W., Lee, Y.-P., Taylor, W., Oygar, A. & Kim, W. K. Use of navigation-assisted fluoroscopy to decrease radiation exposure during minimally invasive spine surgery. *The Spine Journal* **8,** 584–590 (2008).

27. Kakarla, U. K., Little, A. S., Chang, S. W., Sonntag, V. K. & Theodore, N. Placement of percutaneous thoracic pedicle screws using neuronavigation. *World neurosurgery* **74,** 606–610 (2010).

28. Crawford, N. R., Theodore, N. & Foster, M. A. *Surgical robot platform* US Patent 9,782,229. Oct. 2017.

29. Yi, T., Ramchandran, V., Siewerdsen, J. H. & Uneri, A. Robotic drill guide positioning using known-component 3D–2D image registration. *Journal of Medical Imaging* **5,** 021212 (2018).

30. Joskowicz, L. & Hazan, E. J. *Computer aided orthopaedic surgery: incremental shift or paradigm change?* 2016.

31. Grupp, R. B. *et al.* Pose Estimation of Periacetabular Osteotomy Fragments with Intraoperative X-Ray Navigation. *IEEE Transactions on Biomedical Engineering* (2019).

32. Goerres, J. *et al.* Planning, guidance, and quality assurance of pelvic screw placement using deformable image registration. *Physics in Medicine & Biology* **62,** 9018 (2017).

33. Van de Kraats, E. B., van Walsum, T., Kendrick, L., Noordhoek, N. J. & Niessen, W. J. Accuracy evaluation of direct navigation with an isocentric 3D rotational X-ray system. *Medical image analysis* **10,** 113–124 (2006).

34. Matthews, F. *et al.* Navigating the fluoroscope's C-arm back into position: an accurate and practicable solution to cut radiation and optimize intraoperative workflow. *Journal of orthopaedic trauma* **21,** 687–692 (2007).

35. Gebhard, F. T. *et al.* Does computer-assisted spine surgery reduce intraoperative radiation doses? *Spine* **31,** 2024–2027 (2006).

36. Gras, F. *et al.* Screw placement for acetabular fractures: which navigation modality (2-dimensional vs. 3-dimensional) should be used? An experimental study. *Journal of orthopaedic trauma* **26,** 466–473 (2012).

37. Kraus, M., Weiskopf, J., Dreyhaupt, J., Krischak, G. & Gebhard, F. Computer-aided surgery does not increase the accuracy of dorsal pedicle screw placement in the thoracic and lumbar spine: a retrospective analysis of 2,003 pedicle screws in a level I trauma center. *Global spine journal* **5,** 093–101 (2015).

38. Digioia, A. M. *et al.* Surgical navigation for total hip replacement with the use of HipNav. *Operative Techniques in Orthopaedics* **10,** 3–8 (2000).

BIBLIOGRAPHY

39. Jaramaz, B. & Eckman, K. 2D/3D registration for measurement of implant alignment after total hip replacement. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006,* 653–661 (2006).

40. Nikou, C. *et al. POP: Preoperative planning and simulation software for total hip replacement surgery* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (1999), 868–875.

41. Leenders, T., Vandevelde, D., Mahieu, G. & Nuyts, R. Reduction in variability of acetabular cup abduction using computer assisted surgery: a prospective and randomized study. *Computer Aided Surgery* **7,** 99–106 (2002).

42. Widmer, K.-H. & Grützner, P. A. Joint replacement—total hip replacement with CT-based navigation. *Injury* **35,** 84–89 (2004).

43. Haaker, R. G. *et al.* Comparison of conventional versus computer-navigated acetabular component insertion. *The Journal of arthroplasty* **22,** 151–159 (2007).

44. Sato, Y. *et al. Intraoperative simulation and planning using a combined acetabular and femoral (CAF) navigation system for total hip replacement* in *MICCAI* (2000), 1114–1125.

45. Sarin, V. K., Pratt, C. R., Apgar, M. E. & Pratt, W. R. *Non-imaging, computer assisted navigation system for hip replacement surgery* US Patent 6,711,431. 2004.

BIBLIOGRAPHY

46. Kalteis, T. *et al.* Imageless navigation for insertion of the acetabular component in total hip arthroplasty. *Bone & Joint Journal* **88,** 163–167 (2006).

47. Lin, F. *et al.* Limitations of imageless computer-assisted navigation for total hip arthroplasty. *The Journal of arthroplasty* **26,** 596–605 (2011).

48. Taylor, R. H. *et al.* An image-directed robotic system for precise orthopaedic surgery. *IEEE Transactions on Robotics and Automation* **10,** 261–275 (1994).

49. Sugano, N. Computer-assisted orthopaedic surgery and robotic surgery in total hip arthroplasty. *Clinics in orthopedic surgery* **5,** 1–9 (2013).

50. Conditt, M. A. & Roche, M. W. Minimally invasive robotic-arm-guided unicompartmental knee arthroplasty. *JBJS* **91,** 63–68 (2009).

51. Elmallah, R. *et al.* Robotic-Arm Assisted Surgery in Total Hip Arthroplasty. *Surgical technology international* **26,** 283–288 (2015).

52. Schulz, A. P. *et al.* Results of total hip replacement using the Robodoc surgical assistant system: clinical outcome and evaluation of complications for 97 procedures. *The International Journal of Medical Robotics and Computer Assisted Surgery* **3,** 301–306 (2007).

53. Siebert, W., Mai, S., Kober, R. & Heeckt, P. F. Technique and first clinical results of robot-assisted total knee replacement. *The Knee* **9,** 173–180 (2002).

54. Nakamura, N. *et al.* Robot-assisted primary cementless total hip arthroplasty using surface registration techniques: a short-term clinical report. *International journal of computer assisted radiology and surgery* **4,** 157–162 (2009).

55. Yao, J. *et al.* AC-arm fluoroscopy-guided progressive cut refinement strategy using a surgical robot. *Computer Aided Surgery* **5,** 373–390 (2000).

56. Azuma, R. T. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* **6,** 355–385 (1997).

57. Wellner, P. Computer-Augmented Environments back to the real world. *Comm. ACM* **36,** 271–278 (1999).

58. Zlatanova, S. Augmented reality technology. *GISt Report No. 17, Delft, 2002, 72 p.* (2002).

59. Wagner, D., Langlotz, T. & Schmalstieg, D. *Robust and unobtrusive marker tracking on mobile phones* in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008), 121–124.

60. Zhang, X., Fronz, S. & Navab, N. *Visual marker detection and decoding in ar systems: A comparative study* in *Proceedings. International Symposium on Mixed and Augmented Reality* (2002), 97–106.

61. Genc, Y., Riedel, S., Souvannavong, F., Akinlar, C. & Navab, N. *Marker-less tracking for AR: A learning-based approach* in *Proceedings. International Symposium on Mixed and Augmented Reality* (2002), 295–304.

BIBLIOGRAPHY

62. Meola, A. *et al.* Augmented reality in neurosurgery: a systematic review. *Neurosurgical review* **40,** 537–548 (2017).

63. Hallet, J. *et al.* Trans-thoracic minimally invasive liver resection guided by augmented reality. *Journal of the American College of Surgeons* **220,** e55–e60 (2015).

64. Ntourakis, D. *et al.* Augmented reality guidance for the resection of missing colorectal liver metastases: an initial experience. *World journal of surgery* **40,** 419–426 (2016).

65. Wu, J.-R., Wang, M.-L., Liu, K.-C., Hu, M.-H. & Lee, P.-Y. Real-time advanced spinal surgery via visible patient model and augmented reality system. *Computer methods and programs in biomedicine* **113,** 869–881 (2014).

66. Abe, Y. *et al.* A novel 3D guidance system using augmented reality for percutaneous vertebroplasty. *Journal of Neurosurgery: Spine* **19,** 492–501 (2013).

67. Navab, N., Heining, S.-M. & Traub, J. Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications. *IEEE transactions on medical imaging* **29,** 1412–1423 (2010).

68. Heining, S., Wiesner, S., Euler, E. & Navab, N. Pedicle screw placement under videoaugmented fluoroscopic control: First clinical application in a cadaver study. *Int J Computer Assist Radiol Surg* **1,** 189–190 (2006).

69. Gibby, J. T., Swenson, S. A., Cvetko, S., Rao, R. & Javan, R. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. *International journal of computer assisted radiology and surgery* **14,** 525–535 (2019).

70. Paweena, U. *et al.* MR image overlay guidance: system evaluation for preclinical use. *International journal of computer assisted radiology and surgery* **8,** 365–378 (2013).

71. Fischer, G. S. *et al.* MRI image overlay: application to arthrography needle insertion. *Computer Aided Surgery* **12,** 2–14 (2007).

72. Fichtinger, G. *et al.* Image overlay guidance for needle insertion in CT scanner. *IEEE transactions on biomedical engineering* **52,** 1415–1424 (2005).

73. Londei, R. *et al.* Intra-operative augmented reality in distal locking. *International journal of computer assisted radiology and surgery* **10,** 1395–1403 (2015).

74. Wang, H. *et al.* Precision insertion of percutaneous sacroiliac screws using a novel augmented reality-based navigation system: a pilot study. *International orthopaedics* **40,** 1941–1947 (2016).

75. Ogawa, H., Hasegawa, S., Tsukada, S. & Matsubara, M. A pilot study of augmented reality technology applied to the acetabular cup placement during total hip arthroplasty. *The Journal of arthroplasty* **33,** 1833–1837 (2018).

76. Liu, H., Auvinet, E., Giles, J. & y Baena, F. R. Augmented reality based navigation for computer assisted hip resurfacing: a proof of concept study. *Annals of biomedical engineering* **46,** 1595–1605 (2018).

77. Liu, H., Bowyer, S., Auvinet, E. & Baena, F. R. Y. A Smart Registration Assistant for Joint Replacement: Concept Demonstration. *CAOS* **1,** 189–196 (2017).

78. Wang, L. *et al.* Parallax-free intra-operative X-ray image stitching. *Medical Image Analysis* **14,** 674–686 (2010).

79. Fallavollita, P. *et al.* An augmented reality C-arm for intraoperative assessment of the mechanical axis: a preclinical study. *International journal of computer assisted radiology and surgery* **11,** 2111–2117 (2016).

80. Cho, H. S. *et al.* Can augmented reality be helpful in pelvic bone cancer surgery? An in vitro study. *Clinical orthopaedics and related research* **476,** 1719 (2018).

81. Shen, F., Chen, B., Guo, Q., Qi, Y. & Shen, Y. Augmented reality patient-specific reconstruction plate design for pelvic and acetabular fracture surgery. *International journal of computer assisted radiology and surgery* **8,** 169–179 (2013).

BIBLIOGRAPHY

82. Van Duren, B., Sugand, K., Wescott, R., Carrington, R. & Hart, A. Augmented reality fluoroscopy simulation of the guide-wire insertion in DHS surgery: A proof of concept study. *Medical engineering & physics* **55,** 52–59 (2018).

83. Hiranaka, T. *et al.* Augmented reality: the use of the PicoLinker smart glasses improves wire insertion under fluoroscopy. *World journal of orthopedics* **8,** 891 (2017).

84. Yeo, C. T. *et al.* The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections. *IEEE Transactions on Biomedical Engineering* **58,** 2031–2037 (2011).

85. Ponce, B. A., Menendez, M. E., Oladeji, L. O., Fryberger, C. T. & Dantuluri, P. K. Emerging technology in surgical education: combining real-time augmented reality and wearable computing devices. *Orthopedics* **37,** 751–757 (2014).

86. Ponce, B. A. *et al.* Telementoring: use of augmented reality in orthopaedic education: AAOS exhibit selection. *JBJS* **96,** e84 (2014).

87. Condino, S. *et al.* How to build a patient-specific hybrid simulator for orthopaedic open surgery: benefits and limits of mixed-reality using the Microsoft HoloLens. *Journal of Healthcare Engineering* **2018** (2018).

88. Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* **22,** 1330–1334 (2000).

89. Geiger, A., Moosmann, F., Car, Ö. & Schuster, B. *Automatic camera and range sensor calibration using a single shot* in *Robotics and Automation (ICRA), 2012 IEEE International Conference on* (2012), 3936–3943.

90. Newcombe, R. A. *et al. KinectFusion: Real-time dense surface mapping and tracking* in *Mixed and Augmented Reality, IEEE International Symposium on* (2011), 127–136.

91. Rusu, R., Blodow, N. & Beetz, M. *Fast Point Feature Histograms for 3D registration* in *Robotics and Automation, International Conference on* (2009), 3212–3217.

92. Lee, S. C. *et al.* Calibration of RGBD camera and cone-beam CT for 3D intra-operative mixed reality visualization. *International journal of computer assisted radiology and surgery* **11,** 967–975 (2016).

93. Fotouhi, J. *et al. Interventional 3D Augmented Reality for Orthopedic and Trauma Surgery* in *16th Annual Meeting of the International Society for Computer Assisted Orthopedic Surgery (CAOS)* (2016).

94. Fischer, M. *et al.* Preclinical usability study of multiple augmented reality concepts for K-wire placement. *International journal of computer assisted radiology and surgery* **11,** 1007–1014 (2016).

95. Fotouhi, J. *et al.* Pose-Aware C-Arm for Automatic Re-Initialization of Interventional 2D/3D Image Registration. *Accepted for presentation at IPCAI 2017.*

96. Tucker, E. *et al. Towards clinical translation of augmented orthopedic surgery: from pre-op CT to intra-op x-ray via RGBD sensing* in *SPIE Medical Imaging* (2018).

97. Andress, S. *et al.* On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. *Journal of Medical Imaging* **5,** 5 - 5 –12 (2018).

98. Endres, F. *et al. An evaluation of the RGB-D SLAM system* in *Robotics and Automation (ICRA), 2012 IEEE International Conference on* (2012), 1691–1696.

99. Tsai, R. Y. & Lenz, R. K. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on robotics and automation* **5,** 345–358 (1989).

100. Berger, M. *et al.* Marker-free motion correction in weight-bearing cone-beam CT of the knee joint. *Medical Physics* **43,** 1235–1248 (2016).

101. De Silva, T. *et al.* 3D–2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. *Physics in Medicine & Biology* **61,** 3009 (2016).

BIBLIOGRAPHY

102. Markelj, P., Tomaževič, D., Likar, B. & Pernuš, F. A review of 3D/2D registration methods for image-guided interventions. *Medical image analysis* **16,** 642–661 (2012).

103. Hajek, J. *et al. Closing the Calibration Loop: An Inside-out-tracking Paradigm for Augmented Reality in Orthopedic Surgery* in *Medical Image Computing and Computer Assisted Intervention* (2018), to appear.

104. Mezger, U., Jendrewski, C. & Bartels, M. Navigation in surgery. *Langenbeck's archives of surgery* **398,** 501–514 (2013).

105. Beister, M., Kolditz, D. & Kalender, W. A. Iterative reconstruction methods in X-ray CT. *Physica medica* **28,** 94–108 (2012).

106. Gordon, R., Bender, R. & Herman, G. T. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology* **29,** 471–481 (1970).

107. Andersen, A. H. Algebraic reconstruction in CT from limited views. *Medical Imaging, IEEE Transactions on* **8,** 50–55 (1989).

108. Hara, A. K. *et al.* Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *American Journal of Roentgenology* **193,** 764–771 (2009).

109. Scherl, H., Keck, B., Kowarschik, M. & Hornegger, J. *Fast GPU-based CT reconstruction using the common unified device architecture* in *Nuclear Science Symposium Conference Record* **6** (2007), 4464–4466.

110. Oehler, M. & Buzug, T. Statistical image reconstruction for inconsistent CT projection data. *Methods of information in medicine* **46,** 261–269 (2007).

111. Ramamurthi, K. & Prince, J. in *Medical Image Computing and Computer-Assisted Intervention* 134–141 (Springer, 2003).

112. Orth, R. C., Wallace, M. J., Kuo, M. D. & Technology Assessment Committee of the Society of Interventional Radiology. C-arm cone-beam CT: general principles and technical considerations for use in interventional radiology. *Journal of Vascular and Interventional Radiology* **19,** 814–820 (2008).

113. Tam, A., Mohamed, A., Pfister, M., Rohm, E. & Wallace, M. J. C-arm cone beam computed tomographic needle path overlay for fluoroscopic-guided placement of translumbar central venous catheters. *Cardiovascular and interventional radiology* **32,** 820–824 (2009).

114. Hwang, H. S., Chung, M. J., Lee, J. W., Shin, S. W. & Lee, K. S. C-arm cone-beam CT-guided percutaneous transthoracic lung biopsy: usefulness in evaluation of small pulmonary nodules. *American Journal of Roentgenology* **195,** W400–W407 (2010).

115. Klein, L. Anesthetic complications in the horse. *The Veterinary Clinics of North America. Equine Practice* **6,** 665–692 (1990).

116. Bodensteiner, C., Darolti, C., Schumacher, H., Matthäus, L. & Schweikard, A. in *Medical Image Computing and Computer-Assisted Intervention* 177–185 (Springer, 2007).

117. Zhang, Q. *et al.* Correction of motion artifacts in cone-beam CT using a patient-specific respiratory motion model. *Medical physics* **37,** 2901–2909 (2010).

118. Kainz, B., Grabner, M. & Rüther, M. *Fast marker based C-arm pose estimation* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2008), 652–659.

119. Fuerst, B., Fotouhi, J. & Navab, N. *Vision-based intraoperative cone-beam ct stitching for non-overlapping volumes* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015), 387–395.

120. Navab, N., Mitschke, M. & Schütz, O. Medical Image Computing and Computer-Assisted Intervention (eds Taylor, C. & Colchester, A.) 688–697 (1999).

121. Smith, R., Self, M. & Cheeseman, P. in *Autonomous robot vehicles* 167–193 (Springer, 1990).

122. Lange, K., Carson, R., *et al.* EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr* **8,** 306–16 (1984).

123. Newcombe, R. A. *et al. KinectFusion: Real-time dense surface mapping and tracking* in *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on* (2011), 127–136.

124. Sidky, E. Y. & Pan, X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in medicine and biology* **53,** 4777 (2008).

125. Fuerst, B. *et al.* Patient-Specific Biomechanical Model for the Prediction of Lung Motion From 4-D CT Images. *IEEE Transactions on Medical Imaging* **34,** 599–607 (2015).

126. Goitein, M., Abrams, M., Rowell, D., Pollari, H. & Wiles, J. Multi-dimensional treatment planning: II. Beam's eye-view, back projection, and projection through CT sections. *International Journal of Radiation Oncology* Biology* Physics* **9,** 789–797 (1983).

127. Lemieux, L., Jagoe, R., Fish, D., Kitchen, N. & Thomas, D. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Medical physics* **21,** 1749–1760 (1994).

128. Knaan, D. & Joskowicz, L. *Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2003), 351–358.

129. Sadowsky, O., Chintalapani, G. & Taylor, R. H. *Deformable 2D-3D registration of the pelvis with a limited field of view, using shape statistics* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2007), 519–526.

130. Jain, A. & Fichtinger, G. *C-arm tracking and reconstruction without an external tracker* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2006), 494–502.

131. Uneri, A. *et al.* 3D 2D registration for surgical guidance: effect of projection view angles on registration accuracy. *Physics in medicine and biology* **59,** 271 (2013).

132. Varnavas, A., Carrell, T. & Penney, G. Increasing the automation of a 2D-3D registration system. *IEEE transactions on medical imaging* **32,** 387–399 (2013).

133. Gong, R. H., Güler, Ö., Kürklüoglu, M., Lovejoy, J. & Yaniv, Z. Interactive initialization of 2D/3D rigid registration. *Medical physics* **40,** 121911 (2013).

134. Varnavas, A., Carrell, T. & Penney, G. Fully automated 2D–3D registration and verification. *Medical image analysis* **26,** 108–119 (2015).

135. Van der Bom, M. *et al.* Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation. *Medical physics* **37,** 1884–1892 (2010).

136. Rusinkiewicz, S. & Levoy, M. *Efficient variants of the ICP algorithm* in *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on* (2001), 145–152.

137. Bay, H., Tuytelaars, T. & Van Gool, L. *Surf: Speeded up robust features* in *European conference on computer vision* (2006), 404–417.

138. Fischler, M. A. & Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24,** 381–395 (1981).

139. Wein, W. Intensity based rigid 2d-3d registration algorithms for radiation therapy. *Dec* **15,** 100 (2003).

140. Navab, N. *et al. 3D reconstruction from projection matrices in a C-arm based 3D-angiography system* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (1998), 119–129.

141. Glassner, A. S. *An introduction to ray tracing* (Elsevier, 1989).

142. Powell, M. J. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* (2009).

143. Daly, M., Siewerdsen, J., Moseley, D., Jaffray, D. & Irish, J. Intraoperative cone-beam CT for guidance of head and neck surgery: assessment of dose and image quality using a C-arm prototype. *Medical physics* **33,** 3767–3780 (2006).

144. Mirota, D. J., Uneri, A., Schafer, S., *et al. High-accuracy 3D image-based registration of endoscopic video to C-arm cone-beam CT for image-guided skull base surgery* in *SPIE Medical Imaging* (2011), 79640J–79640J.

145. Schafer, S., Nithiananthan, S., Mirota, D., *et al.* Mobile C-arm cone-beam CT for guidance of spine surgery: image quality, radiation dose, and integration with interventional guidance. *Medical physics* **38,** 4563–4574 (2011).

146. Unberath, M., Aichert, A., Achenbach, S. & Maier, A. Consistency-based respiratory motion estimation in rotational angiography. *Medical Physics* **44** (2017).

147. Pauwels, R., Araki, K., Siewerdsen, J. & Thongvigitmanee, S. S. Technical aspects of dental CBCT: state of the art. *Dentomaxillofacial Radiology* **44** (2014).

148. Chang, J., Zhou, L., Wang, S. & Clifford Chao, K. S. Panoramic cone beam computed tomography. *Medical Physics* **39,** 2930–2946 (2012).

149. Emmenlauer, M., Ronneberger, O., Ponti, A., *et al.* XuvTools: free, fast and reliable stitching of large 3D datasets. *Journal of microscopy* **233,** 42–60 (2009).

150. Lamecker, H., Wenckebach, T. & Hege, H.-C. *Atlas-based 3D-Shape Reconstruction from X-Ray Images* in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* **1** (2006), 371–374.

151. Fotouhi, J. *et al.* Pose-aware c-arm for automatic re-initialization of interventional 2d/3d image registration. *International Journal of Computer Assisted Radiology and Surgery,* 1–10 (2017).

152. Fotouhi, J., Fuerst, B., Wein, W. & Navab, N. Can real-time RGBD enhance intraoperative Cone-Beam CT? *International Journal of Computer Assisted Radiology and Surgery,* 1–9 (2017).

153. Kato, H. & Billinghurst, M. *Marker tracking and hmd calibration for a video-based augmented reality conferencing system* in *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on* (1999), 85–94.

154. Yaniv, Z. Localizing spherical fiducials in C-arm based cone-beam CT. *Medical physics* **36,** 4957–4966 (2009).

155. Arun, K. S., Huang, T. S. & Blostein, S. D. Least-squares fitting of two 3-D point sets. *PAMI, IEEE Transactions on,* 698–700 (1987).

156. Sawhney, H. S., Hsu, S. & Kumar, R. *Robust video mosaicing through topology inference and local to global alignment* in *European conference on computer vision* (1998), 103–119.

157. Capel, D. & Zisserman, A. *Automated mosaicing with super-resolution zoom* in *Proc. CVPR* **98** (1998), 885–891.

158. Szeliski, R. & Shum, H.-Y. *Creating full view panoramic image mosaics and environment maps* in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), 251–258.

159. Shum, H.-Y. & Szeliski, R. *Construction and refinement of panoramic mosaics with global and local alignment* in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)* (1998), 953–956.

160. Szeliski, R. *et al.* Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision* **2,** 1–104 (2007).

161. Davis, J. *Mosaics of scenes with moving objects* in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231)* (1998), 354–360.

162. Zhi, Q. & Cooperstock, J. R. Toward dynamic image mosaic generation with robustness to parallax. *IEEE Transactions on Image Processing* **21,** 366–378 (2011).

163. Zhang, F. & Liu, F. *Parallax-tolerant image stitching* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), 3262–3269.

164. Lin, C.-C., Pankanti, S. U., Natesan Ramamurthy, K. & Aravkin, A. Y. *Adaptive as-natural-as-possible image stitching* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), 1155–1163.

BIBLIOGRAPHY

165. Lin, K., Jiang, N., Cheong, L.-F., Do, M. & Lu, J. *Seagull: Seam-guided local alignment for parallax-tolerant image stitching* in *European Conference on Computer Vision* (2016), 370–385.

166. Yaniv, Z. & Joskowicz, L. Long bone panoramas from fluoroscopic X-ray images. *IEEE transactions on medical imaging* **23,** 26–35 (2004).

167. Messmer, P. *et al.* Image fusion for intraoperative control of axis in long bone fracture treatment. *European Journal of Trauma* **32,** 555–561 (2006).

168. Chen, C. *et al.* Ruler based automatic C-arm image stitching without overlapping constraint. *Journal of digital imaging* **28,** 474–480 (2015).

169. Wang, L. *et al. Parallax-free long bone x-ray image stitching* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2009), 173–180.

170. Fotouhi, J. *et al.* Automatic Intra-Operative Stitching of Non-Overlapping Cone-Beam CT Acquisitions. *Medical Physics* (2018).

171. Turbell, H. *Cone-beam reconstruction using filtered backprojection* PhD thesis (Linköping University Electronic Press, 2001).

172. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13,** 600–612 (2004).

173. Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GAN. *arXiv preprint arXiv:1807.00734* (2018).

174. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A. & Bengio, Y. *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation* in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (2017), 11–19.

175. Liu, X. *et al. Extremely Dense Point Correspondences using a Learned Feature Descriptor* 2020. eprint: `arXiv:2003.00619`.

176. Unberath, M. *et al. Deepdrr–a catalyst for machine learning in fluoroscopy-guided procedures* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), 98–106.

177. Unberath, M. *et al.* Enabling machine learning in X-ray-based procedures via realistic simulation of image formation. *International journal of computer assisted radiology and surgery,* 1–12 (2019).

178. Bier, B. *et al. X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), 55–63.

179. Bier, B. *et al.* Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views. *International journal of computer assisted radiology and surgery,* 1–11 (2019).

180. Hartley, R. & Zisserman, A. *Multiple view geometry in computer vision* (Cambridge university press, 2003).

181. Esfandiari, H., Newell, R., Anglin, C., Street, J. & Hodgson, A. J. A deep learning framework for segmentation and pose estimation of pedicle screw implants based on C-arm fluoroscopy. *International journal of computer assisted radiology and surgery* **13,** 1269–1282 (2018).

182. Gao, C., Unberath, M., Taylor, R. & Armand, M. Localizing dexterous surgical tools in X-ray for image-based navigation. *arXiv preprint arXiv:1901.06672* (2019).

183. Belagiannis, V., Rupprecht, C., Carneiro, G. & Navab, N. *Robust optimization for deep regression* in *Proceedings of the ieee international conference on computer vision* (2015), 2830–2838.

184. Barrack, R. L., Lavernia, C., Ries, M., Thornberry, R. & Tozakoglou, E. Virtual reality computer animation of the effect of component position and design on stability after total hip arthroplasty. *Orthopedic Clinics* **32,** 569–577 (2001).

185. Charnley, J. & Cupic, Z. The nine and ten year results of the low-friction arthroplasty of the hip. *Clinical orthopaedics and related research* **95,** 9–25 (1973).

186. D'lima, D. D., Urquhart, A. G., Buehler, K. O., Walker, R. H. & Colwell, C. W. The effect of the orientation of the acetabular and femoral components

on the range of motion of the hip at different head-neck ratios. *JBJS* **82,** 315–21 (2000).

187. Scifert, C. F., Brown, T. D., Pedersen, D. R. & Callaghan, J. J. A finite element analysis of factors influencing total hip dislocation. *Clinical orthopaedics and related research* **355,** 152–162 (1998).

188. Yamaguchi, M., Akisue, T., Bauer, T. W. & Hashimoto, Y. The spatial location of impingement in total hip arthroplasty. *The journal of Arthroplasty* **15,** 305–313 (2000).

189. Lewinnek, G. E., Lewis, J., Tarr, R., Compere, C. & Zimmerman, J. Dislocations after total hip-replacement arthroplasties. *JBJS* **60,** 217–220 (1978).

190. Elkins, J. M., Callaghan, J. J. & Brown, T. D. The 2014 Frank Stinchfield Award: The 'landing zone'for wear and stability in total hip arthroplasty is smaller than we thought: a computational analysis. *Clinical Orthopaedics and Related Research®* **473,** 441–452 (2015).

191. Danoff, J. R. *et al.* Redefining the acetabular component safe zone for posterior approach total hip arthroplasty. *The Journal of arthroplasty* **31,** 506–511 (2016).

192. Esposito, C. I. *et al.* Cup position alone does not predict risk of dislocation after hip arthroplasty. *The Journal of arthroplasty* **30,** 109–113 (2015).

BIBLIOGRAPHY

193. DiGioia III, A. M., Hafez, M. A., Jaramaz, B., Levison, T. J. & Moody, J. E. Functional pelvic orientation measured from lateral standing and sitting radiographs. *Clinical orthopaedics and related research* **453,** 272–276 (2006).

194. Zhu, J., Wan, Z. & Dorr, L. D. Quantification of pelvic tilt in total hip arthroplasty. *Clinical Orthopaedics and Related Research®* **468,** 571–575 (2010).

195. Callanan, M. C. *et al.* The John Charnley Award: risk factors for cup malpositioning: quality improvement through a joint registry at a tertiary hospital. *Clinical Orthopaedics and Related Research®* **469,** 319–329 (2011).

196. Bosker, B., Verheyen, C., Horstmann, W. & Tulp, N. Poor accuracy of free-hand cup positioning during total hip arthroplasty. *Archives of orthopaedic and trauma surgery* **127,** 375–379 (2007).

197. Saxler, G. *et al.* The accuracy of free-hand cup positioning-a CT based measurement of cup placement in 105 total hip arthroplasties. *International orthopaedics* **28,** 198–201 (2004).

198. Investigators, A. T. H. A. C. (*et al.* Outcomes following the single-incision anterior approach to total hip arthroplasty: a multicenter observational study. *Orthopedic Clinics of North America* **40,** 329–342 (2009).

199. Domb, B. G., El Bitar, Y. F., Sadik, A. Y., Stake, C. E. & Botser, I. B. Comparison of robotic-assisted and conventional acetabular cup placement in THA:

a matched-pair controlled study. *Clinical Orthopaedics and Related Research®* **472,** 329–336 (2014).

200. Dorr, L. D., Malik, A., Wan, Z., Long, W. T. & Harris, M. Precision and bias of imageless computer navigation and surgeon estimates for acetabular component position. *Clinical orthopaedics and related research* **465,** 92–99 (2007).

201. Moskal, J. T. & Capps, S. G. Acetabular component positioning in total hip arthroplasty: an evidence-based analysis. *The Journal of arthroplasty* **26,** 1432–1437 (2011).

202. Murphy, S. B., Ecker, T. M. & Tannast, M. THA performed using conventional and navigated tissue-preserving techniques. *Clinical orthopaedics and related research* **453,** 160–167 (2006).

203. Xu, K. *et al.* Computer navigation in total hip arthroplasty: a meta-analysis of randomized controlled trials. *International Journal of Surgery* **12,** 528–533 (2014).

204. Reininga, I. H. *et al.* Minimally invasive and computer-navigated total hip arthroplasty: a qualitative and systematic review of the literature. *BMC musculoskeletal disorders* **11,** 92 (2010).

205. Barrett, W. P., Turner, S. E. & Leopold, J. P. Prospective randomized study of direct anterior vs postero-lateral approach for total hip arthroplasty. *The Journal of arthroplasty* **28,** 1634–1638 (2013).

206. Slotkin, E. M., Patel, P. D. & Suarez, J. C. Accuracy of fluoroscopic guided acetabular component positioning during direct anterior total hip arthroplasty. *The Journal of arthroplasty* **30,** 102–106 (2015).

207. Masonis, J., Thompson, C. & Odum, S. Safe and accurate: learning the direct anterior total hip arthroplasty. *Orthopedics* **31,** 1417–1426 (2008).

208. Rusu, R. B., Blodow, N. & Beetz, M. *Fast point feature histograms (FPFH) for 3D registration* in *2009 IEEE international conference on robotics and automation* (2009), 3212–3217.

209. Ji, W. & Stewart, N. Fluoroscopy assessment during anterior minimally invasive hip replacement is more accurate than with the posterior approach. *International orthopaedics* **40,** 21–27 (2016).

210. Rousseau, M.-A. *et al.* Optimization of total hip arthroplasty implantation: is the anterior pelvic plane concept valid? *The Journal of arthroplasty* **24,** 22–26 (2009).

211. Siewerdsen, J. *et al.* Volume CT with a flat-panel detector on a mobile, isocentric C-arm: Pre-clinical investigation in guidance of minimally invasive surgery. *Medical physics* **32,** 241–254 (2005).

212. Hott, J. S. *et al.* Intraoperative Iso-C C-arm navigation in craniospinal surgery: the first 60 cases. *Neurosurgery* **54,** 1131–1137 (2004).

213. Miller, D. L. *et al.* Occupational radiation protection in interventional radiology: a joint guideline of the Cardiovascular and Interventional Radiology Society of Europe and the Society of Interventional Radiology. *Cardiovascular and interventional radiology* **33,** 230–239 (2010).

214. Theocharopoulos, N. *et al.* Occupational exposure from common fluoroscopic projections used in orthopaedic surgery. *JBJS* **85,** 1698–1703 (2003).

215. Synowitz, M. & Kiwit, J. Surgeon's radiation exposure during percutaneous vertebroplasty. *Journal of Neurosurgery: Spine* **4,** 106–109 (2006).

216. Aagaard, K., Laursen, B. S., Rasmussen, B. S. & Sørensen, E. E. Interaction between nurse anesthetists and patients in a highly technological environment. *Journal of PeriAnesthesia Nursing* **32,** 453–463 (2017).

217. Laverdière, C. *et al.* Augmented reality in orthopaedics: a systematic review and a window on future possibilities. *The Bone & Joint Journal* **101,** 1479–1488 (2019).

218. Hatscher, B. *et al. GazeTap: towards hands-free interaction in the operating room* in *Proceedings of the 19th ACM international conference on multimodal interaction* (2017), 243–251.

219. Mewes, A., Hensen, B., Wacker, F. & Hansen, C. Touchless interaction with software in interventional radiology and surgery: a systematic literature review.

International journal of computer assisted radiology and surgery **12,** 291–305 (2017).

220. Ma, M., Fallavollita, P., Habert, S., Weidert, S. & Navab, N. Device-and system-independent personal touchless user interface for operating rooms: One personal UI to control all displays in an operating room. *International journal of computer assisted radiology and surgery* **11,** 853–861 (2016).

221. Sato, Y. *et al.* Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization. *IEEE Transactions on Medical Imaging* **17,** 681–693 (1998).

222. Navab, N., Bani-Kashemi, A. & Mitschke, M. *Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications* in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (1999), 134–141.

223. Navab, N., Heining, S.-M. & Traub, J. Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications. *IEEE transactions on medical imaging* **29,** 1412–1423 (2009).

224. Tucker, E. *et al. Towards Clinical Translation of Augmented Orthopedic Surgery: From Pre-op CT to Intra-op X-ray via RGBD Sensing* in *SPIE Medical Imaging* (SPIE, 2017), accepted.

BIBLIOGRAPHY

225. Fotouhi, J. *et al.* Plan in 2-D, execute in 3-D: an augmented reality solution for cup placement in total hip arthroplasty. *Journal of Medical Imaging* **5,** 021205 (2018).

226. Müller, F. *et al.* Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging. *The Spine Journal* (2019).

227. Agten, C. A. *et al.* Augmented Reality–Guided Lumbar Facet Joint Injections. *Investigative radiology* **53,** 495–498 (2018).

228. Brun, H. *et al.* Mixed reality holograms for heart surgery planning: first user experience in congenital heart disease. *European Heart Journal-Cardiovascular Imaging* (2018).

229. Pelargos, P. E. *et al.* Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *Journal of Clinical Neuroscience* **35,** 1–4 (2017).

230. Deib, G. *et al.* Image guided percutaneous spine procedures using an optical see-through head mounted display: proof of concept and rationale. *Journal of neurointerventional surgery* **10,** 1187–1191 (2018).

231. Chimenti, P. C. & Mitten, D. J. Google Glass as an alternative to standard fluoroscopic visualization for percutaneous fixation of hand fractures: a pilot study. *Plastic and reconstructive surgery* **136,** 328–330 (2015).

232. Moreta-Martinez, R. *et al.* Augmented reality in computer-assisted interventions based on patient-specific 3D printed reference. *Healthcare technology letters* **5,** 162–166 (2018).

233. Meulstee, J. W. *et al.* Toward holographic-guided surgery. *Surgical innovation* **26,** 86–94 (2019).

234. Ma, L. *et al.* Three-dimensional augmented reality surgical navigation with hybrid optical and electromagnetic tracking for distal intramedullary nail interlocking. *The International Journal of Medical Robotics and Computer Assisted Surgery* **14,** e1909 (2018).

235. Andress, S. *et al.* On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. *Journal of Medical Imaging* **5,** 021209 (2018).

236. Fotouhi, J. *et al.* Co-localized augmented human and X-ray observers in collaborative surgical ecosystem. *International journal of computer assisted radiology and surgery* **14,** 1553–1563 (2019).

237. Fotouhi, J. *et al.* Interactive Flying Frustums (IFFs): spatially aware surgical data visualization. *International journal of computer assisted radiology and surgery,* 1–10 (2019).

238. Alexander, C. *et al.* Augmented Reality for Acetabular Component Placement in Direct Anterior Total Hip Arthroplasty. *The Journal of Arthroplasty* (2020).

239. Hoff, W. & Vincent, T. Analysis of head pose accuracy in augmented reality. *IEEE Transactions on Visualization and Computer Graphics* **6,** 319–334 (2000).

240. Mischke, M. & Navab, N. *Recovering projection geometry: How a cheap camera can outperform an expensive stereo system* in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)* **1** (2000), 193–200.

241. Qian, L. *et al.* Towards virtual monitors for image guided interventions-real-time streaming to optical see-through head-mounted displays. *arXiv preprint arXiv:1710.00808* (2017).

242. Boulay, C. *et al.* Three-dimensional study of pelvic asymmetry on anatomical specimens and its clinical perspectives. *Journal of Anatomy* **208,** 21–33 (2006).

243. Mukherjee, D. P., Zisserman, A. P., Brady, M. & Smith, F. Shape from symmetry: Detecting and exploiting symmetry in affine images. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* **351,** 77–106 (1995).

244. Gao, Y. & Yuille, A. L. *Exploiting symmetry and/or Manhattan properties for 3D object structure estimation from single and multiple images* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), 7408–7417.

245. Hong, W., Yang, A. Y., Huang, K. & Ma, Y. On symmetry and multiple-view geometry: Structure, pose, and calibration from a single image. *International Journal of Computer Vision* **60,** 241–265 (2004).

246. Nagar, R. & Raman, S. *Symmslic: Symmetry aware superpixel segmentation* in *Proceedings of the IEEE International Conference on Computer Vision* (2017), 1764–1773.

247. Loy, G. & Eklundh, J.-O. *Detecting symmetry and symmetric constellations of features* in *European Conference on Computer Vision* (2006), 508–521.

248. Cornelius, H. & Loy, G. *Detecting bilateral symmetry in perspective* in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)* (2006), 191–191.

249. Cornelius, H., Perďoch, M., Matas, J. & Loy, G. *Efficient symmetry detection using local affine frames* in *Scandinavian Conference on Image Analysis* (2007), 152–161.

250. Wang, Z., Tang, Z. & Zhang, X. Reflection symmetry detection using locally affine invariant edge correspondence. *IEEE Transactions on Image Processing* **24,** 1297–1301 (2015).

251. Lee, S. & Liu, Y. Curved glide-reflection symmetry detection. *IEEE transactions on pattern analysis and machine intelligence* **34,** 266–278 (2011).

252. Brachmann, A. & Redies, C. Using convolutional neural network filters to measure left-right mirror symmetry in images. *Symmetry* **8,** 144 (2016).

253. Funk, C. & Liu, Y. *Beyond planar symmetry: Modeling human perception of reflection and rotation symmetries in the wild* in *Proceedings of the IEEE International Conference on Computer Vision* (2017), 793–803.

254. Liu, J. *et al. Symmetry detection from realworld images competition 2013: Summary and results* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2013), 200–205.

255. Mitra, N. J., Pauly, M., Wand, M. & Ceylan, D. *Symmetry in 3d geometry: Extraction and applications* in *Computer Graphics Forum* **32** (2013), 1–23.

256. Xu, K. *et al.* Multi-scale partial intrinsic symmetry detection. *ACM Transactions on Graphics (TOG)* **31,** 181 (2012).

257. Kazhdan, M., Chazelle, B., Dobkin, D., Finkelstein, A. & Funkhouser, T. *A reflective symmetry descriptor* in *European Conference on Computer Vision* (2002), 642–656.

258. Kazhdan, M., Chazelle, B., Dobkin, D., Funkhouser, T. & Rusinkiewicz, S. A reflective symmetry descriptor for 3D models. *Algorithmica* **38,** 201–225 (2004).

259. Podolak, J., Shilane, P., Golovinskiy, A., Rusinkiewicz, S. & Funkhouser, T. *A planar-reflective symmetry transform for 3D shapes* in *ACM Transactions on Graphics (TOG)* **25** (2006), 549–559.

BIBLIOGRAPHY

260. Ovsjanikov, M., Sun, J. & Guibas, L. *Global intrinsic symmetries of shapes* in *Computer graphics forum* **27** (2008), 1341–1348.

261. Kim, V. G., Lipman, Y., Chen, X. & Funkhouser, T. *Möbius transformations for global intrinsic symmetry analysis* in *Computer Graphics Forum* **29** (2010), 1689–1700.

262. Combès, B., Hennessy, R., Waddington, J., Roberts, N. & Prima, S. *Automatic symmetry plane estimation of bilateral objects in point clouds* in *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), 1–8.

263. Ecins, A., Fermuller, C. & Aloimonos, Y. *Detecting reflectional symmetries in 3d data through symmetrical fitting* in *Proceedings of the IEEE International Conference on Computer Vision* (2017), 1779–1783.

264. Nagar, R. & Raman, S. Detecting approximate reflection symmetry in a point set using optimization on manifold. *IEEE Transactions on Signal Processing* **67,** 1582–1595 (2019).

265. Pastor, T., Tiziani, S., Kasper, C. D., Pape, H.-C. & Osterhoff, G. Quality of reduction correlates with clinical outcome in pelvic ring fractures. *Injury* (2019).

266. Verbeek, D. O., van der List, J. P., Villa, J. C., Wellman, D. S. & Helfet, D. L. Postoperative CT is superior for acetabular fracture reduction assessment and reliably predicts hip survivorship. *JBJS* **99,** 1745–1752 (2017).

BIBLIOGRAPHY

267. Pascarella, R. *et al.* Surgical results and factors influencing outcome in patients with posterior wall acetabular fracture. *Injury* **48,** 1819–1824 (2017).

268. Shi, H.-f., Xiong, J., Chen, Y.-x., Wang, J.-f. & Wang, Y.-h. Radiographic analysis of the restoration of hip joint center following open reduction and internal fixation of acetabular fractures: a retrospective cohort study. *BMC musculoskeletal disorders* **15,** 277 (2014).

269. Tornetta, P. & Matta, J. M. Outcome of operatively treated unstable posterior pelvic ring disruptions. *Clinical Orthopaedics and Related Research®* **329,** 186–193 (1996).

270. Schenker, M. L., Mauck, R. L., Ahn, J. & Mehta, S. Pathogenesis and prevention of posttraumatic osteoarthritis after intra-articular fracture. *The Journal of the American Academy of Orthopaedic Surgeons* **22,** 20 (2014).

271. Giannoudis, P., Tzioupis, C., Papathanassopoulos, A., Obakponovwe, O. & Roberts, C. Articular step-off and risk of post-traumatic osteoarthritis. Evidence today. *Injury* **41,** 986–995 (2010).

272. Zhang, W., Ji, Y., Wang, X., Liu, J. & Li, D. Can the recovery of lower limb fractures be achieved by use of 3D printing mirror model? *Injury* **48,** 2485–2495 (2017).

273. Gray, R. J. *et al.* Image-Based Comparison Between the Bilateral Symmetry of the Distal Radii Through Established Measures. *The Journal of hand surgery* (2019).

274. Bao, T. *et al.* Quantitative assessment of symmetry recovery in navigation-assisted surgical reduction of zygomaticomaxillary complex fractures. *Journal of Cranio-Maxillofacial Surgery* **47,** 311–319 (2019).

275. Vannier, M. W., Marsh, J. L. & Warren, J. O. Three dimensional CT reconstruction images for craniofacial surgical planning and evaluation. *Radiology* **150,** 179–184 (1984).

276. Raina, K., Yahorau, U. & Schmah, T. Exploiting bilateral symmetry in brain lesion segmentation. *arXiv preprint arXiv:1907.08196* (2019).

277. Yu, C.-C., Bergeron, L., Lin, C.-H., Chu, Y.-M. & Chen, Y.-R. Single-splint technique in orthognathic surgery: intraoperative checkpoints to control facial symmetry. *Plastic and reconstructive surgery* **124,** 879–886 (2009).

278. Van der Meulen, J. The pursuit of symmetry in cranio-facial surgery. *British journal of plastic surgery* **29,** 85–91 (1976).

279. Preuhs, A. *et al.* Symmetry prior for epipolar consistency. *International journal of computer assisted radiology and surgery,* 1–11 (2019).

BIBLIOGRAPHY

280. Edsander-Nord, A., Brandberg, Y. & Wickman, M. Quality of life, patients' satisfaction, and aesthetic outcome after pedicled or free TRAM flap breast surgery. *Plastic and reconstructive surgery* **107,** 1142–53 (2001).

281. Nahabedian, M. Y. Symmetrical breast reconstruction: analysis of secondary procedures after reconstruction with implants and autologous tissue. *Plastic and reconstructive surgery* **115,** 257–260 (2005).

282. Teo, I. *et al.* Body image dissatisfaction in patients undergoing breast reconstruction: Examining the roles of breast symmetry and appearance investment. *Psycho-oncology* **27,** 857–863 (2018).

283. Pinkall, U. & Polthier, K. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics* **2,** 15–36 (1993).

284. Dey, T. K. & Wang, Y. Reeb graphs: Approximation and persistence. *Discrete & Computational Geometry* **49,** 46–73 (2013).

285. Lipman, Y. & Funkhouser, T. Möbius voting for surface correspondence. *ACM Transactions on Graphics (TOG)* **28,** 72 (2009).

286. Fotouhi, J. *et al. Exploiting Partial Structural Symmetry for Patient-Specific Image Augmentation in Trauma Interventions* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), 107–115.

BIBLIOGRAPHY

287. Huber, P. J. in *International Encyclopedia of Statistical Science* 1248–1251 (Springer, 2011).

288. Tile, M. Acute pelvic fractures: I. Causation and classification. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* **4,** 143–151 (1996).

289. Chintalapani, G. *et al. Statistical atlas based extrapolation of CT data* in *Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling* **7625** (2010), 762539.

290. Van der Schatte Olivier, R. v., van't Hullenaar, C., Ruurda, J. & Broeders, I. Ergonomics, user comfort, and performance in standard and robot-assisted laparoscopic surgery. *Surgical endoscopy* **23,** 1365 (2009).

291. Jayakumaran, J. *et al.* Robotic-assisted laparoscopy in reproductive surgery: a contemporary review. *Journal of robotic surgery* **11,** 97–109 (2017).

292. Juza, R. M., Lyn-Sue, J. R. & Pauli, E. M. in *Laparoscopic and Robotic Incisional Hernia Repair* 103–115 (Springer, 2018).

293. Liu, H., Kinoshita, T., Tonouchi, A., Kaito, A. & Tokunaga, M. What are the reasons for a longer operation time in robotic gastrectomy than in laparoscopic gastrectomy for stomach cancer? *Surgical Endoscopy* **33,** 192–198 (2019).

294. Qian, L., Deguet, A., Wang, Z., Liu, Y.-H. & Kazanzides, P. *Augmented reality assisted instrument insertion and tool manipulation for the first assistant in*

robotic surgery in *2019 International Conference on Robotics and Automation (ICRA)* (2019), 5173–5179.

295. Qian, L., Zhang, X., Deguet, A. & Kazanzides, P. *ARAMIS: Augmented Reality Assistance for Minimally Invasive Surgery Using a Head-Mounted Display* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2019), 74–82.

296. Qian, L., Deguet, A. & Kazanzides, P. ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery. *Healthcare technology letters* **5,** 194–200 (2018).

297. Mohareri, O. *et al. Ultrasound-based image guidance for robot-assisted laparoscopic radical prostatectomy: initial in-vivo results* in *International Conference on Information Processing in Computer-Assisted Interventions* (2013), 40–50.

298. Pessaux, P. *et al.* Towards cybernetic surgery: robotic and augmented reality-assisted liver segmentectomy. *Langenbeck's archives of surgery* **400,** 381–385 (2015).

299. Reiners, D., Stricker, D., Klinker, G. & Müller, S. Augmented reality for construction tasks: Doorlock assembly. *Proc. IEEE and ACM IWAR* **98,** 31–46 (1998).

BIBLIOGRAPHY

300. Henderson, S. J. & Feiner, S. K. *Augmented reality in the psychomotor phase of a procedural task* in *2011 10th IEEE International Symposium on Mixed and Augmented Reality* (2011), 191–200.

301. Feiner, S., Macintyre, B. & Seligmann, D. Knowledge-based augmented reality. *Communications of the ACM* **36,** 53–62 (1993).

302. Caudell, T. P. & Mizell, D. W. *Augmented reality: An application of heads-up display technology to manual manufacturing processes* in *Proceedings of the twenty-fifth Hawaii international conference on system sciences* **2** (1992), 659–669.

303. Nuernberger, B., Ofek, E., Benko, H. & Wilson, A. D. *Snaptoreality: Aligning augmented reality to the real world* in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), 1233–1244.

304. Martin-Gomez, A., Eck, U. & Navab, N. *Visualization Techniques for Precise Alignment in VR: A Comparative Study* in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), 735–741.

305. Milgram, P. & Drascic, D. *Perceptual effects in aligning virtual and real objects in augmented reality displays* in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **41** (1997), 1239–1243.

306. Willemsen, P., Gooch, A. A., Thompson, W. B. & Creem-Regehr, S. H. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments* **17,** 91–101 (2008).

307. Moakher, M. Means and averaging in the group of rotations. *SIAM journal on matrix analysis and applications* **24,** 1–16 (2002).

308. Chang, C., Steinberg, Z., Shah, A. & Gundeti, M. S. Patient positioning and port placement for robot-assisted surgery. *Journal of endourology* **28,** 631–638 (2014).

309. Iranmanesh, P. *et al.* Set-up and docking of the da Vinci® surgical system: prospective analysis of initial experience. *The international journal of medical robotics and computer assisted surgery* **6,** 57–60 (2010).

310. Safeea, M. & Neto, P. Kuka sunrise toolbox: Interfacing collaborative robots with matlab. *IEEE Robotics & Automation Magazine* **26,** 91–96 (2019).

311. Dünser, A. & Billinghurst, M. in *Handbook of augmented reality* 289–307 (Springer, 2011).

312. Tuceryan, M., Genc, Y. & Navab, N. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators & Virtual Environments* **11,** 259–276 (2002).

313. Maloney, W. J. & Keeney, J. A. Leg length discrepancy after total hip arthroplasty 1. *The Journal of arthroplasty* **19,** 108–110 (2004).

Vita

Javad Fotouhi earned his BSc degree in Electrical Engineering from the University of Tehran, MSc degree in Biomedical Computing from the Technical University of Munich, and MSE degree in Robotics from Johns Hopkins University. He then joined the Laboratory for Computational Sensing and Robotics (LCSR) at Johns Hopkins and worked towards his PhD degree at Johns Hopkins University. His PhD research was co-advised by Nassir Navab and Mehran Armand.

Javad co-authored over 50 publications and received multiple paper awards for his contributions from MICCAI, IPCAI, the Medical Physics Journal, etc. During his PhD, he was selected as a Siebel Scholar that recognizes the top students from the world's leading graduate schools for their academic excellence and demonstrated leadership. Javad's research focus includes the applications of augmented reality, machine learning, and robotics in interventional medicine. Javad is excited to join Philips Research North America in Cambridge, MA as a Research Scientist.