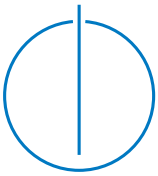
FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Master's Thesis in Biomedical Computing

Integration of RGBD Camera for Mobile C-arms — Calibration, Accuracy and Application

Sing Chun Lee



Department of Informatics

Technische Universität München

Master's Thesis in Biomedical Computing

Integration of RGBD Camera for Mobile C-arms — Calibration, Accuracy and Application

Integration einer RGBD Kamera für Mobile C-Bögen — Kalibrierung, Genauigkeit und Anwendungen

Author:	Sing Chun Lee
Supervisor:	Prof. Dr. Nassir Navab
Advisor:	Bernhard Fuerst
Thesis handed in on:	May 15, 2016

I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

Munich, on May 15, 2016 Sing Chun Lee

Acknowledgments

I would like to thank my supervisor Professor Dr. Nassir Navab for the opportunity of an internship at Johns Hopkins University and his valuable advice. I am immensely grateful to my advisor Bernhard Fuerst for his trust, guidance and continuous supports to my research. Special thanks go to Risto Kojcev, Marius Fischer and Javad Fotouhi, who helped me and took their time to discuss problems with me, not limited to scientific but also philosophic.

I would like to show my deepest gratitude to my mum and brothers for their unlimited and unconditional support. None of this would have been possible without their love and care, and I am thankful to my friends in Hong Kong, Munich and Baltimore. We talked, listened and played together. Last but not least, many thanks to Ben, Fabian, Jennifer, Jie Ying and Zerui for proof-reading and commenting on the grammar mistakes, word collocation and presentation format.

Abstract

Nowadays, orthopedic trauma surgery is performed in a minimal invasive way by using X-Ray imaging technique with mobile C-arms. In order to improve accuracy and reduce radiation dose, numerous solutions were proposed. These solutions usually introduce new hardware into the operation room (OR), at the same time introducing additional constraints on the surgical workspaces. For example, line-of-sight problem imposed by optical tracking systems; therefore, they require radical changes to the surgical setup and workflow. To overcome this limitation, Camera augmented mobile C-arm (CAMC) is proposed, which provides an intuitive overlay of an X-Ray and video images for a robust guidance solution (accuracy of less than 1 mm) with ideally only one single X-ray image. CAMC achieves the aforesaid goals with minimal impact on existing setup and workflow. However, depth information assessment in the OR remains an open topic for computer scientists.

In this thesis, we integrate a commercially available RGBD camera (e.g. Intel RealSense) into the C-arm system. After an one-time calibration routine, the depth camera space is registered with the cone beam computed tomography (CBCT) volume space. A surface reconstruction technique (e.g. KinectFusion) is used to generate the target surface; while digital reconstruction radiography (DRR) is used for simulating the X-Ray image at any desired angles. They are overlaid to provide an intuitive interface for surgical guidance, similar to CAMC; yet, instead of providing only one static view, multiple views at any desired angles can be provided. It allows for a mixed reality visualization platform that can provide depth information intra-operatively.

Experiments are designed to evaluate the accuracy and repeatability of our method. A pre-clinical study is conducted to compare the proposed system with conventional methods and CAMC. We conclude the proposed method achieves a reasonable accuracy of 2.58 mm and has clear advantages in terms of time spent, radiation dose and surgical work load, compared to conventional methods.

Contents

Outline of the Thesis

Part I: Introduction and Background Theory

CHAPTER 1: INTRODUCTION

An overview of the thesis, current state of art and its problem statement are presented.

CHAPTER 2: BACKGROUND THEORY

The background theory and tools utilized in this thesis are described.

Part II: System Design, Calibration Method and Mixed Reality Visualization

CHAPTER 3: SYSTEM DESIGN

A brief comparison of depth cameras, the system design and the calibration phantom design are discussed.

CHAPTER 4: CALIBRATION METHOD

The calibration method, its procedure and limitations are explained.

CHAPTER 5: MIXED REALITY VISUALIZATION

The visualization techniques and the mixed reality platform provided by the proposed method are demonstrated.

Part III: Experimental Assessment and Usability Evaluation

CHAPTER 6: EXPERIMENTAL ASSESSMENT

The experimental setup and results, which assesses the accuracy and repeatability of the method; and the influence of different point cloud densities and phantom choices are reported.

CHAPTER 7: USABILITY EVALUATION AND CONCLUSION

The pre-clinical study, which evaluates the system together with conventional C-arm and Camera augmented mobile C-arm, is presented and followed by a conclusion.

Part I: Introduction and Background Theory

1 Introduction

Contents

(a) (b) (c)

Figure 1.1: Anatomical structure drawing by Leonardo da Vinci (a), modern X-Ray image (b) and fused visualization of different imaging techniques (PET and CT) (c).
Sources: images are taken from [9, 2, 14].

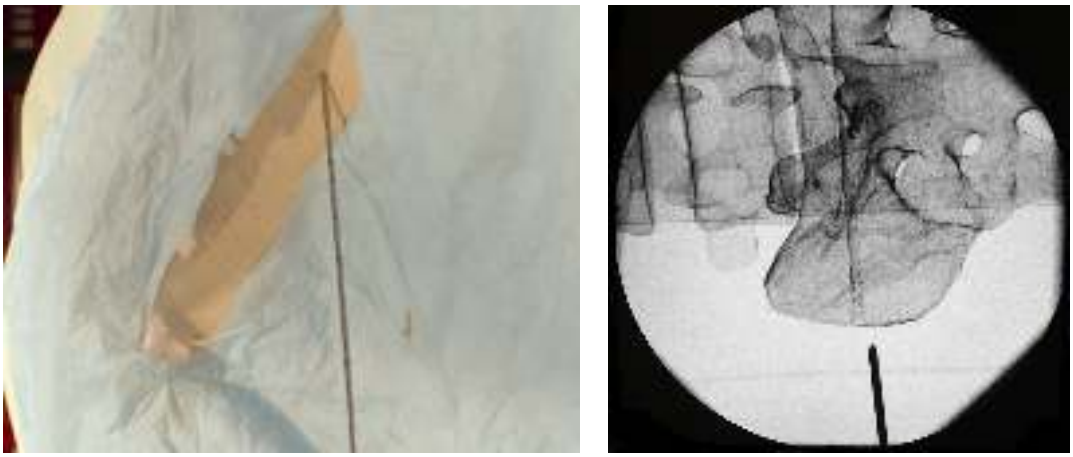
Before the discovery of X-Ray imaging by Wilhelm Conrad Röntgen in the 19th century, physicians had long been looking for a deeper understanding of the human body anatomy. In the 15th century, Leonardo da Vinci pioneeringly studied the human anatomy by drawing and detailing human skeletons, muscles and sinews (Fig. 1.1a), which greatly contributed to the advancement of medical science by revealing the mystery of our anatomy. The discovery of X-Ray imaging (Fig. 1.1b) transformed conventional procedures by allowing physicians to see-through patient's anatomy directly without opening the body,

thereby transforming the conventional open procedures to minimally invasive procedures.

Nowadays, minimally invasive medical procedures are mainly computer-assisted by different imaging techniques. A sophisticated way to facilitate physicians to access this imaging information is by data fusion and visualizing it as intuitively as possible. It helps physicians to quickly and correctly perceive the information, such as in Fig. 1.1c. In this case, it fuses the positron emission tomography (PET) and computed tomography (CT) to allow physicians identify the anatomical structure and tumor tissues at the same time. In this thesis, we are motivated to integrate a red-green-blue-depth (RGBD) camera for mobile C-arms and to develop better intra-operative visualization of CT data. We focus our interests on the application in orthopedic and trauma surgery.

1.1 State of the Art

In orthopedic and trauma surgery, X-ray imaging is frequently used for locating the entry point and determining the insertion angle, such as percutaneous ilio-sacral and pedicle screw placements in spine surgeries as illustrated in Fig. 1.2. The task is challenging as 2D X-Ray images lack depth information for localization in 3D space. It demands the surgeons to mentally align the surgical tool and patient's body among X-Ray images at different perspectives. In order to prevent unnecessary damages to the soft tissues and nervous system around the vertebra, a considerable number of X-Ray images has to be taken for a single screw or wire placement. The surgeons may even undergo several failed attempts before reaching an acceptable placement orientation [24, 29]. Therefore, the procedure duration is typically long and involves high radiation exposures.

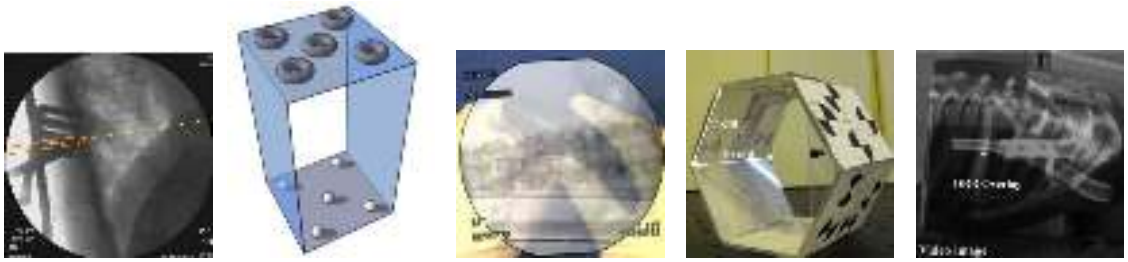


(a) (b)

Figure 1.2: In a minimally invasive orthopedic procedure, surgeons are trying to locate the entry point and determine the insertion angle with the help of X-Ray imaging. (a) shows the surgeon's view while (b) shows the X-Ray image. Surgeons are required to mentally align their view to the X-Ray image.

To address the problem, optical-based image-guided navigation systems are introduced to the operation room (OR) for tracking surgical tools and the patient, to recover the spatial relationship between them and the X-Ray images. These systems (e.g. [15, 17, 33]) provide sub-millimeter accuracy for surgical tool navigations in X-Ray images (Fig. 1.3a). They essentially replace the current navigation habit that requires a considerable number of X-Ray images, thus reducing the radiation dose significantly. However, these systems do not help remarkably shorten operation time. They are outside-in tracking systems that occupy extra space in the OR, and tracking results are limited by its line-of-sight. The additional device setup also increases the complexity of the procedure.

An alternative approach is to directly integrates video cameras or optical-based tracking devices to the mobile C-arms. For instance, Camera augmented mobile C-arm (CAMC) involves a video camera and an X-Ray transparent mirror construction, which are rigidly mounted on the gantry of C-arms. This design allows the X-Ray imaging and video camera to share the same view by calibration [20]. The X-Ray imaging and video camera views are calibrated with a phantom consisting of radio-opaque and optical markers (Fig. 1.3b), so that by registering the markers in both views, their optical centers and image planes are aligned. This design provides an augmentation of X-Ray images onto a live video as illustrated in Fig. 1.3c. The system was evaluated and used in orthopedics and trauma surgeries. [7, 19] concluded that it improves the localization of an incision entry point and orientation, and further reduces radiation exposures. However, limited by the projection models of the X-Ray imaging and video camera, the system enforces the C-arm to be in an



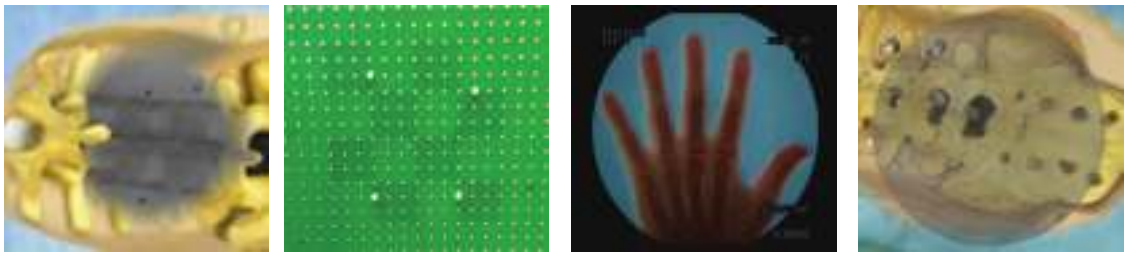
(a) (b) (c) (d) (e)

Figure 1.3: (a) shows the augmentation of the surgical tools such as drills in the X-Ray images using outside-in tracking systems. (b) is the bi-planer calibration phantom for aligning X-Ray imaging and video camera optical centers, whose results allows augmentation of X-Ray images on the live video camera as illustrated in (c). (d) is another calibration phantom for registering the CBCT volume and video camera, which allows augmentation of the DRR from CBCT on the live video camera as shown in (e). **Sources: images are taken from [17, 20, 25].**

up-side down setting for the sake of aligning the physical models. The mirror construction also reduces the surgical workspace and the system may not provide sufficient depth information in one single 2D-2D augmentation view. [32] extended the design by introducing a second camera to provide supplementary view which is orthogonal to the CAMC view. With optical-based tools tracking, it augments the surgical tool in this orthogonal view to supplement the depth information.

On the other hand, Tracker-on-C integrates the optical-based tracking device into C-arms [25]. This work steps further from the 2D-2D augmentation to an interactive 2D-3D augmentation by registering a Cone-beam CT (CBCT) and the tracking device camera. An hex-face calibration phantom with radio-opaque and optical markers (Fig. 1.3d) is designed and paired-point registrations of markers is performed to recover the projection of X-Ray images during CBCT scanning relative to the camera origin. Using these calibrated projections, similar to CAMC, an augmentation of X-Ray images on the live video can be computed. Since a CBCT volume is acquired, a Digitally Reconstructed Radiographs (DRR) at a given projection can also be computed. Therefore, the system provides an interactive 2D-3D augmentation view from different projection angles as shown in Fig. 1.3e. The major drawbacks of this system are the involvement of marker tracking on the surgical sites and the limitation of the augmentation view by the tracking device position, which results in frequent C-arm re-positioning for obtaining new desired views. Nevertheless, both CamC and Tracker-on-C eliminate the outside-in tracking system, reduce X-Ray exposures and achieve sub-millimeter accuracy. However, providing sufficient depth information intuitively in the OR remains a challenge.

In the last decade, with a more mature and commercially available depth sensing technology, depth cameras such as Kinect, Intel RealSense, PrimeSense etc. are studied so as



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

Figure 1.4: (a) shows the augmentation of X-Ray images with a reconstructed surface. (b) depicts the calibration phantom for registering depth camera and X-Ray imaging. (c) and (d) show the similar augmentation under different designs in [35] and [13].
Sources: images are taken from [12, 35, 13]

to resolve the depth perception challenge in clinical applications. The concept of [20, 25] is extended by integrating RGBD cameras into C-arms such as in [12, 13, 35]. [12] replaced the video camera of CAMC with a RGBD camera and studied the accuracy of the depth measurement under the mirror construction design. It was ascertained that the mirror does not exert a major influence on the depth measurement. Given the calibration result of the RGB camera and X-Ray imaging system (using the same method as in [20]) and the calibration result of RGB and depth cameras, the system can compute augmentation of X-Ray images onto a reconstructed surface from the depth camera (Fig. 1.4a). On the other hand, [35] proposes to calibrate the attached RGBD camera to 2D X-Ray imaging device directly without using the mirror construction. A calibration phantom with radio-opaque and optical markers is designed (Fig. 1.4b) and the calibration is done by computing the projection of the 3D point cloud of the markers acquired by the depth camera to the corresponding 2D points of the markers on the X-Ray image plane. Similar to [12], an augmentation of X-Ray images onto the reconstructed surface is obtained (Fig. 1.4c). The calibration method reached an error of 0.54 mm (RMS) $\pm$ 1.40 mm. [13] further demonstrated the feasibility of using multiple depth cameras to achieve the same augmentation with better depth information (Fig. 1.4d). These works step further from the 2D-2D augmentation to a 3D-2D argumentation, and attempted to improve depth perception by utilizing 3D reconstructed surfaces. The main limitation of these works is the nature of 2D projection of the X-Ray imaging. Although the depth camera provides a 3D reconstructed surface, the augmentation is physically correct only when the viewpoint is the same as the X-Ray source.

1.2 Problem Statement

For orthopedic and trauma surgery, the ultimate goals are always to shorten the operation time, reduce the radiation dose, avoid unnecessary tissue damages without altering the existing procedure. Since the procedure fundamentally relies on X-Ray imaging, the main difficulty to achieve the aforesaid goals is the depth perception from multiple 2D X-Ray images. To resolve this challenge, an advanced visualization technique is desired, which should facilitate surgeons to perceive the depth information in the OR quickly and accurately. An important cue for depth perception in human vision system is motion parallax. By observing movements in a sequence of 2D images or multiple static 2D images in different angles, we can easily distinguish the foreground and background, thereby perceiving depth information of the observed target. Therefore, providing an interactive visualization with multiple perspective views is essential for the depth understanding in the OR.

Inspired by the literature, this thesis proposes a system designed to tackle the depth perception challenge by integrating an RGBD camera for C-arms and calibrating the RGBD camera to the CBCT volumes. It steps further from the 2D-3D or 3D-2D augmentation to a 3D-3D augmentation. It enables an intuitive 3D visualization with the augmentation of both physical and anatomical information, and provides multiple arbitrary views simultaneously. On top of this, live flying point clouds of the surgical tools or hands are also augmented in the scene, so surgeons are able to interact with the visualization and quickly understand the relationship among the tools, the surgical sites and the targets in this 3D mixed augmented reality.

1.3 Thesis Structure

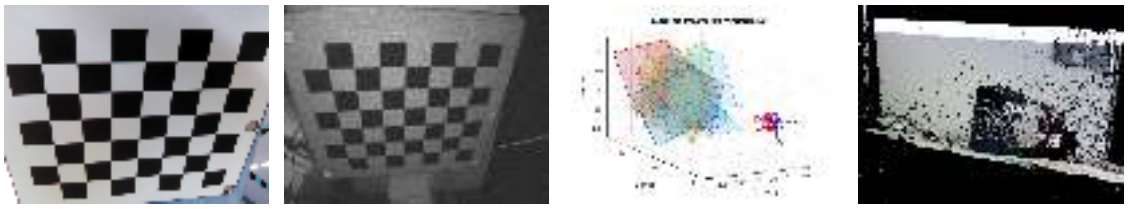
The thesis is organized as follows: Part I contains Chapter 1 which gives the introduction, the overall picture of the problem that we are trying to resolve, and the ultimate goal that we would like to achieve; and Chapter 2 which outlines the theories, tools and techniques that are involved in the thesis. Part II describes the approach that we propose to resolve the problem, which is separated into three chapters. Chapter 3 discusses the system design — the choice of RGBD cameras, the integration to mobile C-arms, and the calibration phantom design. Chapter 4 details the calibration method for the RGBD camera and the CBCT acquired by the mobile C-arm, and Chapter 5 sketches out a mixed visualization concept using 3D-3D augmentation provided by the calibration result. Lastly, Part III reports the experimental assessment results in Chapter 6 and draws a conclusion with a pre-clinical usability study in Chapter 7.

2 Background Theory

Contents

In this chapter, we are going to sketch out the background theories that are involved in this thesis, and acknowledge the tools that we used for the realization and implementation of these theories.

2.1 Camera Parameter Estimation - Checkerboard Calibration



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

Figure 2.1: (a) and (b) are corresponding checkerboard images in RGB and depth camera. (c) shows the extrinsic calibration results using Matlab calibration toolbox and (d) illustrates the color point clouds using the calibration results.

In order to compute the 3D points from depth camera and project the color from RGB camera to depth camera space, we need to calibrate the RGBD camera to recover its intrinsic and extrinsic parameters. A standard checkerboard stereo camera calibration described

in [38] using the checkerboard shown in Fig. 2.1 is performed on the RGB and depth cameras. The RGB and depth cameras are modeled as pin-hole cameras and their intrinsic parameters $\mathbf{K}_{RGB}$, $\mathbf{K}_{Depth}$ and extrinsic parameters ${}^{Depth}\mathbf{T}_{RGB}$ are expressed as follow:

$$\mathbf{K}_{RGB} = \begin{bmatrix} f_x^{RGB} & 0 & c_x^{RGB} \\ 0 & f_y^{RGB} & c_y^{RGB} \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_{Depth} = \begin{bmatrix} f_x^{Depth} & 0 & c_x^{Depth} \\ 0 & f_y^{Depth} & c_y^{Depth} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

where f_x^{RGB}, f_y^{RGB} and f_x^{Depth}, f_y^{Dpeth} are the focus lengths of the RGB and depth camera; c_x^{RGB}, c_y^{RGB} and c_x^{Depth}, c_y^{Depth} are the principal points of the RGB and depth camera.

$${}^{RGB}\mathbf{T}_{Depth} = \begin{bmatrix} {}^{RGB}\mathbf{R}_{Depth} & {}^{RGB}\mathbf{t}_{Depth} \\ \mathbf{0} & 1 \end{bmatrix} \tag{2.2}$$

where ${}^{RGB}\mathbf{R}_{Depth}$ and ${}^{RGB}\mathbf{t}_{Depth}$ are the rotation matrix and translation vector from depth camera to RGB camera.

A sequence of corresponding RGB and depth camera images are obtained by placing the checkerboard at different poses within the shared view (an example is shown in Fig. 2.1a and 2.1b). The parameters are estimated by using Matlab calibration toolbox (Fig. 2.1c). With the calibration results, the depth measurement from the depth camera and the color information from RGB camera, a set of colored point clouds $\mathcal{P} := \{\mathbf{p}_i = [x_i\ y_i\ z_i\ r_i\ g_i\ b_i]^T | i = 0...N\}$, where N is the total number of points, can be computed as follow:

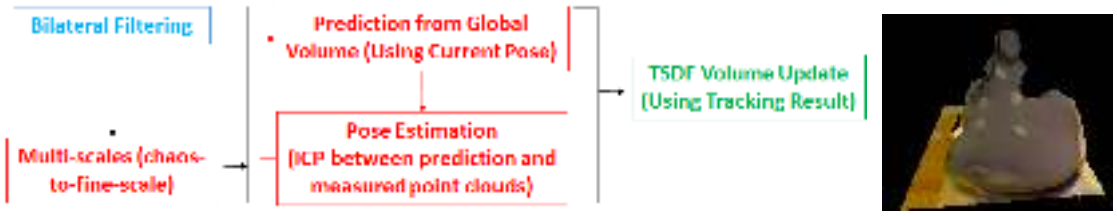
$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = d_i \mathbf{K}_{Depth}^{-1} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}, \begin{bmatrix} cu_i \\ cv_i \\ 1 \end{bmatrix} = \frac{1}{z_i} \mathbf{K}_{RGB} \left({}^{RGB}\mathbf{R}_{Depth} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + {}^{RGB}\,\mathbf{t}_{Depth} \right) \tag{2.3}$$

and

$$\begin{bmatrix} r_i \\ g_i \\ b_i \end{bmatrix} = \begin{cases} \text{RGB value at } (cu_i, cv_i) & \text{if } (cu_i, cv_i) \text{ within field of view of RGB camera} \\ [0\ 0\ 0\ ]^T & \text{Otherwise} \end{cases} \tag{2.4}$$

where u_i, v_i are the pixel coordinates of the depth camera space, d_i is the depth measurement at u_i, v_i; x_i, y_i, z_i are the 3D coordinate in depth camera space computed from the depth measurement and its intrinsic parameters; cu_i, cv_i are the projected pixel coordinate on the RGB image plane using the intrinsic and extrinsic parameters, and r_i, g_i, b_i are the corresponding color information. r_i, g_i, b_i are set to white if the projected point cu_i, cv_i are out of view of the RGB camera. An example of a computed colored point cloud is shown in Fig. 2.1d.

2.2 Surface Reconstruction Technique - KinectFusion



(a) (b)

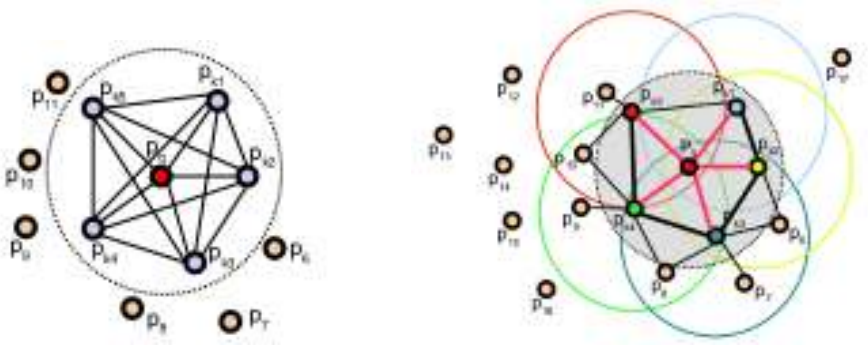
Figure 2.2: (a) briefly describes the KinectFusion reconstruction algorithm and a reconstruction result is shown in (b).

Given a live sequence of 3D point clouds computed from the depth camera, real-time simultaneous localization and mapping (SLAM) can be performed and used to reconstruct an object surface. KinectFusion is an remarkable tool to reconstruct object surface in this manner [21]. KinectFuson utilizes bilateral filtering to remove noise in the depth measurements, and a course-to-fine iterative closest points (ICP) algorithm to estimate the transformation between two consecutive frames. This thesis adapted the open source version of KinectFusion[1] to reconstruct colored object surfaces.

After KinectFusion is initialized, a 3D global volume in truncated signed distance function (TSDF) is created in the global coordinate. For each consecutive frame, it is filtered by bilateral filtering for noise removal, and downsampled with different scale. At each scale level, a measured point cloud is computed from the downsampled frame. Using the current tracking result, the 3D global volume is transformed to a predicted camera coordinate and then projected to the image plane to obtain a predicted point cloud. Now, in the camera coordinate, potential corresponding points are found by comparing the difference of the surface normal and distance between the measured and predicted point clouds. Given these corresponding points, ICP is performed to estimate the new transformation, which also gives a new prediction. This process is iteratively applied at each scale level and continue in a course-to-fine manner, until it is converged. At last, the TSDF volume is updated by merging the current point cloud using the final tracking result. The reconstruction algorithm is summarized in Fig. 2.2a and an example of the reconstructed surface is shown in Fig. 2.2b. KinectFusion allows us to reconstruct the surface and perform tracking with depth camera in real time. For more mathematical and technical details, readers are referred to [21].

[1] open source code is available at https://github.com/Nerei/kinfu_remake

2.3 3D Features - Fast Point Feature Histogram (FPFH)



(a) (b)

Figure 2.3: (a) and (b) are the graphical relationship of point feature histogram and fast point feature histogram. **Sources: images are taken from [23, 22]**

In classic computer vision, we typically extract features of an object of interest from images and match them so that we can register the object in different images. It can be used for tracking, object recognition, image registration, picture stitching etc. When coming to the 3D world, we would also like to extract 3D features from different measured point clouds, such that we can also do tracking, registration, stitching etc.. [34] proposed surflet-pair-relation histograms which extracts 4D geometric features from all surflet pairs of a point cloud and their normals into a histogram. This feature is a generalization of curvature and is invariant to rotation and translation. It is robust and able to distinguish arbitrary shapes in 3D. [27] extended this work for point cloud density invariance. And in this thesis, we used the speed-up version Fast Point Feature Histogram to describe our 3D features [28], which is implemented in the PCL library [26].

Given a sets of points $\mathcal{P} := \{p_i\}$ and a defined radius r, for each point p_i, we can define a neighborhood Ω_i centered at point p_i with radius r. For all pairs $\{p_j, p_k\} \in \Omega_i$, a point feature histogram (PFH) can be computed with the 4D geometric features ϕ_{jk} defined in [34]:

$$PFH_i = H\left(\{\phi_{jk}\}\right) \tag{2.5}$$

where $H(.)$ is an operator to put the features into histogram, and ϕ_{jk} is a set of features of angular variations computed from the two points p_i, p_j and their normals. Fast point feature histogram (FPFH) is then computed as a weighted PFH in Ω_i, where w_j is the Euclidean distance between point pairs $\{p_i, p_j\}$:

$$FPFH_i = PFH_i + \frac{1}{k}\sum_{j\in\Omega_i}\frac{1}{w_j}PFH_j \tag{2.6}$$

FPFH can be computed in $O(nk^2)$ time, where n is the total number of points and k is the total number of neighbor; and provide robust 3D features for template matching, tracking, 3D-3D point cloud registration etc.. An graphical illustrations of PFH and FPFH are shown in Fig. 2.3.

2.4 3D Point Cloud Registration - Iterative Closest Points (ICP)

Figure 2.4: Summary of ICP algorithm

Iterative closest points algorithm is typically used for registering two sets of points. The algorithm estimates the transformation between two sets of points by minimizing a cost function in a least square sense. The main drawback of the algorithm is the potential trap of a local minimum. Therefore, it should be used with caution that either the optimization is guaranteed to be strictly convex or there is a good initialization that ensures the expected solution in the initial neighborhood. In this thesis, we perform ICP with a good initialization and with the cost function below using the PCL interactive ICP implementation [26].

Given two sets of points $\mathcal{P}^{target} := \{p_i^{target}\}$ and $\mathcal{P} := \{p_i\}$, we try to obtain the transformation $\mathbf{T}_{min}$ such that:

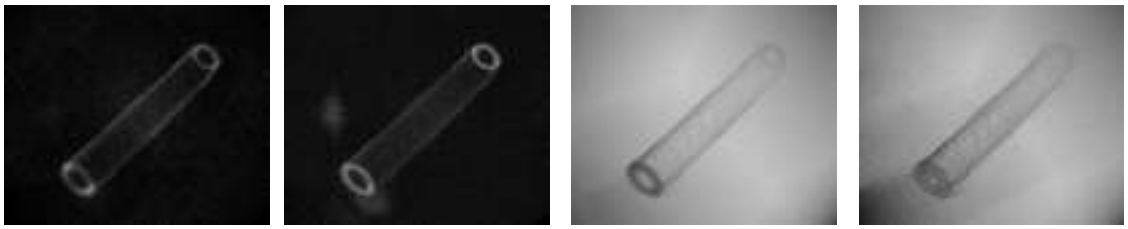
$$\mathbf{T}_{min} = \min_{\mathbf{T},\mathcal{S}} \sum_{(i,j)\in\mathcal{S}} \|p_i^{target} - \mathbf{T}p_j\|_2^2 \tag{2.7}$$

where $S := \{(i,j)|j = \min_k \|p_i^{target} - \mathbf{T}p_k\|_2^2\}$ is the set containing the corresponding indices between the target and the current point clouds which is found by treating the closest point as the correspondence. The ICP algorithm is summarized in Fig. 2.4.

2.5 Synthetic X-Ray - Digitally Reconstructed Radiograph (DRR)

Digitally reconstructed radiograph (DRR) treats a CBCT volume as an attenuation model to generate synthetic X-Ray images at a given X-Ray source position. Current DRR rendering utilizes modern computer graphics rendering technique with GPU acceleration — direct volume rendering with ray casting passing through the attenuation model. Multiple transfer functions can be defined so that particular ranges of attenuation values can be emphasized with different colors and opacities. Some typical GPU-accelerated DRR rendering pipelines could be found in [1, 8]. In this thesis, we adapted the direct volume rendering implemented in ImFusion SDK[2] with customized transfer function for synthetic

[2]ImFusion SDK http://imfusion-tech.com/8-english/23-imfusion-sdk



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

Figure 2.5: Examples of DRR generated synthetic X-Ray images of a tube phantom

X-Ray image generation. Examples of synthetic X-Ray with different transfer functions are illustrated in Fig. 2.5.

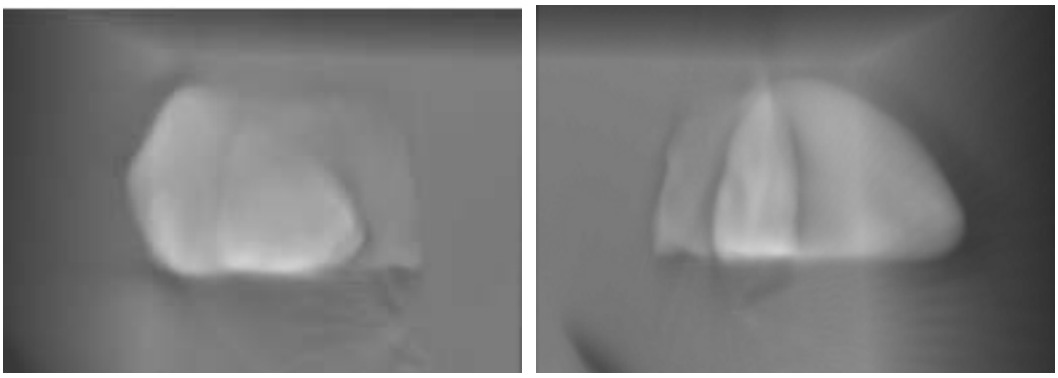
2.6 Depth Perception - Occlusion and Motion Parallax

(a) (b)

Figure 2.6: (a) and (b) show the overlay of the bone model and body surface with and without using the depth cue - occlusion **Sources: images are taken from [4]**

In order to enhance the depth perception in the visualization, two main depth perception cues are applied — occlusion and motion parallax [5]. We would like to overlay the object surface with the synthetic X-Ray image, which raises a problem of how users can perceive the correct order of the objects in the overlay. We need to ensure the object surface in the overlay is preserved perceptually on the top of the synthetic X-Ray image. In a word, we need to avoid the effect of floating object as shown in Fig. 2.6a. To achieve this, we apply the similar technique as in [4], a virtual window is created and synthetic X-Ray

image is only overlaid within the window such that the occlusion of the object surface outside the window helps users perceive the correct order within the overlay region. An example of the overlay with and without the window are shown in Fig. 2.6a and 2.6b. More sophisticated perceptual visualization technique such as contextual anatomic mimesis [3], which uses the curvature, angle of incidence factor and fade out windows can be used to further improve the overlay. In this thesis, we applied the occlusion depth cue with fade out windows.



(a) (b)

Figure 2.7: (a) and (b) demonstrates that multiple views help us to perceive the target depth.

On the other hand, even though depth order can be perceived better with occlusion, it is still hard to perceive the distance between two objects in a single 2D image. For example, Fig. 2.7a shows an example of 2D synthetic X-Ray generated by DRR. It consists of two stones placed next to each other. It is hard to tell how far the two stones are apart. To provide better depth understanding in this case, motion parallax is applied. Multiple simultaneous views at different view angles show different occlusion levels of the two objects. Our vision system can interpret and estimate the distance between the objects from them. For instance, if we look at both Fig. 2.7a and 2.7b together, we are able to understand which one is foreground and which one is background, and can imagine a motion parallax as if we are changing our view from left to right, and hence estimate the distance between them.

Even better and more intuitive, when we are allowed to interactively change the view by rotating it and observe a sequence of dynamic motion changes, as if we are rotating a transparent container and estimating the object depth inside. In this thesis, the system proposed consists of multiple interactive views with fade out windows overlaid, which aims to provide an intuitive visualization of CT and patient data for improving depth understanding in the OR.

2.7 Error Measurement - Target Registration Error (TRE)

There are different error measurements for calibration accuracy evaluation such as fiducial localization error (FLE), fiducial registration error (FRE) and target registration error (TRE) proposed by [16]. [18] did a mathematical analysis and concluded that different error measurement algorithms in fact converge to a general maximum likelihood solution and also demonstrated experimentally that TRE can predict FLE even in the presence of anisotropic noise. In this thesis, we chose to use TRE for our error measurement. We positioned our landmarks in a way that is as much non-collinear, non-coplanar and uniformly distributed as possible, so that a small rotational and translational error is reflected in the TRE measurement. Provided the calibration result, we have the transformation $\mathbf{T}$ between two spaces. Given a set of corresponding landmark positions in both spaces $\mathcal{L} := \{(m_i^{target}, m_i) | i = 0...n\}$ where n is the total number of landmarks, the TRE is computed as follow:

$$TRE = \frac{1}{n} \sum_{(m_i^{target}, m_i) \in \mathcal{L}} \| m_i^{target} - \mathbf{T} m_i \|_2^2 \tag{2.8}$$

Part II: System Design, Calibration Method and Mixed Reality Visualization

3 System Design

Contents

In the following sections, we describe the components of the system design. We briefly compare different types of depth cameras regarding their advantages and disadvantages, depict the system setup and discuss the calibration phantom design for the depth camera integration for C-arms.

3.1 Depth Camera Comparison

A verity of methods for depth measurement exists, such as depth-from-focus, depth-from-motion, depth-from-shape, stereo vision, structured light and time-of-flight (ToF) cameras [30]. The first three, as their names suggested, are based on estimation from the changes in focus, motion, and shape, which in general uses multiple consecutive images to find their correspondences and thereby estimate the depth. It is computationally expensive and may produce ambiguous results when there are no good corresponding features in the moving image sequences. On the other hand, the last three uses triangulation on stereo images, deformation of structured patterns and modulated time-of-flight measurement to estimate depth information. The data is acquired in a single time frame and they generally give more precise information.

As discussed in [11], stereo vision requires textured surfaces to find correspondence in stereo images in order to generate an accurate triangulation result. Its depth measurement effective range depends on the baseline of the stereo camera. The wider the baseline is, the deeper the depth information can be measured. Since in clinical application the observable surfaces are frequently untextured and the working spaces are limited, stereo vision is not suitable for acquiring depth information. Conversely, both structured light and ToF cameras do not require textured surfaces for depth measurement. Both techniques utilize infrared for fast depth measurement. ToF camera shoots out modulated infrared and measures the flight time of the infrared light traveling to the surface and bouncing back to the camera and hence calculate the traveled distance as depth measurement. On the other

hand, a structured light camera projects a structured infrared pattern onto the surface and observes the deformation of the pattern. Based on the deformation, the depth information by interpolating the curvature changes is recovered. Since both cameras are using infrared, if there are additional infrared sources, they suffer from interference. Due to the wave modulation, ToF is less susceptible to interference effect. However, ToF cameras underlie several limitations: Since ToF camera directly measures infrared intensity, the measurement becomes materials dependent. Materials having different infrared reflectance therefore yield to different depth measurements. In addition, the infrared sensors typically require around 20 minutes warm-up time for a stable measurement. Furthermore, depth distortion (namely Wiggling effect) occurs in ToF cameras. An additional depth distortion rectification is required to obtain accurate depth information.

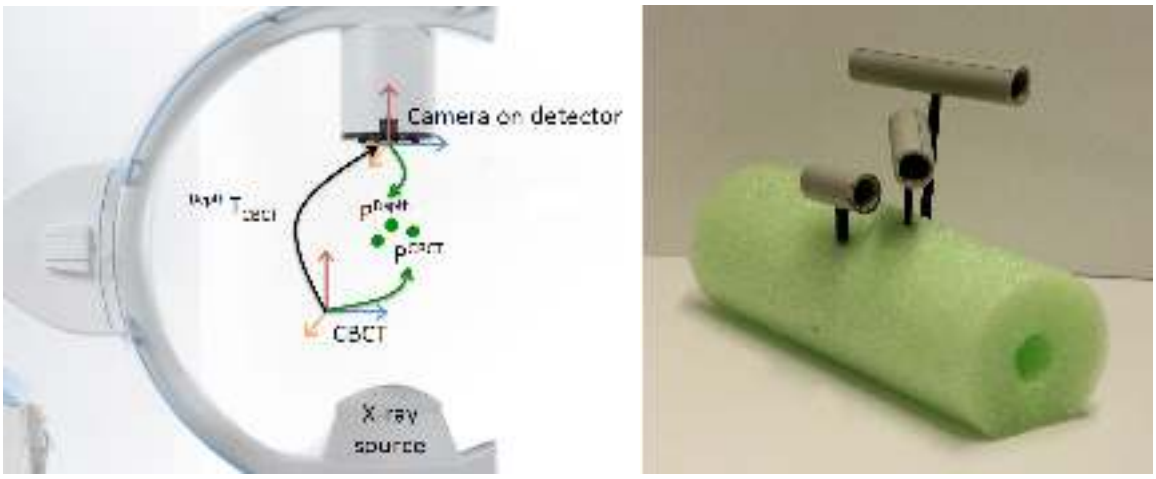
On the contrary, depth measurement by structured light camera is not material dependent and does not require warm-up time nor depth distortion rectification. It can provide relatively reliable depth information. Its limitation is mainly the interference by other infrared sources, which also hinders the use of multiple structured light cameras. Nevertheless, among the aforementioned depth measurement techniques, structured light is more suitable in clinical application. In this thesis, we chose to use Intel RealSense F200 close range structured light camera for the integration. We also compared the result with Microsoft Kinect 360 structured light camera.

3.2 System Setup

A close-range structured light depth camera, such as Intel RealSense F200, is rigidly mounted next to the detector of a C-arm as illustrated in Fig. 3.1a. The camera is positioned such that its view covers the whole CBCT volume. With this setup, the transformation between the CBCT volume and the depth camera center can be modeled as a rigid body transformation ${}^{Depth}\mathbf{T}_{CBCT}$ in $SE(3)$.

$$ {}^{Depth}\mathbf{T}_{CBCT} = \begin{bmatrix} {}^{Depth}\mathbf{R}_{CBCT} & {}^{Depth}\mathbf{t}_{CBCT} \\ \mathbf{0} & 1 \end{bmatrix} \tag{3.1} $$

We designed a calibration phantom as described in the following section, and proposed to obtain the surface point clouds of this phantom in both CBCT space ($\mathcal{P}^{CBCT}$) and depth camera space ($\mathcal{P}^{Depth}$) to perform point clouds registration, so as to recover ${}^{Depth}\mathbf{T}_{CBCT}$, which will be discussed in Chapter 4.



(a) (b)

Figure 3.1: (a) depicts the system setup of the depth camera integration and the transformation relationship and (b) shows the calibration phantom

3.3 Calibration Phantom

The phantom design is shown in Fig. 3.1b. It is composed of three pipes and a cylindrical foam base. Theoretically, any object, whose surface is distinguishably visible in CBCT volume scan and clearly visible in infrared images for surface reconstruction, is suitable for the calibration. Therefore, we chose the materials such that their surfaces are visible in both the CBCT and in depth camera. The pipes have higher radiation absorption than the foam base, which allows us to easily segment it out to obtain their surfaces in CBCT volume. In addition, the round shape provided by the pipes is good for the depth camera measurement since it lowers the corner reflection effect of the infrared; meanwhile the foam base acts as additional feature for depth camera tracking and surface reconstruction.

The phantom is constructed with special geometric properties for the sake of avoiding ambiguity. The pipes have different lengths and are positioned at different heights and orientations, such that it provides an unique 3D shape profile. The unique shape is essential and ensures the uniqueness of the rigid transformation to be recovered from CBCT volume and depth camera. The design aims to provide stable and reliable surface point clouds acquired in both CBCT and depth camera space.

3.4 Proposed Workflow

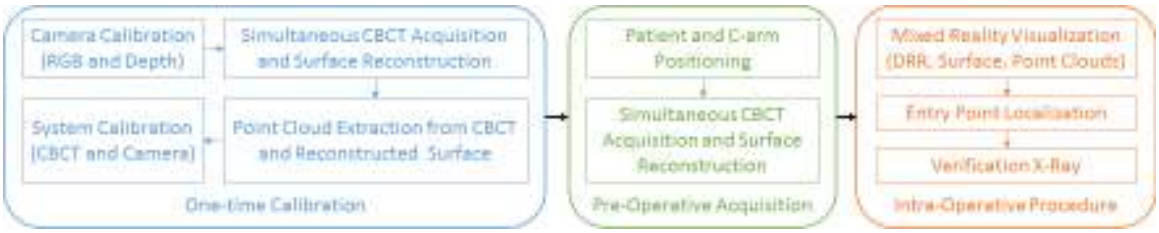


Figure 3.2: Workflow with using new system design

With this system design, the camera is rigidly mounted on the C-arm, thereby it only requires an one-time calibration to recover the spatial relationship between the CBCT volume and the camera. Fig. 3.2 describes the proposed workflow. Before the procedure starts, patient and C-arm are positioned and pre-operative CBCT data and patient surface are acquired. Then, using the one-time system calibration result, the system can provide mixed reality visualization for surgeons to locate the entry point, ideally without taking any other X-Ray images. The one-time calibration will be described in Chapter 4 while the mixed reality visualization will be illustrated in Chapter 5.

4 Calibration Method

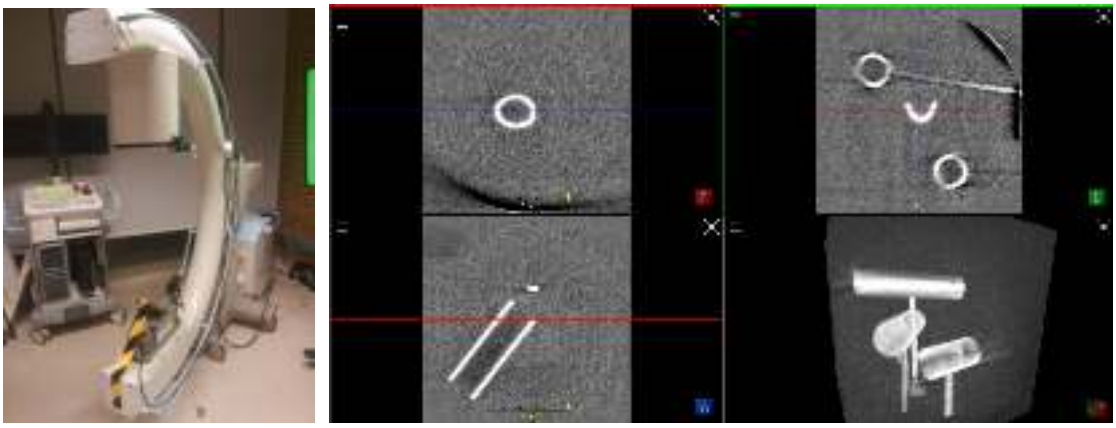
Contents

This chapter explains the one-time calibration method in Fig.3.2. The RGBD camera is attached on the C-arm as described in Section 3.2 is calibrated by using stereo camera checkerboard calibration described in Section 2.1. Therefore, the intrinsic parameters $\mathbf{K}_{RGB}$, $\mathbf{K}_{Depth}$ and the extrinsic parameters ${}^{Depth}\mathbf{T}_{RGB}$ of the camera are known. Hence, we can obtain the colored point clouds $\mathcal{P}$ from the camera with Eq. 2.3 and Eq. 2.4. In the following sections explain how to utilize the phantom discussed in Section 3.3 to calibrate the camera to the CBCT volume space. The calibration method is divided into three phases — data acquisition in both CBCT and camera spaces (Section 4.1), point clouds extraction (Section 4.2), and 3D-3D point clouds registration (Section 4.3).

4.1 Data Acquisition

The calibration phantom (Fig. 3.1b) is positioned with the laser guidance of C-arm, such that all the pipes are visible in the CBCT and the depth camera view. A CBCT volume scan is performed with simultaneous surface reconstruction by KinectFusion. After the scanning, we obtain the calibration phantom data in two different modalities — CBCT and surface reconstruction.

4.1.1 CBCT Volume Scan

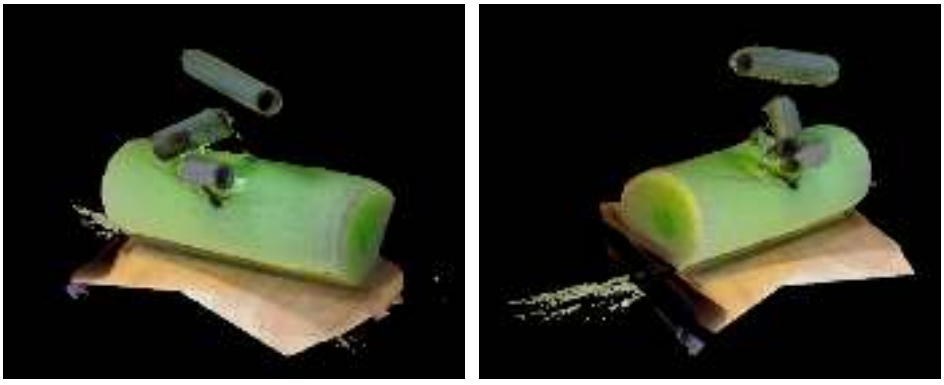


(a) (b)

Figure 4.1: (a) The mobile C-arm device, and (b) the CBCT scanning result visualized in ImFusion SDK.

A SIEMENS ARCADIS Orbic 3D mobile C-arm (Fig. 4.1a) is used to perform the CBCT volume scan. The CBCT scan involves 100 X-Ray images spread over 190 degree projection angles. The resulting volume size is $12cm^3$ and the volume is stored in DICOM format. Fig. 4.1b shows the CBCT volume scan result of the calibration phantom.

4.1.2 Surface Reconstruction with KinectFusion



(a) (b)

Figure 4.2: (a) and (b) show the surfaces at different angels reconstructed by KinectFusion.

During the scanning, simultaneous surface reconstruction is performed using KinectFusion discussed in Section 2.2. We modified the KinectFusion to suit our application. We defined the volume box to 30cm^3, which is large enough to cover the CBCT volume and also the foam base, which provide enough depth information for reliable tracking. $\mathbf{K}_{RGB}$, $\mathbf{K}_{Depth}$ and ${}^{Depth}\mathbf{T}_{RGB}$ are used to compute and merge the colored point clouds at each tracked position. Reconstructed calibration phantom surfaces are shown in Fig. 4.2 and it is stored in truncated signed distance function (TSDF) in the software memory.

4.2 Point Clouds Extraction

After acquiring the raw data — CBCT in DICOM format and TSDF in the software memory as mentioned in previous section, we can extract the surface point clouds of the pipes, which are in common in both spaces and therefore can be used for registration.

4.2.1 Point Clouds Extraction from CBCT Volume

(a) Simple thresholded CBCT Volume (b) Extracted Surface Mesh (c) Further thresholding by ambient occlusion (d) Extracted Point Clouds

Figure 4.3: Point Clouds Extraction from CBCT Volume.

Point clouds extraction from CBCT volume is performed in 4 steps. First, by design, the X-Ray attenuation index of the pipes are higher than other materials in use, therefore, a simple thresholding can be performed to segment the pipes as shown in Fig. 4.3a. Next, based on the segmentation result, a greedy triangulation algorithm [6] is used to generate the surface mesh of the pipes (Fig. 4.3b). After that, it is further thresholded based on the ambient occlusion values assigned to each vertex of the mesh grid so as to eliminate the pipe's inner surface (Fig. 4.3c). The last step is to extract all the vertices to produce the resulting surface point clouds as shown in Fig. 4.3d. The extracted point cloud $\mathcal{P}_{Raw}^{CBCT}$ will be further processed to eliminate the outliers as much as possible.

4.2.2 Point Clouds Extraction from Reconstructed Surface

(a) TSDF rendering of reconstructed surface (b) Extracted Colored Point Clouds (c) Extracted Point Clouds

Figure 4.4: Point Clouds Extraction from Reconstructed Surface.

On the other hand, since the surface reconstruction is stored in TSDF, it can also be discretized and be used to generate point cloud representation of the surface. An example of extracted point cloud with and without colors is shown in Fig. 4.4a and Fig. 4.4c. The extracted point cloud $\mathcal{P}_{Raw}^{Depth}$ will be further processed to extract the pipes that we are interested in, for the registration.

4.2.3 Data Pre-processing

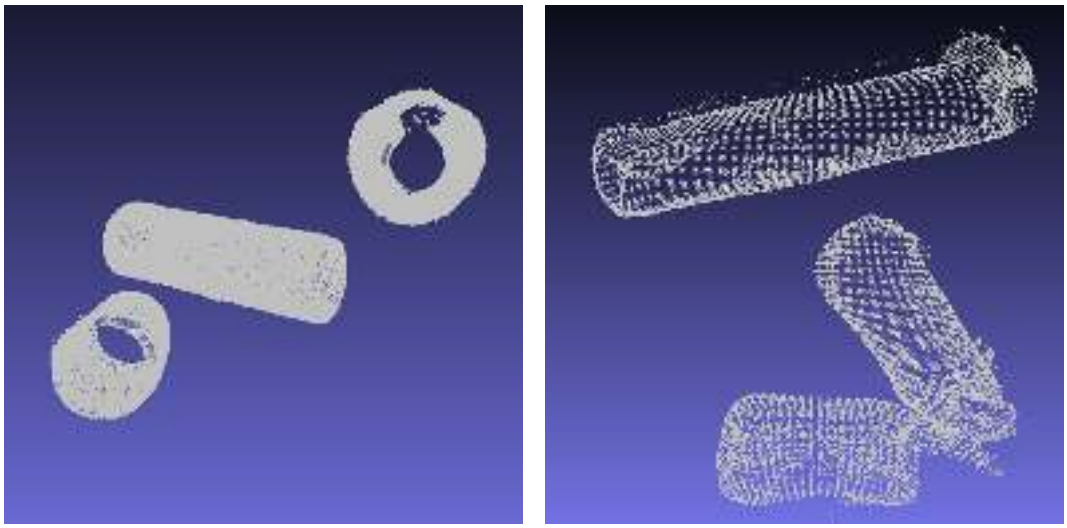
As shown in Fig. 4.3d and Fig. 4.4c, the raw point clouds $\mathcal{P}_{Raw}^{CBCT}$ and $\mathcal{P}_{Raw}^{Depth}$ contain outliers other than the pipes that we are interested in. One may manually select the pipes from the raw point clouds. Another automated method to extract the pipes from the raw point clouds is by cylinder fitting. As by design, we know the diameters and lengths of the pipes, we can perform least square cylinder filling for the three pipes.

Given a radius r and a length l of a cylinder, we can fit the cylinder in our raw point clouds by minimizing the residual function $R(.)$ with M-estimator SAmple and Consensus (MSAC) [31]:

$$E(\mathbf{c}, \mathbf{u}) = \min_{\mathbf{c},\mathbf{u}\in\mathbb{R}^3} \sum_{\substack{\mathbf{p}\in S_j, \\ |\mathbf{u}\cdot(\mathbf{p}_i-\mathbf{c})|\leq l/2}} R(\mathbf{p}, \mathbf{c}, \mathbf{u})^2 \tag{4.1}$$

where $\mathbf{c}, \mathbf{u} \in \mathbb{R}^3$ are the center and the orientation of the principle axis of the cylinder, S_j is the sampling set of the point cloud ($\mathcal{P}_{Raw}^{CBCT}$, $\mathcal{P}_{Raw}^{Depth}$) at the j-th MSAC iterations, and $R(\mathbf{p}, \mathbf{c}, \mathbf{u})$ is the residual function which measures the distance of a point p from the cylinder centered at $\mathbf{c}$ with orientation $\mathbf{u}$, defined as below, where $\mathbf{I}$ is the identity matrix in $\mathbb{R}^3$:

$$R(\mathbf{p}, \mathbf{c}, \mathbf{u})^2 = (\mathbf{p}-\mathbf{c})^\intercal(I - \mathbf{u}\mathbf{u}^\intercal)(\mathbf{p}-\mathbf{c}) - r^2 \tag{4.2}$$



(a) Final Extracted Point Clouds from CBCT Volume

(b) Final Extracted Point Clouds from Reconstructed Surface

Figure 4.5: Final Extracted Point Clouds from CBCT and depth camera spaces.

After fitting the cylinders in $\mathcal{P}_{Raw}^{CBCT}$ and $\mathcal{P}_{Raw}^{Depth}$, we can obtain the surface point cloud representations of the pipes in CBCT and depth camera spaces, denoted by $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$ and illustrated in Fig. 4.5, by extracting the fitted model's neighborhood with a distance d. i.e. Given a fitted cylinder model parameterized by $\mathbf{c}$ and $\mathbf{u}$, for a given point cloud $\mathcal{P}$, we can extract the neighborhood $\mathcal{P}(d)$:

$$\mathcal{P}(d) = \{\mathbf{p} \in \mathcal{P} | R(\mathbf{p}, \mathbf{c}, \mathbf{u}) < d\} \tag{4.3}$$

The cylinder fitting can be done in MatLab. $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$ are thus the concatenation of their three fitted cylinder models' neighborhoods. They will then be used for 3D-3D point cloud registration as mentioned in the next section.

4.3 Point Clouds to Point Clouds Registration

To register $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$ and to obtain the transformation from CBCT volume space to depth camera space ${}^{Depth}\mathbf{T}_{CBCT}$, we propose to use ICP described in Section 2.4, with a good initial guess obtained by SAmple Consensus Initial Alignment (SAC-IA) with 3D features as described in Section 2.3.

4.3.1 FPFH Initialization

(a) Two point clouds before alignment

(b) Two point clouds are initially aligned by SAC-IA with FPFH

(c) Two point clouds are registered by ICP

Figure 4.6: Two point clouds are registered using SAC-IA with FPFH as initialization and ICP as refinement.

Given $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$, we can compute its 3D features $\mathbf{FPFH}^{CBCT}$ and $\mathbf{FPFH}^{Depth}$ by Eq. 2.5 and Eq. 2.6. Using SAC-IA to perform iterative alignment between two point clouds gives us an initial estimation ${}^{Depth}\mathbf{T}^{\text{init}}_{CBCT}$ of the transformation from CBCT to depth camera. The result of the initial estimation is illustrated in Fig. 4.6b. Since the modalities are fundamentally different, the point cloud density, distribution and geometry of $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$ are different. It is challenging to register across different modalities. Nevertheless, using SAC-IA with FPFH features gives a good initial alignment with suitable parameter settings. Its result is then used as an initial guess for ICP for estimation refinement.

4.3.2 ICP Refinement

With a good initial guess ${}^{Depth}\mathbf{T}^{\text{init}}_{CBCT}$, the result can be improved by using ICP where the local minimal is generally the expected global solution. The final registration result (Fig. 4.6c) is thus obtained by:

$$ {}^{Depth}\mathbf{T}_{CBCT} = \min_{\mathbf{T},\mathcal{S}} \sum_{(i,j)\in\mathcal{S}} \|\mathbf{p}_i^{Depth} - \mathbf{T}\mathbf{p}_j^{CBCT}\|_2^2 \tag{4.4} $$

${}^{Depth}\mathbf{T}_{CBCT}$ is fixed once it is recovered and therefore can be used for intra-operative visualization between CBCT and depth camera data as described in the next chapter.

5 Mixed Reality Visualization

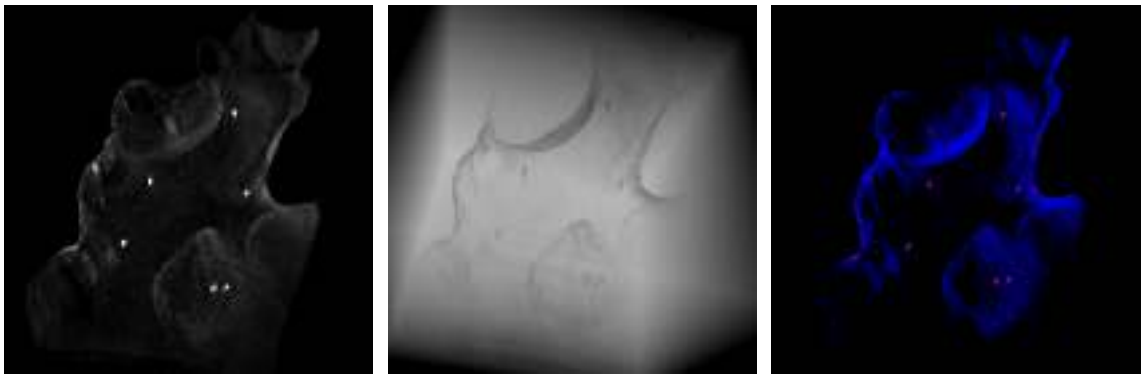
Contents

5.1 Intra-operative Visualization

This chapter describes the intra-operative visualization method in Fig.3.2. By the method described in Chapter 4, we recovered the transformation between CBCT and depth camera space ${}^{Depth}\mathbf{T}_{CBCT}$, which is fixed as the camera is rigidly mounted and is barely movable. After the patient data acquisition as described in Fig.3.2, we also have the reconstructed patient surface and corresponding CBCT volume data. Therefore, we can apply ${}^{Depth}\mathbf{T}_{CBCT}$ to overlay the surface and CBCT data for a better fused data visualization. Certain depth cues as discussed in Section 2.6 are applied to provide better depth perception, and DRR as described in Section 2.5 is used for generating synthetic X-Ray image from CBCT volume at any desired angle. All together, they provide a mixed reality visualization which aims to speed up the operation time, reduce radiation dose and make the entry point localization task easier. The components of the visualization are described in the following subsections.

5.1.1 Dynamic DRR and Fade Out Windows

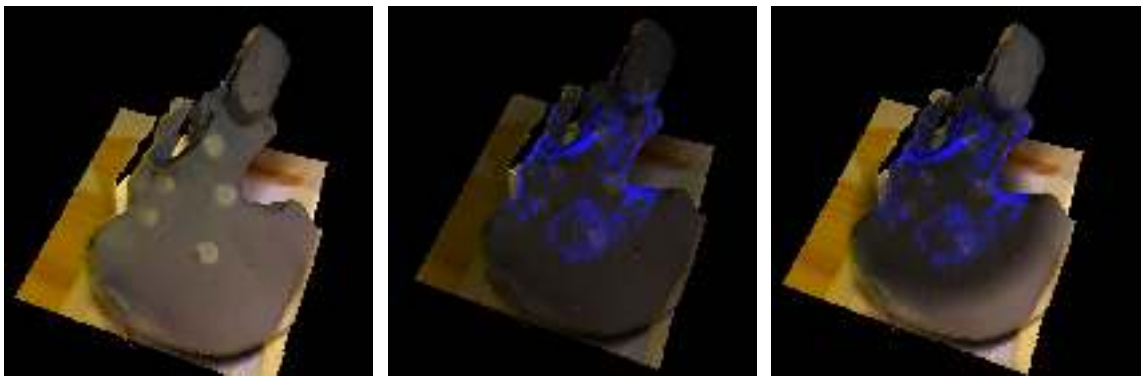
When overlaying/blending two objects, it is important to preserve their depth ordering. Otherwise, the resulting image will lead to depth confusion (e.g. the floating object on surface as illustrated in Fig. 2.6a). For a better depth perceptual visualization, we use DRR to generate synthetic X-Ray images with customizable transfer function for emphasizing the anatomy of interest and a fade out window to include occlusion effect between two objects which help in depth understanding.



(a) (b) (c)

Figure 5.1: Examples of synthetic X-Ray with different emphasis.

Fig. 5.1 shows DRR rendering with different transfer functions. Fig 5.1c demonstrates the feasibility to color different anatomy using DRR rendering technique, which makes the visualization highlight the important anatomy structures.



(a) (b) (c)

Figure 5.2: Examples of overlay using fade out windows.

Fig. 5.2a shows the rendering of the reconstructed surface, Fig. 5.2b and Fig. 5.2c demonstrates the effect of overlay with and without fade out windows. The fade out window helps the overlay better preserve the depth order between the surface and the DRR.

5.1.2 Multiple Views with Desired Angles

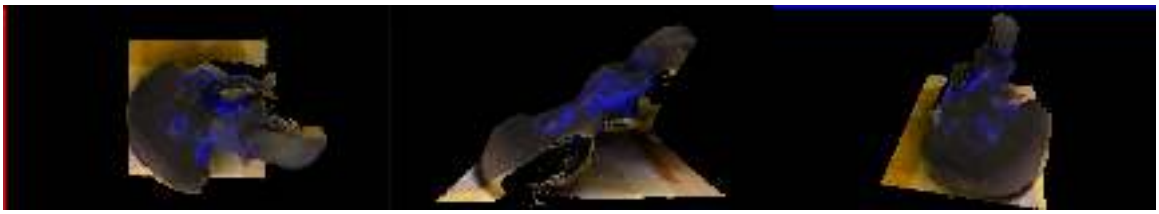


Figure 5.3: Example of Multiple Views of Mixed Reality Visualization - overlaying of CBCT volume and reconstructed surface with fade out windows and live flying point clouds for user interaction.

Another depth cue to improve depth perception visualization is motion parallax as mentioned in Section 2.6. We achieve this by providing multiple interactive views as illustrated in Fig. 5.3. The view angles are changeable with the mouse so that surgeons can change between multiple viewing angles.

5.1.3 Live Flying Point Clouds

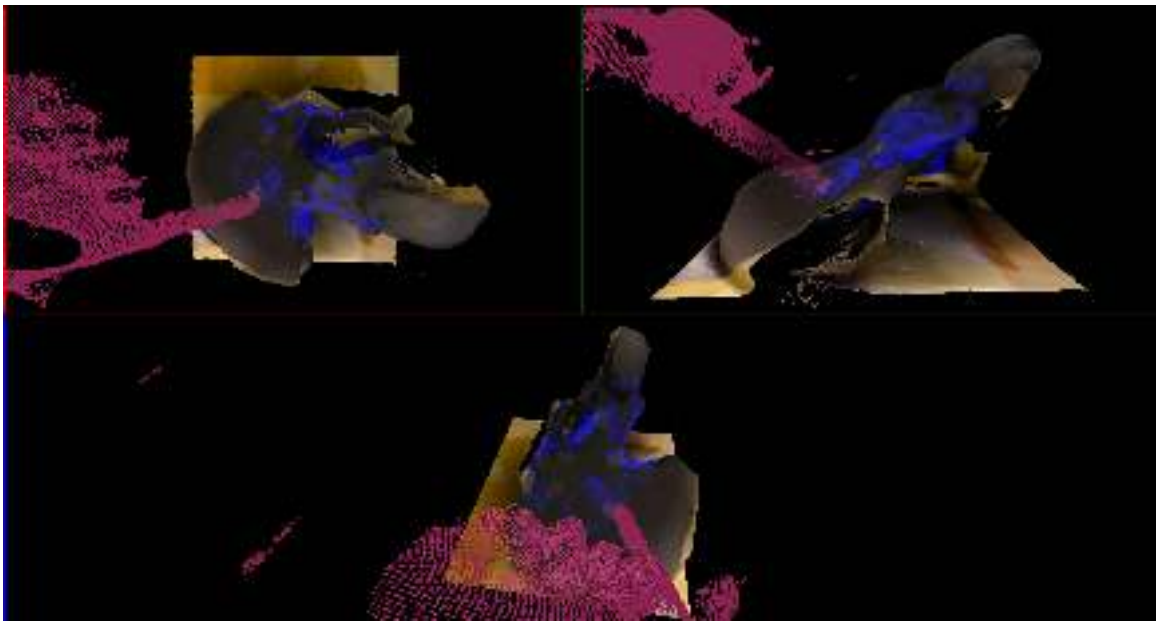
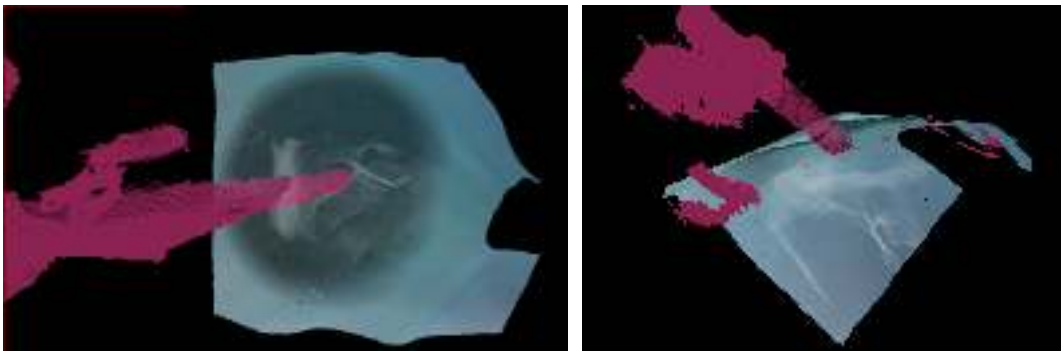


Figure 5.4: Example of Multiple Views of Mixed Reality Visualization - overlaying of CBCT volume and reconstructed surface with fade out windows and live flying point clouds for user interaction.

The last component of the visualization is live flying point clouds. Since the camera attached is calibrated to the C-arm and thus the visualization space, the live RGBD information can be utilized to generate live flying point clouds of the moving objects in the scene. To achieve this, a simple background subtraction is performed by subtracting a reference frame which is recorded at the beginning of the visualization. Combining all components, the system can provide a mixed visualization as shown in Fig. 5.4.

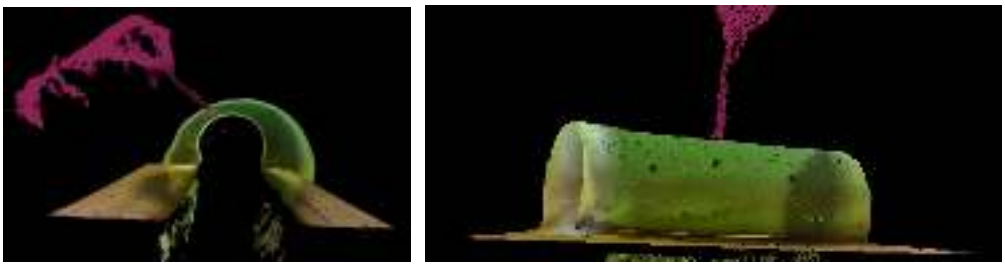
5.2 Applications

The system enables a real-time visualization of flying point clouds of the hands and/or surgical tools with overlaid on the patient surface and the DRR generated from CBCT. It is useful for orthopedic interventions such as guide wire placement and shrapnel removal. An illustration of using the system for guide wire placement and shrapnel removal is shown in Fig. 5.5 and Fig. 5.6 below.



(a) (b)

Figure 5.5: Demonstration of guide wire application.



(a) (b)

Figure 5.6: Demonstration of shrapnel removal application.

Part III: Experimental Assessment and Usability Evaluation

6 Experimental Assessment

Contents

Experiments are designed and conducted to assess the repeatability, accuracy and influence of noise, point cloud density, and choice of phantom. In this chapter, the experimental setup and results are discussed.

6.1 Repeatability Assessment

(a) Pose 1 (b) Pose 2 (c) Pose 3 (d) Pose 4 (e) Pose 5

Figure 6.1: Different poses for repeatablility test.

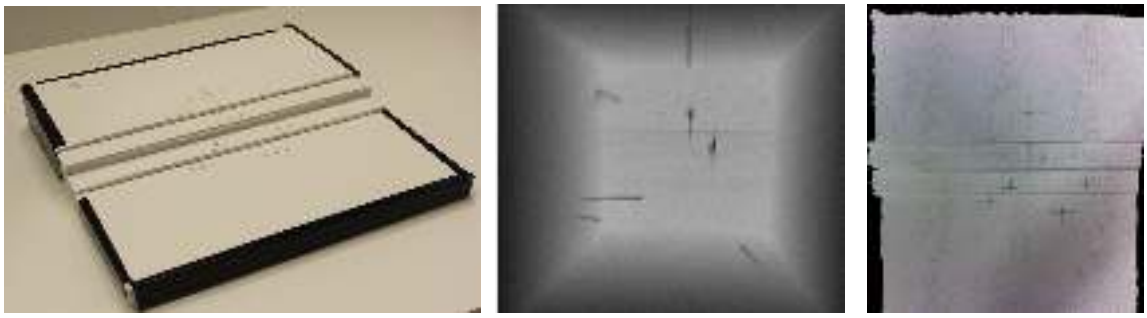
The calibration phantom in 3.1b is used in the repeatability test. The phantom is placed at different poses as shown in Fig.6.1, where the pipes are visible by both CBCT scan and RGBD camera. This gives 5 calibration results. The calibration results in terms of rotation Euler angles α, β, and γ, and the translation component x, y and z, and their standard deviation are shown in Table 6.1.

The results show the calibration result only has small deviation (< 1 mm) along x, y, and z axis and also small deviation in the rotation Euler angle α, β, and γ, which means the calibration method is repeatable for the phantom and the camera.

Table 6.1: The calibration results of the repeatability test in term of Euler angles α, β, γ of ${}^{Depth}\mathbf{R}_{CBCT}$, the x, y, z components of ${}^{Depth}\mathbf{t}_{CBCT}$, and $\|{}^{Depth}\mathbf{t}_{CBCT}\|_2$, of the calibration results ${}^{Depth}\mathbf{T}_{CBCT}$. The last row is their standard deviation (SD).

	α (rad)	β (rad)	γ (rad)	x (mm)	y (mm)	z (mm)	$\|{}^{Depth}\mathbf{t}_{CBCT}\|_2$
Pose 1	3.1125	-0.0538	0.9931	-21.9796	-52.9218	391.9107	396.0780
Pose 2	3.1095	-0.0601	1.0055	-20.7848	-52.5036	390.8201	394.8785
Pose 3	3.1089	-0.0696	0.9802	-19.4545	-52.1404	390.6211	394.5655
Pose 4	3.1050	-0.0588	0.9860	-20.6831	-52.4334	392.4686	396.4952
Pose 5	3.1172	-0.0633	0.9614	-20.6490	-52.6556	390.8562	394.9273
SD	0.0045	0.0059	0.0163	0.8942	0.2877	0.8093	0.8442

6.2 Accuracy Assessment



(a) TRE test phantom (b) CBCT volume scan (c) Surface reconstruction

Figure 6.2: Accuarcy test phantom, and its CBCT scan and surface reconstruction.

Next, we would like to assess the accuracy of the calibration method. Given the reasons stated in Section 2.7, we use TRE to measure the accuracy. An accuracy test phantom is designed with radio-opaque and optical landmarks which are uniformly, non-coplanarly and non-colinearly distributed. There are 8 landmarks in total. We position the phantom differently three times and acquired the CBCT volume scan surface reconstruction of it. The landmarks in both spaces are then manually extracted and Eq. 2.8 is used to compute the TRE. The landmark positions of the three tests and the TRE are shown in Table 6.2, Table 6.3 and Table 6.4.

Table 6.2: TRE Test 1: The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 1	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	50.7500	-41.2500	-5.3274	65.6164
Landmark 1 (Depth)	52.3190	-41.5090	-5.6389	67.0229
Landmark 1 ($\|\delta\|_2$)	1.5690	0.2590	0.3115	1.4066
Landmark 2 (CBCT)	28.2500	-40.4429	31.2500	58.3974
Landmark 2 (Depth)	30.2690	-37.8000	32.9990	58.6002
Landmark 2 ($\|\delta\|_2$)	2.0190	2.6429	1.7490	0.2029
Landmark 3 (CBCT)	21.2500	-35.2500	-5.6019	41.5392
Landmark 3 (Depth)	22.0490	-37.5730	-6.6189	44.0647
Landmark 3 ($\|\delta\|_2$)	0.7990	2.3230	1.0170	2.5255
Landmark 4 (CBCT)	8.2500	-35.7500	-16.5877	40.2651
Landmark 4 (Depth)	8.7223	-36.8970	-17.1460	41.6107
Landmark 4 ($\|\delta\|_2$)	0.4723	1.1470	0.5583	1.3456
Landmark 5 (CBCT)	-12.2500	-26.0343	-56.2500	63.1816
Landmark 5 (Depth)	-13.8740	-27.1660	-57.0170	64.6639
Landmark 5 ($\|\delta\|_2$)	1.6240	1.1317	0.7670	1.4823
Landmark 6 (CBCT)	-16.7500	-26.1449	16.2500	35.0454
Landmark 6 (Depth)	-17.2130	-22.1760	15.7950	32.2109
Landmark 6 ($\|\delta\|_2$)	0.4630	3.9689	0.4550	2.8345
Landmark 7 (CBCT)	-29.2500	-21.4330	37.7500	52.3450
Landmark 7 (Depth)	-28.4820	-17.5520	37.7030	50.4065
Landmark 7 ($\|\delta\|_2$)	0.7680	3.8810	0.0470	1.9385
Landmark 8 (CBCT)	-39.2500	-20.7500	-30.0412	53.6060
Landmark 8 (Depth)	-41.6030	-20.9480	-32.9870	57.0769
Landmark 8 ($\|\delta\|_2$)	2.3530	0.1980	2.9458	3.4709
TRE	**1.26 $\pm$ 0.73**	**1.94 $\pm$ 1.50**	**0.98 $\pm$ 0.95**	**2.91 $\pm$ 1.10**

Table 6.3: TRE Test 2: The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 2	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	1.6569	-41.2500	-53.2500	67.3786
Landmark 1 (Depth)	-0.6363	-44.8360	-54.8740	70.8649
Landmark 1 ($\|\delta\|_2$)	2.2932	3.5860	1.6240	3.4863
Landmark 2 (CBCT)	36.7500	-41.1212	-30.2500	62.9013
Landmark 2 (Depth)	36.9960	-43.0810	-32.0180	65.1907
Landmark 2 ($\|\delta\|_2$)	0.2460	1.9598	1.7680	2.2894
Landmark 3 (CBCT)	0.7500	-36.5727	-23.7500	43.6141
Landmark 3 (Depth)	0.2039	-40.7530	-26.5500	48.6390
Landmark 3 ($\|\delta\|_2$)	0.5461	4.1803	2.8000	5.0249
Landmark 4 (CBCT)	-10.2500	-36.5208	-10.7500	39.4258
Landmark 4 (Depth)	-10.5290	-37.2860	-11.4610	40.4037
Landmark 4 ($\|\delta\|_2$)	0.2790	0.7652	0.7110	0.9779
Landmark 5 (CBCT)	-49.2500	-23.2500	9.2596	55.2437
Landmark 5 (Depth)	-49.0520	-24.7550	8.5544	55.6065
Landmark 5 ($\|\delta\|_2$)	0.1980	1.5050	0.7052	0.3628
Landmark 6 (CBCT)	21.7500	-25.9073	15.7500	37.3137
Landmark 6 (Depth)	21.7730	-23.8270	15.5780	35.8394
Landmark 6 ($\|\delta\|_2$)	0.0230	2.0803	0.1720	1.4743
Landmark 7 (CBCT)	43.7500	-21.1250	27.2500	55.7036
Landmark 7 (Depth)	45.2180	-18.6510	26.7020	55.7272
Landmark 7 ($\|\delta\|_2$)	1.4680	2.4740	0.5480	0.0236
Landmark 8 (CBCT)	-25.2500	-20.7500	38.1079	50.2030
Landmark 8 (Depth)	-24.5800	-17.6260	37.5080	48.1840
Landmark 8 ($\|\delta\|_2$)	0.6700	3.1240	0.5999	2.0189
TRE	**0.72 $\pm$ 0.78**	**2.46 $\pm$ 1.12**	**1.12 $\pm$ 0.87**	**2.91 $\pm$ 1.37**

Table 6.4: TRE Test 3: The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 3	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	-0.7500	-41.2500	45.0440	61.0826
Landmark 1 (Depth)	0.5919	-39.9840	45.4780	60.5584
Landmark 1 ($\|\delta\|_2$)	1.3419	1.2660	0.4340	0.5242
Landmark 2 (CBCT)	-35.6489	-40.7500	21.7500	58.3478
Landmark 2 (Depth)	-36.0440	-40.4990	22.2150	58.5905
Landmark 2 ($\|\delta\|_2$)	0.3951	0.2510	0.4650	0.2427
Landmark 3 (CBCT)	1.2500	-36.2500	14.9365	39.2266
Landmark 3 (Depth)	0.8929	-36.1970	15.4430	39.3638
Landmark 3 ($\|\delta\|_2$)	0.3571	0.0530	0.5065	0.1372
Landmark 4 (CBCT)	12.7500	-35.7500	2.5535	38.0414
Landmark 4 (Depth)	11.8150	-36.7850	1.9105	38.6831
Landmark 4 ($\|\delta\|_2$)	0.9350	1.0350	0.6430	0.6417
Landmark 5 (CBCT)	51.2500	-26.4935	-17.2500	60.2165
Landmark 5 (Depth)	53.0000	-25.9950	-19.4260	62.1459
Landmark 5 ($\|\delta\|_2$)	1.7500	0.4985	2.1760	1.9293
Landmark 6 (CBCT)	-20.2500	-26.2500	-23.0600	40.3843
Landmark 6 (Depth)	-19.7520	-26.0130	-24.7810	40.9990
Landmark 6 ($\|\delta\|_2$)	0.4980	0.2370	1.7210	0.6147
Landmark 7 (CBCT)	-41.7500	-21.5209	-35.7500	59.0277
Landmark 7 (Depth)	-42.9480	-22.8340	-38.0930	61.7819
Landmark 7 ($\|\delta\|_2$)	1.1980	1.3131	2.3430	2.7541
Landmark 8 (CBCT)	27.7500	-21.2500	-45.7412	57.5663
Landmark 8 (Depth)	0.1450	1.1460	2.8748	2.7928
Landmark 8 ($\|\delta\|_2$)	0.1450	1.1460	2.8748	2.7928
TRE	**0.83 $\pm$ 0.57**	**0.72 $\pm$ 0.52**	**1.40 $\pm$ 1.00**	**1.92 $\pm$ 0.98**

The average TRE of the three tests shows the calibration method achieved an average accuracy of 2.58 mm. Between each test the phantom is rotated by 90 degrees. The landmarks are distributed uniformly, non-coplanarly and non-colinearly. It thus gives different distance between the landmarks and the depth camera in each test, which results in different errors. The setting that has more landmarks far away from the depth camera introduces more error. Therefore, the error mainly comes from the depth measurement. In orthopedic surgery, it is typically required to place a 2 to 7.3 mm wide guide wire or screw into a bone structure that varies in thickness between 5 mm (vertebra) and 12 mm (superior pubic ramus) in diameters. Therefore, the calibration in general achieve reasonable accuracy for

this intervention.

6.3 Influence of Noise and Point Cloud Densities

The calibration method involves data acquired from CBCT and surface reconstruction. By our phantom design, the CBCT data can be segmented with a simple thresholding which greatly prevents noisy data acquisition from CBCT volume scan; on the other thand, the surface reconstruction technique KinectFusion utilizes bilateral filtering as preprocessing step to eliminate noisy depth data. Therefore, the acquired data is in principle less noisy. Furthermore, the chosen methods SAC-IA with FPFH and ICP are both tolerant to outliers. As a result, the calibration result is also tolerant to outliers.

Typically, the point clouds acquired from the CBCT volume data and the surface reconstruction is dense, which means that it is computational expensive. Although for a one-time calibration method, computation efficiency is not a main issue, we conduct a test on the point cloud densities and see the feasibility of speeding up. The point clouds $\mathcal{P}^{CBCT}$ and $\mathcal{P}^{Depth}$ are downsampled using voxelgrid sampling algorithm with different grid size before performing the calibration. TRE is measured to evaluate the accuracy of the downsampled calibration results, which is reported in Table 6.5.

Table 6.5: The original data acquired from the CBCT volume and the surface reconstruction contains 94547 and 25226 points. The TRE for calibrations with downsampled point clouds are shown with mean $\pm$ standard deviation.

Grid Size (mm)	# of pts (CBCT)	# of pts (Depth)	δx	δy	δz	$\|\delta\|_2$
-	94547	25226	0.83 ± 0.57	0.72 ± 0.52	1.40 ± 1.00	1.92 ± 0.98
0.5	51831	21014	0.95 ± 0.46	0.74 ± 0.63	1.58 ± 1.25	2.18 ± 1.12
1.0	18684	16221	0.90 ± 0.59	0.82 ± 0.50	1.52 ± 1.32	2.19 ± 1.08
1.5	9016	8536	0.88 ± 0.52	0.95 ± 0.50	1.35 ± 0.83	2.05 ± 0.62
2.0	5238	5183	0.89 ± 0.79	0.70 ± 0.56	1.37 ± 1.06	1.96 ± 1.13

The results show small variation in the transformation results. Therefore, the method also achieve reasonable accuracy even though the point clouds are downsampled. However, when the point cloud density of both data sets is below grid size 2.5 mm (fewer than 3000 points), the initialization with FPFH failed to provide good estimation, and thus the calibration failed or require manual adjustment for the ICP initialization.

6.4 Influence of Camera and Phantom Choices

The last assessment we conducted is to compare the calibration results from different depth camera and phantoms. Similarly, we mounted a Microsoft Kinect 360 rigidly onto the C-

arm. It is rigidly mounted on the gantry instead of the detector because of the depth range limitation, which needs at least 50 cm apart. We perform the same repeatability test as mentioned in Section 6.1 with the Kinect 360 camera and compare it with the RealSense F200 camera. The results are summarized in Table 6.6 and Table 6.7.

Table 6.6: The calibration results of the repeatability test in term of Euler angles α, β, γ of ${}^{Depth}\mathbf{R}_{CBCT}$, the x, y, z components of ${}^{Depth}\mathbf{t}_{CBCT}$, and $\|{}^{Depth}\mathbf{t}_{CBCT}\|_2$, of the calibration results ${}^{Depth}\mathbf{T}_{CBCT}$.

	α (rad)	β (rad)	γ (rad)	x (mm)	y (mm)	z (mm)	$\|{}^{Depth}\mathbf{t}_{CBCT}\|_2$
Pose 1	-1.5663	-0.3262	1.5603	9.7328	-132.1384	739.4741	751.2504
Pose 2	-1.5634	-0.3233	1.5619	10.0696	-132.4561	739.5164	751.3524
Pose 3	-1.5627	-0.3278	1.5686	9.9723	-130.8103	738.7425	750.3008
Pose 4	-1.5591	-0.3276	1.5666	9.9238	-131.1044	739.1235	750.7266
Pose 5	-1.5540	-0.3217	1.5484	9.2921	-132.2286	739.5430	751.3285
SD	0.0047	0.0027	0.0079	0.3083	0.7380	0.3450	0.4632

Table 6.7: Comparsion between Intel RealSense F200 and Microsoft Kinect 360 - Repeatability Test in standard deviation

	α (rad)	β (rad)	γ (rad)	x (mm)	y (mm)	z (mm)	$\|{}^{c}\mathbf{t}_{CBCT}\|_2$
F200	0.0045	0.0059	0.0163	0.8942	0.2877	0.8093	0.8442
Kinect	0.0047	0.0027	0.0079	0.3083	0.7380	0.3450	0.4632

The results show that there is little variation between two cameras, which means the repeatability of the calibration method is independent of the camera choice.

Similarly, the same accuracy test, as mentioned in Section 6.2, is performed with the Kinect 360 camera. The results are summarized in Table 6.8, Table 6.9, and Table 6.10 and compared with the RealSense F200's in Table 6.11.

Table 6.8: TRE Test 1 (Kinect): The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 1	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	-0.7500	-38.7500	44.0664	58.6854
Landmark 1 (Depth)	-0.9072	-44.4080	35.9020	57.1126
Landmark 1 ($\|\delta\|_2$)	0.1572	5.6580	8.1644	1.5728
Landmark 2 (CBCT)	-35.2500	-38.7500	21.2868	56.5443
Landmark 2 (Depth)	-36.9600	-41.6930	15.6700	57.8783
Landmark 2 ($\|\delta\|_2$)	1.7100	2.9430	5.6168	1.3340
Landmark 3 (CBCT)	0.2500	-34.7500	14.9109	37.8148
Landmark 3 (Depth)	-0.2805	-40.1340	9.5371	41.2526
Landmark 3 ($\|\delta\|_2$)	0.5305	5.3840	5.3738	3.4377
Landmark 4 (CBCT)	10.8294	-36.2500	2.7500	37.9328
Landmark 4 (Depth)	10.1360	-40.1300	-3.3381	41.5247
Landmark 4 ($\|\delta\|_2$)	0.6934	3.8800	6.0881	3.5918
Landmark 5 (CBCT)	-20.2500	-26.7500	-23.8884	41.1859
Landmark 5 (Depth)	-20.4840	-27.6010	-27.8510	44.2390
Landmark 5 ($\|\delta\|_2$)	0.2340	0.8510	3.9626	3.0531
Landmark 6 (CBCT)	50.7500	-26.7500	-17.9049	60.0975
Landmark 6 (Depth)	51.7920	-27.3290	-25.2280	63.7631
Landmark 6 ($\|\delta\|_2$)	1.0420	0.5790	7.3231	3.6656
Landmark 7 (CBCT)	-42.7500	-20.7500	-36.4606	59.8957
Landmark 7 (Depth)	-42.6060	-23.1490	-39.2020	62.3534
Landmark 7 ($\|\delta\|_2$)	0.1440	2.3990	2.7414	2.4576
Landmark 8 (CBCT)	26.7500	-21.7500	-46.1628	57.6162
Landmark 8 (Depth)	26.6350	-23.0510	-51.7890	62.6328
Landmark 8 ($\|\delta\|_2$)	0.1150	1.3010	5.6262	5.0166
TRE	**0.57 $\pm$ 0.56**	**2.87 $\pm$ 1.97**	**5.61 $\pm$ 1.72**	**6.54 $\pm$ 1.10**

Table 6.9: TRE Test 2 (Kinect): The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 2	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	1.7500	-41.7500	-53.2027	67.6510
Landmark 1 (Depth)	-0.1362	-44.3050	-58.8210	73.6401
Landmark 1 ($\|\delta\|_2$)	1.8862	2.5550	5.6183	5.9891
Landmark 2 (CBCT)	37.7500	-40.7500	-30.0245	63.1435
Landmark 2 (Depth)	36.8100	-44.3810	-38.0310	69.0725
Landmark 2 ($\|\delta\|_2$)	0.9400	3.6310	8.0065	5.9290
Landmark 3 (CBCT)	1.8125	-36.2500	-23.2500	43.1035
Landmark 3 (Depth)	0.6876	-41.2130	-29.5180	50.6981
Landmark 3 ($\|\delta\|_2$)	1.1248	4.9630	6.2680	7.5946
Landmark 4 (CBCT)	-9.2500	-36.2500	-10.3379	38.8136
Landmark 4 (Depth)	-9.9854	-38.9670	-17.8820	44.0216
Landmark 4 ($\|\delta\|_2$)	0.7354	2.7170	7.5441	5.2080
Landmark 5 (CBCT)	-48.5786	-25.2500	10.2500	55.7001
Landmark 5 (Depth)	-47.7670	-28.2530	3.3104	55.5957
Landmark 5 ($\|\delta\|_2$)	0.8116	3.0030	6.9396	0.1045
Landmark 6 (CBCT)	22.2500	-26.0645	15.7500	37.7158
Landmark 6 (Depth)	21.2420	-29.4880	6.7674	36.9670
Landmark 6 ($\|\delta\|_2$)	1.0080	3.4235	8.9826	0.7488
Landmark 7 (CBCT)	44.9395	-21.2500	28.2500	57.1768
Landmark 7 (Depth)	44.4550	-23.6140	19.6380	54.0326
Landmark 7 ($\|\delta\|_2$)	0.4845	2.3640	8.6120	3.1442
Landmark 8 (CBCT)	-23.8566	-20.2500	38.2500	49.4193
Landmark 8 (Depth)	-25.6120	-23.1170	31.7980	46.9200
Landmark 8 ($\|\delta\|_2$)	1.7554	2.8670	6.4520	2.4993
TRE	**1.09 $\pm$ 0.49**	**3.19 $\pm$ 0.83**	**7.30 $\pm$ 1.19**	**8.11 $\pm$ 1.02**

Table 6.10: TRE Test 3 (Kinect): The positions of the landmarks x, y, and z, and their differences $\|.\|_2$ are shown. The last row shows the TRE shown as δx, δy, δz, and $\|\delta\|_2$ in Euclidean distances (mean $\pm$ standard deviation).

Test 3	x	y	z	$\|.\|_2$
Landmark 1 (CBCT)	49.7500	-41.6937	-3.7500	65.0191
Landmark 1 (Depth)	51.8570	-45.0490	-12.5240	69.8241
Landmark 1 ($\|\delta\|_2$)	2.1070	3.3553	8.7740	4.8050
Landmark 2 (CBCT)	26.2500	-40.4205	31.7500	57.7143
Landmark 2 (Depth)	27.4080	-43.3260	23.4220	56.3643
Landmark 2 ($\|\delta\|_2$)	1.1580	2.9055	8.3280	1.3500
Landmark 3 (CBCT)	20.2500	-36.5254	-4.2500	41.9789
Landmark 3 (Depth)	18.6250	-38.8700	-12.7810	44.9569
Landmark 3 ($\|\delta\|_2$)	1.6250	2.3446	8.5310	2.9780
Landmark 4 (CBCT)	8.2500	-36.2500	-15.9487	40.4535
Landmark 4 (Depth)	7.1992	-40.1190	-23.1830	46.8915
Landmark 4 ($\|\delta\|_2$)	1.0508	3.8690	7.2343	6.4380
Landmark 5 (CBCT)	-12.7500	-26.2500	-55.3691	62.5888
Landmark 5 (Depth)	-14.3640	-27.5470	-60.8890	68.3567
Landmark 5 ($\|\delta\|_2$)	1.6140	1.2970	5.5199	5.7678
Landmark 6 (CBCT)	-18.7500	-25.7500	14.8985	35.1652
Landmark 6 (Depth)	-17.1540	-26.0800	8.4261	32.3330
Landmark 6 ($\|\delta\|_2$)	1.5960	0.3300	6.4724	2.8321
Landmark 7 (CBCT)	-31.2500	-20.7500	36.4012	52.2702
Landmark 7 (Depth)	-29.2350	-21.6840	29.8170	47.0525
Landmark 7 ($\|\delta\|_2$)	2.0150	0.9340	6.5842	5.2177
Landmark 8 (CBCT)	-40.2500	-21.1774	-31.2500	55.1825
Landmark 8 (Depth)	-40.3190	-21.8690	-36.6750	58.7276
Landmark 8 ($\|\delta\|_2$)	0.0690	0.6916	5.4250	3.5451
TRE	**1.40 $\pm$ 0.65**	**1.97 $\pm$ 1.33**	**7.11 $\pm$ 1.33**	**7.60 $\pm$ 1.55**

Table 6.11: Comparision between Intel RealSense F200 and Microsoft Kinect 360 - Accuracy Test in TRE.

	δx	δy	δz	$\|\delta\|_2$
TRE 1 (F200)	1.26 ± 0.73	1.94 ± 1.50	0.98 ± 0.95	2.91 ± 1.10
TRE 1 (Kinect)	0.58 ± 0.56	2.87 ± 1.97	5.61 ± 1.72	6.54 ± 2.04
TRE 2 (F200)	0.72 ± 0.78	2.46 ± 1.12	1.12 ± 0.87	2.91 ± 1.37
TRE 2 (Kinect)	1.09 ± 0.49	3.19 ± 0.83	7.30 ± 1.19	8.11 ± 1.02
TRE 3 (F200)	0.83 ± 0.57	0.72 ± 0.52	1.40 ± 1.00	1.92 ± 0.98
TRE 3 (Kinect)	1.40 ± 0.65	1.97 ± 1.33	7.11 ± 1.33	7.60 ± 1.55

The results show that the error mainly comes from the z-axis. Therefore, the calibration quality indeed depends on the quality of the depth measurement provided by the camera. The further the object is, the more the error in depth measurement there is. Intel RealSense camera has an average accuracy of 2.58 mm while Kinect camera only achieve an averages accuracy of 7.42 mm.

(a) Pose 1 (b) Pose 2

Figure 6.3: Two arbitrary calibration phantoms.

Theoretically, arbitrary objects whose surface is visible in both CBCT volume and depth camera is suitable for the calibration. To examine and evaluate this hypothesis, two arbitrary objects — a stone and a spines phantom (Fig. 6.3) are used for evaluation. TRE is used to evaluate the accuracy of the calibration using arbitrary objects. The results are shown in Table 6.12 below.

Table 6.12: Compaision of calibration using arbitrary objects using TRE.

TRE	δx	δy	δz	$\|\|\delta\|\|_2$
Calibration Phantom	0.83 ± 0.57	0.72 ± 0.52	1.40 ± 1.00	1.92 ± 0.98
Spine	1.03 ± 0.47	1.53 ± 1.27	1.88 ± 1.14	2.87 ± 1.28
Stone	1.00 ± 0.57	3.35 ± 2.39	2.46 ± 1.35	4.77 ± 2.05

We found that arbitrary objects like the spine phantom which have unambiguous 3D shape and sufficient visibility in CBCT and depth camera can achieve reasonable accuracy of 2.87 mm, although the computation time is significantly longer due to the fact that the phantom contains more geometric structures and hence requires more points to represent. On the contrary, the stone object gives poor infrared reflectance, which yields poor depth information, and therefore results in poor calibration.

7 Usability Evaluation and Conclusion

Contents

In the last chapter, a pre-clinical study for evaluating the system and comparing with existing systems is presented. It is followed by a conclusion for summarizing the work, as well as future work for improving the system.

7.1 Pre-Clinical Study

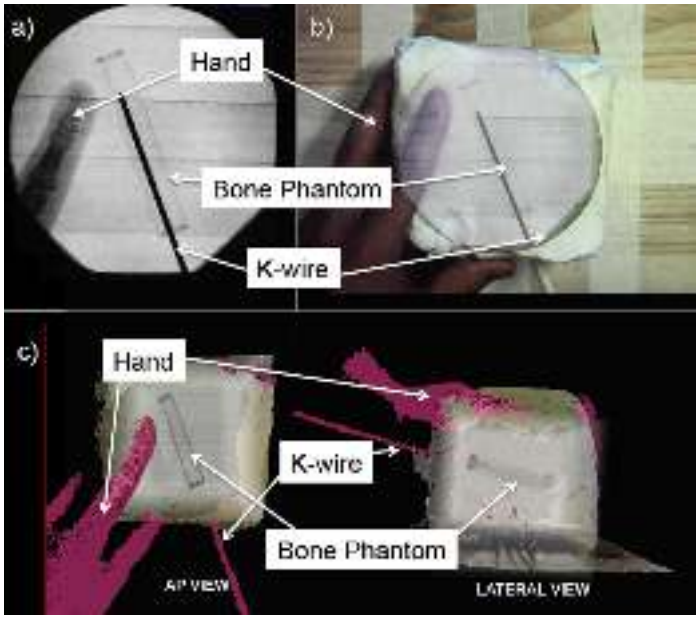


Figure 7.1: The three visualization compared in the preclinical study. (a) Conventional X-Ray systems, (b) Camera augmented mobile C-arm system, and (c) Mixed reality visualization system proposed in the thesis **Sources: images are taken from [10]**

A pre-clinical study comparing the system proposed in this thesis with the conventional X-Ray system and Camera augmented mobile C-arm system was conducted [10]. In the study, 7 surgeons were invited to perform a k-wire placement task using the three systems and the evaluation was measured by surgical efficiency (total time spent, number of X-Ray images acquired, total radiation dose, k-wire placement accuracy) and surgical task load (surgical task load index defined by [37])). Table 7.1 and Table 7.2 below summarizes all the measurements in the study for the three systems.

Table 7.1: Preclinical study evaluation results

	Participants						
	1	2	3	4	5	6	7
	S1: Conventional X-Ray system						
Total Time Spent (sec)	937	686	617	464	636	388	432
Number of X-Ray Images	80	47	44	33	32	21	29
Radiation Dose (cGycm^2)	7.68	1.73	3.54	4.38	5.62	2.69	5.38
Placement Error (mm)	3.08	7.88	11.43	3.01	1.87	2.27	2.72
Surgical Task Load	76	25.67	41.67	17.67	53.33	19.33	70.67
	S2: Camera augmented mobile C-arm system						
Total Time Spent (sec)	360	431	521	295	436	691	768
Number of X-Ray Images	19	13	20	13	18	20	30
Radiation Dose (cGycm^2)	3.07	1.3	1.57	1.92	1.42	2.38	5.56
Placement Error (mm)	7.92	2.69	3.85	4.23	4.88	3.44	1.74
Surgical Task Load	60.33	10	20	21.67	26	22.33	62.33
	S3: Mixed reality visualization system						
Total Time Spent (sec)	182	180	380	181	190	254	339
Number of X-Ray Images	1	2	2	2	2	3	3
Radiation Dose (cGycm^2)	1.76	1.9	1.48	1.44	1.55	1.47	1.59
Placement Error (mm)	7.38	6.39	8.45	6.53	1.39	2.31	3.48
Surgical Task Load	20.33	5	24.33	23	11.33	8.67	30.33

Table 7.2: The standard deviation of Table 7.1.

	S1	S2	S3
Total Time Spent (sec)	594 ± 188	500 ± 172	243 ± 84
Number of X-Ray Images	40.86 ± 19.38	19.00 ± 5.72	2.14 ± 0.69
Radiation Dose (cGycm^2)	4.43 ± 2.00	2.46 ± 1.50	1.60 ± 0.17
Placement Error (mm)	4.61 ± 3.62	4.11 ± 1.97	5.13 ± 2.72
Surgical Task Load	43.48 ± 24.03	31.81 ± 20.76	17.57 ± 9.33

The evaluation shows that the system (S3) significantly speeds up the total time spent, and significantly reduces the number of X-Ray image required, compared to S1 and S2. Furthermore, S3 and S2 significantly reduces the radiation dose compared to S1; while there is no significant difference between S2 and S3 found in terms of radiation dose. Last but not least, there is no significant difference in placement error among the three systems. On the other hand, the surgical task load also shows a significant difference between S1 and S3, which indicates that the proposed system significantly helps to reduce the surgical task load. Therefore, the evaluation supports that the purposed system has clear advantages over the existing systems while maintaining necessary placement accuracy.

7.2 Conclusion

In this thesis, we explored the state of the art of visualization system for orthopedic and trauma surgery, from the 2D-2D augmented reality visualization, to 2D-3D/3D-2D augmented reality visualization. We are amazed by the radiation exposure reduction by using augmented reality for intra-operative visualization, and realize the current difficulty of depth perception in medical images in the OR. Therefore, we proposed a new system, which provides 3D-3D mixed reality visualization for the interventions.

The calibration makes use of KinectFusion for a reliable surface reconstruction during CBCT scan, calibration phantom surface point clouds extraction, and point clouds to point clouds registration by SAC-IA with FPFH and ICP. Our experiment results showed the method is repeatable and the system achieved reasonable accuracy of 2.58 mm. With the calibration result, the system provides multiple interactive views at any desired angles, which gives a mixed reality visualization of the patient surface, the DRR of CBCT volume, and flying point clouds of moving objects. A pre-clinical study was conducted to evaluate the system compared to conventional X-Ray system and CAMC system. The study showed our system reduces the radiation dose, and more importantly significantly reduces the total operation time and decreases the surgical workload.

In a nutshell, we stepped further from the literature, from 2D/3D-3D/2D visualization to 3D/3D visualization. We proposed a method to calibrate two 3D imaging systems and utilized depth cues to provide better depth perceptual visualization. The method is repeatable and the system has clear advantage over existing systems.

7.3 Future Work

Although the proposed system enhanced the intervention in terms of time, radiation dose and surgical workload, there are certain limitations, which can be further improved. In this last section, the limitations and further improvements are discussed.

The camera is calibrated to the CBCT volume when the C-arm is in its upright position. When the C-arm is moved, the visualization is out-dated. To compensate, patient or C-arm tracking is needed, so that the system is able to update the visualization based on the

new patient or C-arm position. There are many methods to achieve it, such as marker-base tracking, which attaches a marker onto patient surface and track it with the camera. Another way is marker-less tracking, which uses the patient surface obtained by the depth camera to perform frame-by-frame ICP-base tracking. More sophisticated method to combine RGB and depth information can be used to provide marker-less patient tracking.

In this thesis, the phantom design is only a proof of concept. A more advanced phantom, which has more reliable features in CBCT and in depth camera, can be designed to achieve better accuracy. For example, a model with fiducial markers visible in CBCT with a known surface can be designed, which can give accurate marker positions in CBCT while its surface can be reconstructed by depth camera and matched with the model surface for getting the known marker positions, and hence perform paired points registration for a better calibration. By this means, it reduces the error in surface segmentation in CBCT. Furthermore, it would be nice if the method can not only recover the transformation between CBCT and depth camera, but also all the CBCT projections relative to the depth camera position. It can also be achieved by a better phantom design. Such a phantom can be used to validate or re-verify the calibration result, which is required in a clinical practice.

On the other hand, depth camera based tool tracking can be used for augmenting the surgical tool into the scene, so that the system can also serve as a navigation guiding system. Furthermore, with tool tracking, the visualization can be further improved. For example, the interactive 3D visualization technique in [36] can be utilized to display an interactive X-Ray image according to the current tool position inside the body, which can help for the tool navigation.

7.4 Publications

The system presented in this master thesis was presented and evaluated in the following journal papers, accepted at the "Information Processing for Computer Assisted Interventions (IPCAI)", Heidelberg, Germany, June 2016 and published in "International Journal of Computer Assisted Radiology and Surgery" by Springer, March and April 2016.

Marius Fischer, Bernhard Fuerst, **Sing Chun Lee**, Javad Fotouhi, Severine Habert, Simon Weidert, Ekkehard Euler, Greg Osgood, and Nassir Navab: "Preclinical usability study of multiple augmented reality concepts for K-wire placement". *International Journal of Computer Assisted Radiology and Surgery*, 2016, doi:"10.1007/s11548-016-1363-x"

Sing Chun Lee, Bernhard Fuerst, Javad Fotouhi, Marius Fischer, Greg Osgood, and Nassir Navab: "Calibration of RGBD camera and cone-beam CT for 3D intra-operative mixed reality visualization", *International Journal of Computer Assisted Radiology and Surgery*, 2016, doi:"10.1007/s11548-016-1396-1"

And the extended work with tool tracking was submitted to International Conference

on Medical Image Computing and Computer Assisted Interventions (MICCAI), Athens, Greece, October 2016.

Bibliography

[1] M. Abdellah, A. Eldeib, and M. I. Owis. Gpu acceleration for digitally reconstructed radiographs using bindless texture objects and cuda/opengl interoperability. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 4242–4245, Aug 2015.

[2] ACVS. American college of veterinary surgeons. https://www.acvs.org/sites/default/files/images/Hip%20Dysplasia%20Figure%204.jpg. Online; accessed 19-March-2016.

[3] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 129–138, Nov 2007.

[4] Christoph Bichlmeier, Tobias Sielhorst, Sandro M. Heining, and Nassir Navab. *Bildverarbeitung für die Medizin 2007: Algorithmen – Systeme – Anwendungen Proceedings des Workshops vom 25.–27. März 2007 in München*, chapter Improving Depth Perception in Medical AR, pages 217–221. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[5] J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: the integration, relative potency and contextual use of different information about depth. In W. Epstein and S. Rogers, editors, *Handbook of perception and Cognition.*, volume 5: Perception of Space and Motion, pages 69–117. 1995.

[6] Matthew T. Dickerson, Robert L. Scot Drysdale, Scott A. McElfresh, and Emo Welzl. Fast greedy triangulation algorithms. In *Proceedings of the Tenth Annual Symposium on Computational Geometry*, SCG '94, pages 211–220, New York, NY, USA, 1994. ACM.

[7] Benoit Diotte, Pascal Fallavollita, Lejing Wang, Simon Weidert, Peter-Helmut Thaller, Ekkehard Euler, and Nassir Navab. Radiation-free drill guidance in interlocking of intramedullary nails. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 18–25. Springer, 2012.

[8] O. M. Dorgham, S. D. Laycock, and M. H. Fisher. Gpu accelerated generation of digitally reconstructed radiographs for 2-d/3-d image registration. *IEEE Transactions on Biomedical Engineering*, 59(9):2594–2603, Sept 2012.

[9] Daniel Marx Emil Krén. Web gallery of art. http://www.wga.hu/preview/l/leonardo/10anatom/8torso2.jpg. Online; accessed 19-March-2016.

[10] Marius Fischer, Bernhard Fuerst, Sing Chun Lee, Javad Fotouhi, Severine Habert, Simon Weidert, Ekkehard Euler, Greg Osgood, and Nassir Navab. Preclinical usability study of multiple augmented reality concepts for k-wire placement. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–8, 2016.

[11] S. Foix, G. Alenya, and C. Torras. Lock-in time-of-flight (tof) cameras: A survey. *Sensors Journal, IEEE*, 11(9):1917–1926, Sept 2011.

[12] Severine Habert, Jose Gardiazabal, Pascal Fallavollita, and Nassir Navab. RGBDX: first design and experimental validation of a mirror-based RGBD Xray imaging system. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, Fukuoka, October 2015.

[13] Severine Habert, Ma Meng, Wadim Kehl, Xiang Wang, Federico Tombari, Pascal Fallavollita, and Nassir Navab. Augmenting mobile c-arm fluoroscopes via stereo-rgbd sensors for multimodal visualization. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, Fukuoka, October 2015.

[14] Vital Element. Inc. Sovereign health imaging. http://www.imagingcenteratcarlstadt.com/images/pet-ct-images.jpg. Online; accessed 19-March-2016.

[15] Felix Matthews, Dominik J Hoigne, Manfred Weiser, Guido A Wanner, Pietro Regazzoni, Norbert Suhm, and Peter Messmer. 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(10):687–692, 2007.

[16] C.R. Maurer, J.J. McCrory, and J.M. Fitzpatrick. Estimation of accuracy in localizing externally attached markers in multimodal volume head images. *SPIE*, 1898:43–54, 1993.

[17] Philippe Merloz, Jocelyne Troccaz, Hervé Vouaillat, Christian Vasile, Jérôme Tonetti, Ahmad Eid, and Stéphane Plaweski. Fluoroscopy-based navigation system in spine surgery. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 221(7):813–820, 2007.

[18] Mehdi Hedjazi Moghari, Burton Ma, and Purang Abolmaesumi. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part II*, chapter A Theoretical Comparison of Different Target Registration Error Estimators, pages 1032–1040. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[19] Nassir Navab, Tobias Blum, Lejing Wang, Aslı Okur, and Thomas Wendler. First deployments of augmented reality in operating rooms. *Computer*, (7):48–55, 2012.

[20] Nassir Navab, Sandro-Michael Heining, and Joerg Traub. Camera augmented mobile c-arm (camc): calibration, accuracy study, and clinical applications. *Medical Imaging, IEEE Transactions on*, 29(7):1412–1423, 2010.

[21] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136, Oct 2011.

[22] PCL. Fast point feature histograms (fpfh) descriptors. http://pointclouds.org/documentation/tutorials/_images/fpfh_diagram.png. Online; accessed 03-April-2016.

[23] PCL. Point feature histograms (pfh) descriptors. http://pointclouds.org/documentation/tutorials/_images/pfh_diagram.png. Online; accessed 03-April-2016.

[24] James A Pfaff and Gregory P Moore. Reducing risk in emergency department wound management. *Emergency medicine clinics of North America*, 25(1):189–201, 2007.

[25] S Reaungamornrat, Yoshito Otake, Ali Uneri, Sebastian Schafer, DJ Mirota, Sajendra Nithiananthan, J Webster Stayman, Gerhard Kleinszig, A Jay Khanna, Russell H Taylor, and Jeffrey H Siewerdsen. An on-board surgical tracking and video augmentation system for c-arm image guidance. *International journal of computer assisted radiology and surgery*, 7(5):647–665, 2012.

[26] Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 1–4. IEEE, 2011.

[27] Radu Bogdan Rusu, Zoltan Csaba Marton, Nico Blodow, Michael Beetz, Intelligent Autonomous Systems, and Technische Universität München. Persistent point feature histograms for 3d point clouds. In *In Proceedings of the 10th International Conference on Intelligent Autonomous Systems (IAS-10*, 2008.

[28] R.B. Rusu, N. Blodow, and M. Beetz. Fast Point Feature Histograms (FPFH) for 3D registration. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 3212–3217, 2009.

[29] AJ Starr, AL Jones, CM Reinert, and DS Borer. Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the actabulum. *Injury*, 32:45–50, 2001.

[30] E. Stoykova, A. A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis. 3-d time-varying scene capture technologies: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11):1568–1586, Nov 2007.

[31] P.H.S. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138 – 156, 2000.

[32] Joerg Traub, Tim Hauke Heibel, Philipp Dressel, Sandro Michael Heining, Rainer Graumann, and Nassir Navab. A multi-view Opto-Xray imaging system: development and first application in trauma surgery. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 10(Pt 2):18–25, 2007.

[33] Everine B van de Kraats, Theo van Walsum, Lance Kendrick, Niels J Noordhoek, and Wiro J Niessen. Accuracy evaluation of direct navigation with an isocentric 3d rotational x-ray system. *Medical image analysis*, 10(2):113–124, 2006.

[34] E. Wahl, U. Hillenbrand, and G. Hirzinger. Surflet-pair-relation histograms: a statistical 3d-shape representation for rapid classification. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 474–481, Oct 2003.

[35] Xiang Wang, Severine Habert, Ma Meng, Xiang Wang, Chun-Hao Huang, Pascal Fallavollita, and Nassir Navab. Rgb-d/c-arm calibration and application in medical augmented reality. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, Fukuoka, Japan, October 2015.

[36] Matthias Wieczorek, André Aichert, Pascal Fallavollita, Oliver Kutter, Ahmad Ahmadi, Lejing Wang, and Nassir Navab. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011: 14th International Conference, Toronto, Canada, September 18-22, 2011, Proceedings, Part I*, chapter Interactive 3D Visualization of a Single-View X-Ray Image, pages 73–80. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[37] Mark R. Wilson, Jamie M. Poolton, Neha Malhotra, Karen Ngo, Elizabeth Bright, and Rich S. W. Masters. Development and validation of a surgical workload measure: The surgery task load index (surg-tlx). *World Journal of Surgery*, 35(9):1961–1969, 2011.

[38] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, November 2000.

List of Figures

List of Tables