Informatics

Real-time motion correction in CBCT reconstructions using a mobile CBCT device

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medieninformatik und Visual Computing

eingereicht von

Felix Ginzinger, BSc
Matrikelnummer 01527276

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: PD Dipl.-Ing. Dr.techn. Martin Kampel

Wien, 13. Jänner 2024

Felix Ginzinger | Martin Kampel

Informatics

Real-time motion correction in CBCT reconstructions using a mobile CBCT device

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics and Visual Computing

by

Felix Ginzinger, BSc
Registration Number 01527276

to the Faculty of Informatics

at the TU Wien

Advisor: PD Dipl.-Ing. Dr.techn. Martin Kampel

Vienna, 13th January, 2024

Felix Ginzinger Martin Kampel

Erklärung zur Verfassung der Arbeit

Felix Ginzinger, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 13. Jänner 2024

Felix Ginzinger

Danksagung

Diese akademische Reise anzutreten und diese Masterarbeit abzuschließen war eine lohnende und bereichernde Erfahrung, und ich bin all jenen sehr dankbar, die bei ihrer Realisierung eine entscheidende Rolle gespielt haben.

In erster Linie möchte ich mich bei meinem Betreuer Martin Kampel bedanken, der mit seinem aufschlussreichen Feedback und seiner konstruktiven Kritik die akademische Qulität dieser Arbeit maßgeblich geprägt hat. Noch wichtiger war jedoch seine nahezu unendliche Geduld, seine pragmatische und angenehme Art, die entscheidend dazu beitrugen, die komplizierte Reise zur Fertigstellung dieser Arbeit zu meistern.

Ich bin Phil Steininger zu tiefstem Dank verpflichtet, dessen Anleitung, Fachwissen und unerschütterliche Unterstützung für dieses Forschungsvorhaben entscheidend waren. Dein Mentorat hat nicht nur die grßartigen Ergebnisse dieser Arbeit geprägt, sondern auch wesentlich zu meinem persönlichen und beruflichen Wachstum beigetragen.

Darüber hinaus möchte ich Heinz Deutschman für die großartige Unterstützung danken und dafür, dass er mir die Möglichkeit gegeben hat, diese Arbeit zusammen mit medPhoton zu schreiben. Außerdem bin dankbar für die Unterstützung durch meine Kollegen bei medPhoton, die eine Schlüsselrolle in den einzelnen Entwicklungschritten gespielt haben.

Acknowledgements

Embarking on this academic journey and completing this master's thesis has been a rewarding and enriching experience, and I am deeply grateful to those who have played pivotal roles in its realization.

First and foremost, I extend my sincere gratitude to my thesis advisor, Martin Kampel, who has not only shaped the academic rigor of this thesis with his insightful feedback and constructive critique, which have undoubtedly elevated the quality of this work significantly. But even more important was his near-endless patience, pragmatical, and pleasant nature, which were pivotal in navigating the intricate journey of completing this thesis.

I extend my deepest appreciation to Phil Steininger, whose guidance, expertise, and unwavering support have been instrumental for this research endeavor. Your mentorship has not only shaped the great results of this thesis but has also contributed significantly to my personal and professional growth.

Furthermore, I want to thank Heinz Deutschman for the great support and for providing me with the possibility to do this thesis together with medPhoton. I am grateful for all the support from my colleagues at medPhoton, who played a key role in the development process.

Kurzfassung

Cone-Beam-Computertomographie (CBCT) ist eine leistungsstarke medizinische Bildgebungstechnik, jedoch sehr anfällig für Bewegungsartefakten. Unfreiwillige Patientenbewegungen wie das Atmen können zu unscharfen und verzerrten Bildern führen, wodurch auch die diagnostische Genauigkeit beeinträchtigt wird. Die Bewältigung dieser Herausforderung erfordert ein Echtzeit-Bewegungskompensationssystem, das in der Lage ist, sowohl Patienten- als auch systeminduzierte Bewegungen während CBCT-Scans zu erkennen und zu korrigieren.

In dieser Masterarbeit wird die Entwicklung und Implementierung eines Stereo-Vision-Systems zur Erkennung und Kompensation von Bewegungen in CBCT-Scans vorgestellt. Die Arbeit beginnt mit der Untersuchung verschiedener Kameratechnologien und identifiziert Infrarot (IR)-Bildgebung in Kombination mit IR-Markern am Patienten als optimale Lösung. Dieser Ansatz bietet Robustheit, hohe Zuverlässigkeit und Echtzeitfähigkeit.

Das vorgeschlagene System erreicht die Echtzeitsynchronisation von getrackten IR Markern, die mit bis zu 240 Hz erfasst werden, mit CT-Projektionen. Diese synchronisierten Daten fügen sich nahtlos in die CBCT-Rekonstruktionspipeline ein und ermöglichen eine bewegungskompensierte Volumenrekonstruktion, wodurch die Notwendigkeit einer weiteren Aufnahme entfällt und so die Strahlenexposition des Patienten reduziert wird.

Die Evaluierungsphase zeigt die Fähigkeit des Systems, in verschiedenen Testszenarien scharfe Bildrekonstruktionen auch bei großen Bewegungsamplituden zu erzeugen. Experimente mit dem CatPhan-Phantom bestätigen, dass das System trotz erheblicher Bewegung während der Aufnahme eine Bildqualität liefert, die mit Scans ohne Bewegung vergleichbar ist.

Darüber hinaus zeigt eine neuartige geometrische Evaluierungsmethode, die Systembewegung und Deep-Learning-Segmentierung einsetzt, eine Trackinggenauigkeit im Submillimeterbereich. Diese Ergebnisse zeigen das hohe Potential des Systems und erweitern den Anwednungsbereich weit über den der Patientenbewegungskompensation hinaus.

Abstract

Cone-beam computed tomography (CBCT) is a powerful medical imaging technique, but it is prone to motion that can compromise image quality and diagnostic accuracy. Involuntary patient motions such as breathing can result in blurry and distorted images. Addressing this challenge requires a real-time motion compensation system capable of detecting and correcting both patient and system-induced motion during CBCT scans.

This master's thesis introduces the development and implementation of a stereo vision system designed to detect and compensate for motion in CBCT acquisitions. The study begins by investigating various camera technologies and identifies infrared (IR) imaging paired with IR markers on the patient as the optimal solution. This approach offers robustness, high reliability, and real-time capabilities.

The proposed system achieves real-time synchronization of tracking data, captured at up to 240Hz, with CT projections. This synchronized data seamlessly integrates into the CBCT reconstruction pipeline, enabling motion-compensated volume reconstruction without the need for reacquisition, thus minimizing patient radiation exposure.

The evaluation phase demonstrates the system's ability to produce sharp image reconstructions across diverse setups, even in the presence of substantial motion amplitudes. Experiments conducted with the CatPhan phantom confirm that the system maintains overall image quality comparable to scans without motion, despite significant motion during acquisition.

Furthermore, a novel geometric evaluation method, employing system motion and deep learning segmentation, reveals sub-millimeter tracking accuracy. This assessment enhances the system's reliability and unveils its potential clinical applications, extending far beyond patient motion compensation to encompass various medical imaging challenges.

Contents

CHAPTER 1

Introduction

1.1 Motivation

Due to the deliberate slow rotation pace of Cone Beam Computed Tomography (CBCT) devices, the potential for patient motion introduces a significant challenge to maintaining image quality. Historically, most CBCT imaging systems, including the IRm, necessitate strict patient immobilization during the acquisition process [WPS+19]. Nonetheless, given the typical scan duration ranging from 30 seconds to a minute, ensuring absolute patient stillness can be an arduous task.

A main application of IRm lies in spine surgery, where anesthesiologists must meticulously control patient respiration to obtain optimal image quality during scanning. Regrettably, proper breath inhibition is not consistently achieved, resulting in motion artifacts that variably impact different regions of the acquired images. Subsequently, surgeons are faced with the dilemma of either repeating the scan, incurring additional patient radiation dose, or proceeding with less-than-ideal images, thereby elevating the risk of procedural errors.

The field of spine surgery underscores the critical significance of consistently delivering dependable image quality. Surgeons frequently rely on precise 3D volumetric data, especially when operating on the spine or placing screws. Errors during spinal procedures can yield dire consequences for patients. The increasing number of image quality complaints at medPhoton, attributed to motion artifacts stemming from inadequate patient preparation prior to scanning, underscores the pressing need to develop a solution that accommodates patient motion during the scan.

Implementing such a solution would not only alleviate the burdens on surgical staff but also mitigate the peril of procedural mishaps.

1.2 Problem Statement

The issue of motion during CBCT acquisitions is recognized as a prominent factor contributing to the emergence of imaging artifacts within CBCT imaging [KNKF11] [KMBS21] [SNCS+18] [SHG+11] [SF08] [SNKH+20]. While conventional strategies aim to mitigate or halt patient motion, such as breath-holding techniques, numerous diverse approaches have been proposed to address this challenge [WPS+19]. Among these, two prevailing methodologies stand out: those leveraging CBCT image data directly for motion compensation and those relying on supplementary camera systems.

Image-based solutions necessitate pre-existing image data, often in the form of CT scans, or computationally intensive calculations prior to CBCT acquisition. However, these approaches frequently encounter limitations with larger ranges of motion, particularly affecting image-based solutions. Their advantage lies in their superior accuracy and absence of additional hardware requirements [BUG+17] [KBM+18] [BMA+16] [MBC+15].

Conversely, supplementary camera systems can detect motion concurrently with CBCT acquisitions, offering potential improvements in reconstruction speed due to their capacity to recognize motion parallel to imaging. However, these methods tend to exhibit reduced accuracy, necessitate external equipment within the surgical environment, and are susceptible to room-specific variables like lighting conditions and draped surfaces, introducing challenges related to robustness and dependability [BRU+18] [FFWN17] [KNKF11] [MNH+19].

Although solutions for motion compensation in CBCT imaging already exist, none of them meet the comprehensive requirements for both accuracy and speed necessary to seamlessly integrate with the IRm system. This gap in the existing landscape underscores the need for a tailored solution to ensure optimal performance with the IRm.

1.3 Aim of this Work

This work aims to conceptualize and develop a cutting-edge system that can swiftly identify and rectify patient motion in real-time during the reconstruction phase of Cone Beam Computed Tomography. Moreover, this system seamlessly integrates stereo-vision technology into an existing CBCT apparatus, rendering it adaptable for installation onto current hardware configurations. Additionally, the system will be harmoniously compatible with the software infrastructure of the device, enabling efficient processing of motion-related data.

The following high-level requirements delineate the system's scope:

- The system seamlessly functions within the IRm's gantry, independent of its spatial orientation. The design is modular, facilitating its augmentation on older IRm devices.

- The system's core competency is the real-time detection of patient motion with sub-millimetre accuracy, aligning with the IRm's acquisition rate of at least 30 Hz. This matches the IRm's maximum image acquisition frequency and accommodates for even faster motion scenarios.
- The system has the ability to capture patient position data even when the patient is draped, a necessity for scenarios such as spine surgeries where patients are typically fully covered during image capture.
- To ensure accurate compensation, the system includes mechanisms to synchronize motion data with each acquired projection frame. This synchronization is stored within frame data to permit reconstructions at later stages.
- The system enables motion compensation in real-time during CBCT volume reconstruction, while seamlessly integrating with medPhoton's imaging pipeline without slowing down the reconstruction process significantly.

By addressing these critical aspects, this work seeks to establish a novel system that not only mitigates the impact of motion artifacts on CBCT imaging but also enhances the overall precision, efficiency, and versatility of the IRm system.

1.4 Methodology

The outlined goals are accomplished through a systematic methodology, structured as follows:

As a first step, based on a critical evaluation of existing literature, three imaging technologies are identified as prime candidates. With these three candidates, a comprehensive feasibility study is initiated to assess their suitability for detecting patient motion. This study includes small-scale test setups, to empirically evaluate the performance of these technologies, with a keen focus on 3D position detection accuracy and processing speed. An in-depth comparison of the evaluation results allows for the selection of the most fitting imaging technology to be integrated into the IRm's motion compensation system.

Upon selecting the optimal imaging technology, a comprehensive software and hardware concept for the modular motion compensation system is devised and implemented. The resulting system will be seamlessly integrated into medPhoton's imaging pipeline, capable of synchronizing patient motion data with projection frames and effecting real-time motion correction during the reconstruction process.

As the developed system functions within an autonomous coordinate framework, an effective calibration strategy will be formulated to seamlessly align disparate coordinate systems. This calibration approach will enable the swift conversion of motion data into the IRm's imaging coordinate system.

This process results in a functional prototype, capable of being installed on an IRm within medPhoton's test lab. Extensive testing will then be conducted to assess the prototype's performance, showcasing its capabilities in compensating for motion artifacts and enhancing imaging quality.

1.5 Contributions

This work offers the following contributions to the state of the art:

- A feasibility study of different imaging technologies regarding the use of motion compensation in a CBCT device, proving IR-Imaging as the best option, which requires the use of markers but provides a 3D accuracy a factor of ten better than the compared options.
- The software and hardware design of a modular stereo-vision system that can be fully integrated into an existing CBCT device and used to detect patient motion data.
- The first fully in a CBCT-device integrated optical tracking system, capable of correcting even large rigid motion up to 10cm with sub-millimeter accuracy and without slowing down the reconstruction process. To allow live reconstruction, the motion data is acquired with 30 Hz, synchronized, and stored with the single X-ray projections as soon as they are acquired. Quantitative analysis of 50 different points in a motion-corrected CBCT image showed only a mean motion error of 0.27mm, for a system yaw of 10. This system is used as a starting point for further research in motion-related topics of CBCT devices, for example, to reduce Cone-Beam artifacts.

By adhering to this methodical progression, this work aims to culminate in a robust and effective motion compensation system, meaning a tracking accuracy below 1mm compared with a real-time capability of 30Hz, that seamlessly integrates with the IRm, advancing the capabilities of the IRm and enabling more reliable and precise surgical procedures with it.

1.6 Structure of the Work

The structure of the thesis correlates to the chronological order of the development process. In addition to that, Chapter 2 provides the reader with a foundational understanding of CBCT imaging, encompassing its fundamental principles, reconstruction methodologies, and key imaging artifacts. Additionally, a succinct overview of the IRm system is presented, as the forthcoming motion compensation system is seamlessly integrated into it.

Chapter 3 presents the current state of the art of motion compensation approaches in CBCT images. The focus of this chapter is on providing the reader with an overview of the different available concepts used to solve the problem of motion in CBCT. To provide the reader with a proper idea of the basic concepts, the available techniques are sorted into different categories and are described shortly to give the reader a better understanding of how such methods work. Furthermore, one approach per category is described shortly to give the reader a better understanding of the limitations and capabilities of techniques of this category.

After this step, three imaging technologies best suited for detecting patient motion are selected and a feasibility study is performed in Chapter 4. During the feasibility study, small test setups of the different technologies are built, where the cameras are mounted the same way as they would be in the IRm, and their performance with a focus on processing speed and accuracy is evaluated. The evaluation results are compared to each other to allow the selection of the best-fitting technology for the motion compensation system, where the main criterion is a tracking accuracy below 1mm.

In Chapter 5 the design and development of the motion compensation system are described in detail. This Chapter describes the hardware concept as well as the software concept and the integration in the IRm. Furthermore, an approach to calibrate the camera coordinate system to the imaging ring coordinate system is proposed and described. Additionally, the used synchronization approaches are described in more detail since it is a crucial component for the functionality of the system.

Chapter 6 gives an extensive overview of the performance of the prototype resulting from the previous Chapters. To gain a better understanding of the system's accuracy the reconstruction performance on data sets containing complex motion in the cm range is evaluated using different phantoms. In addition to this, a geometrical evaluation approach is developed, where system motion is used to induce even larger motion up to 15cm to the projection data.

The last Chapter concludes the thesis and describes the potential extensions and shortcomings of the proposed system. Furthermore, it gives a short overview of the areas where the system is already used for research purposes.

CHAPTER 2

Background on CBCT imaging

An understanding of the fundamental principles of CBCT imaging is important to follow the approaches described in this work. If the reader already has this knowledge, the chapter can be left out.

2.1 Imaging Hardware

At the heart of CBCT devices lies imaging hardware, crucial for shaping the quality and accuracy of diagnostic outcomes. The X-ray tube, gantry, and detector constitute integral components of a CBCT device, each playing a pivotal role in the imaging process.

2.1.1 X-ray Tube

The X-ray beam originates within an X-ray tube composed of an intricate electrical circuit housing a cathode and anode, separated by a vacuum environment. Illustrated in Figure 2.1, this apparatus functions by subjecting the anode and cathode to an electric current. The anode contains a filament, which, upon heating, induces thermionic emission, releasing electrons into the vacuum space. Employing high voltage (kV), these liberated electrons accelerate toward the anode, engaging in high-speed collisions at a focal spot. The focal spot's dimensions exert influence not only on image sharpness but also on the tube's thermal characteristics. During these electron collisions, a fraction of the kinetic energy transforms into heat, while a smaller portion converts to X-ray photons through a phenomenon referred to as Bremsstrahlung. Bremsstrahlung materializes as the electrons decelerate within the anode's high-density material. Consistent with the principle of energy conservation, kinetic energy reduction results in the emission of X-ray photons, where the energy of these photons hinges on the applied voltage. This voltage manipulation is pivotal for image quality control and dose management, given that photon absorption by the patient is correlated with their energy. The X-rays are emitted in various directions, yet due to the anode's slightly tilted surface, the bulk of the X-ray beam is directed toward the exit window of the tube. Consequently, a perpendicular X-ray beam emerges from the tube, as other directions are absorbed by the tube housing [PAST15].

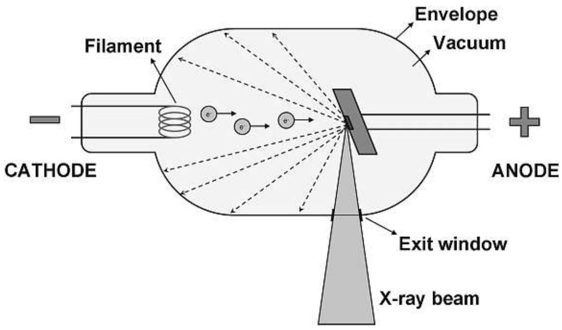

Figure 2.1: Simplified schematic of an X-ray tube [PAST15]

To exercise control over the system's field of view (FOV) and to confine patient exposure solely to the region of interest, a collimator is maneuvered to partially obscure the exit window of the tube. By strategically obstructing extraneous X-rays, the collimator limits the irradiated area to the relevant anatomical region. It's essential to note that the energy spectrum of Bremsstrahlung radiation is continuous, resulting in the emission of undesired low-energy photons. These photons, due to their likelihood of full absorption by the patient, fail to enhance image quality but contribute to an increased applied dose. To counteract this effect, an additional metallic filter is introduced, designed to impede the exit of most of these undesirable low-energy photons. In shaping the overall radiation dose and the number of photons emitted from the tube, both tube current and exposure time play pivotal roles. By manipulating these parameters, the system exercises precise control over the radiation dose while ensuring the acquisition of diagnostically valuable images [PAST15].

2.1.2 Gantry

The central function of the gantry lies in its facilitation of rotational movement for the X-ray source and detector, both around and, in certain configurations, along the patient's anatomy. An imperative aspect for achieving accurate back projection is the reproducibility of the rotation orbit. It's important to note that while the source and detector need not adhere to a strictly circular trajectory, ensuring a reproducible orbit is vital for precise imaging. Any deviations from a perfect circle can be rectified through meticulous geometric calibration procedures. CBCT gantries come in diverse shapes, with one of the most common configurations featuring a C-arm design. In this configuration, the source and gantry are positioned opposite to each other, connected via a rotatable C-shaped platform. A comprehensive exploration of the various gantry shapes can be found in Abramovitch et al.'s work [AR14].

2.1.3 Detector

X-ray detectors serve the pivotal role of converting the X-ray photons emitted by the source into electrical signals, exerting a profound influence on the resultant image quality. Beyond resolution, the efficacy and speed of this conversion process stand as vital attributes of detectors. Various types of detectors are available, with a prevalent choice among contemporary CBCT units being flat panel detectors (FPDs). FPDs confer several advantages, such as reduced distortion and heightened dose efficiency, when compared to the image intensifiers that were more commonly employed in older systems [AR14].

FPDs operate by employing a slender layer of X-ray absorptive material, either deposited on an electronic active-matrix array within a hydrogenated amorphous silicon film or integrated onto a complementary metal-oxide-semiconductor panel. Depending on the type of absorptive material utilized, FPDs can be categorized into two distinct groups: indirect-conversion (x-ray scintillator-based) and direct-conversion (x-ray photoconductor-based) detectors [FJSS21].

These detectors play a critical role in capturing and converting the X-ray signal, ultimately contributing to the precision and fidelity of the reconstructed images, since there is no longitudinal movement during a CBCT scan, meaning the final resolution of the volume depends on the resolution of the detector.

2.2 Reconstruction

Image reconstruction is a fundamental technique employed to synthesize 3D images from a multitude of 2D projections captured by the CBCT scanner across various angles. While numerous systems collect projections spanning a full 360° rotation of the source and detector (full-scan), a rotation of 180° in conjunction with beam angles (short-scan) suffices for accurate reconstruction of the entire field of view if the initial data set is weighted properly, as described by Parker [Par82].

The Feldkamp–Davis–Kress (FDK) algorithm stands as the most prominent due to its simplicity and rapid reconstruction capabilities [PAST15]. The FDK algorithm [FDK84] operates on the premise of the filtered back projection (FBP) principle, a cornerstone in the field of computed tomography. For a more comprehensive understanding of this algorithm, the work by Turbell et al. [Tur01] offers an in-depth exploration.

The FBP principle underscores the reconstruction process: A projection image is essentially a summation of the attenuation values encountered by X-rays traversing the patient's anatomy along specific paths. By exploiting this principle, FDK algorithm first applies a filtering step on each projection to enhance image quality. Subsequently, these filtered projections are back-projected into the volume space, with each pixel accumulating the contributions from multiple projections. This summation process generates a 3D volume image that encapsulates the internal structures of the imaged subject.

Back projection Multiple projections of a full-/ short-scan are required for a proper reconstruction. The basic principle of a back projection can be outlined like this:

If the intensity I_s emitted by the source is known, then the intensity I behind the absorbing object can be described by the Lambert-Beer law:

$$I = I_s * exp[-\int_l \mu dl] \tag{2.1}$$

Where μ describes the absorption coefficient of the object material, dl is the distance the X-ray moves through it, and l stands for the line over which the attenuation is integrated. The goal of the volume reconstruction is now to recover the density function μ of the object with the help of the projection data. Since the projections are not continuous Equation 2.2 is needed to describe the discrete case:

$$\ln \frac{I_d}{I_s} = -\sum_i^N \mu_i d_i \tag{2.2}$$

μ_i describes the absorption of the object at a sample along the projection line at a position i, where the line contains N different sample positions. Since the projection geometry of the imaging system is known, all projection lines between the source and the detector cells can be constructed. The attenuation value $-\ln \frac{I_d}{I_s}$, where I_d describes the intensity measured at one detector cell, is now back-projected toward the source along the constructed line. This is done by smearing the value $-\ln \frac{I_d}{I_s}$ through all voxels of the object's image from point B to A, see Figure 2.2.

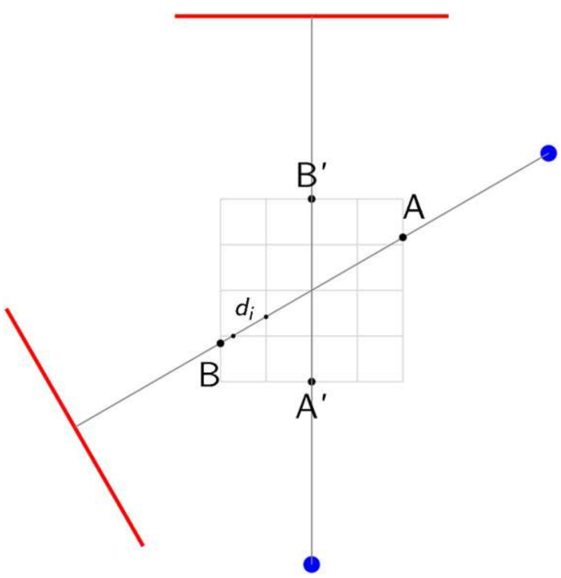

Figure 2.2: Simplified visualization of the back projection process[SHG+11]

A way to accomplish this is to calculate the distance d_i, see Figure 2.2, the X-ray travels through a single voxel and computes the distance contribution of the X-ray to the grey value of the particular voxel. The object of interest is not homogeneous and it is therefore impossible to decide how each voxel has contributed to the attenuation so no correct grey value can be computed. To be able to estimate it, the back-projection is performed for multiple projections from different view angles[SHG+11].

The image quality depends on the amount and angle of the projection data. As visualized in Figure 2.3, 2 projections cause undersampling of the object and the introduction of streaks along the direction of the back-projected rays, but when increasing this number, the resulting image gets sharper [PAST15].

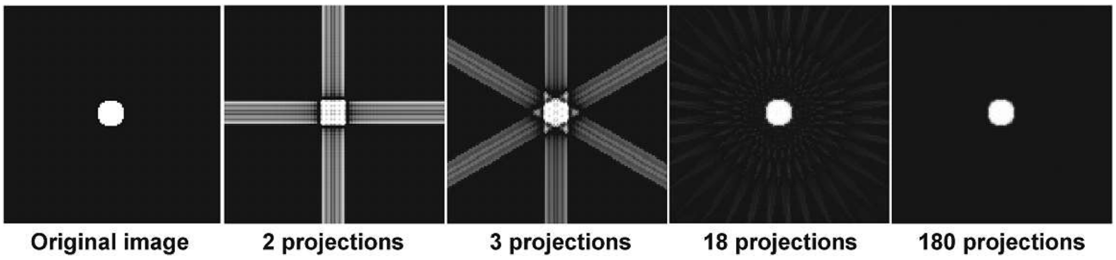

Figure 2.3: Image reconstructed using filtered back projection with a different number of projections[PAST15]

For filtered back projection algorithms like FDK, the projections have to be processed with a ramp filter. This corrects the inherent blur introduced by the back projection process. Additionally, a smoothing filter is applied to reduce high-frequency noise, which the ramp filter tends to amplify, prior to performing the back projection [PAST15].

Advancements in hardware capabilities and computational power have prompted renewed interest in another reconstruction approach, namely iterative reconstructions, where particularly noteworthy are the outcomes achieved by combining iterative reconstruction with deep learning techniques [WYMF18] [FJSS21]. These advancements herald a potential paradigm shift in the CBCT reconstruction landscape.

2.3 Image Artefacts

Image artifacts manifest as observable structures within the reconstructed volume that lack correspondence to the imaged physical object[SHG+11].

These artifacts predominantly arise due to disparities between the real-world attributes of the CBCT scanner setup—such as its configuration, the object's behavior, or its positioning—and the simplified mathematical approximations employed during volume reconstruction [SHG+11].

2.3.1 Motion artefacts

Motion artifacts encompass a spectrum of image distortions arising from motion during the image acquisition process [SHG+11]. These artifacts can manifest in diverse forms, such as double contours, streaking, blurring, or even structural deformation [KMBS21]. The crux of motion artifacts lies in the divergence between the anticipated stationary geometry, assumed during normal back-projection reconstruction, and the actual motion of the imaged object. This mismatch results in a misalignment of the attenuation lines recorded during acquisition with the lines utilized for back-projection, thus giving rise to motion artifacts. A pivotal determinant of motion artifacts is the volume's resolution. Smaller voxel dimensions render structures within the image susceptible to displacement caused by even minor movements, underscoring the influence of resolution on the manifestation of these artifacts [SHG+11].

Examples of motion artifacts in CBCT images can be seen in Figure 2.4.

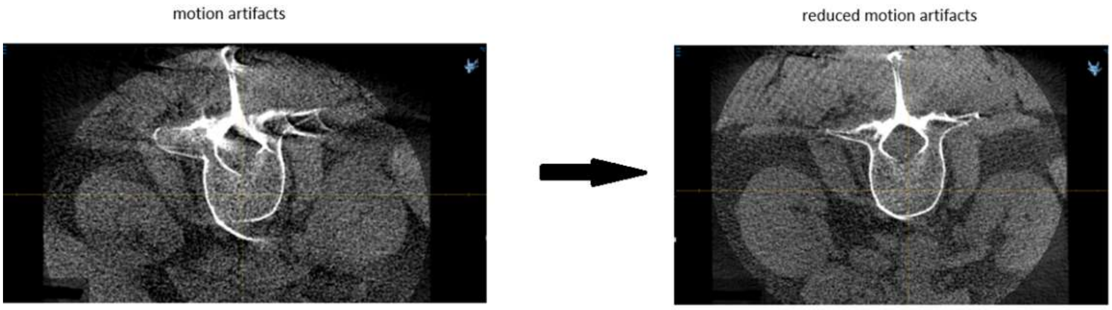

Figure 2.4: Image of a spine CBCT volume containing motion the same data set where the motion is compensated.

Various investigations underscore that motion artifacts substantially impair the diagnostic efficacy of CT/CBCT images [KNKF11] [KMBS21] [SNCS+18] [SHG+11] [SF08] [SNKH+20]. These artifacts pose a more pronounced challenge for CBCT systems due to their lengthier scan times, making it inherently difficult to mitigate rigid and non-rigid patient motion over extended periods [KMBS21].

2.3.2 Beam hardening artefacts

Beam hardening artifacts, also known as metal artifacts, emerge from the polychromatic nature of the X-ray spectrum emitted by the X-ray source [SF08]. This phenomenon leads to an elevation in the mean energy of the X-ray beam due to the differential absorption of lower energy photons compared to higher energy ones when traversing an object [SF08].

As an object becomes denser and the material it's composed of possesses a higher atomic number, a greater number of wavelengths are absorbed. When the emitted beam incorporates more low-energy rays than the detector records after it interacts with a highly absorbing material, a non-linear error is introduced. This occurs because, in the beam path beyond the material, an excessive amount of energy is registered. Subsequently, this non-linear error is incorporated into the volume during reconstruction, yielding two distinct artifacts: cupping artifacts, which distort metallic structures, and streak artifacts, manifesting as dark bands between dense objects [SHG+11].

To mitigate beam hardening artifacts, it is recommended to reduce the field of view (FOV) to avoid scanning excessively dense objects. This can be accomplished through collimation or patient positioning strategies. If such measures are impractical, various software algorithms are available to ameliorate these artifacts [SF08]. Employing these techniques aids in achieving more accurate and diagnostically valuable images, especially when dealing with high-density materials or metal implants.

2.3.3 Scatter artefacts

While the fundamental reconstruction principle solely considers photons following a direct path from the source to the detector, real-world scenarios encompass photon diffraction from their original trajectory due to interactions with matter, resulting in scatter. These divergent scattered rays induce an augmentation in the recorded pixel intensities on the detector. Upon back-projection, these elevated intensities lead to an overestimation of voxel values along the path, causing an underestimation of absorption. The magnitude of the reconstruction error is directly proportional to the extent of scatter present and is contingent upon the imaged object's properties [SHG+11].

Owing to the expansive surface area of flat panel detectors utilized in CBCT devices, the likelihood of scattered X-rays impacting pixels is substantially higher compared to conventional highly collimated CT machines. As a consequence, scatter artifacts manifest more prominently in CBCTs than in CTs. These artifacts materialize as streaks, closely resembling those engendered by beam hardening [SHG+11].

Two distinct categories of scatter correction methods are available: hardware-based and software-based approaches. The former necessitates the integration of supplementary hardware within the CBCT machine, an exemplar being an anti-scatter grid placed atop the detector. Conversely, the latter employs software algorithms—such as Monte Carlo simulations—to estimate and subsequently mitigate scatter artifacts [XYZL18]. Employing these correction techniques is instrumental in enhancing image accuracy and diagnostically relevant information.

2.3.4 Cone-beam artefacts

Cone-beam artifacts manifest within the peripheral extents of the scan volume, giving rise to image distortions, streaking artifacts, and heightened noise levels. The acquisition process involves each pixel of the detector collecting data while circumnavigating the patient. This accumulation corresponds to the cumulative attenuation recorded within the planar projections. However, the cone shape of the X-ray beam and the dimensions of the detector contribute to a reduction in information within the peripheral regions. Specifically, the outer rows of detector pixels exhibit lesser attenuation recording compared

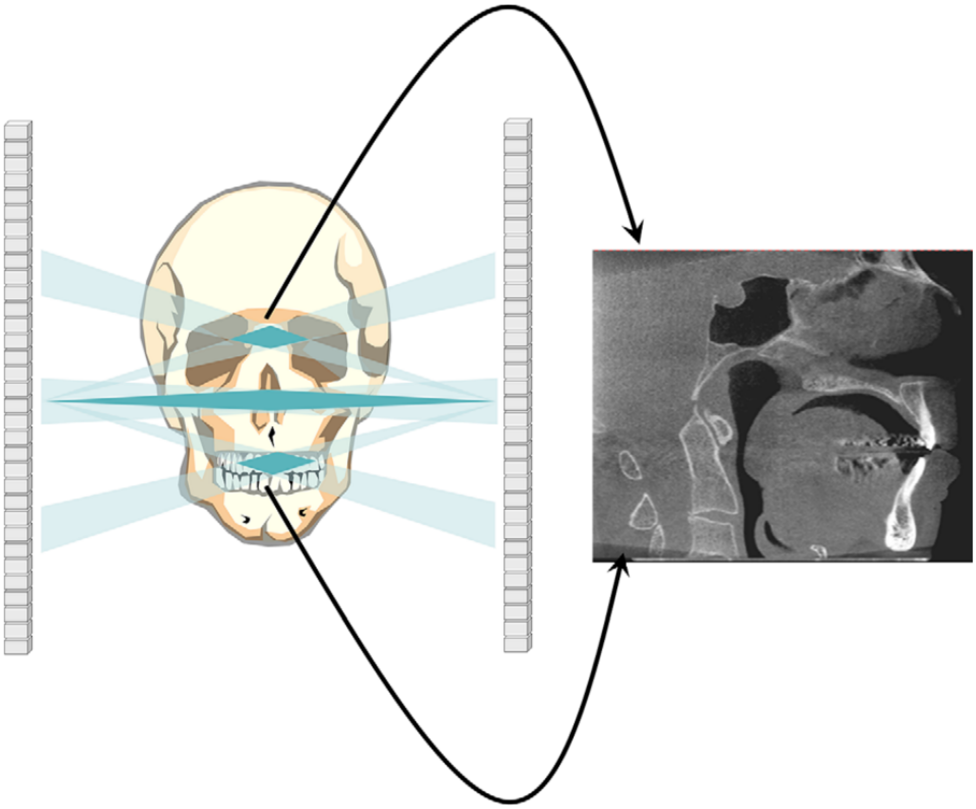

Figure 2.5: The cause of the cone-beam artifact is visualized with three X-ray beams and two detector positions opposite to each other. The overlap of the X-rays corresponds to the data used for the reconstruction and it shows that the data in the central slice is maximal and gets less with the distance to it. The result of this effect is shown in the midsagittal section images on the right side [SF08].

to their inner counterparts. Figure 2.5 visually elucidates this quandary across divergent detector positions, where the solid volume delineates the data utilized for reconstruction. Consequently, data availability is maximal at the central slice and progressively diminishes as the distance from this central slice increases [SF08].

There have been two primary avenues explored to alleviate these artifacts. One involves modifying the data acquisition geometry to mitigate missing data in the peripheral zones. Alternatively, efforts can be directed towards enhancing the reconstruction algorithm to counteract these artifacts [CKB18].

The proposed system provides also a way to reduce cone-beam artifacts by employing saddle trajectories utilizing the optical system developed within this study. This strategy is poised to ameliorate the limitations posed by cone-beam artifacts, offering a pathway to enhanced image quality and diagnostic utility.

2.4 Imaging Ring m

The Imaging Ring m (IRm) stands as a pioneering fully mobile CBCT device, developed by medPhoton GmbH. It distinguishes itself from conventional CBCT devices through a distinctive set of features, namely the closed gantry, individual rotatable source/detector, and the large FOV of up to 51cm, that amplify its capabilities and versatility. Notably, the IRm encompasses the ability for independent rotation of both the detector and the source, thus facilitating non-isocentric acquisitions—a departure from standard practice. Additionally, the imaging center can be freely adjusted along six degrees of freedom (DOF), endowing the IRm with unparalleled adaptability. The device can be seamlessly controlled via a remote control panel, allowing it to operate autonomously even in diverse surgery rooms, independent of the main power supply.

A significant attribute of the IRm is its expansive gantry, coupled with robotic control and a lightweight design, which synergistically enables exceptional flexibility in patient positioning. This amalgamation of features underscores the IRm's capability to accommodate diverse clinical scenarios and patient needs, promoting precise and optimized imaging procedures.

For a more tangible understanding of the system, Figure 2.6 visually encapsulates the key components and design of the IRm.

Figure 2.6: Image of the mobile CBCT device IRm developed by medPhoton.

CHAPTER 3

State of the Art

To impart a coherent grasp of the fundamental principles used for motion compensation in CBCT imaging, the available techniques are organized into distinct categories and succinctly elucidated, thereby facilitating a deeper comprehension of the operational mechanisms of such methods. This comprehensive exploration serves to equip readers with a comprehensive understanding of the evolving landscape of motion compensation methodologies in CBCT imaging.

3.1 Image-based systems

Image-based systems harness CBCT image data to compute and mitigate patient motion occurring during the scanning process. These systems fall into two distinct categories: purely image-based and fiducial marker-based. In the latter category, fiducial markers are placed in the FOV of the device, which are visible in CBCT projection images. These markers can be segmented in the projections and the motion between the frames, depending on the marker locations is calculated. On the other hand, purely image-based systems don't use any markers, therefore they rely solely on the information inherent in the CBCT images to detect and correct motion artifacts.

3.1.1 Purely image-based

Purely image-based motion compensation systems offer correction mechanisms devoid of any reliance on external data sources, such as fiducial markers or tracking hardware. This category encompasses well-established methods, with optimization-based motion estimation, 2D/3D registration algorithms, and, more recently, deep learning approaches emerging as prominent avenues of exploration. The subsequent paragraphs delineate the underlying principles of these three distinct approaches.

2D/3D registration: One viable avenue for addressing motion in CBCT images involves the utilization of 2D/3D registration algorithms, which leverage a motion-affected scan alongside a motion-free prior CT or CBCT dataset.

A notable instance of this approach is presented by Berger et al., who proposed marker-free motion correction in weight-bearing CBCT scans of the knee joint [BMA+16]. The weight-bearing scenario allows for a more authentic evaluation of joint health under realistic conditions; however, it introduces heightened motion artifacts [BUG+17]. The motion correction process hinges on the 2D/3D registration of adjacent bones within the knee. Precursor to motion correction, bones are segmented within a motion-free 3D volume obtained in the supine position (patient lying on their back). The methodology unfolds through two steps. In the initial step, segmented bones are aligned to the motion-affected volume acquired under weight-bearing conditions (patient standing or squatting). Subsequently, rigid 2D/3D registration is executed for all volume frames, estimating six degrees of freedom (6x4) motion parameters for each frame. These parameters are then used to compute global motion fields employing thin-plate-spline extrapolation. The calculated motion field is incorporated into the volume reconstruction's back-projection stage, culminating in a motion-corrected CT volume [BMA+16].

Other instances of 2D/3D registration-based motion correction are proposed by Ouadah et al. [OJS+17] and Klugmann et al. [KBM+18]. It's worth noting that while these marker-free motion correction algorithms display promise, they can occasionally exhibit performance inferior to marker-dependent approaches, and their efficacy might hinge on the availability of requisite prior information [BMA+16].

This facet of 2D/3D registration elucidates the potency of leveraging existing datasets and spatial relationships to mitigate motion-induced artifacts, even in the absence of external markers. However, it also underscores the potential challenges linked with acquiring accurate prior information and achieving robust registration outcomes.

Optimization-based motion estimation: Optimization-based motion correction algorithms within CBCT aim to derive an optimal motion model from image data by optimizing a metric that assesses image sharpness in the reconstructed 3D image. Notably distinct from 2D/3D registration approaches, optimization-based methods circumvent the need for prior image data, and they are particularly amenable to focusing reconstruction within a predefined volume of interest. This flexibility is attributed to the absence of the necessity for global motion estimation, permitting localized correction of complex motions involving multiple rigid bodies [SSY+17][SSC+16].

An illustrative example of such a purely image-based approach, devoid of fiducials or prior images, is proposed by Capostagno et al. [CSS+21]. The method employs an iterative algorithm for motion model estimation, striving to maximize the image sharpness criterion alongside a regularization term that promotes smooth motion trajectories while penalizing abrupt movements. The optimization of the non-convex cost function is carried out by sampling potential motion trajectories from a randomized distribution. For each trajectory, the sharpness criterion and regularization term are computed, leading to a distinct cost function value. Following motion model estimation, it is applied during the FDK reconstruction [RND04], culminating in a motion-corrected volume [CSS+21].

Noteworthy advancements in optimization-based motion estimation materialized some years ago, with methods sharing akin functional principles to the aforementioned approach, albeit with variations in the region of interest and cost function [NST+19][HBA+16] [SSY+17] [WKKK13] [JKKR18] [STS+19]. These algorithms underscore the potency of harnessing iterative optimization to derive accurate motion models directly from image data, offering precise correction without the reliance on external tracking hardware or previous imaging information.

Deep Learning-based motion correction Deep neural networks (DNNs) are composed of multiple layers that collectively approximate an unknown mapping function between input and output data[KMBS21]. Each layer can be trained to perform various operations, such as feature extraction, modeling non-linearities, or altering data dimensions. The stacking of multiple layers enables the development of a powerful hierarchical representation, facilitating the comprehension of intricate semantic relationships between input and output data. Deep learning-based algorithms excel in computer vision tasks and have proven successful in modeling non-linear and complex decision processes[KMBS21]. Given that motion artifacts in CBCT volumes exhibit highly non-linear characteristics, neural networks can be effectively employed for image-based motion correction in CBCT [KMBS21].

Ko et al. [KMBS21] propose a real-time technique for compensating rigid and non-rigid patient motion in CBCT data using a deep residual network enhanced by a self-attention mechanism. The architecture encompasses two principal components: the residual block and the attention block. The residual block integrates a skip connection, two convolution layers, and an interposed rectified linear unit. Meanwhile, the attention block augments the network's modeling capabilities by selectively amplifying or attenuating feature maps based on their significance. The objective function involves a combination of two loss terms: mean absolute error, which computes pixel-level loss, and perceptual loss, which gauges feature distribution through a pre-trained feature extractor to preserve image details and content [KMBS21].

Several other approaches harness deep learning for CBCT motion correction, as demonstrated by methodologies presented by Jiang et al. [JZG+19], Lossau et al. [LNW+19], and Sisniega et al. [SHZ+21]. One key advantage of these approaches is their rapid evaluation times, enabling real-time applications. Moreover, they require only motion-corrupted image data as input, underscoring their practical utility [KMBS21].

3.1.2 Fiducial marker-based

Fiducial markers are physical objects placed within the imaging field of view that are visible in the produced images[BLH+11]. In the realm of CBCT imaging, these markers typically consist of small metal objects attached to or implanted within the patient's body. One common application of fiducial markers is in Image-Guided Radiotherapy, where these markers aid in accurately locating tumors for precise therapy delivery, ensuring high radiation doses are directed at the tumor while sparing healthy tissue [BLH+11][1].

Beyond their role in Image-Guided Radiotherapy, fiducial markers can be exploited for correcting patient motion in CBCT, as proposed by Müller et al. [MBC+15]. The approach is summarized below. The process commences with the automatic detection of marker positions within 2D images using the method introduced by Berger et al. [BFS+14]. Detected 2D marker positions are then back-projected to estimate corresponding 3D marker positions. To achieve this, marker positions are matched with their respective 3D locations by minimizing the reprojection error between the projected 3D points and detected 2D points. To approximate rigid motion, the unknown parameters of a transformation matrix $M \in \mathbb{R}^{4x4}$ applied to the projection matrix are estimated by further minimizing the reprojection error [MBC+15].

The accuracy of motion estimation hinges on precise initial estimation of 3D marker positions and correspondences between points. However, significant patient motion can negatively impact these factors [SBB+17]. To address this, Syben et al. propose an extension to the aforementioned approach. Their methodology involves iteratively refining the initial 3D marker positions and incorporating a self-calibration component to enhance the accuracy of motion estimation [SBB+17].

3.2 Optical systems

Motion compensation using optical systems, due to their non-invasive nature and well-established technology, has been successfully used to compensate motion in multiple CBCT / CT applications [BRU+18][FFWN17][KNKF11][MNH+19].

Optical systems work independently of the CBCT system. Therefore the surface data/marker positions are in a different coordinate system than the imaging data. To correct the motion during the reconstruction of the volume, a calibration between the optical coordinate system and the CT coordinate system is required to determine a transformation between them.

The optical systems can be divided into two different categories, based on the basic functionality. The first category is modeling the surface of the patient and estimating the patient's motion by tracking the changes in the surface. The benefit of it is that the patient doesn't have to be equipped with markers, which requires time and proper training of the hospital staff. The second one is tracking external markers placed on the patient and the motion of the markers is assumed to be the patient's motion. Here the benefit is in the more dependable and accurate tracking results because motion estimation parallel to the rotation axis can be unreliable with surface tracking methods [BUG+17].

[1] https://www.radiologyinfo.org/en/info/fiducial-marker/, accessed 2022-02-17

3.2.1 Surface tracking

The most commonly used optical systems to extract the patient surface during CBCT are Red-Green-Blue-Depth (RGBD) sensors, because they are in contrast to the time of flight cameras less dependent on temperature, lightning, or materials, provide in most cases a higher resolution and color information. Approaches using RGBD cameras to compensate motion have been shown the best results in the last years and have been proposed by multiple authors, like Bier et al. [BRU+18], Moghari et al. [MNH+19], or Fotouhi et al.[FFWN17].

In their proposal, Bier et al. [BRU+18] have an RGBD sensor mounted to the image intensifier of a CBCT-enabled C-arm, which allows the simultaneous acquisition of CBCT data and RGBD surface data. To use the patient surface data during the reconstruction of the volume, a calibration of the RGBD camera and the CBCT coordinate systems is required. To accomplish this calibration, a planar checkerboard pattern phantom with radiopaque markers is used to perform a stereo calibration of the C-arm and the RGBD camera. The rigid motion of the patient during a scan is then observed using the RGBD sensor and the transformation is estimated using RGBD-based Simultaneous Localization and Mapping (SLAM). It augments 2D extracted color features with co-calibrated depth information, which increases the accuracy of surface detection. The resulting transformations are further refined using the iterative closed point algorithm and the motion trajectory is optimized using a pose graph solver with a non-linear energy function. The resulting trajectory is then matched with the CBCT geometry parameters and trajectory and in case of dissimilarities, it is used to correct for patient movement during the reconstruction of the volume.

This approach showcases how RGBD sensors can effectively provide motion information for motion compensation in CBCT imaging. By integrating color and depth information, the RGBD-based SLAM method enhances the accuracy of surface tracking, allowing for precise motion correction during volume reconstruction.

3.2.2 Optical marker based

Optical markers can be used to increase the accuracy and consistency of the position tracking at specific locations on the patient surface, but they require also proper positioning before the acquisition. There are multiple kinds of optical markers, the most popular examples are flat surface markers, detectable by RGB-Cameras as used by Burstroem et al. [BNH+20] and Infrared(IR)-Markers, detectable by IR-Cameras as proposed by McDonald et al. [MLS+20] or Kim et al. [KNKF11].

An example of an optical marker-based approach is the method proposed by Kim et al. [KNKF11]. In this method, an optical tracking system called Polaris Hybrid Spectra is utilized to monitor patient motion using IR light reflected by passive retro-reflective markers. The system tracks the position of these markers in real-time during the CBCT scan.

Since the coordinate systems of the tracking system and the CT scanner are distinct, a calibration process is required to transform the tracking data into the scanner's coordinate system. This initial calibration ensures that the motion data collected by the optical tracking system can be accurately applied to the CT scanner's images.

Once the transformation is established, the tracked motion data is used to compensate for patient motion in the CBCT images. This is achieved by applying the motion data to the acquired images after the motion has occurred. This correction process helps generate motion-free image data, allowing for improved image quality and diagnostic accuracy.

Overall, optical marker-based methods offer a reliable way to capture and correct patient motion during CBCT imaging. By using specific markers and optical tracking systems, these methods provide accurate motion information that can be used to enhance the quality of reconstructed images and minimize motion artifacts.

3.3 Inertial measurements

Inertial measurements provide an alternative approach to compensate for motion in CBCT images without relying on image data or optical systems [MNC+21]. This method involves measuring specific forces and angular rates at one or multiple points on the patient's body. By analyzing these measurements, the transformation of the motion at the measurement point can be calculated. This transformation can then be applied during or after the reconstruction of the CBCT volume to correct for patient motion.

A notable example of this approach is the work by Maier et al. [MNC+21], where inertial measurement units (IMUs) are attached to the patient's leg for motion estimation during CBCT imaging. IMUs are devices that consist of accelerometers and gyroscopes, allowing them to measure accelerations and angular velocities. The authors demonstrate that their proposed method, which relies on IMU measurements, is capable of achieving improvements in the structural similarity index measure of 24-35%, where a motion-free reference image is compared to motion-corrected images and non-motion-corrected ones.

In this method, the IMUs attached to the patient's leg continuously measure the accelerations and angular velocities of the leg during the CBCT scan. Using this motion information, the transformation of the leg motion is estimated. This estimated transformation is then applied to the acquired CBCT images to compensate for the detected motion, resulting in motion-corrected images.

The advantage of using IMUs for motion compensation is that they are unobtrusive, and can be attached to different parts of the patient's body, in case of the knee on the thigh and the shine. This provides a non-invasive and efficient way to estimate and correct for patient motion during CBCT imaging. While there might be challenges in calibrating and synchronizing the IMUs with the imaging system, the results from studies like Maier et al.'s work show the potential effectiveness of this approach in motion compensation for CBCT images.

3.4 Gating methods

Gating methods are specialized techniques developed for compensating for periodic movements, such as those caused by heartbeat or breathing, during the reconstruction process of CBCT images. These methods take advantage of prior knowledge about the periodicity of the motion to synchronize the image data acquisition with the cycles of the movement, effectively reducing motion artifacts. Gating methods are applicable for rigid and non-rigid motions and can be either prospective or retrospective in nature [NST+19].

Prospective gating involves limiting data acquisition to specific intervals within the movement cycle. This can be achieved in various ways. For instance, during respiratory motion, the scan can be triggered to start at a specific phase of the breathing cycle. Alternatively, the patient might be required to hold their breath at a certain point, allowing the scan to be acquired during that phase. However, a potential drawback of prospective gating is that image data is captured only at a single phase of the movement cycle.

In retrospective gating, data is acquired continuously, and then the acquired data is sorted, grouped, or correlated based on the phase of the movement cycle during reconstruction. This method requires more images to be acquired, depending on the number of different phase intervals used for reconstruction. A balance must be struck between the number of intervals and individual image sets, as the patient's radiation dose cannot be increased excessively [MB12]. [2]

Several advancements have been made to improve the basic principle of retrospective gating and reduce patient dose while maintaining image quality. Guo et al. [GCO+19] proposed a method involving motion modeling based on single interval reconstruction, enabling the generation of new volumes at any point within the motion cycle. Another approach suggested by Bergner et al. [BBO+09] involves dividing the projections into regions affected by motion and regions at rest. The regions at rest are shared among different interval reconstructions, leading to an overall reduction in artifacts and noise.

The success of motion correction using gating methods heavily depends on the accurate detection of the periodic movement cycle. An illustrative instance involves the identification of the patient's respiratory cycle, a topic that will be further explored in the following:

External optical markers: A well-established system using external optical markers for the detection of the patient respiratory cycle, is the Real-Time Position Management (RPM) system, by Variant Medical System, Palo Alto, CA. It uses a box with six retro-reflective IR markers, usually played on the mid-line of the patient. The markers are illuminated by an external IR light source and tracked by a CCD camera. Out of the motion pattern of the marker box, the respiratory cycle can then be extracted [MB12] [GCO+19].

[2] `https://www.mriquestions.com/gating-methods.html`, accessed 2022-03-19

Optical surface tracking: Optical surface tracking methods, such as the approach proposed by Fassi et al., utilize optical systems to measure the displacement of the external surface of the patient. By analyzing the measured data, a patient-specific breathing model can be estimated and synchronized with the image data to compensate for motion [FSF+14].

Spirometer based systems: Another way to monitor the respiration of patients is the use of Spirometer-based systems. They can measure and sometimes even control the airflow of the patient's respiration. It allows applications in breath-hold CT, where a gating of the CBCT is not needed anymore, but it comes with an increased discomfort of the patient [MB12].

Back pressure measurements: Zhang et al. proposed a novel approach that involves placing a sensor beneath the patient's back to measure pressure changes during the respiratory cycle. The change in pressure is correlated with breathing motion, providing an alternative way to monitor respiration and body movement. This approach is less invasive and may improve treatment comfort while providing valuable information about the patient's motion [ZTS+20].

These methods highlight the variety of techniques available for accurately detecting the patient's respiratory cycle and enabling effective gating-based motion compensation strategies in CBCT imaging.

CHAPTER 4

Feasibility study

The feasibility study conducted in this chapter aims to evaluate three different imaging technologies for their suitability in detecting patient motion, as specified by the requirments in Section 1.3: RGB stereo cameras, structured light RGB-D cameras, and IR stereo cameras. The study involves building small test setups for each technology and assessing their performance in terms of processing speed and accuracy.

In conducting this feasibility study, the objective is to select the imaging technology that can effectively and accurately detect patient motion. This choice is crucial for ensuring the success and reliability of the motion compensation system.

4.1 RGB stereo cameras

As a starting point for the feasibility study, the use of RGB cameras to detect motion during acquisition is studied. This system is chosen due to the fact that the IRm is already equipped with four RGB cameras located in the detector arm and the gantry, see image 4.1. Two of the cameras are placed in the gantry of the system and provide a static view of the scene. The other two are mounted on the detector arm and move therefore around the imaging center during an acquisition. Each pair of cameras has different advantages, the detector pair enables more possible applications in the field of photogrammetry but the patient can't be seen at all times during full scans. Since it is important for motion compensation, based on the RGB cameras, that the patient and therefore a possible motion can be seen during the whole acquisition time the gantry camera pair is used for further experiments.

If the cameras are activated, the images are streamed over a network to the medPhoton Controls (mPc) and stored in a circular buffer for each individual camera. Due to hardware limitations, the streaming of images with the highest quality is restricted to only 12 frames per second(fps). Furthermore, there is no possibility for a hardware

	Detector	Gantry
Max resolution	1080p	1080p
Pixel Size	2.9x2.9 µm	3x3 µm
Framerate (max/at max resolution)	50fps/50fps	120fps/30fps
Focal length	2.8mm	3.6mm
Opening Angle	100°	80°

Table 4.1: Technical specifications of the RGB camera types used in the IRm.

Figure 4.1: The locations of the RGB cameras mounted in the IRm. Two cameras are placed in the gantry of the system, providing a static view of the patient, and the other two are mounted on the detector providing different views during a scan.

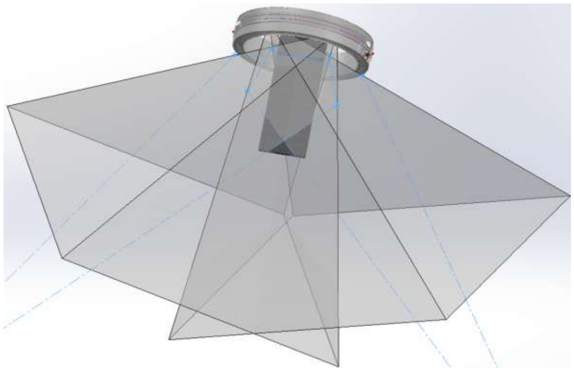

Figure 4.2: Visualization of the FOV of the cameras mounted in the Gantry of the IRm. The blue line shows the overlapping FOV area of the two cameras at the high of the isocenter.

synchronization of the cameras which leads to the max. synchronization offsets below 100ms. Since it is assumed that the usual patient motion is not very fast, this offset is acceptable for the first experiments. Figure 4.2 visualizes the FOV of the Gantry cameras, where the blue line shows the overlapping area of the FOVs in the isocenter. It shows that this area is large enough that the patient's surface is visible to both cameras.

The first goal of the feasibility study is to use these cameras to reconstruct the 3D surface of patients during an acquisition. The surface information can then be used to detect and compensate for patient motion.

4.1.1 Stereo Vision

Stereo vision is a technique that involves extracting three-dimensional (3D) information from images captured of a scene from two different points at the same horizontal level [MPHG91] [Aya91]. This is achieved by identifying corresponding points visible in both images and calculating the relative depth information through the measured disparity[MPHG91]. The result is a disparity map that encodes the distance of the scene points from the cameras, with each pixel's location corresponding to a specific depth value.

In an ideal stereo vision setup, two cameras are positioned with a horizontal offset, resulting in coplanar image planes for the left and right cameras. This configuration simplifies the process of finding matching points, as the search is confined to the same horizontal line in both images[BK08].

Assuming the cameras are ideal, the pinhole camera model can be applied to derive an equation for calculating the depth of a point in the 3D scene. This model establishes a relationship between a point on the image plane and its corresponding point in the scene. This relationship can be expressed as follows:

$$x = -f * \frac{X}{Z} \tag{4.1}$$

where x defines the point in the image plane, X in the scene, f the focal length of the lens, and Z the distance of the point in the 3D scene to the camera. Equation 4.1 is now used in the stereo vision setup to extract the distance of the point in the 3D scene with the help of similar triangles resulting in:

$$\frac{B-(x_l - x_r)}{Z-f} = \frac{B}{Z} \quad \Longrightarrow \quad Z = f * \frac{B}{D}, \qquad where \quad D = x_l - x_r \tag{4.2}$$

B is the baseline of the stereo vision setup, D is called disparity, the distance between corresponding points in the left and right image, f is the focal length and Z defines the distance between the 3D point and the baseline. There is a nonlinear relationship between the depth and the disparity. This means that if the disparity is close to zero, small disparity differences result in large depth differences. Due to this, stereo vision systems have only a high depth resolution for objects in close proximity [BK08].

As shown, it is possible to extract the Z value for this ideal stereo vision setup. The issue is that for real-world scenarios, the cameras will almost never be aligned as required for the calculations. The issue can be bypassed, by finding image projections and distortion maps rectifying the left and right camera images into a geometry that resembles the ideal camera arrangement. It is beneficial for the stereo setup, if the cameras are as close as possible to a horizontal alignment, with an exception for applications where a high resolution at close range is required where a slight tilt of the cameras towards each other has proven as beneficial [BK08].

The camera arrangement is already defined for this work and can be seen in Figure 4.3. Since the baseline of the cameras is 712.77mm and the tilt angle is 35 degrees, the mathematical alignment may produce one hand more distortions. But on the other hand, it could also prove beneficial for the close range accuracy, which is the use case of this work.

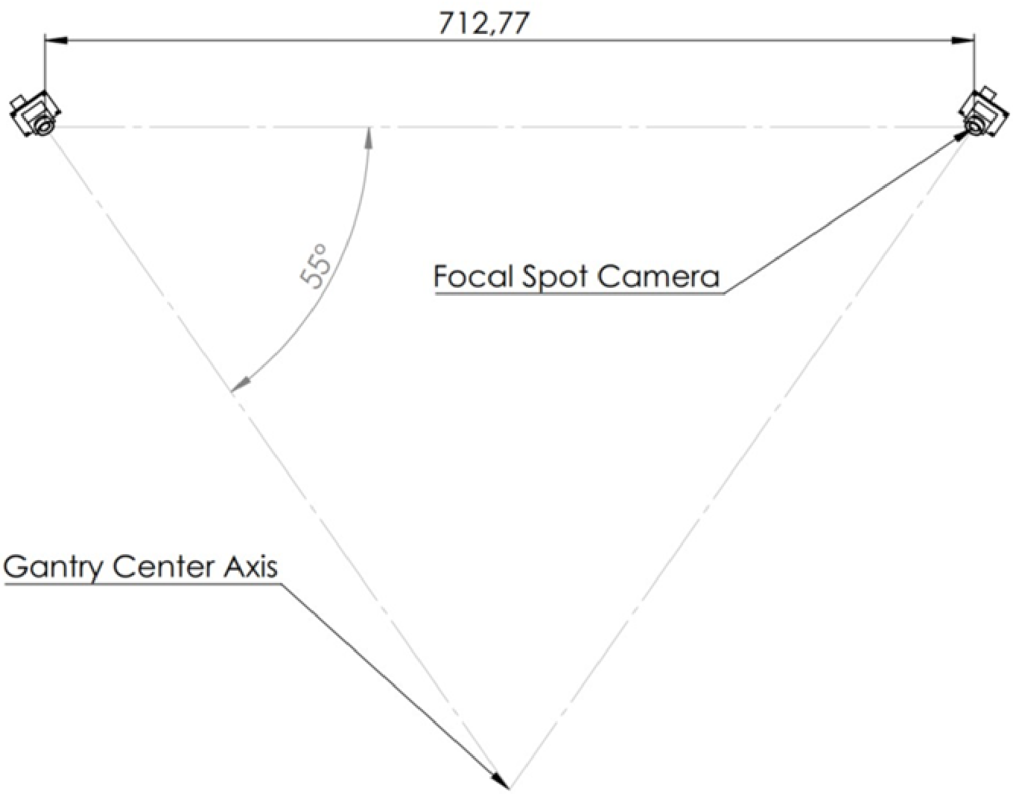

Figure 4.3: The geometrical setup of the two RGB cameras mounted in the Gantry of the IRm.

To achieve the rectification of the cameras, the camera parameters of the setup are first determined via a stereo calibration algorithm.

4.1.2 Stereo Calibration

Using stereo vision for depth calculation requires accurate calibration of the cameras [WCH+92] [Gen79][CF98]. Calibration involves determining both the intrinsic parameters (such as focal length, principal point, and lens distortion) and extrinsic parameters (such as the relative rotation and translation) of the cameras. These parameters define the geometrical relationship between the cameras and are essential for accurate depth calculations[WCH+92].

To perform stereo calibration, the Stereo Camera Calibrator App provided by MATLAB is utilized. This choice is made due to the user-friendly interface the app offers. It allows for easy loading of image data and the refinement of calibration data by discarding problematic image pairs. The app simplifies the calibration process and streamlines the extraction of intrinsic and extrinsic parameters, as depicted in Figure 4.4.

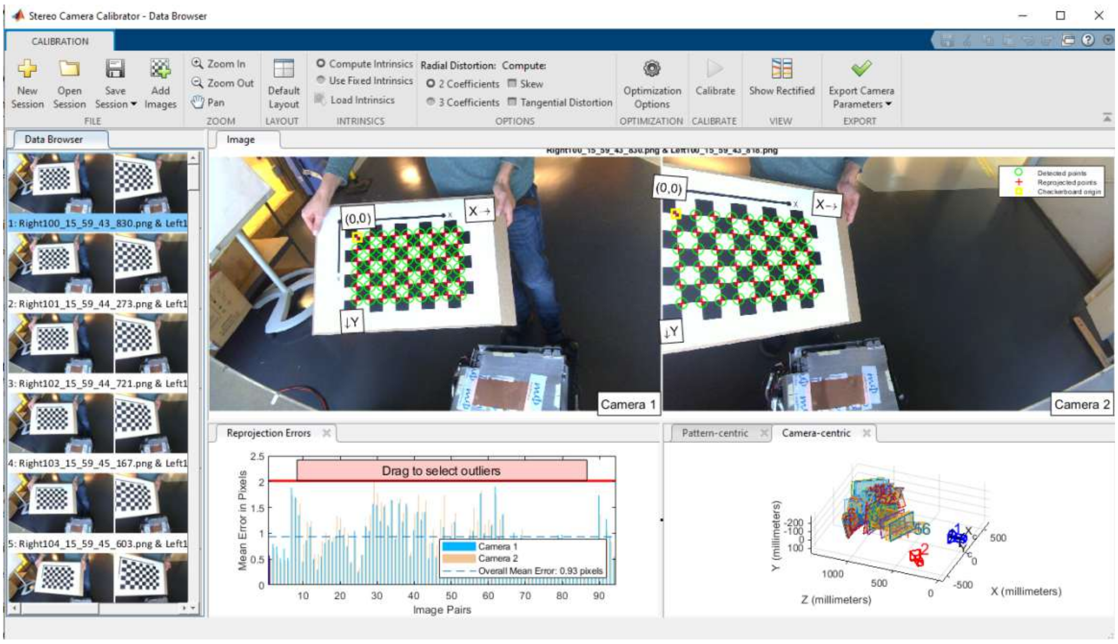

Figure 4.4: An illustration of the RGB camera calibration process using the Stereo Camera Calibrator App in MATLAB. The app aids in extracting intrinsic and extrinsic camera parameters by facilitating the loading of image data and the elimination of undesirable image pairs.

The Stereo Camera Calibrator App is based on the well-established camera calibration algorithm, by Zhang et al. [Zha00]. This algorithm relies on multiple sample images of a planar calibration board with easily detectable structures with known geometry, for example, a chessboard pattern [1] used in this work, which was printed on a Din A3 paper with the geometrical requirements shown in Figure 4.5. It detects the feature points in the images and estimates the extrinsic and intrinsic camera parameters, by means of a closed-form solution. The resulting parameters are then refined by non-linear optimization with the Levenberg-Marquardt algorithm.

After the extrinsic parameters of the left and right camera are known, the rotation matrix R and translation vector T between the cameras can be computed by following simple relations:

$$
\begin{aligned}
R &= R_r (R_l)^T \\
T &= T_r - R T_l
\end{aligned}
\tag{4.3}
$$

where R_r / T_r are the extrinsic parameters of the right camera and R_l / T_l for the left one.

[1] https://de.mathworks.com/help/vision/ug/calibration-patterns.html, accessed 2024-01-13

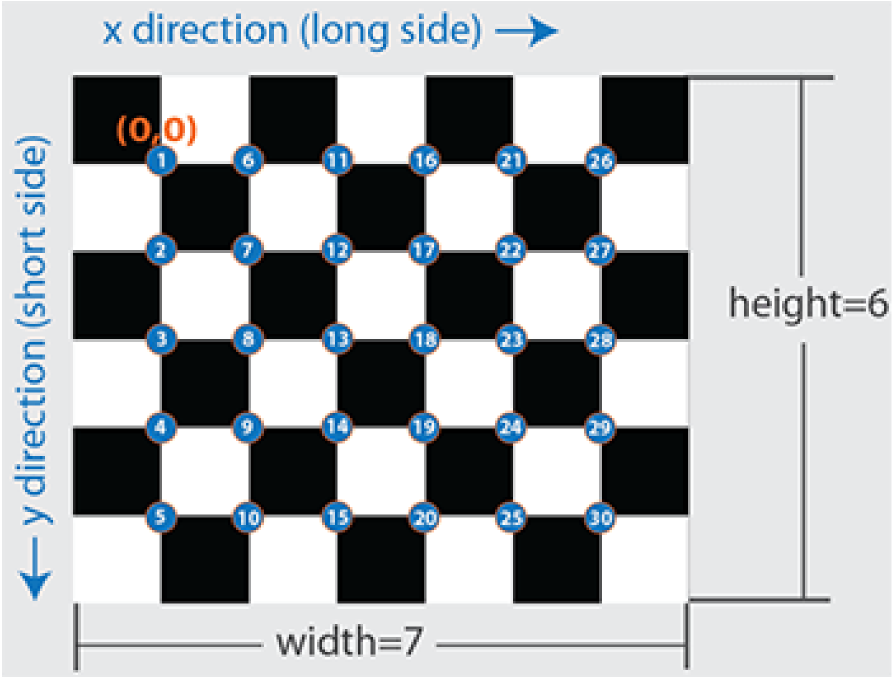

Figure 4.5: An illustration of the geometrical requirements of the chessboard used for the stereo calibration .

Since the software framework of the IRm is written in C++, a small program, capable of importing the results of the Stereo Camera Calibrator App to it is written. After a successful calibration of the stereo setup, where the overall error of the reprojected points is below 0.1 pixels, the rectification of the image data has to be performed.

4.1.3 Rectification

The goal of the rectification process is to align the image rows of the two cameras in order to allow efficient computation of the stereo correspondences, which means detecting the corresponding points between the two images [Aya91]. The reason why this computation is efficient after proper rectification is that the stereo correspondence search problem is reduced to a one-dimensional problem (horizontal rows) [Aya91][BK08]. There are different algorithms to perform the rectification of stereo images, but since the rotation and translation between the cameras are available, the well-established Bouguet's algorithm is used [BK08].

Bouguet's algorithm tries to minimize the reprojection distortion, by splitting the rotation matrix R, received via the stereo calibration and defining the rotation between the two camera image planes, in half. The resulting rotation matrix is be applied to the left camera and the inverted matrix to the right one. This results in the camera's image planes being in a coplanar alignment. To achieve also row alignment, a rotation matrix, which takes the epipoles to infinity is required:

$$R_r align = \begin{bmatrix} (d_1)^T \\ (d_2)^T \\ (d_3)^T \end{bmatrix} \tag{4.4}$$

The first vector of $R_r align$ is defined by the normalized direction of the epipole, which is the same as the translation vector T between the centers of the left and right cameras resulting in:

$$d_1 = \frac{T}{||T||} \tag{4.5}$$

The next direction vector has to be orthogonal to d_1. A good choice to achieve this is by using the cross product of d_1 and the principal ray and normalizing it:

$$d_2 = \frac{[-T_y T_x 0]^T}{\sqrt{T_y^2 + T_x^2}} \tag{4.6}$$

The last vector has to be then orthogonal to d_1 and d_2 so it can be simply calculated by:

$$d_3 = d_1 \times d_2 \tag{4.7}$$

The final row alignment of the cameras is then accomplished by combining the result rotation matrix $R_r align$ with the previous one:

$$R_l = R_r align R_(1/2) R_r = R_r align R_(1/2)^-1 \tag{4.8}$$

These resulting matrices define now how both cameras have to be rotated to achieve an ideal stereo setup. To calculate now the rectified images needed for the search of the stereo correspondences, the image points for both cameras have to be rotated and re-projected accordingly [BK08].

Figure 4.6 shows the calibration images from Figure 4.4 rectified with the discussed algorithm and the visualized epipolar lines show that despite the large rotation between the cameras, the rectification provides reasonable results.

4.1.4 3D point calculation

Utilizing the rectified images obtained through Bouguet's algorithm, the extraction of depth values using Equation 4.2 becomes feasible. A critical step is to locate the corresponding 3D points in both images along the epipolar lines. It's essential for these 3D points to be visible from both cameras, otherwise, two different points are compared resulting in wrong depth values. For this reason, camera setups with closely parallel arrangements are advantageous.

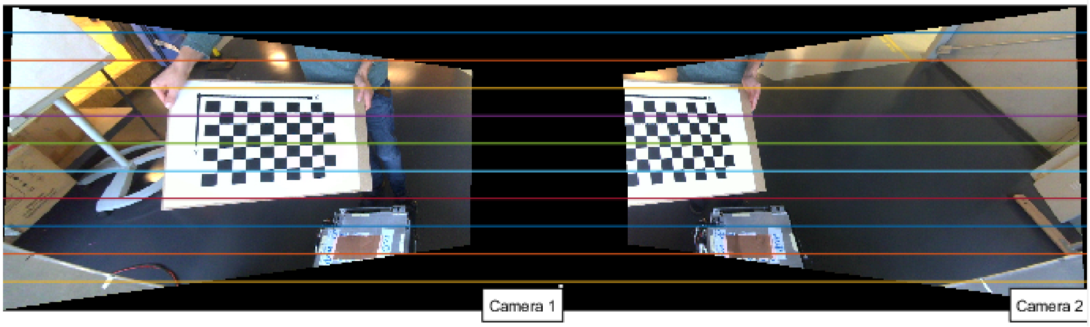

Figure 4.6: Calibration images seen in Figure 4.4 rectified.

To achieve this, the Semiglobal Matching (SGM) method proposed by Hirschmueller [Hir07] is employed in this work. SGM is a hierarchical approach that calculates the matching cost using mutual information and aggregates cost through pathwise optimizations from multiple image directions, approximating a global energy function. The disparity calculation is then determined by a winner-takes-all approach, with potential refinements such as subpixel interpolation and consistency checking available [Hir07].

The decision to use SGM for this project stems from its established reliability and easy integration in C++ through OpenCV. After applying SGM to compute the stereo correspondences, the resulting depth map is calculated and visually represented, offering insights into the spatial distribution of depths within the images.

4.1.5 Results

Different objects at different locations in the FOV of the cameras were imaged to get a rough idea of how well the stereo system is performing and to show the results of the depth maps acquired with the proposed stereo setup.

Figure 4.7: Results of a person squatting in front of the IRm. The first and second images show the images of the two cameras of the stereo setup and the right image the corresponding depth map of them.

Figure 4.7 provides a visual representation of a person squatting in front of the IRm, while Figure 4.8 showcases a wooden phantom placed on a table at a height similar to that of patient treatments. Figure 4.9 depicts a small box with imprinted letters, also

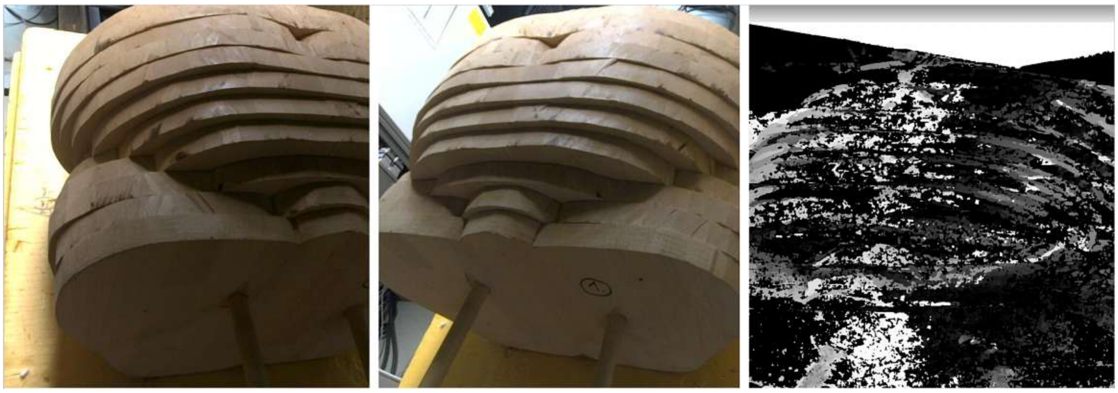

Figure 4.8: Results of the wooden phantom, simulating the size and position of a real patient. The first and second parts show the images of the two cameras of the stereo setup and the right image the corresponding depth map of them.

Figure 4.9: Results of a with letters imprinted box. The first and second parts show the images of the two cameras of the stereo setup and the right image the corresponding depth map of them.

situated on the same table. For improved visualization of the depth maps, a min-max normalization process is applied to enhance the clarity of the results.

From the images, it's evident that none of the datasets yield high-quality depth maps, because the generated depth maps exhibit pronounced noise, numerous regions with missing data, and instances where closer surfaces possess higher depth values than more distant ones. Despite the exploration of various parameters for the SGM matching algorithm and even testing the block matching algorithm developed by Konolige [Kon98], there has been no substantial improvement in the outcomes.

Considering the potential impact of image noise on depth map quality, attempts are made to mitigate this by generating an average depth map across over 100 frames. However, even with this approach, no significant enhancement in the results can be observed, as seen in Figure 4.10.

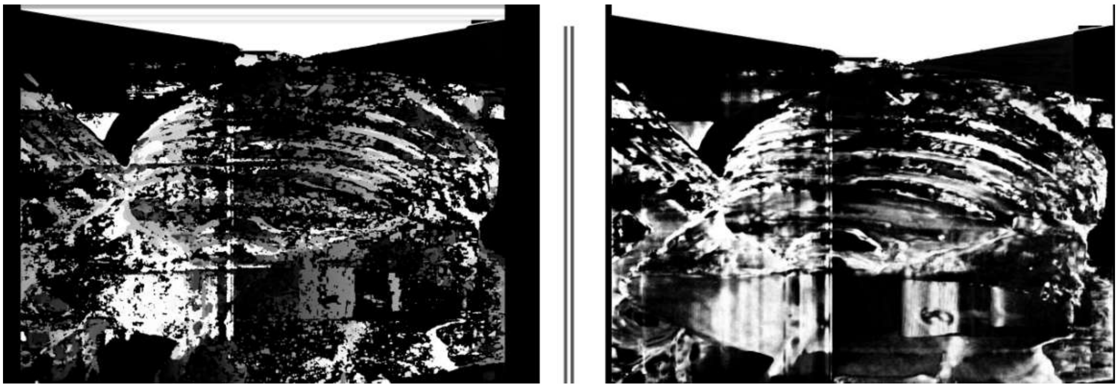

Figure 4.10: The left image shows the depth map of one image pair. The right image is the average depth map of 115 image pairs.

The bad quality of the resulting depth maps, suggests that the geometry of the stereo system presents challenges for the used algorithms due to the wide baseline, large camera angles, and close distance, as shown in Figure 4.3. The obtained results underscore the fact that the existing camera positions and orientations are not optimal for this specific use case. The angle between the cameras leads to a 50 percent overlapping image area, and the close distance exacerbates the differences in perspective between the two cameras, further reducing the overlapping area.

Since changing the geometry of the camera is not an option, the following main options to improve the results could be further investigated:

- **Illumination Enhancement:** Provide the stereo setup with an additional light source for more uniform illumination. Since the light source is above the IRm, it can cast a shadow on the object/patient, which has a negative influence on the stereo correspondence search.

- **Contrast enhancement:** Add a projector to the stereo setup to project patterns on the scene, which can then be detected in both images, and the depth could be calculated for these patterns accurately. Depending on the use case it could also be sufficient to use, for example, a hospital gown containing a clearly visible line at the body center to enable accurate tracking for this region.

- **Improve algorithms:** Improve the current algorithms by adding proper pre- / post-porocessing steps. Investigate other approaches that are more suitable for this kind of stereo setup or try deep learning approaches like the one proposed by Luo et al. [LSU16].

4.2 Structured light RGB-D cameras

The next objective of the feasibility study is to identify a more robust replacement for the current RGB cameras installed in the IRm. This replacement needs to have the capability of producing depth data. Given that the IRm relies on RGB data for various applications, the new solution has to consistently provide high-quality RGB images in addition to depth data. Therefore, it is imperative to locate a suitable RGB-D camera for this purpose.

In addition to image quality, the size of the RGB-D camera posed a significant constraint. It needed to be a suitable replacement for the existing compact 38x38mm RGB camera. Although the dimensions of the RGB-D camera could exceed this size, practical considerations came into play. The integration of a RGB-D camera into the IRm's detector and gantry presented spatial limitations. As a result, a larger RGB-D camera necessitates more extensive mechanical modifications to the overall system.

Another crucial characteristic of the new camera is its "plug-and-play" capability. This means that the camera must come with the necessary software framework for controlling and receiving both RGB and depth data. This decision is driven by the focus of this phase of the feasibility study, which revolves around exploring the potential utilization of an RGB-D camera within the IRm, rather than developing such a system from scratch.

After proper research for a suitable camera only two options, the Intel RealSense Depth Camera D435[2] and the Orbbec Astra Stereo S U3(OAS)[3] remained, both based on the structured light principle. Cameras based on other principles like, for example, time-of-flight (TOF) had to be, despite their higher depth accuracy, discarded due to their lower resolution, which is required to track the FOV of the IRm.

The RGB-D camera from Orbbec, shown in Figure 4.11, has some advantages over the one from Intel, namely a smaller construction form of 65.3mm x 22.5mm x 12.3mm compared to the 90 mm × 25 mm × 25 mm and an around 30% cheaper price.

Intel on the opposite provides the possibility of hardware synchronization and a way more detailed documentation.

.

The camera from Orbbec is chosen since the smallest one from Intel would require unreasonable changes in IRm design.

The OAS camera can be used as a stereo vision setup since it contains also one RGB camera, but it is also a structured light camera using IR providing a depth map without the need for a second camera. In the OAS two IR cameras and an IR projector are included. This projector is used to project a structured light pattern, here an irregular pattern of IR dots into the FOV, see Figure 4.12.

[2]https://www.intelrealsense.com/depth-camera-d435/, accessed 15.09.2022
[3]https://shop.orbbec3d.com/Orbbec-Astra-Stereo-S-U3, accessed 15.09.2022
[4]https://shop.orbbec3d.com/Orbbec-Astra-Stereo-S-U3, accessed 15.09.2022

Figure 4.11: The Orbbec Astra Stereo S U3 camera[4].

This pattern is then used to simplify the stereo-matching problem. Furthermore, it allows the OAS camera to generate valid depth maps in low-textured areas, since the stereo-matching is done on the projected pattern and not on the object texture.

The OAS cameras were then fixed on a wooden rig, simulating the same geometry as in the IRm, and then evaluated for two different use cases discussed in the following.

4.2.1 Surface detection

The primary use case for the Orbbec Astra Stereo S U3 cameras is to capture depth data in conjunction with CT data, which is used to create a patient's surface model. This model is then synchronized and aligned for each projection in Cone Beam CT (CBCT). This synchronization allows for the detection of patient movements between successive projections. These movements are corrected during the CBCT reconstruction process.

To ensure accurate corrections, highly precise motion data is crucial. Hence, the initial step involved confirming the general depth accuracy of the cameras. Depth maps are generated for known objects placed at distances ranging from 0.3m to 1.5m, representing real-world scenarios. To prevent unwanted reflections, a smooth wooden block is used for these measurements.

Two different test measurements are performed for this purpose. For the first one, the object is placed at a certain distance from the camera and then moved for a known distance and the difference in depth values is compared to it. Table 4.2 shows the results of the first measurement, where *base position* and *position* describe the depth value of the two different positions and *diff* their difference. *diff measured* is the difference between the two object positions measured with a fine measuring tape, which is used here as ground truth, and *error* describes the deviation of depth values to it.

For the second measurement, the object is placed at a known distance (*measured distance*) to the camera, and the depth values (*depth value*) are compared to it. *error* describes again the deviation between the two values.

Figure 4.12: The structured light pattern projected by the OAS camera.

The results of these measurements show that the depth error of the data generated with the OAS is not below 1mm, which is required to use the difference directly during the reconstruction of the CBCT images, because the error can't be larger than the volume voxel size without assigning wrong projection data to the voxel during reconstruction. Nevertheless, this level of accuracy suffices for various other applications. In cases where motion follows a consistent pattern, such as respiratory motion, these cameras can still be employed. One possible approach is to develop a model of the respiratory cycle based on the camera data and then align the CBCT projections accordingly. This methodology would enable the generation of motion-free reconstructions for specific phases of the cycle, facilitating even the implementation of 4D imaging techniques. To be able to do this, the surface of the patient has to be constructed beforehand. The OAS provides for this the 3D point cloud of the scene, which is already computed on the camera and can be requested per acquired frame. With the provided point cloud, the Visualization Toolkit(VTK)[5] is used to visualize it and reconstruct the surface mesh out of it, see Figure 4.13.

[5]https://vtk.org/, version 7.1.1

base position[mm]	position2[mm]	diff[mm]	diff measured[mm]	error [mm]
597	646	49	50	1
646	724	78	80	2
498	579	81	84	3
264	348	84	84	<1
584	664	80	84	4
502	789	294	287	7
366	828	462	473	11

Table 4.2: Accuracy test data of the Orbbec Astra Stereo S U3 of object movement at different distances.

depth value[mm]	measured distance[mm]	error [mm]
347	350	3
493	497	4
644	653	9
838	851	13
975	1000	25
1409	1470	61

Table 4.3: Accuracy test data of the Orbbec Astra Stereo S U3 at different distances.

To refine the patient surface, the idea is to combine the point cloud of the two OAS cameras before constructing the mesh. The first idea is to register the point clouds of the two cameras to each other using the Iterative Closest Point (ICP) [SHT09] algorithm provided by matlab[6] and then use the resulting transformation to transform the point cloud of one camera to the coordinate system of the other. This idea did not result in a refined patient surface, since the transformation still produced an Offset of several millimeters between the two point clouds.

The second idea is to calibrate the IR cameras of the two OAS cameras using the stereo camera calibration discussed in the previous section. To use the same chessboard pattern as previously, the projector of the OAS cameras had to be covered since the pattern made the detection of the chessboard impossible. Furthermore, an external IR light in the form of natural light together with a contrast enhancement as an image post-processing step is needed to use the available stereo calibration pipeline. Even with the transformation received by the stereo calibration, it is not possible to align the point clouds properly. One reason for this could be that the 3D point cloud is not based on the position of the available IR camera.

[6]https://de.mathworks.com/help/vision/ug/3-d-point-cloud-registration-and-stitching.html, accessed 15.09.2022

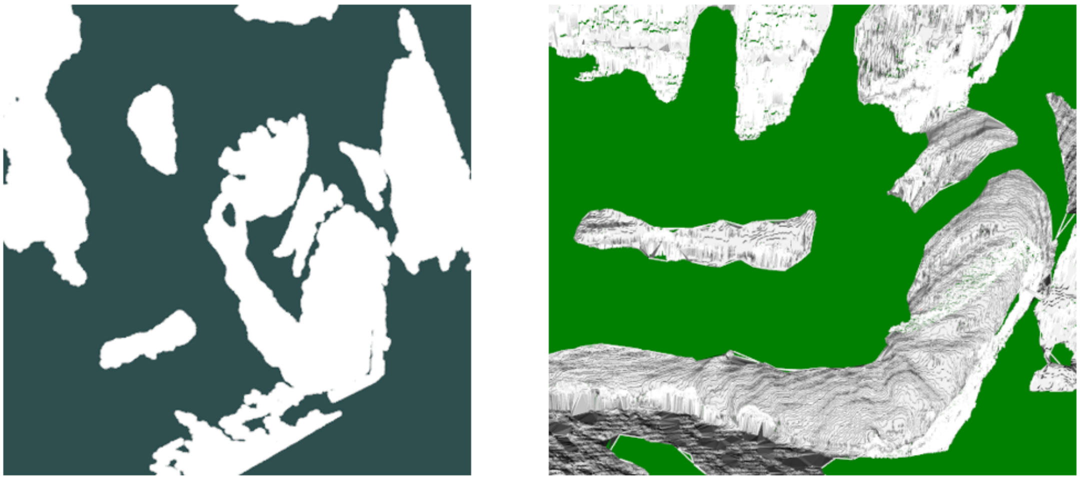

Figure 4.13: The left image shows a visualization of a point cloud acquired with the OAS camera. The right images show a surface mesh generated by such a point cloud.

4.2.2 Marker detection

One main use case of the IRm is intra-operative imaging. During surgical procedures, maintaining a sterile environment necessitates that the patient is covered with a sterile sheet, leaving only the specific area of interest exposed. In this context, utilizing the approach to detect the patient's surface might not be practical. This is because the visible patient surface is minimal due to the coverage, and the sterile sheet might not be closely fitted to the patient's contours, making it unable to accurately capture all potential movements.

For most of the intra-operative use cases, the IRm is used together with a navigation system provided by Brainlab[7]. This navigation system requires a construction with IR marker spheres to be fixed directly onto the patient.

Since the OAS cameras are capable of capturing IR images, the next step is to implement an algorithm to track the markers in the 3D space.

As the first step, the IR markers had to be detected in the images of the two cameras. Since the markers have a spherical shape, a circle detection algorithm, using the Hough transform, is applied to the images using Matlab[8]. This resulted in the circle centers for each image frame and visualization to evaluate the results, see Figure 4.14.

As the next step, out of the circle centers, detected in the two images, the corresponding 3D position had to be estimated. This is done via the Linear-Least square triangulation method described by Hartly et al. [HS97]. Here the camera matrix P is used to derive 2

[7]https://www.brainlab.com/surgery-products/overview-neurosurgery-products/cranial-navigation/, accessed 15.09.2022

[8]https://de.mathworks.com/help/images/ref/imfindcircles.html, accessed 15.09.2022

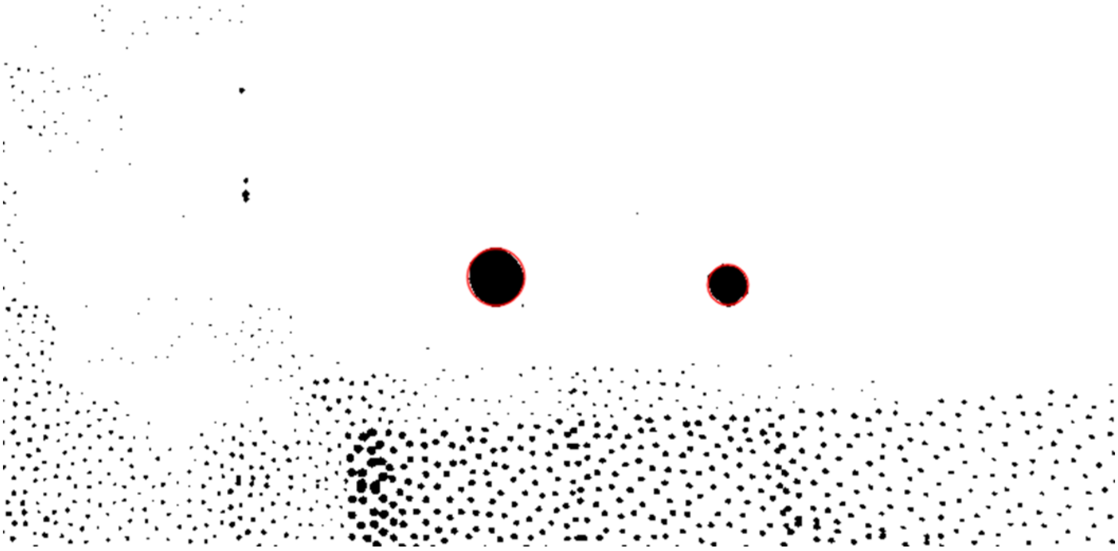

Figure 4.14: The result of the circle detection in the IR images.

linear equations per image, which are then written in the form $Ax = 0$ resulting in:

$$\begin{bmatrix} u\boldsymbol{p_3^T} - \boldsymbol{p_1^T} \\ v\boldsymbol{p_3^T} - \boldsymbol{p_2^T} \\ u'\boldsymbol{p_3'^T} - \boldsymbol{p_1'^T} \\ v'\boldsymbol{p_3'^T} - \boldsymbol{p_2'^T} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \\ \boldsymbol{1} \end{bmatrix} = 0 \tag{4.9}$$

where u and v are coordinates of the observed marker centers, p_i^T defines the i-th row of the camera matrix and $'$ defines the data of the second image. The least-square solution to this problem can then be found by using the Singular Value Decomposition(SVD).

The evaluation of the 3D position revealed instances where the error exceeded 1cm. This substantial error can be attributed to the illumination conditions of the scene, particularly the way it is illuminated by the Orbbec Astra Stereo S U3 (OAS) cameras using the projected point pattern. This method of illumination leads to degradation in the accurate reflection of the marker spheres. Consequently, depending on the marker positions within the field of view, the depiction of the marker spheres in the images deviates from their expected circular shape, as illustrated in Figure 4.15.

When the marker spheres are not accurately depicted as circles in the images, it becomes challenging to correctly detect the center points required for triangulation. In order to enhance the accuracy of position estimation, it's imperative to disable the projector and consider incorporating an additional IR light source. This added light source should provide uniform illumination across the scene. By doing so, the quality of the position estimation can be significantly improved. This change in setup would mitigate the issues arising from the irregular reflections caused by the projected point pattern and lead to more reliable and accurate results.

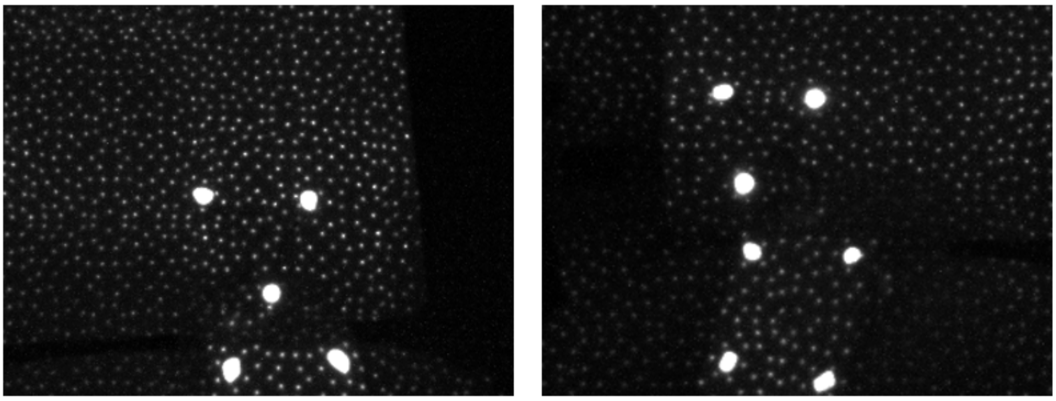

Figure 4.15: Example of marker spheres depicted not circular anymore in the images due to the non-uniform illumination.

4.3 IR stereo cameras

As the last step of the feasibility study, a camera system focused on high-precision measurements is evaluated namely Optitrack Primex 13 cameras[9].

Unlike the previous two systems, the primary objective with the Primex 13 cameras is not the detection of the patient's surface. Instead, the focus is on precisely tracking 3D IR markers within the field of view. The choice of the Primex 13 cameras is motivated by the higher accuracy promised by Optitrack for this tracking functionality in comparison to other similar systems. The paramount purpose of this accurate tracking is to facilitate the compensation of motion in 3D reconstructions. Given that precise information about patient motion is essential for this compensation, the capabilities of the Primex 13 cameras align well with the requirements of the task at hand.

The Primex 13 camera is an infrared camera equipped with a ring of 10 infrared LEDs operating at 850 nm for ensuring uniform illumination within the FOV. It also incorporates an 850 nm band-pass filter to minimize interference from external light sources. In terms of size, the Primex 13 camera is larger compared to the previously evaluated cameras, with dimensions of 6.9 cm x 6.9 cm x 5.3 cm, see Figure 4.16. It's worth noting that the camera is designed to be powered using Power over Ethernet (PoE), allowing for convenient hardware synchronization of multiple cameras.

With an impressive high framerate of 240 Hz, the Primex 13 camera can perform on-board segmentation of markers and marker centers within the captured images. This eliminates the need for extensive image processing on the workstation side. Consequently, data transmission over the network is significantly reduced, as only the marker center positions and sizes need to be transmitted, rather than transmitting the entire image data.

[9]https://optitrack.com/cameras/primex-13/, accessed 20.09.2022

Resolution	Framerate	3D Accuracy	Range	FOV
1280x1024	240Hz	$+/-0.20mm$	16m	56° horizontal and 46° vertical

Table 4.4: Specifications of the Optitrack PrimeX13 camera.

Key specifications of the Primex 13 camera are summarized in Table 4.4. The camera has the capability to detect various shapes of IR reflective markers, but it achieves the highest precision with marker spheres coated with 3M 7610 Reflective surface coating. This combination of features and capabilities makes the Primex 13 camera suitable for the accurate tracking of 3D IR markers in the FOV, fulfilling the specific requirements of motion compensation in 3D reconstructions.

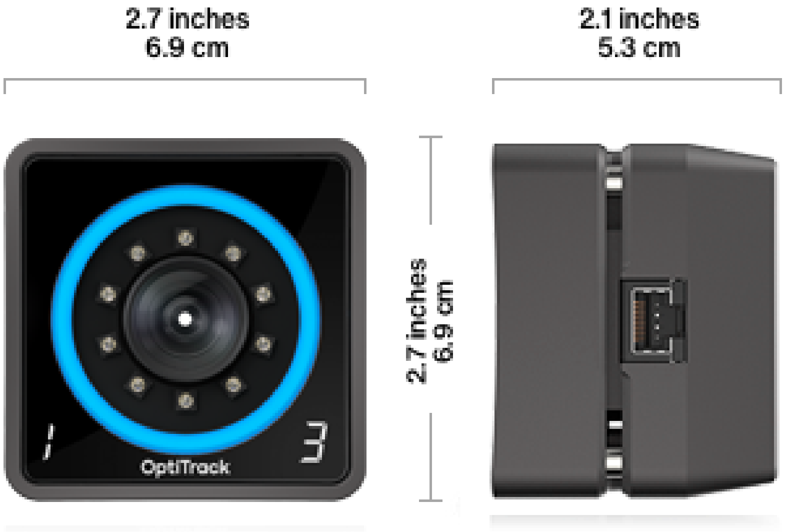

Figure 4.16: The dimensions of the Optitrack Primex 13 camera.

The stated 3D tracking accuracy of the Primex 13 is estimated by Optitrack for a 9mx9m tracking area with multiple cameras. In the scope of this usability study, a stereo setup with a geometry shown in Figure 4.3 must be used. Therefore the first step is to build such a setup with the Primex 13 and evaluate its 3D accuracy. To be able to do this the Primex 13 had to be calibrated first. Optitrack offers the software Motive to communicate with, synchronize, and calibrate the Primex 13 and to calculate the 3D marker information.

4.3.1 Calibration

Using Motive, the calibration of the Primex 13 stereo setup can be done by performing the following steps:

- If possible remove all objects reflecting IR light from the FOV of the camera setup.

[9]https://optitrack.com/software/motive/, accessed 20.09.2022

- Apply a mask to ignore the remaining objects during calibration and tracking. The pixel information in the masked areas will be then completely filtered out from the 2D data, which is why it is important to calibrate the setup in an environment with as few extraneous reflections as possible.
- Wave a calibration wand in front of the cameras. It is important to wave it slowly and with varying orientations through the whole capture volume. The calibration wand consists of 3 marker spheres with known geometry, see Figure 4.17, which is then used by Motive to calculate the 3D position and Orientation of the cameras out of the 2D image samples.
- After enough data samples are generated for the calibration, indicated by a green camera LED ring, the camera positions are calculated and a calibration report is shown, see Figure 4.18. In this dialog especially the parameter "Overall Wand Error" is important, because it indicates if the waving of the wand is done properly.
- After applying the result, the last step of the calibration is to put a calibration square in the volume. Motive will detect it automatically and depending on its orientation and location it will span up the camera coordinate system accordingly.

Figure 4.17: The Optitrack calibration wand CW-250 [10].

4.3.2 Results

Following a successful calibration of the stereo setup, the tracking accuracy of a 3D marker sphere is assessed. This evaluation process is executed using a similar methodology as employed with the Orbbec Astra Stereo S U3 cameras. A marker is positioned within the capture volume and then displaced over a predetermined distance.

One notable distinction between this measurement and the previous one lies in the evaluation approach. In the prior case, the assessment focuses solely on the distance between the camera and the object. Conversely, in this instance, the evaluation is performed in the 3D position within the camera coordinate system of the stereo setup. Consequently, the Euclidean distance is computed for each movement and subsequently

[10] https://optitrack.com/accessories/calibration-tools/, accessed 20.09.2022

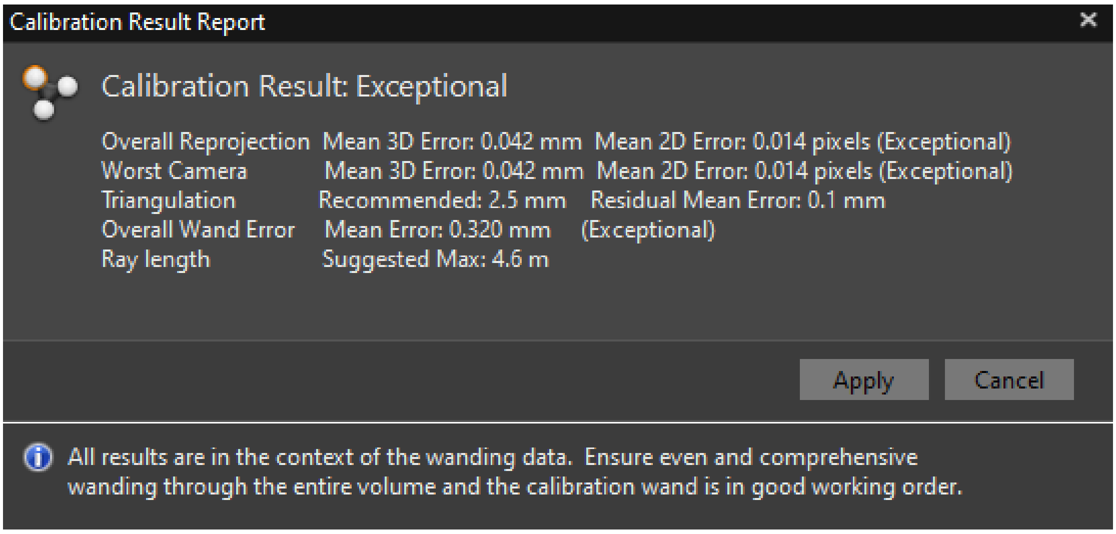

Figure 4.18: The result of an Optitrack calibration.

euclidean distance CCS [mm]	distance measured[mm]	error [mm]
100.2	100	0.2
300.4	300	0.4
199.7	200	0.3
500.3	500	0.3
399.6	400	0.4
299,8	300	0.2

Table 4.5: Accuracy test data of the Optitrack Primex13 cameras at different distances.

compared to the measurements obtained using a measuring tape. This method allows for a comprehensive assessment of the tracking accuracy of the Primex 13 cameras, providing insights into the reliability of their 3D marker sphere tracking capabilities. Figure 4.5 shows the results of these measurements.

The results of this test show an improvement compared to any of the previously examined setups. It's worth highlighting that the measuring tape utilized for the comparison had a measuring accuracy of 1mm, introducing an additional source of inaccuracy to the measurements. As such, the actual accuracy of the system could potentially be even better than what is shown in the evaluation results.

To be able to assess the accuracy of the system further, data of a static marker object is acquired, and the noise of the signal is measured. Figure 4.19 shows this data. The distance of the marker to the cameras is 3m and the 3D accuracy for distance to the cameras is still $+/-0.19mm$.

Even with these already very promising results, the data is not acquired in perfect circumstances. For one thing, the cameras are mounted on two Walimex lamp tripods [10], see Figure 4.20, which are not perfectly rigid and vibrations of the floor, caused by people passing by, degraded the calibration of the setup over time. Furthermore, the Optitrack IR cameras work best if no additional light source illuminates the marker spheres, but there is no way to prevent daylight from entering the room of the test setup.

Indeed, the obtained results provide a strong indication that the Primex 13 cameras perform exceptionally well under favorable conditions. This level of accuracy is particularly encouraging considering the intended application for motion compensation during CBCT acquisition. Given the demonstrated tracking accuracy, it is reasonable to anticipate that the Primex 13 cameras would effectively serve their purpose in compensating for patient motion during the CBCT process.

Despite the promising results, the price of the Primex 13 is with 2.500$[11] is around ten times higher than the OAS camera and it is not able to provide RGB images. Furthermore, the bigger size of the camera would also result in the need for significant mechanical changes for integration in the IRm.

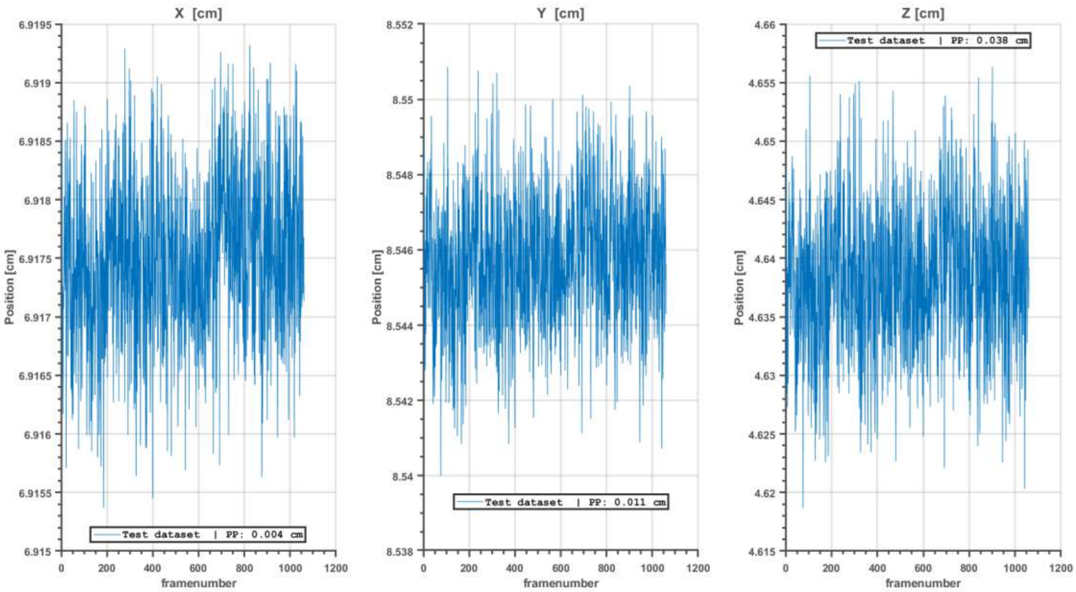

Figure 4.19: Noise in the tracking data of a static marker object.

4.4 Outcome

The primary objective of this feasibility study is to explore various camera options that can be integrated into the IRm system to enable motion tracking during CBCT acquisitions.

[10] https://www.walimex.biz/Walimex-pro-Set-of-4-WT-806-Lamp-Tripods-256cm, accessed 20.09.2022
[11] https://optitrack.com/cameras/primex-13/, accessed 20.09.2022

Figure 4.20: Test setup of the Primex 13 cameras.

In pursuit of this goal, three distinct camera types are subjected to validation for the specific use case outlined earlier. Based on the results obtained from this validation, a final decision is made to select one of these camera types for integration within the IRm.

Below, the advantages and disadvantages of the investigated cameras are summarized. This summary aims to provide a concise comparison between these cameras, aiding in the decision-making process:

RGB cameras: The primary advantage of using RGB cameras within the IRm system is their low price of only 60 euros. However, no depth map can be generated with this stereo vision setup with the used tools. This is attributed to factors such as the short distance, wide baseline, and notably, the substantial skew between the cameras. This skew renders the detection of stereo correspondences unfeasible using conventional algorithms.

To address these issues, one potential approach involves refining the algorithms to better suit this unique stereo vision geometry. Additionally, the introduction of a projector into the setup, as suggested by Konolige et al. [Kon10], could enhance the detection of stereo correspondences by improving contrast. Even with these enhancements, the question of whether the depth estimation would be accurate enough for direct motion compensation remains uncertain, although it would undoubtedly demand significant development efforts.

An additional challenge arises from the processing constraints. The RGB image processing needs to be performed on the same industrial PC (IPC) responsible for processing the CBCT imaging pipeline. Consequently, the computer vision algorithms utilized must be optimized for efficiency, ensuring that they do not interfere with the CBCT generation process.

Furthermore, calibrating the RGB cameras with the imaging coordinate system of the IRm presents additional complexity compared to the other two camera options. The capability of the other options to segment IR marker spheres, albeit to a certain extent, simplifies the calibration process for those cases.

Advantages:

- Hardware already integrated into the IRm
- Inexpensive, 65 euro [12]
- Only a projector has to be added in the gantry of the IRm
- RGB data available together with the depth data

Disadvantages:

- Patient surface has to be visible, f.e. blankets used during operations could impair the tracking results
- 3D accuracy > 1mm
- Calibration between CCS and ICS of the IRm required
- All computations are performed on the IPC
- High development effort
- No hardware synchronization available

In summary, while the use of existing RGB cameras offers convenience, the challenges associated with stereo vision performance, depth estimation accuracy, and integration within the IPC environment pose significant hurdles. Careful algorithm optimization, potential hardware enhancements, and a comprehensive calibration approach would be necessary to make this option feasible and accurate for motion compensation within the CBCT process.

[12] https://www.amazon.de/stores/ELPUSBKamera/ELP-SUSB1080P01Series_2MegapixelUSBKamera/page/4804BCFB-A077-4DCC-90E3-F7445083FBA2, accessed 2024-01-13

RGB-D cameras: The Orbbec Astra Stereo S U3 cameras stand out as the most versatile option within this study. They are capable of providing RGB, depth, and IR image streams, along with the depth map converted into a 3D point cloud. These computations are performed directly on the camera, which has benefits in terms of reducing the potential workload on the IPC. However, this onboard processing results in significant heat generation, potentially posing a challenge for integration into the IRm system. The potential for overheating, especially if the cameras are covered in the IRm setup, needs to be carefully considered to prevent any adverse effects on surrounding electronics.

Long-term use of these cameras revealed reliability issues, as they would intermittently turn off and only respond after a period of being idle. This behavior suggests a possible link between the high temperature generated and these disruptions in performance.

If the existing RGB cameras were to be replaced with the Orbbec Astra Stereo S U3 cameras, certain challenges would arise. Continuously streaming RGB, IR, or depth data isn't feasible due to the IRm's specific workflows that require RGB information. Moreover, the small size of the OAS cameras would enable their integration with minor changes to the IRm's gantry. However, to ensure accurate marker detection, an additional IR light source would need to be incorporated.

An additional drawback of the OAS cameras is their provided software development kit (SDK)[13], which lacks extensive options and documentation. This limitation can potentially impede the development and customization of software interfaces to meet specific IRm requirements.

Advantages:

- RGB-, depth- and IR image data available
- 3D point cloud calculated on-board
- Inexpensive, 169.99$ per camera[14]
- No major mechanical changes needed

Disadvantages:

- Not reliable during long-term usage
- Cannot stream multiple different image types at the same time
- For IR marker detection an additional IR light source required
- Bad software interface

[13] https://orbbec3d.com/index/download.html, accessed 20.09.2022
[14] https://shop.orbbec3d.com/Orbbec-Astra-Stereo-S-U3, accessed 20.09.2022

- High temperature
- No hardware synchronization available

In conclusion, while the Orbbec Astra Stereo S U3 cameras offer impressive versatility, the challenges related to temperature, reliability, and compatibility with the IRm's workflows and infrastructure must be carefully addressed. The decision to integrate these cameras would necessitate resolving these issues to fully leverage their capabilities for motion tracking within the CBCT process.

IR cameras: The Optitrack Primex 13 cameras present several noteworthy advantages. The most prominent among these is their exceptional 3D tracking accuracy, which addresses the core requirement for motion tracking within the CBCT process. Additionally, the Optitrack Motive software provides a user-friendly interface for effortless camera calibration, marker management, and seamless communication with external software applications.

However, these benefits come with significant associated costs, with the purchase price of the Primex 13 cameras being over 10 times that of the Orbbec Astra Stereo S U3 cameras. Furthermore, the substantial physical size of the Primex 13 cameras makes direct integration into the IRm unfeasible. This challenge would necessitate the development of a new external mounting concept, as modifying the existing gantry to accommodate these larger cameras has proven impractical.

Moreover, incorporating the Primex 13 cameras into the IRm setup would also require the addition of a Power over Ethernet (PoE) switch. This switch would be essential for providing the power supply, communication, and synchronization capabilities to the cameras, adding further complexity and infrastructure requirements to the integration process.

Advantages:

- On-board segmentation of the markers
- High 3D accuracy
- Good software interface
- High frame rates possible
- Hardware synchronization available

Disadvantages:

- Expensive, 2.500$[15] per camera

[15] https://optitrack.com/cameras/primex-13/, accessed 20.09.2022

- Significant mechanical changes needed
- Additional hardware needed
- High temperature

In summary, while the Primex 13 cameras offer exceptional tracking accuracy and a robust software interface, the high purchase costs, size constraints, and infrastructure demands must be considered when making the decision to integrate them into the IRm. A comprehensive assessment of the benefits and challenges associated with this option is essential to ensure its suitability for motion tracking in the CBCT context.

Selection The decision has been made to proceed with the development of a computer vision system that utilizes the Optitrack Primex 13 cameras for tracking patient motion within the IRm.

This choice is primarily motivated by the combination of the Primex 13 cameras exhibiting the lowest development risk while also achieving the highest accuracy among the options. Additionally, the estimated development time for this solution is notably shorter compared to the alternative possibilities. This efficiency stems from the reliability and ease of integration of the provided software into the existing software framework of medPhoton.

Furthermore, the decision is influenced by the realization that the IRm's main use case involves intra-surgical imaging. In these scenarios, IR markers are often directly attached to the patient, and the sterile environment typically requires the patient to be covered by sheets, making surface detection more complex.

CHAPTER 5

Implementation

The formulated crucial high-level system requirements serve as a foundation for guiding the design and implementation process and are used to derive the following requirements:

- Fully integrated into the IRm
- Retrofitting capability
- No restriction of other IRm functionalities
- 3D accuracy in the sub-millimeter range
- Self-contained functionality
- Closed construction
- Ability to capture patient position data even when draped
- Synchronization of motion data with projection frames

Fully integrated into the IRm: The mP tracking module must be seamlessly integrated into the IRm system. This entails a robust mechanical connection with the gantry of the IRm, and its functionality should be fully incorporated within the mP software framework. This integration ensures smooth communication between the various components and the tracking module.

Retrofitting capability: The module should be adaptable for integration into both new and older models of the IRm. The integration process should be straightforward, facilitating retrofits even in environments where specialized tools might not be available.

No restriction of other IRm functionalities: The mP tracking module must not hinder any existing IRm functionalities. It should not interfere with the CBCT imaging pipeline or obstruct the field of view (FOV). Furthermore, the module's dimensions must adhere to the existing parameters, ensuring that the IRm can navigate through the same spaces as before.

3D accuracy in sub-millimeter range: The system's core competency must be the real-time detection of patient motion with sub-millimeter accuracy, aligning with the IRm's acquisition rate of at least 30 Hz. This matches the IRm's maximum image acquisition frequency and accommodates for even faster motion scenarios.

Self-contained functionality: The module must possess the capability to perform all necessary functions for delivering an accurate 3D position of the IR marker. This mandates that all computational tasks are carried out by the module itself, guaranteeing the uninterrupted availability of the IPC for the CBCT imaging pipeline.

Closed construction: Considering the module's application in sterile environments, particularly during surgical procedures, the design necessitates a closed construction. This closed design ensures that all surfaces of the tracking module can be effectively sterilized without compromising the system's integrity.

Ability to capture patient position data even when draped: A crucial requirement for the mP tracking module is its capability to capture patient position data even when the patient is fully draped. This functionality is particularly vital for scenarios like spine surgeries, where patients are typically covered by surgical drapes during image capture. Ensuring accurate motion tracking in such scenarios is essential for maintaining the precision and reliability of the motion compensation process during procedures like spine surgeries.

Synchronization of motion data with projection frames: For effective motion compensation, it's imperative that the system incorporates mechanisms to synchronize motion data with each acquired projection frame. This synchronization data must be securely stored within the frame data itself, allowing for reconstructions at later stages. By tightly aligning motion data with projection frames, the system ensures that accurate compensation can be applied during image reconstruction, maintaining the fidelity and precision of the final images.

Based on these outlined requirements, a comprehensive hardware, mechanical, and software concept is developed.

5.1 Hardware concept

The main electric components of the tracking module are two Primex 13W cameras, a small PC, a POE switch, and a DC-DC converter. The Primex 13W cameras are network cameras, which require an individual camera network in the tracking module. Their data communication as well as power supply is done over Ethernet cables connected to a PoE switch. Each camera requires 15.4 watt input power therefore the switch and cables have to conform to the IEEE 802.3at standard [1] .

For all the build-in components a main requirement besides the functionality is a small dimension since there is only limited space available for the tracking module. For this reason a small PC capable of performing all the processing required for the tracking module. The best compromise of size and power is found in an Intel NUC[2].

The NUC acts on one side as an interface between the IRm software over the IRms internal WiFi. On the other side, it is running Motive to communicate with the Primex 13W cameras, calculate 3D tracking data, and manage marker objects. One problem is that there is only a 48V power supply available on the Gantry, which is ideal for the PoE switch but not for the NUC, which requires 24V. For this reason, a DC-DC converter is needed for the tracking module to ensure a proper power supply for the NUC.

Figure 5.1 shows a visualization of the different hardware components required for the mP tracking module.

5.1.1 Mechanical design

All these components must be contained in a housing that can be mechanically and electrically attached to the top area of the gantry's rigid middle section, as shown in the right image of Figure 5.2. Therefore, proper housing must be designed, and the gantry cover and underlying structure must be modified to comprise the required electrical and mechanical interface. All the mechanical constructions of the tracking module were done by Joachim Spielbichler from MedPhoton and the author only helped in finding the proper requirements and needed improvements during the prototype development process.

These improvements are necessary due to the closed construction requirement. Because of this, a translucent cover has to be in front of the Primex 13W cameras. Since the IR light source is also behind this cover, parts of the light rays get reflected by the cover and deteriorate the tracking quality, see the left image of Figure 5.3. This issue is solved by the research and use of a special anti-reflective acrylic cover plate[3]. The result after the change of the cover plate can be seen in the right image of Figure 5.3.

[1] IEEE Std 802.3-2015 (Revision of IEEE Std 802.3-2012), 2016

[2] https://www.intel.com/content/www/us/en/products/details/nuc/mini-pcs/products.html, accessed 22.09.2022

[3] https://www.go-ttv.com/de/materialien/antiflex/antiflex-mc-st/, accessed 22.09.2022

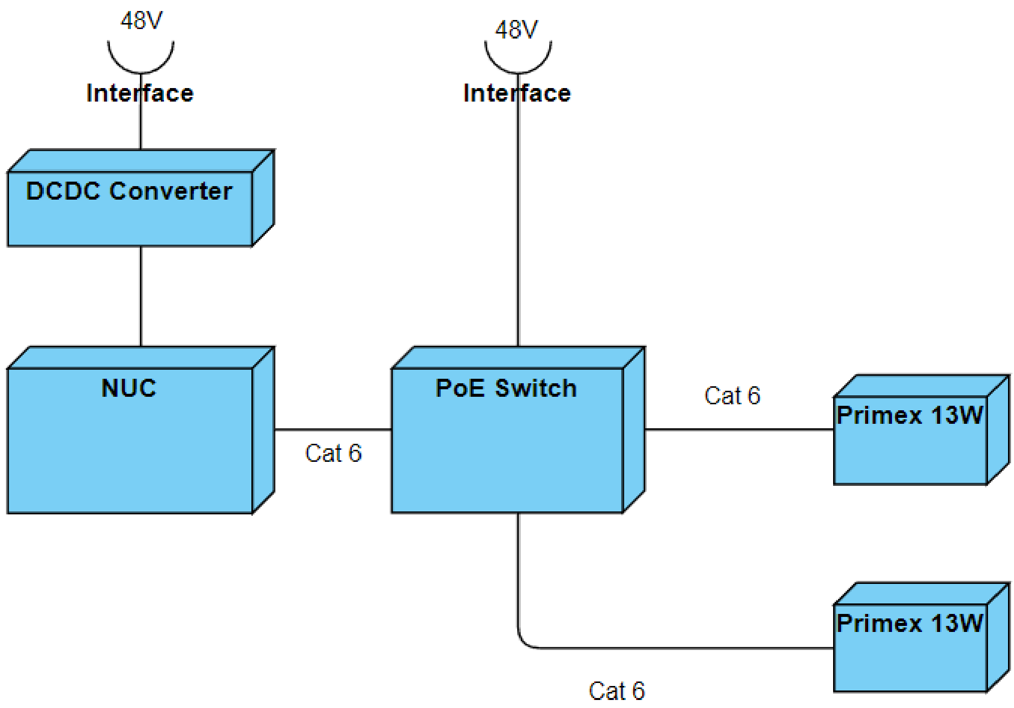

Figure 5.1: Hardware design of the mP tracking module.

5.2 Software design

A rough overview of the basic software design of the mP tracking module is given in the following:

It involves six different main components, the Motive server, mPc, TCS, RTSS, FPPI, and CBCT-M. The Motive server runs on the NUC in the tracking module and is provided by Optitrack. It is responsible for providing the tracking data to the MedPhotonControls(mPc) through a WiFi connection. The tracking data is received at a configurable speed at mPc, immediately transformed into the proper coordinate system and posted to TCS, which is responsible for the exchange of properties in the mP software framework. mPc is also responsible for activating/deactivating the tracking and setting different tracking properties. When the tracking is active, mPc also updates the Imaging Protocol, containing all relevant setting for the acquisition, and send it to the RTSS. RTSS forwards this protocol to FPPI and CBCT-M, allowing the control of the tracking functionality on their side.

During the acquisition of a CBCT image, the FPPI, responsible for the communication with the detector, then searches for each acquired projection with the best fitting synchronized tracking data and stores it into the metadata of the projection.

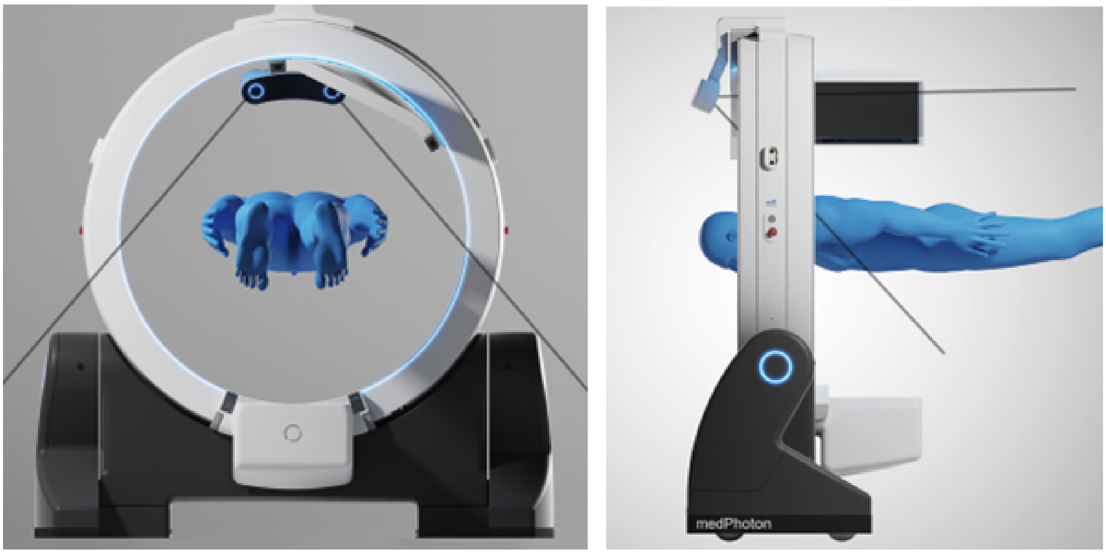

Figure 5.2: Mechanical design of the mP tracking module.

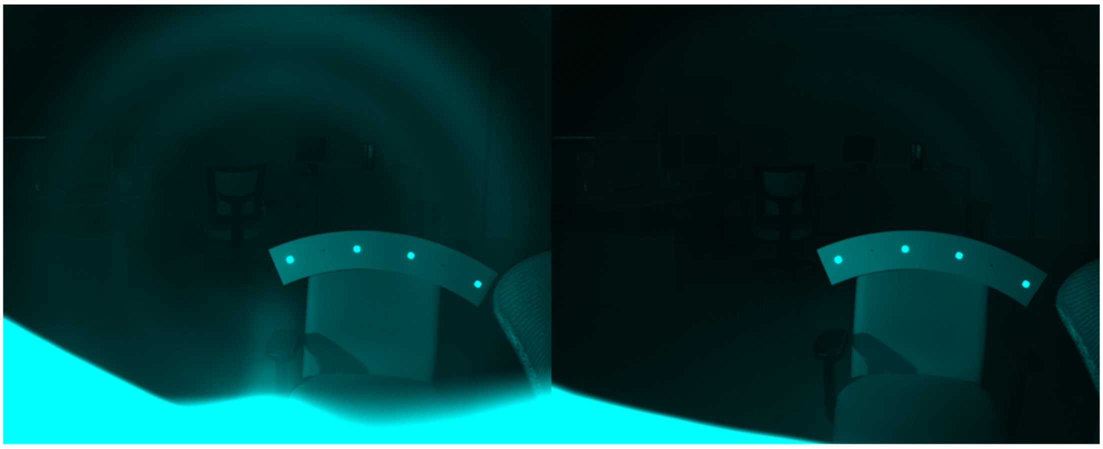

Figure 5.3: Mechanical design of the mP tracking module.

The metadata of the projection is then passed to the CBCT-M where the reconstruction of the volume is done. Here the tracking data is read and the detected motion is compensated during the reconstruction process.

5.3 Software

The motion compensation process involves five key components of the mP software framework, each with specific functionalities that contribute to this endeavor. The following is a brief overview of their roles within this context:

TCS: TCS serves as a central online database for properties from connected software applications. It facilitates coordinated registration of software applications on property events, as well as reading and writing properties. It also offers a configurable property history length, enabling access to older property values.

mPc: mPc encompasses various user interface functionalities and interfaces with external devices like RGB cameras.

RTSS: RTSS functions as the workflow engine of the mP software framework. It manages workflow orchestration by coordinating different software components.

FPPI: FPPI is responsible for image acquisition using the IRm's flat panel detector. It synchronizes acquired images with metadata retrieved from the Imaging Ring Control System.

CBCT-M: CBCT-M facilitates the generation of volumetric images from planar projection images acquired by the IRm. The creation of CBCT volumes is achieved through reconstruction using a convolution back-projection approach.

These implementations can be categorized into five distinct work packages namely, data interface, calibration, synchronization, reconstruction pipeline, and user interface, discussed in the following.

5.3.1 Data interface

The tracking module is equipped to stream tracking data over a network using the User Datagram Protocol (UDP). UDP operates atop the Internet Protocol (IP) and serves as a mechanism for applications to transmit information to other programs with minimal protocol overhead. Unlike some other protocols, UDP lacks error checking and extensive control mechanisms, focusing on simplicity and speed [Pos80].

The streaming of the tracking data can be activated in the motive server by setting a proper network, in case of the mP tracking module this is the IRm WiFi. Using the NatNet SDK[4] the mPc is extended to connect as a client to the motive server and receive the data packets containing real-time tracking information. Furthermore, the functionality to send remote commands to the server is also added to allow basic remote control of motive from the client.

The NatNet client is run within mPc as a separate thread. This architectural choice is made to alleviate the workload of the main thread, thereby ensuring that mPc can efficiently execute all of its various tasks without compromising performance. This separation of threads is particularly crucial due to the potential high frame rate of the Primex 13 cameras. Running the NatNet client in a separate thread prevents it

[4]https://optitrack.com/software/natnet-sdk/, accessed 22.09.2022

from overwhelming the main thread with processing tasks and guarantees that mPc can maintain its functionality without degradation.

Upon receiving a data packet from the NatNet client, a synchronized time stamp is generated. This time stamp serves to ensure temporal alignment between the tracking data and other components of the system, a critical factor for accurate motion compensation.

Subsequently, the received data is subjected to filtering to extract the 3D positions and orientations of the tracked marker objects. These filtered values are then transformed into the Imaging Coordinate System to ensure consistency with the rest of the system.

The filtered and transformed data, along with the synchronized time stamp, is then published to the TCS. This communication enables the seamless integration of the tracking data within the software framework, ensuring that it is available for subsequent stages of the motion compensation process.

During any acquisition involving the active tracking module, the FPPI undertakes a search in the property history of the tracking data properties stored within the TCS. The goal of this search is to identify the most optimally synchronized values. Once these values are identified, they are added to the metadata of the acquired projections. This integration guarantees that the synchronized tracking data is appropriately associated with each projection, enabling accurate motion compensation during the subsequent image reconstruction phase.

5.3.2 Calibration

The calibration process for the mP tracking module is divided into two distinct components.

The initial part involves the calibration of the Optitrack cameras themselves. This step primarily entails calibrating the geometric relationship between the cameras and establishing the internal camera coordinate system. The specific procedures required to accomplish this calibration are elaborated upon in Section 4.3.1.

Once the cameras have been effectively calibrated, they are capable of producing accurate 3D tracking data. However, this data is not yet ready for use in motion compensation during the CBCT reconstruction process. To make this data usable for such compensation, an additional step is necessary: the transformation of the tracking data into the appropriate coordinate system. This transformation ensures that the tracking data aligns properly with the coordinate system utilized within the CBCT imaging process.

Imaging ring coordinate system transform

The CBCT volume is reconstructed in the imaging ring coordinate system and therefore the tracking information also must be available in this coordinate system. To be able to accomplish this, another calibration step of the tracking module is required.

The requirement for such a calibration is that an object can be detected accurately by both systems. This means it has to contain an IR marker, needed for the tracking module, which must be connected via known geometrical relation to the object, containing material easily detectable by X-ray images, like for example metal balls. It is important to consider that it has to consist of at least 3 independent spheres, which is required to resolve the orientation between the coordinate systems.

This means it is also required, that the CBCT system can detect the marker objects. For this purpose, a few X-ray images were acquired to assess if the spheres can be segmented in them. Figure 5.4 shows such images and it can be seen that the spheres are clearly separable from the background. As a first attempt, the Optitrack calibration square is used for this purpose, but due to its metallic frame, it is impossible to segment the markers in the X-ray images. Therefore, a replica of the calibration square, consisting of acrylic resin instead of metal, is made. It enables an easy segmentation of the spheres, see Figure 5.5, and can also be used for the calibration of the CCS.

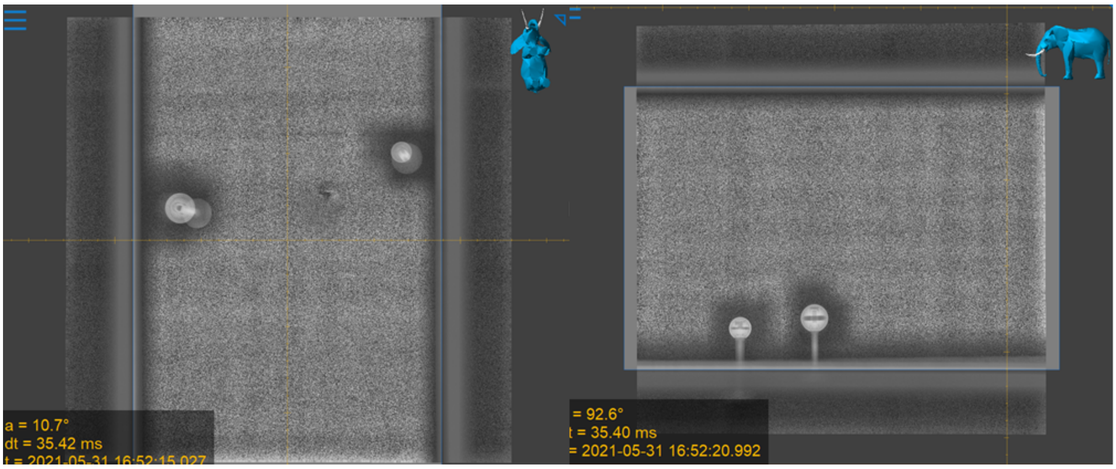

Figure 5.4: Anteroposterior(AP) and lateral(LAT) images of IR marker spheres.

After this step, a concept for the calibration is developed and its single steps are listed in Table 5.1.

As the first step, it is important to place the object properly in the FOV of the IRm. This means that it should be close to the isocenter, and the side with the shortest distance between the spheres has to be aligned with the x-axis of the IRm and pointed in the negative direction of the y-axis. This alignment is crucial since the y- and z-axes of the CSS and ICS are transposed. Therefore, a transformation is applied to the point cloud data of the tracking module to account for this and improve the results of the Iterative Closest Point (ICP) algorithm.

The next step involves acquiring two X-ray images. It's essential to consider that there should be around a 90° angular difference between the source and detector, enhancing the quality of the 3D position calculation in the ICS. Furthermore, a lower X-ray dose needs

	Calibration Steps
1	Place a calibration object in the FOV of the IRm
2	Acquire CBCT image data containing tracking information
3	Segment the IR marker spheres
4	Calculate the 3D position of the centers of all IR markers from the segmentation
5	Extract the tracking information from the images
6	Apply a basic transform to the tracking data
7	Apply the ICP algorithm to the two-point clouds
8	Provide the calibration transform to mPc

Table 5.1: Steps required for the calibration of the transformation between ICS and CCS.

Figure 5.5: Replica of the calibration square, consisting out of acrylic resin.

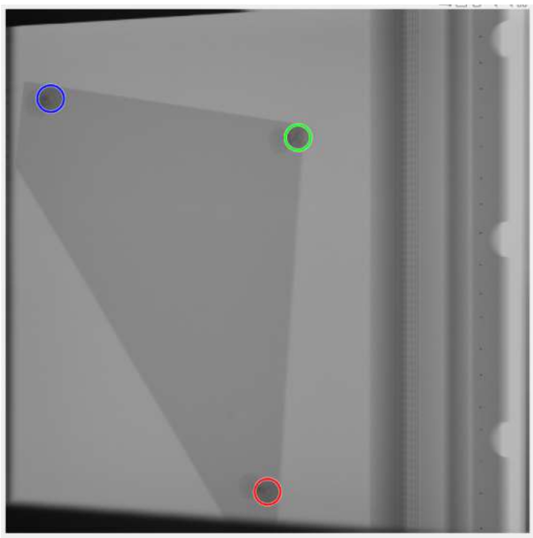

Figure 5.6: Segmented spheres in an X-ray image of the calibration object.

to be used to enhance the contrast of the marker spheres. Activating the tracking module during image acquisition ensures that the necessary position of the marker spheres in the CCS is saved in the metadata.

In these images, the marker spheres, depicted as circles, are segmented using the imfindcircles[5] function provided by Matlab. This function returns already the detected circle centers, which are then ordered depending on the distances among themselves to ensure the same spheres in different images are compared.

The objective of the next step is to project the circle centers in each X-ray image back to the X-ray source. By performing this for each center, it becomes possible to calculate the closest point to an intersection between the corresponding lines, resulting

[5] https://de.mathworks.com/help/images/ref/imfindcircles.html, accessed 23.09.2022

in the 3D position of the marker sphere in the Imaging Coordinate System (ICS). This process is illustrated in Figure 5.7, where the green point represents intersections of the lines. To facilitate this, the circle centers must first be transformed into the ICS. This transformation can be achieved by multiplying the pixel values by the detector's spacing factor. The resulting values can then be transformed using the detector's origin and orientation information, all of which are available in the image's metadata.

Given that lines typically do not intersect directly in 3D space, it is necessary to calculate the point with the smallest distance between two lines.

To solve this problem, as the first step, the line equations of these two lines can be defined as seen in Equation 5.1. The points $I1$ and $I2$ define the points, where the distance or a line between the two lines is the shortest. Now the fact that such a connecting line can only be the shortest if it is perpendicular to the two lines can be used to determine $I1$ and $I2$. Using the dot product to write Equation 5.2 and substituting Equation 5.1 in it allows to solve for m_I1 and m_I2. Knowing now the intersection points on the two lines, the 3D position of the marker can then be estimated by calculating the point in the middle of $I1$ and $I2$.

$$\begin{aligned} I1 &= M1 + m_I1(S1 - M1) \\ I2 &= M2 + m_I2(S2 - M2) \end{aligned} \tag{5.1}$$

$$\begin{aligned} (I1 - I2)dot(S1 - M1) = 0 \\ (I1 - I2)dot(S2 - M2) = 0 \end{aligned} \tag{5.2}$$

The method of detecting the 3D marker positions in the ICS using only two X-ray images revealed some minor inaccuracies in certain calibrations. This is because relying solely on two images makes the process susceptible to noise and geometric imperfections of the IRm. To enhance the accuracy, Daniel Kellner from medPhoton introduced an algorithm that directly segments the IR markers within the CBCT volume, which is outlined in the following:

To facilitate this algorithm, an existing calibration phantom is adapted with the inclusion of five marker spheres. The algorithm's general approach involves initially detecting the cylindrical shape of the phantom, as shown in the left image of Figure 5.9, which subsequently narrows down the regions where the search for marker spheres is conducted. Within these specified regions, the algorithm identifies and locates the spheres. The process then includes matching surface samples of an ideal model sphere with the detected structures, as indicated by the blue dot in Figure 5.10. This process ensures a precise and dependable identification of the marker sphere centers, as demonstrated in the right image of Figure 5.9.

Subsequent to obtaining the marker positions in both coordinate systems, the only remaining task is to determine a transformation that aligns these two point clouds. This transformation is accomplished using the ICP algorithm [BM92], which minimizes the

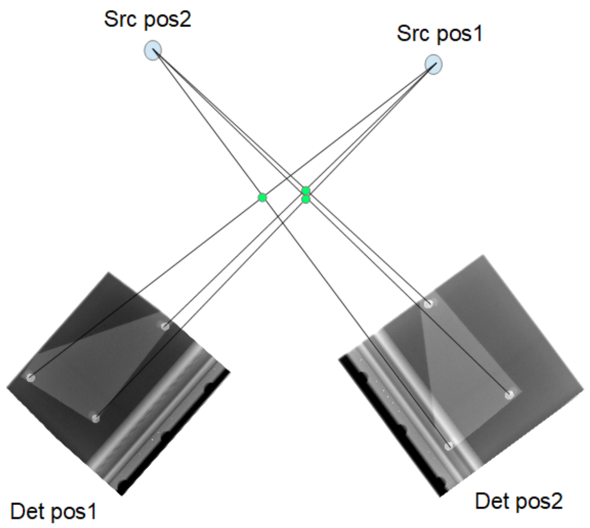

Figure 5.7: Illustration of the intersection of the back-projected marker line to estimate their 3D position in the ICS.

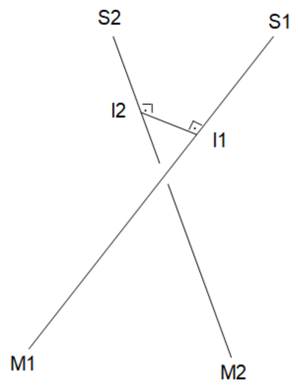

Figure 5.8: Illustration of an intersection of two lines in the 3D space.

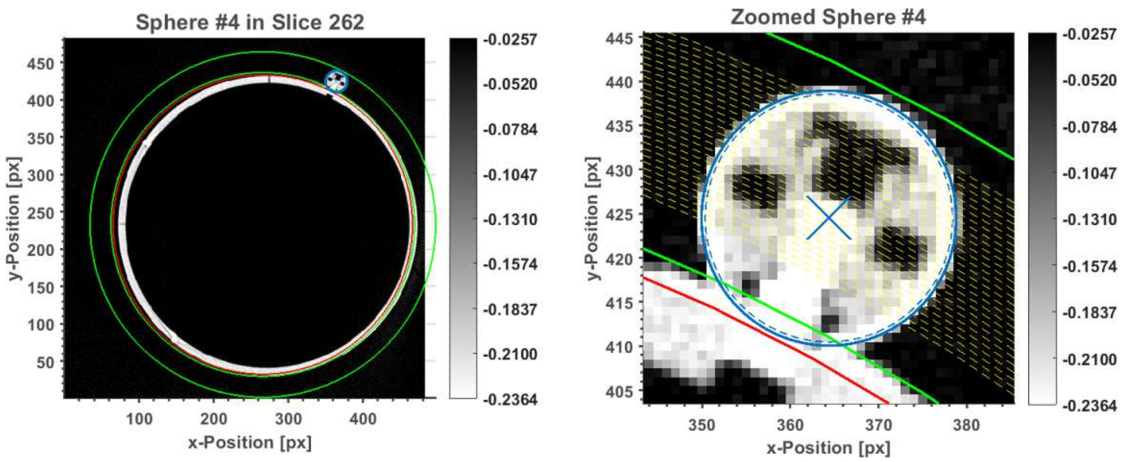

Figure 5.9: Visualization of the CBCT volume slice, containing the segmented marker sphere center.

error function depicted in Equation 5.3. In this equation, p_i represents a marker position in the ICS, while q_i signifies a position in the CCS.

To ensure the ICP algorithm's convergence, the CCS point cloud is initially transformed by means of a basic transformation. This transformation compensates for known differences between the coordinate systems and aligns their centers.

The outcome of applying the ICP algorithm to the calibration data is illustrated in Figure 5.11.

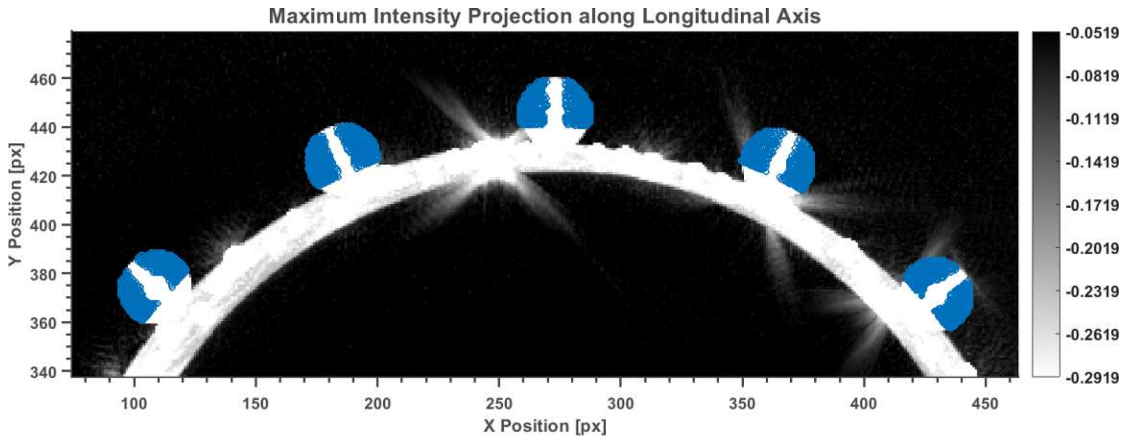

Figure 5.10: Maximum intensity projection of the advanced marker detection in the CBCT volume.

$$E(R,t) = \frac{1}{N}\sum_{i=0}^{N} ||(p_i - (Rq_i + t)||^2 \tag{5.3}$$

Once the transformation is determined, it needs to be exported into a configuration file, allowing accessibility by mPc.

To enable the execution of these calibrations from mPc, all its components needed integration within the mP software framework and automation of the process.

5.3.3 Synchronisation

Enabling motion compensation in CBCT imaging requires, besides a sub-milimeter motion tracking accuracy also a precise synchronization between image and tracking data. This is particularly important for scenarios involving significant motion, such as the saddle trajectory where translations up to 15 cm are not unusual. As an example, for a short scan, the normal scan time is about 30s with a frame rate of 12 FPS. If the synchronization is now off by only one frame, this would introduce an additional error of about 0.5mm.

Synchronization involves coordinating three distinct hardware systems, each with its own clock times: the programmable logic controller (PLC), the IPC, and the NUC. The PLC is a real-time system responsible for controlling all the hardware components of the IRm. Meanwhile, the mP software components operate on the IPC and utilize a command interface to communicate with the PLC.

The PLC provides the system properties, including motor positions and the current timestamp, to the TCS based on the hardware trigger signal from the source/detector. On the other hand, the FPPI reads image data from the detector interface. To accurately

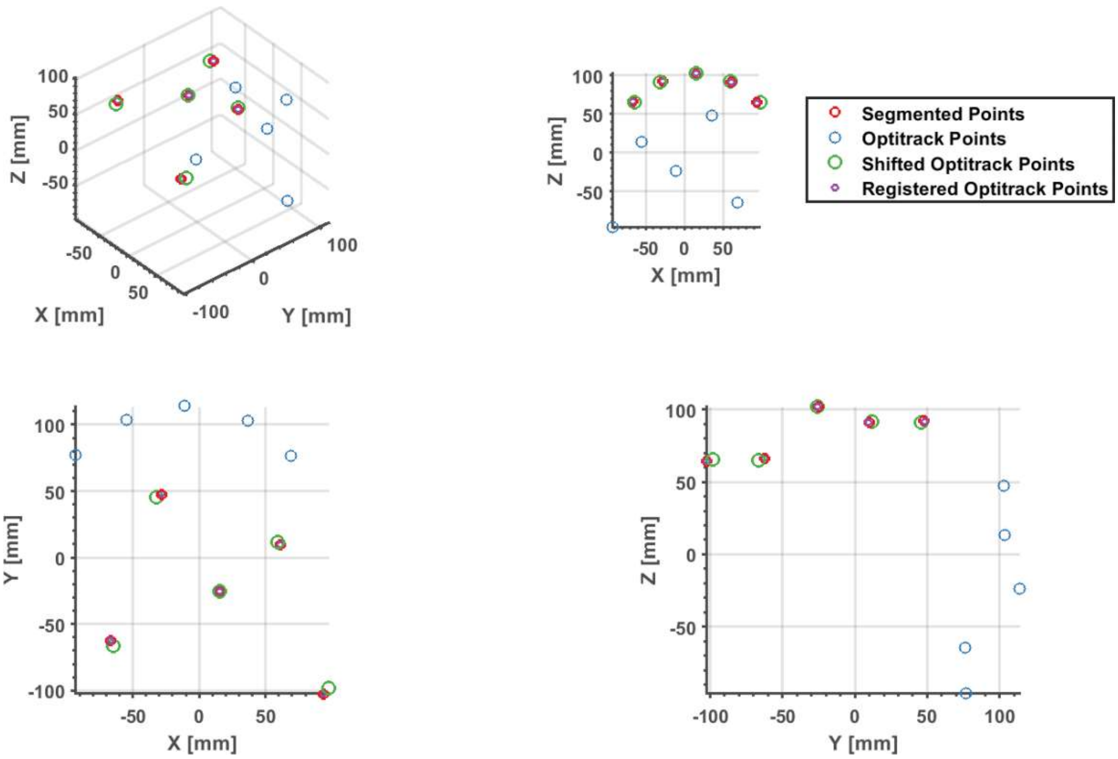

Figure 5.11: A result of the ICP algorithm applied on the calibration data, where Segemented Points refer to the marker in ICS, Optitrack Points to the ones in CCS, Shifted Optitrack points to the previous ones after applying the basic transformation and Registered Optitrack Points to the CCS markers after the transformation resulting from the ICP algorithm is applied to them.

associate these properties with the corresponding image data, it's essential to synchronize the PLC and IPC clocks.

To accomplish this, the Network Time Protocol (NTP) is used, where the PLC acts as the NTP server and the IPC as a client. Figure 5.12 shows the basic concept of the synchronization between the server and the client using NTP, where a UDP message is sent from the client to the server and a response vice versa [Mil85]. The response contains, in the end, four different timestamps, which are then used to calculate the clock offset between server and client by following equation[6]:

$$offset = 0.5 * ((T_2 - T_1) + (T_3 - T_4)) \tag{5.4}$$

T_1 refers to the client clock time when the request is sent. T_2 to the server clock time when the request is received. T_3 to the server clock time when the response is sent. T_4 to the client clock time when the response is received.

[6]https://www.timelinkmicro.info/ntp-basic-how-it-works/, accessed 13.10.2022

To get a proper estimation of the real offset and to compensate for possible fluctuations, this process is performed continuously, and the average value of the last samples is built.

After this, the image data can be synchronized with the required hardware properties.

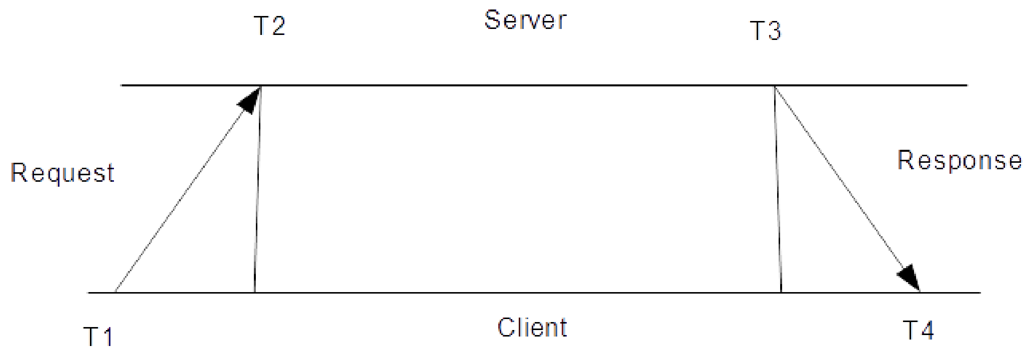

Figure 5.12: NTP synchronization between PLC and ICP.

The synchronization between the NUC and the IPC, to be then also able to correlate the tracking data to the proper images, is done with a similar but more simplistic algorithm proposed by Cristian et al. [Cri89]. This algorithm is beneficial for low-latency networks where speed is of essence as it is in the case of this work since the Optitrack cameras can run at 240 FPS.

The basic principle for the synchronization between server and client via Cristian's algorithm is visualized in Figure 5.13.

Using this algorithm, the synchronized client clock time T_C is defined by:

$$T_C = T_2 + (T_1 - T_0)/2 \tag{5.5}$$

Where T_0 refers to the client clock time at the time the request is sent. T_1 to the clock time of the server returned in the response and T_1 to the client clock time when the response is received.

Using this approach, the tracking data is posted to TCS with this synchronized timestamp, and with the calculated PLC-ICP NTP offset it can be assigned to the proper image data.

5.3.4 Reconstruction Pipeline

The IRm employs a modified version of the FDK algorithm. This algorithm necessitates the acquisition of multiple 2D projections from various viewing angles. Importantly, the

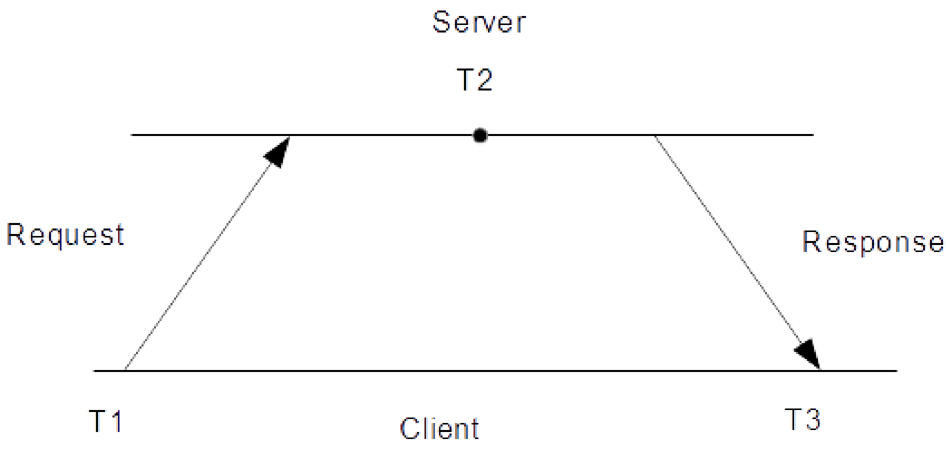

Figure 5.13: Basic principle to synchronize server and client via Cristian's algorithm.

algorithm assumes that the imaged object remains static—its location and orientation shouldn't change during data acquisition.

When the object being imaged is in motion during acquisition, the back-projected rays from its projections may not intersect at the correct locations. This can result in blurring or double contours for minor movements, and for larger motions, it might render the reconstructed volume unusable. Figure 5.14 offers a simple illustration of this issue: on the left side, you can see the projection generation, and on the right side, the back-projection into the reconstruction volume. In this scenario, the imaged object undergoes a transformation T during data acquisition, leading to artifacts in the reconstruction volume.

If the object's motion can be accurately detected, it becomes possible to compensate for the motion during the back-projection step of the FDK algorithm. This ensures that the reconstructed volume remains as accurate as possible even in the presence of motion artifacts.

Figure 5.15 illustrates the concept of this approach using two frames of the acquisition. In this visualization, you can observe how the position and orientation of the reconstruction volume are adjusted for each back projection based on the detected motion of the object. This adaptation ensures that the impact of object motion on the reconstructed volume is minimized, resulting in more accurate and reliable CBCT images.

To accomplish the proper transformation of the reconstruction volume, the relative change in orientation and position of the object has to be calculated. To prevent an accumulation of errors in the volume position and orientation, which could occur by calculating the difference between consecutive projections and transforming the volume accordingly, in this work always the first frame is used as a reference. Therefore, the

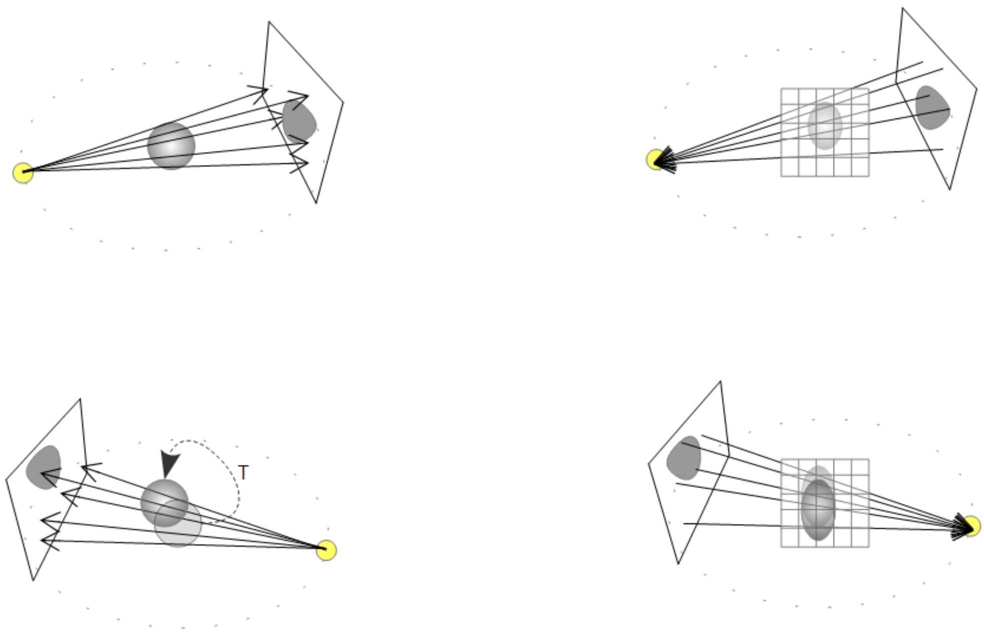

Figure 5.14: Sketch of the reconstruction process of a circular acquisition trajectory containing uncorrected object motion, resulting in motion artifacts in the volume.

object motion between the current projection and the first one is calculated and the initial volume position/orientation is then transformed by the resulting transformation for each projection before back-projecting it into the volume.

The required transformation T of the reconstruction volume for each projection is described by Equation 5.6. This transformation is formed by $4x4$ transformation matrices of the first (T_{R0}, T_{P0}) and current projection(T_{Pf}, T_{Rf}), namely the position of the marker object in the ICS, see Equation 5.7, and the rotation matrix defining the current orientation of the marker object. The orientation of the marker object is defined around its centroid and given as quaternion $q = (qw, qx, qy, qz)$. Therefore it has to be converted into a transformation matrix by Equation 5.8.

$$T = T_{Pf} * T_{Rf} * {T_{R0}}^{-1} * {T_{P0}}^{-1} \tag{5.6}$$

$$T_{Pi} = \begin{bmatrix} 1 & 0 & 0 & x_i \\ 0 & 1 & 0 & y_i \\ 0 & 0 & 1 & z_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.7}$$

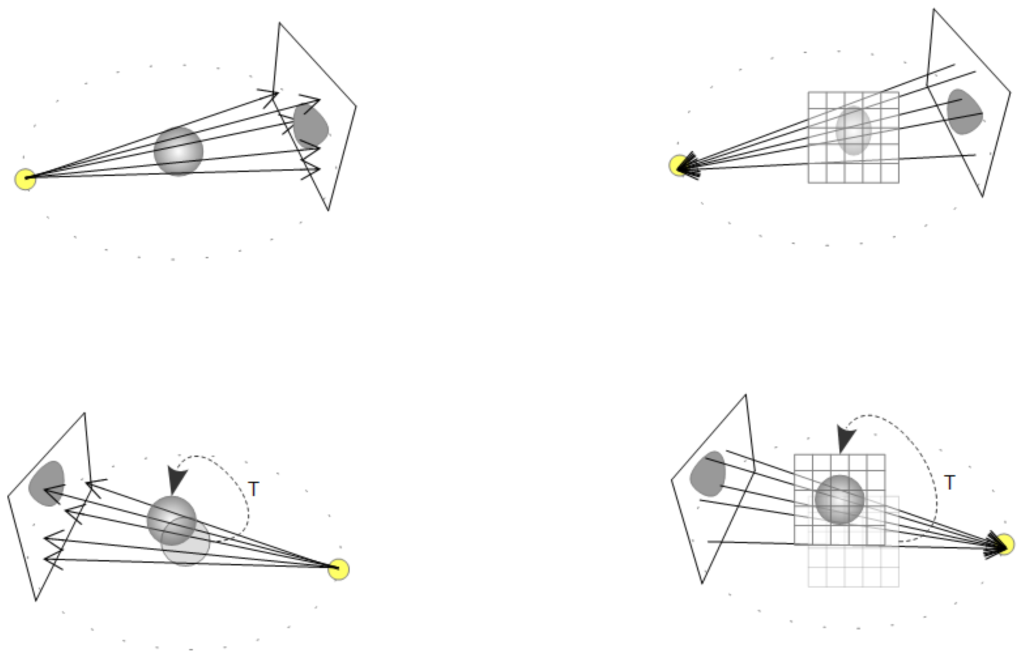

Figure 5.15: Sketch of the reconstruction process of a circular acquisition trajectory containing corrected object motion. The reconstruction volume follows the object motion for each back-projection step in the reconstruction algorithm, resulting in a motion artifact-free volume.

$$T_{Ri} = \begin{bmatrix} 1-2*qy*qy-2*qz*qz & 2*qx*qy-2*qz*qw & 2*qx*qz+2*qy*qw & 0 \\ 2*qx*qy+2*qz*qw & 1-2*qx*qx-2*qz*qz & 2*qy*qz-2*qx*qw & 0 \\ 2*qx*qz-2*qy*qw & 2*qy*qz+2*qx*qw & 1-2*qx*qx-2*qy*qy & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.8}$$

5.3.5 User Interface

Besides the integration of all the described concepts into the mP software framework, also a user interface(UI) to control the motion correction pipeline is required. The main requirement for this is to keep it as simple as possible without major changes in the previous workflow interactions. To achieve this the following concept, visualized in Figure 5.19, is developed: A new drop-down menu is added to the patient setup, where all the basic information of the acquisition, like patient position or imaged body region, are set. This menu contains all available marker objects usable for motion compensation. The user is able to select one object during the patient setup, which is always performed before the acquisition, and with this, the tracking module is activated. Figure 5.16 shows the patient setup menu with an example of the tracking object drop-down menu.

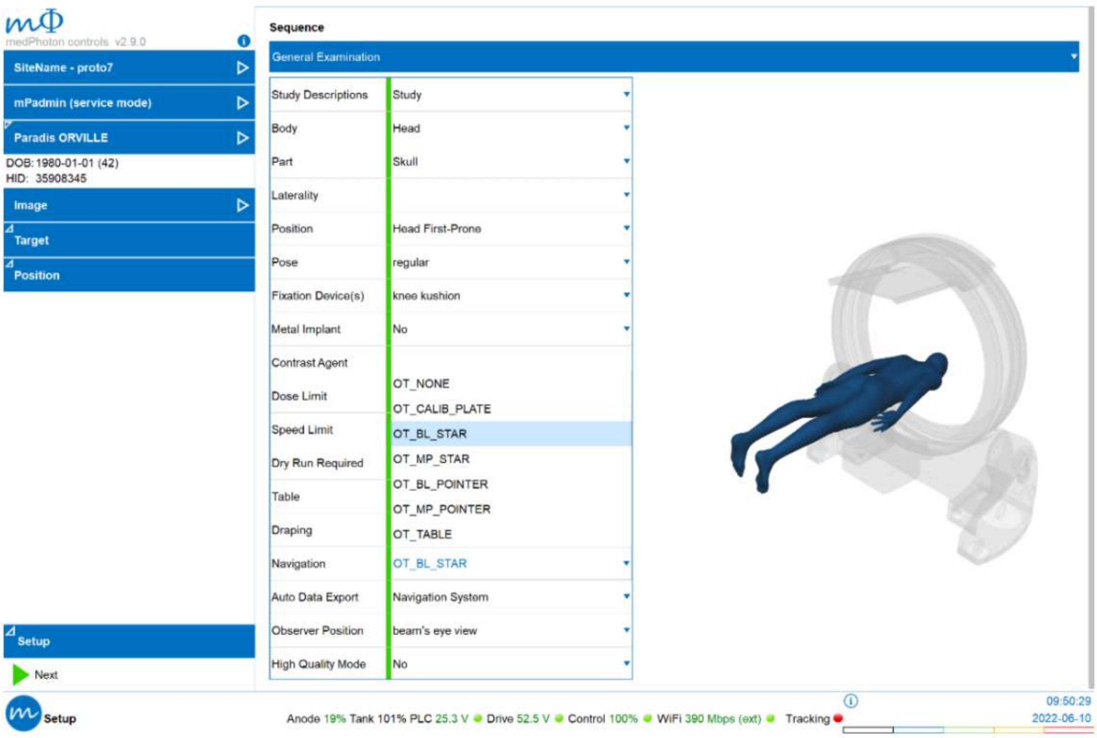

Figure 5.16: The patient setup dialog in mPc, where all the general settings of the acquisition, like patient position or imaged body part, are configured. The Navigation drop-down menu allows the selection of a configured tracking marker object.

Once the tracking system is activated, the selected marker object's ID is included in the imaging protocol, which contains all the relevant acquisition parameters. This protocol is then forwarded to the various software components as part of the standard workflow. In the FPPI, the marker object ID is read, and based on its value, the export of the tracking data to the projection metadata files is enabled. In the CBCT component, the ID triggers the motion compensation process. Additionally, the RTSS component includes a marker object watchdog.

The watchdog in the RTSS ensures that the marker object with the specified ID remains visible throughout the entire acquisition process. If, at any point, the marker object becomes invisible, the acquisition is halted, and a message is displayed to the user. This mechanism helps ensure the accuracy of motion compensation by guaranteeing that the tracking object remains within view during the entire imaging process.

Before initiating the imaging workflow, it's crucial to ensure that the corresponding marker object is correctly positioned within the field of view (FOV) of the tracking module. To aid in accurate placement, the motion app, a component of mPc responsible for visualizing the current ring position and controlling ring motion, also displays the marker object's positions relative to the ring position.

In Figure 5.17, an example of a marker object with three marker spheres is shown. The spheres are color-coded to assist users in quickly identifying any errors in the marker setup.

By providing visual feedback through color coding, the motion app helps users ensure that the correct marker object is positioned accurately for successful motion compensation during the imaging process.

Another helpful feature to guide users in placing the marker object accurately is the LED ring on the cameras of the tracking module. When the imaging workflow starts, the LED ring is activated and follows a similar color coding to that of the marker spheres in the motion app. However, there's a distinction: when no marker object is detected, the LED ring emits a red light. The different LED ring states are illustrated in Figure 5.18.

The LED ring's colors not only assist users in correctly positioning the marker object but also provide information about marker object detection issues during the imaging process. If the marker becomes lost during an acquisition, the LED ring instantly turns red, alerting the user to the situation. This immediate visual feedback allows users to respond to marker detection problems without needing to divert their attention from the IRm's display.

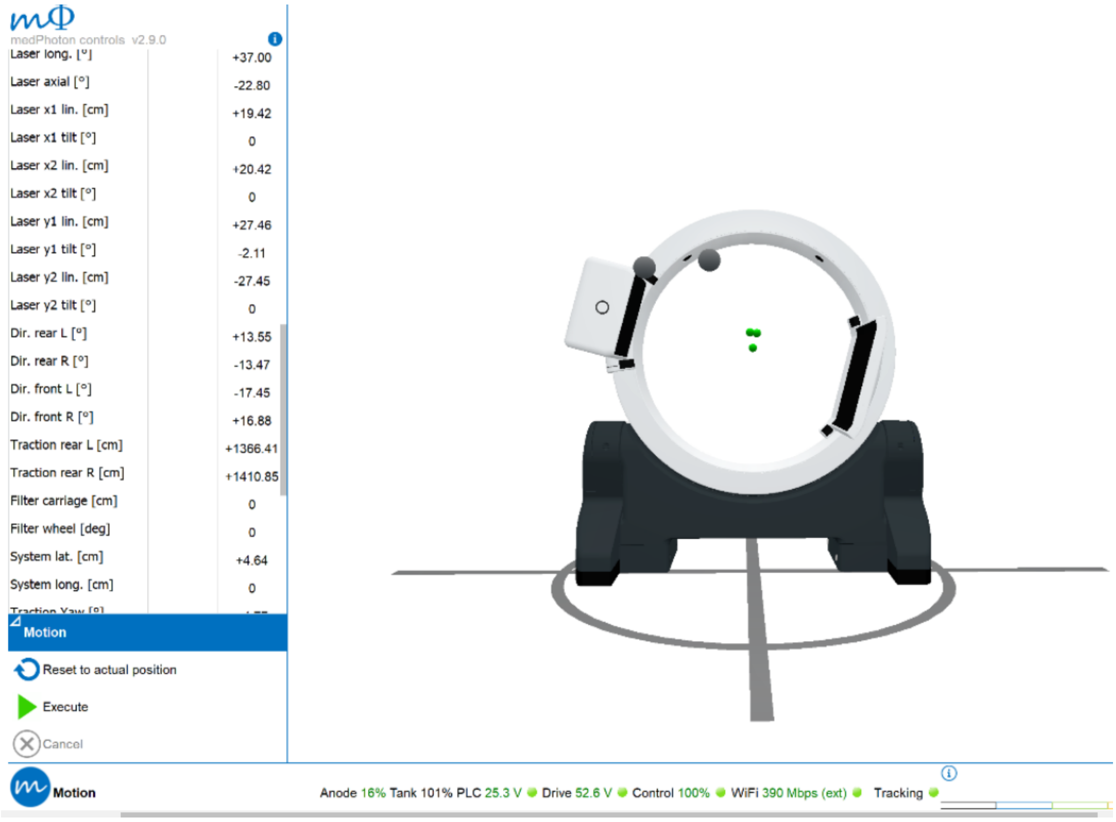

Figure 5.17: Screenshot of the motion app, visualizing the current ring position together with all detected marker object spheres. The sphere colors indicate the following state: Green: The selected marker object is detected and properly positioned within the FOV. Yellow: A configured marker object is detected in the FOV, but the user has selected a different one. Grey: Single marker spheres that appear grey cannot be assigned to a known marker object. This can occur if the marker object is partially occluded.

Figure 5.18: The different colors of the LED ring on the cameras of the tracking module.

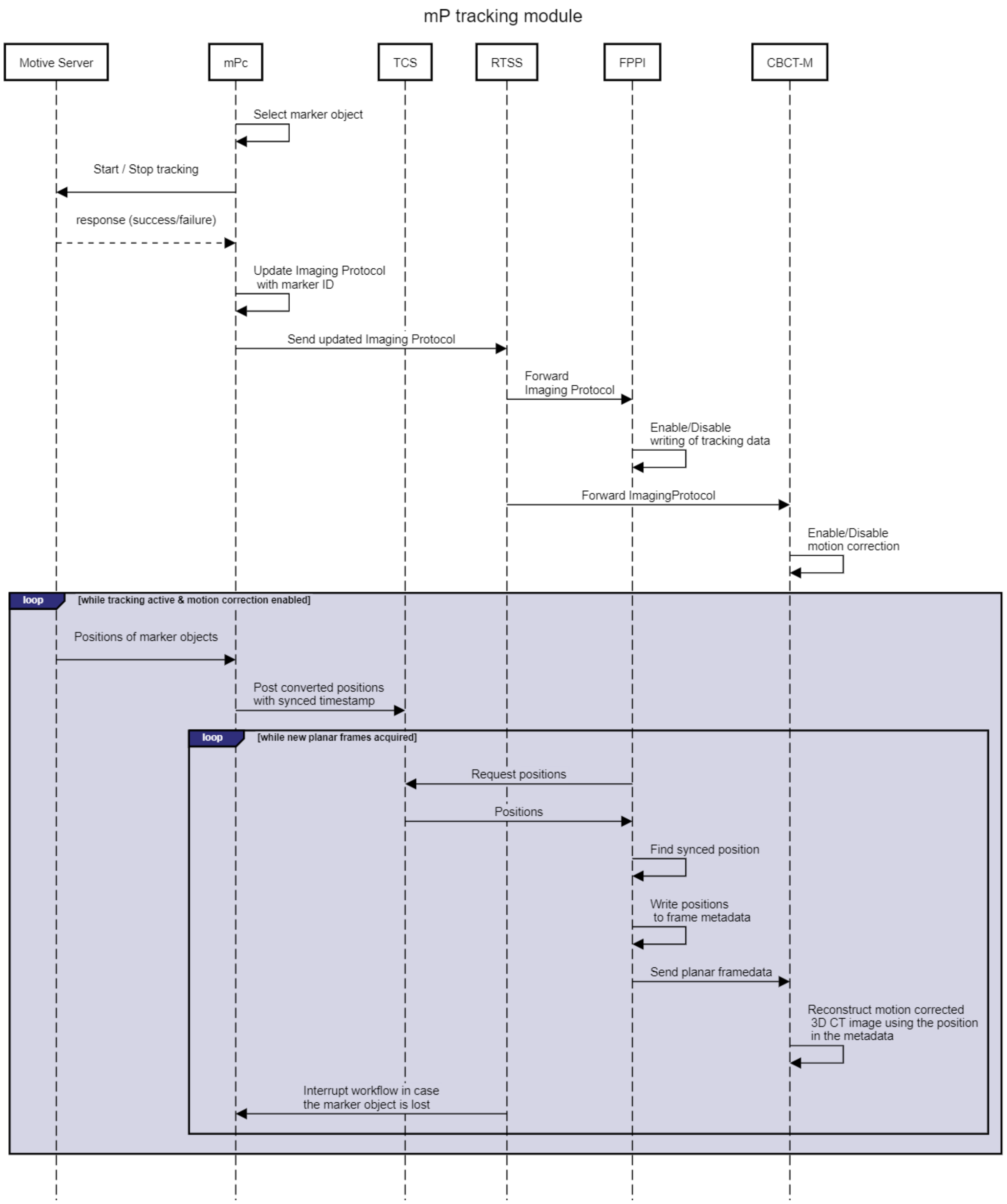

Figure 5.19: Software design of the mP tracking module.

CHAPTER 6

Evaluation and Results

After implementing and integrating the approaches described in the preceding chapter, an evaluation technique is required to assess the performance of the tracking module. The most straightforward evaluation involves generating CBCT volumes with motion correction applied using the tracking module and comparing their image quality against volumes reconstructed without utilizing the tracking information.

This evaluation strategy effectively assesses the functionality and effectiveness of all the different components within the motion correction pipeline simultaneously. If any component within the pipeline fails to function as intended, it could result in inaccurate reconstructed volumes. By comparing motion-corrected CBCT volumes to those without motion correction, the evaluation process validates the effectiveness of the entire motion correction pipeline.

To simulate motion during CBCT acquisitions, two distinct approaches are presented in this work. The first approach involves directly applying motion to the object being imaged, which can be achieved using devices like controllable air pads. The second approach is specific to the IRm system and is referred to as "saddle trajectory acquisitions." In saddle trajectories, coordinated movements of the IRm's three primary axes—yaw, tilt, and longitudinal—are executed during the CBCT scan. These movements follow a periodic function that is dependent on the view angle, as defined by a set of configuration parameters. This range of motion aims to simulate realistic patient movements and variations encountered during imaging scenarios.

The motion compensation algorithm implemented in the mP tracking module is agnostic to whether the object being imaged is moved or if the imaging system itself is moved. This versatility opens up a wide range of potential applications in which controlled motion can be deliberately introduced during acquisitions to mitigate image artifacts. One notable example of such artifacts is the presence of metal or cone-beam artifacts.

In a study conducted by Johanna Albrecht [Alb22], the effectiveness of motion compensation using saddle trajectories with the mP tracking module is demonstrated. The study finds that saddle trajectories, when corrected for motion using the mP tracking module, exhibit significantly fewer cone-beam artifacts compared to acquisitions without motion compensation. An illustration of this improvement can be observed in Figure 6.1, where the left side depicts larger cone-beam artifacts away from the central slice, indicated by the blue line. The induced motion shifts the central slice's position throughout the scan, effectively reducing the occurrence of cone-beam artifacts in the resulting images. This highlights the potential of deliberate motion to enhance image quality by mitigating specific artifacts that commonly arise in static acquisitions.

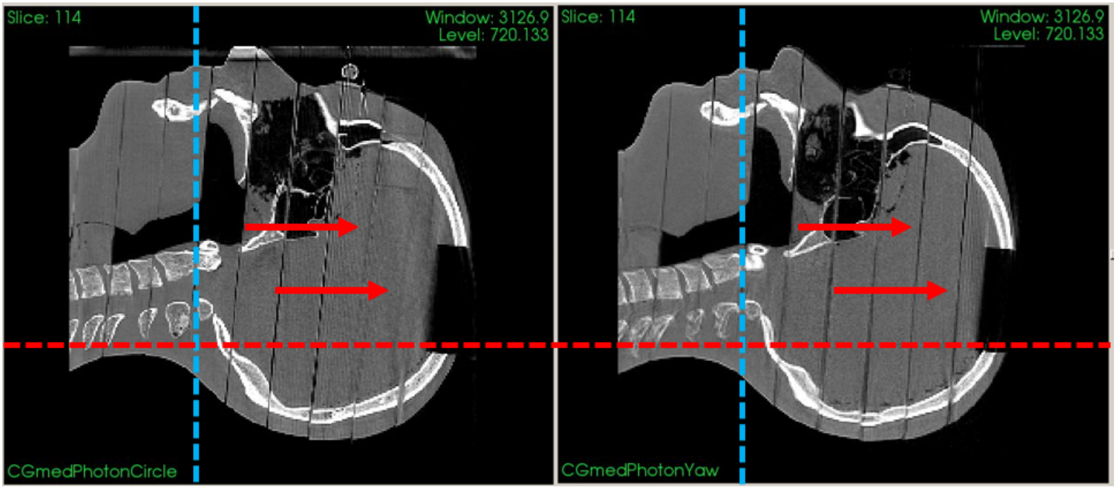

Figure 6.1: Results of a 10° yaw saddle trajectory(right) compared to a scan without motion(left) [Alb22].

6.1 Evaluation via image data

To evaluate the performance of the system, two distinct methods are employed to introduce motion to the phantom during the CBCT acquisitions. The initial approach involves utilizing a stepper motor to facilitate controlled forward and backward movement or rotation of the phantom in a predefined direction. However, this setup exhibits limitations as the rotation is constrained to a single direction, which does not accurately simulate real-world scenarios. Expanding the setup to incorporate multiple motors for enabling multi-directional motion would have incurred significant costs and complexity.

To address this challenge, a simple and cost-effective method for inducing multi-directional motion is adopted. This involves utilizing an inflatable air pad positioned underneath the phantom. Various configurations of the air pad's placement beneath the phantom are experimented with, leading to the selection of a setup that effectively balances safety considerations with motion complexity. This setup is outlined in the schematic diagram

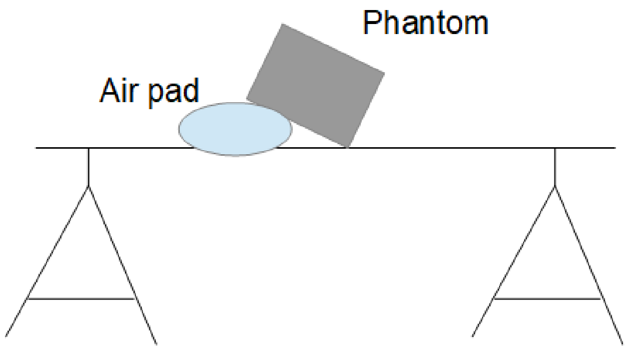

Figure 6.2: Sketch of the test setup used to induce motion to the phantom during the acquisition.

presented in Figure 6.2. Given that the air pad is positioned beneath only one side of the phantom, both translation and rotation effects are introduced. The inherent instability of the air pad's surface further contributes to motion complexity, causing the phantom to move in the y and z axes in addition to translation and rotation. The motion induced by the air pad is genuinely multi-directional, as demonstrated by the motion plot depicted on the right side of Figure 6.3.

To assess the tracking module's performance, motion is introduced to three distinct phantoms:

- Straw phantom: A Phantom consisting of polyvinyl chloride (PVC) pipes with a diameter of 1cm, glued together to form a square. This phantom is used to verify the first results because it allows a verification of the motion compensation in all different axes. An image of the phantom is shown on the right side of Figure 6.3.
- Boar phantom: This phantom is the frozen torso of a wild boar. This phantom is used because it contains a spine and is therefore close to the real use case scenario.
- Catphan phantom: The Catphan[1] is a tool used to characterize the imaging performance of CT / CBCT systems.

The straw phantom is specifically employed to demonstrate the accuracy of motion compensation across various primary directions. Its design is tailored to such use cases, with its pipes aligned along the primary axes, as depicted on the left side of Figure

[1]https://www.phantomlab.com/catphan-phantoms, accessed 05.11.2022

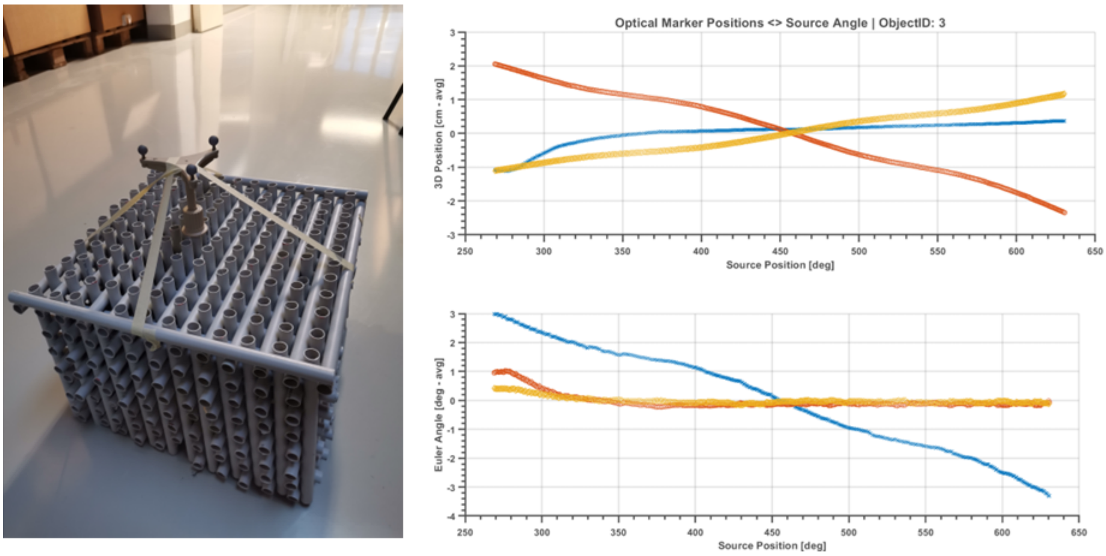

Figure 6.3: The left side shows the straw phantom with the mounted tracking marker object. The right side shows plots of the translation and rotation of the acquisition using the straw phantom together with the air pad (red is the z-axis, blue is the x-axis, and yellow is the y-axis).

6.3. Consequently, thin circular structures are visible within the main directions of the acquired volumes. The clarity of these structures serves as an indicator of accurate motion compensation. Any misalignment of the coordinate system or significant inaccuracies in the tracking data would manifest as unclear or distorted circular structures.

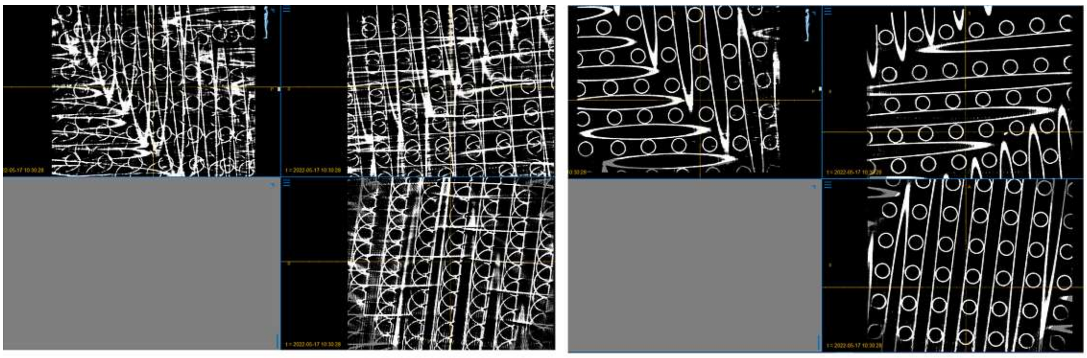

Figure 6.4: Result of a motion compensated acquisition of the straw phantom. The motion is induced via the air pad with an extension of 6cm.

Figure 6.4 presents an acquisition of the straw phantom with introduced motion. On the left side, the reconstructed volume is displayed without the motion compensation strategy. In this image, the thin structures of the pipes within the main views are not discernible. On the right side, the motion-compensated volume from the same acquisition is depicted. This illustration showcases a substantial enhancement in image quality, as the circular structures are distinctly visible with well-defined edges in all three views. This outcome underscores the efficacy of the motion compensation approach.

Figure 6.5: Result of a motion compensated acquisition of the boar phantom. The motion is induced via the air pad with an extension of 5cm.

The initial primary clinical application of the IRm, incorporating the proposed motion compensation approach, is within the realm of spine surgeries. Consequently, it is pivotal to validate its functionality using datasets closely resembling real-world scenarios. To achieve this, data is gathered using the boar phantom, which is essentially the preserved torso of a wild boar.

Figure 6.5 presents the outcomes of a boar phantom acquisition involving motion. The left side showcases the reconstructed volume devoid of motion compensation, while the right side displays the volume with motion compensation applied. The disparity is evident: the uncompensated volume lacks adequate quality in all slices, while the motion-compensated volume distinctly reveals the bony structures of the boar torso in all perspectives. This contrast underscores the remarkable improvement brought about by motion compensation.

Figure 6.6 provides an enlarged view of the sagittal perspective of the boar phantom acquisition. This view primarily captured the main motion during the acquisition, allowing for a clear demonstration of the motion compensation's quality enhancement. On the left side, the uncompensated image exhibits only indistinct bone structures within the spine, marked by multiple overlapping contours. Conversely, the right side, displaying the motion-compensated image, showcases the boar's vertebrae with precise delineation and devoid of any duplications.

The CatPhan 600[2] phantom is utilized by medPhoton to assess the maximum imaging capabilities of the IRm, ensuring the calibration of new imaging systems and their readiness for clinical applications. It comprises various modules designed to evaluate

[2]https://www.phantomlab.com/catphan-600, accessed 22.07.2023

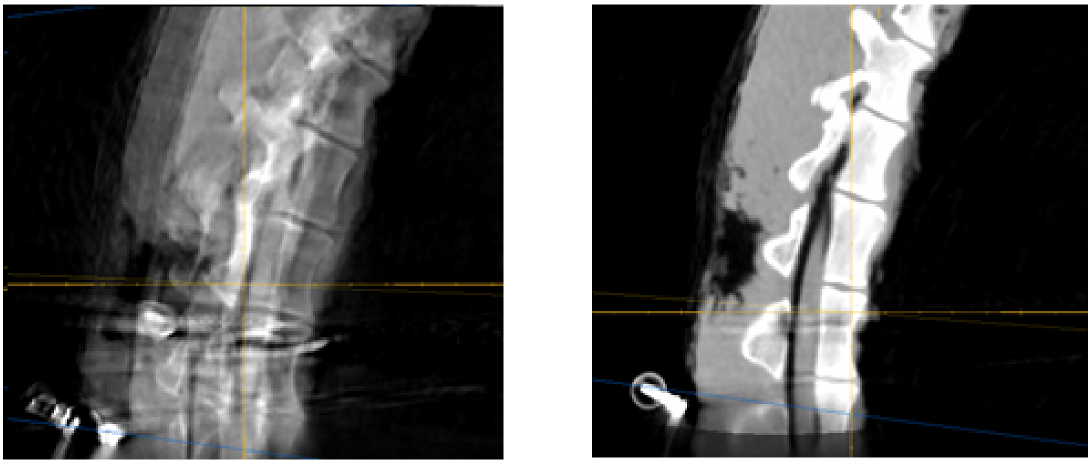

Figure 6.6: Result of the sagittal view of a motion compensated acquisition of the boar phantom. The motion is induced via the air pad with an extension of 6cm.

noise, contrast, and the resolution of CT/CBCT systems. To evaluate the results of motion compensation, the high-resolution measurement module is particularly suitable for assessing the impact of motion on overall accuracy.

This module includes 21 sets of line pairs, ranging from 1 line per centimeter to 21 lines per centimeter. This setup allows for the evaluation of resolution as the lines and spacing between them decrease in size, making them progressively more challenging to resolve in the reconstructed volume.

To induce motion during the CatPhan acquisition, the same approach as before is applied. Figure 6.7 b) displays the outcome of a slice from the high-resolution module, as well as the slice geometry and sensitometry module of an acquisition that involved motion. The image illustrates that none of the line pairs are reconstructed sharp in this scenario.

In contrast, Figure 6.7 c) presents the same slice and volume, but with the proposed motion compensation applied. In this case, the slice geometry and sensitometry module demonstrate that circular symmetry is well-maintained with motion compensation. Furthermore, the line pairs of the high-resolution module can be resolved nearly as effectively as in the absence of motion.

For comparison, Figure 6.7 a) provides a zoomed-in snippet around line pair 18 in a volume without motion, allowing for a direct comparison to Figure 6.7 c). The analysis of the image data indicates that the substantial motion occurring during the acquisition only results in a reduction of the number of line pairs that can be accurately resolved by 1.

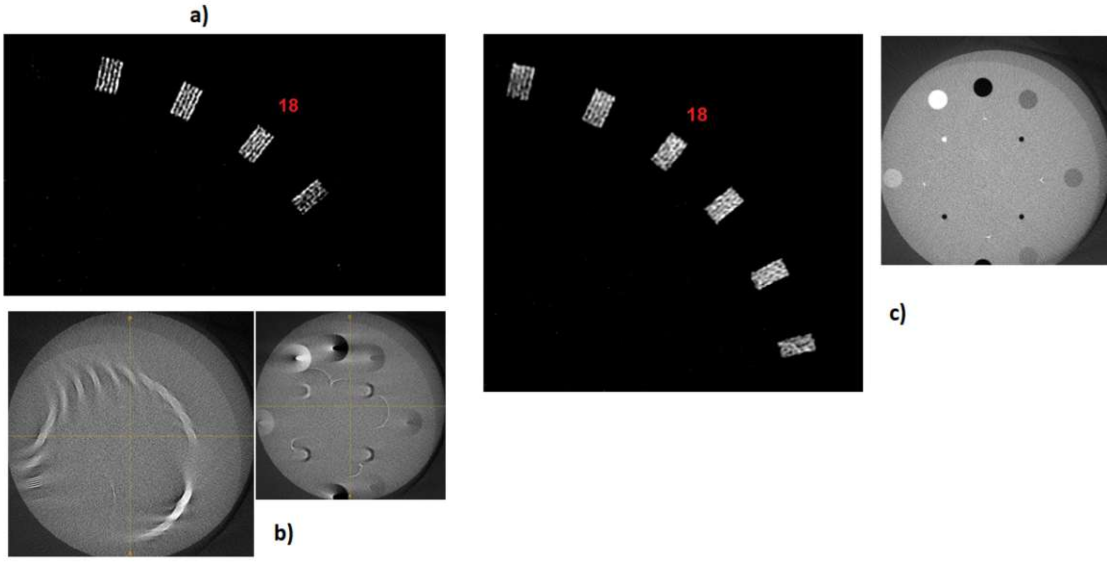

Figure 6.7: Evaluation of the motion compensation via the CatPhan phantom. a) shows a slice of the high-resolution module reconstructed without motion. b) Shows a slice of the high-resolution module and the slice geometry and sensitometry module containing motion but without the motion compensation applied. c) shows the same data set as b, but with the proposed motion compensation applied.

6.2 Evaluation via the flexmap phantom

Up to this point, the accuracy of the motion compensation has been primarily assessed by the qualitative improvement in imaging quality. While these results indicate that the tracking module allows CBCT reconstruction containing motion, a more quantifiable approach is needed. The objective is to develop a method for estimating the residual error boundaries of the tracking module when fully integrated into the X-ray imaging pipeline. This enables an understanding of potential use cases, and limitations, and facilitates verification of the calibration process. Initially, only the residual error of the ICP algorithm is used to validate the system's accuracy. However, due to the limited insight provided by this value, a more rigorous verification process is necessary to ensure a robust system setup for clinical applications.

The proposed process can be outlined as follows:

- Data acquisition: Acquire a saddle trajectory of the Flexmap phantom, containing IR markers.
- Update image metadata: Adapt the source and detector position and orientation depending on the

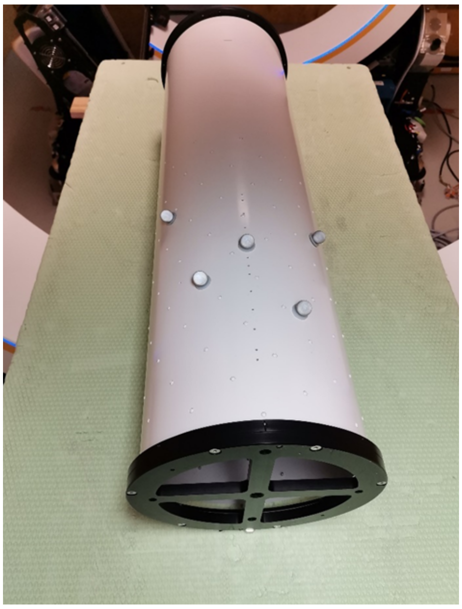

Figure 6.8: MP flexmap phantom containing the tracking marker spheres.

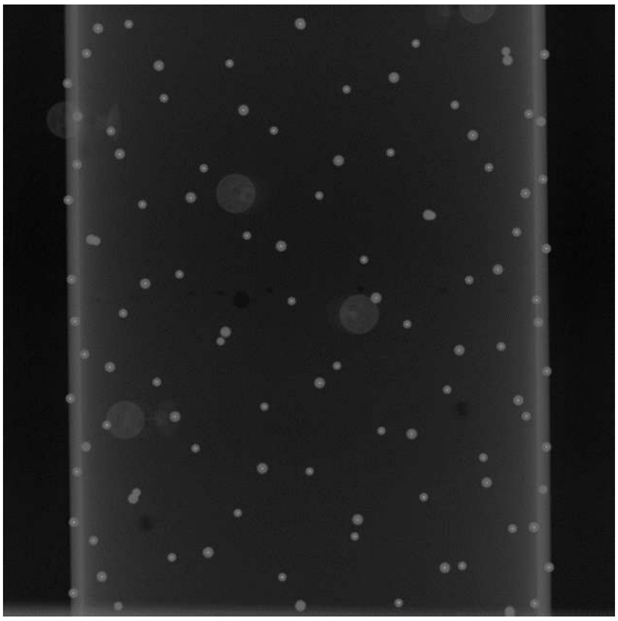

Figure 6.9: An example result image of the mP aiMD marker segmentation of the flexmap phantom.

- Transformation, detected via the tracking module, in the metadata of the projections.
- Marker detection: Run the mP AI marker detection on the updated projections.
- Evaluation: Back-project the rays of the individual markers, detected in the projections, and intersect the rays with each other to get an estimation of the residual error of the tracking.

Each of these steps will be elaborated on in the following paragraphs.

The foundation of the system evaluation involves conducting a saddle trajectory acquisition using the Flexmap phantom, as depicted in Figure 6.8. It's crucial to confirm that the five IR markers are properly affixed to the phantom before initiating the acquisition. Additionally, the tracking module needs to be activated through the UI, and the appropriate ID (1) for reconstruction must be chosen. This preliminary setup ensures that the tracking module is ready to capture and process the motion data effectively.

Furthermore, the following acquisition parameters have to be changed to ensure proper image quality:

- Yaw parameters: Amplitude=10, Phase shift=90
- Power settings: 80kV, 10ms, 10mA
- Detector Settings: 12 fps Gain 5

For evaluation, the Flexmap phantom is positioned near the iso-center of the IRm, and the FOV is selected to maximize the visibility of as many steel markers as possible within the acquired volume. Once these adjustments are made, the acquisition process can be initiated. Throughout the acquisition, it's essential to monitor that the cameras of the tracking module remain illuminated in green.

It's worth noting that the optimal scenario involves conducting this acquisition after a successful flexmap calibration. Alternatively, a flexmap verification must be performed before the measurement to ensure the accuracy of the device's geometric calibration. This guarantees that the flexmap phantom model used for marker detection is up to date and that imaging precision is maintained.

After successful data acquisition, the subsequent step involves adjusting the source/detector position and orientation based on the movement observed during the saddle trajectory. The method employed here is akin to the motion compensation utilized during volume reconstruction. In this approach, the relative motion between the first frame and all subsequent frames is calculated as follows:

$$T = (T_{Pf} * T_{Rf} * {T_{R0}}^{-1} * {T_{P0}}^{-1})^{-1} \tag{6.1}$$

which is built the same way as Equation 5.6, but the result has to be inverted due to the fact that we want to move the ring instead of the reconstruction volume.

The resulting transformation matrix is then applied to adjust the source/detector position and the detector orientation, both of which are read from the metadata files of the projections. Subsequently, all the metadata files need to be updated with the new positions and stored.

The subsequent step involves detecting all the steel markers that are integrated into the flexmap phantom. These markers are present in the projections of the acquisition. The mP marker detection tool aiMD is employed for this purpose. aiMD utilizes deep learning in combination with a 2D-3D registration approach to identify the marker centers in the projection images. To execute the marker segmentation, aiMD.exe can be run using a command line call.

To ensure that the executable is working, the required versions of CUDA dlls, which are provided with the IRm have to be added to the environment of the executing

PC. Furthermore, it has to be ensured that a proper model of the flexmap phantom(NominalStructureSet.csv) is located at `"PATH_TO_SERIES\ MD\NominalStructureSet.csv"` .

The result of the aiMD can be verified by ensuring the detected marker centers are matching with the markers in the images provided in `"PATH_TO_SERIES \MD\01_MD_CW\images"`, as shown in Figure 6.9.

The exact locations of the detected marker centers are stored by aiMD in the `"PATH_TO_SERIES\MD\detected-markers.ini"` file. For each projection, the file contains a section `[Markers2D-ProjID]`, where `ProjID` stands for the number of the current projection. In this section the path to the projections metadata file, the source/detector angle, the information if a marker is detected in this particular projection, and the location of it in the image can be found.

All the marker positions are given in millimeters and the great benefit of this marker detection approach is that each 2D marker location can be related to the correct 3D marker ID, which allows the proper assignment throughout all the projections.

To evaluate the data generated in the previous step, the position of each detected 2D marker needs to be transformed into the 3D imaging coordinate system, and a line connecting this location with the source position must be calculated. Since the positions of the 2D markers are already encoded in the x/y direction along the surface of the detector, they can be easily transformed into the imaging coordinate system using the formula $D_{\mathrm{dir}} \cdot [x, y, 0] + D_{\mathrm{orig}}$, where D_{dir} is the direction matrix of the detector, and x and y are the 2D positions obtained from the aiMD results. D_{orig} represents the detector origin in the imaging coordinate system.

With the source position defined, a 3D line can be established. A similar process is performed for a second projection, where the line is approximately orthogonal to the first one, minimizing errors stemming from intersection calculations. The appropriate pairing between the two lines is determined by identifying the projection where the difference in view angles from the previous projection is approximately +90 or -90 degrees. Once the correct pairing is determined, the midpoint of the shortest line segment between the two lines is calculated. This midpoint serves as an intersection point, as the lines typically don't directly intersect in 3D space.

This intersection point is calculated and recorded for each detected marker in every projection, facilitating a statistical evaluation of the tracking module's accuracy.

Figure 6.10 provides a visual representation of this approach. It's crucial to note that the source and detector positions, as well as orientations, need to be adjusted to align with the path of the saddle trajectory. The system's accurate positional information is available for a circular trajectory of the source and detector without additional ring movements. As such, the system assumes a static position of the gantry. However, in the case of saddle trajectories, each position of the source and gantry along this static

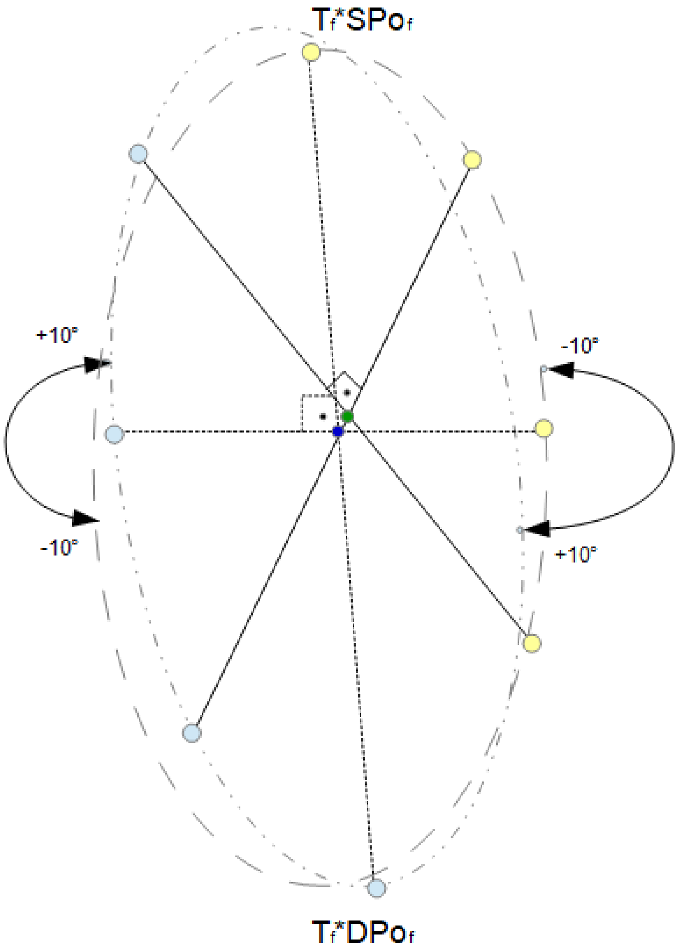

Figure 6.10: Sketch of the source and detector trajectory during a 10° yaw saddle trajectory.

circular trajectory (SPo_f and DPo_f) must be transformed according to the detected system motion, as described by Equation 6.1.

The calculated intersection points form point clouds for each marker and its dimensions allow an assessment of the residual error of the tracking module. To do this, the distance of each point to the centroid of the intersection point cloud is calculated, and statistical evaluations, like the 95 % quantile, the median, or the mean are calculated. The result of this is shown in Table 1 for three datasets, where "No Movement" stands for a normal CBCT, "New Calib." is a 10° yaw saddle trajectory after a fresh calibration of the tracking module, and "Older Calib." is a 10° yaw saddle trajectory with an older calibration.

The "New Calib." dataset demonstrates that the accuracy of the tracking module, including the proper calibration of the imaging coordinate system, is well within 1mm. The low mean deviation of only 0.27mm concerning the centroid is also evident in the imaging

	No Movement	New Calib.	Older Calib.
central dist. to intersections, q95 [mm]:	0.12106	0.780673	1 .07722
central dist. to intersections, mean [mm]:	0.0558796	0.274652	0.338801
central dist. to intersections, median [mm]:	0.0519456	0.247096	0.300531
central X overall dist. to intersections, q95 [mm]:	0.0979942	0.576421	0.742661
central Y overall dist. to intersections, q95 [mm]:	0.0649811	0.59091	0.513176
central Z overall dist. to intersections, q95 [mm]:	0.0937887	0.505722	0.880938

Table 6.1: Result of the tracking cam evaluation, central to intersections.

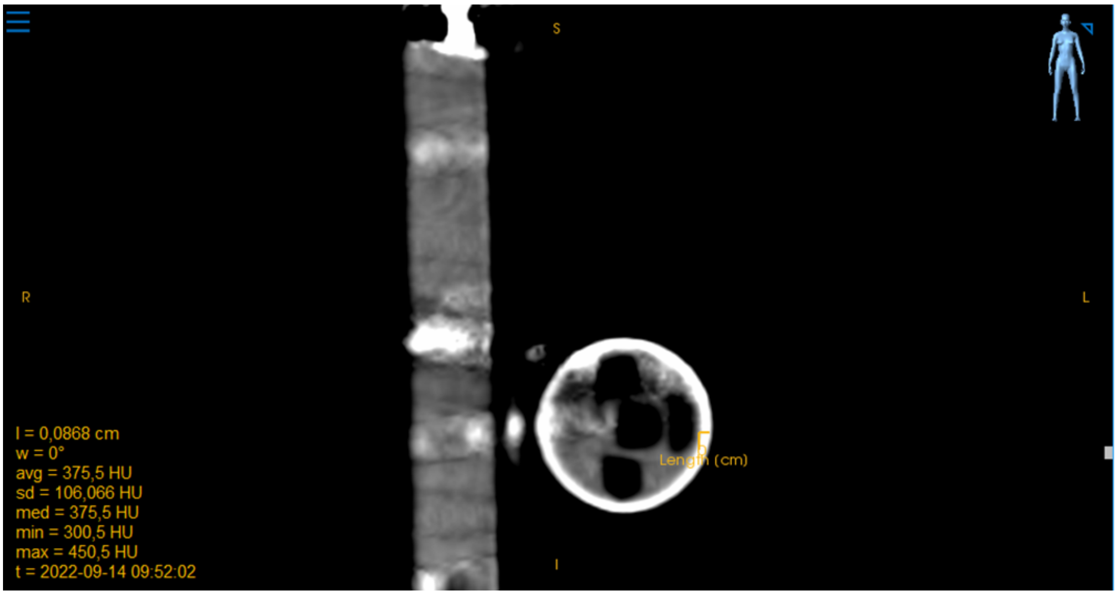

Figure 6.11: Measurement of the IR marker coating in the HQ reconstruction of the dataset “New Calib” .

data, where even small structures such as the reflective coating of an IR marker sphere are clearly depicted, as shown in Figure 6.11. Additionally, Figure 6.12 illustrates that the size of the IR marker sphere accurately matches the specified diameter of 14mm from the producer.

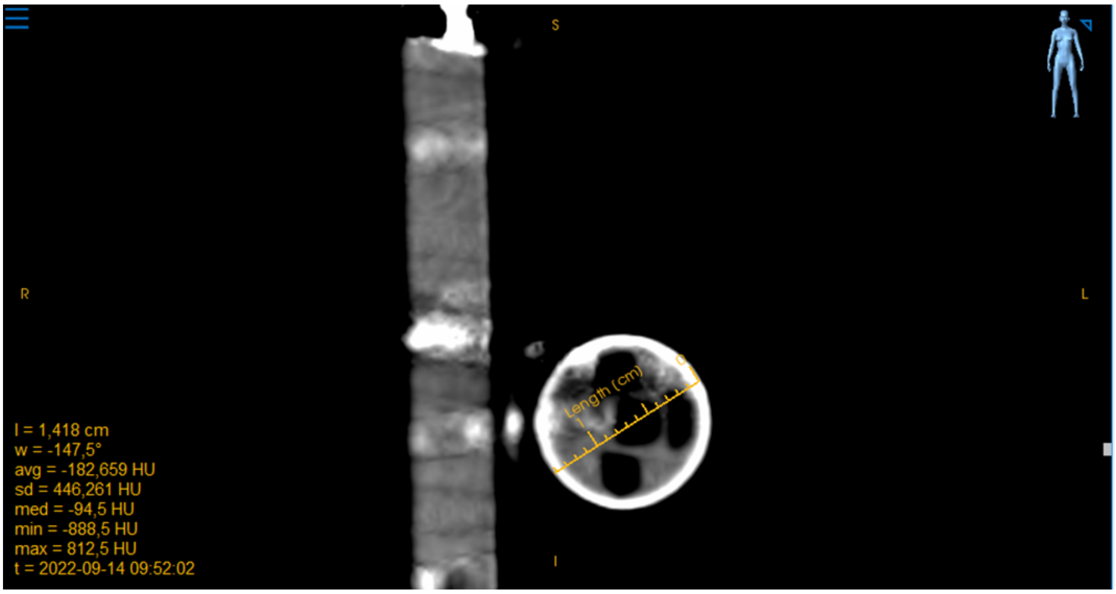

Figure 6.12: Measurement of the IR marker diameter in the HQ reconstruction of the dataset "New Calib".

Moreover, these measurements are conducted in the coronal plane of the volume, where most of the motion artifacts can be anticipated, especially in the case of yaw saddle trajectories.

6.3 Conclusion and Further Work

This thesis chronicles the exploration and development of a stereo vision system designed to detect and compensate for motion during CBCT acquisitions. The investigation entailed the evaluation of diverse camera technologies in small-scale experimental setups. Among these technologies, infrared (IR) imaging in conjunction with IR markers affixed to the patient emerged as the most suitable choice for achieving real-time motion compensation. This approach offers not only robustness across different environments but also reliability and high frame rates, crucial for real-time applications.

The resultant system is capable of generating tracking data at rates of up to 240Hz while simultaneously synchronizing this data with individual CT projections during acquisition. The tracking data is consistently stored alongside the CT projections. Additionally, the motion compensation pipeline is designed to enable volume reconstruction without motion compensation, ensuring that patient treatment can proceed safely during the clinical trial phase of the system.

The evaluation phase of the system demonstrated sharp image reconstructions across a variety of setups, even in the presence of motion amplitudes spanning several centimeters. Further validations using the CatPhan phantom highlighted that despite significant motion induced during the scan, the overall quality of the final volume remains comparable to scans conducted without motion.

To further substantiate the accuracy of the system, a more reproducible geometric evaluation approach involving system motion and a deep learning segmentation method is devised. This approach illustrated that the system's accuracy lies within the sub-millimeter range.

The stereo vision system is not confined solely to motion detection. During its development, a range of potential applications emerged. Two major areas of practical utility include compensation for system motion and the tracking of medical tools within the IRm's 3D space. Collaborative investigations with Ludwig-Maximilians-University Munich unveiled the system's potential in reducing cone-beam artifacts in CBCT scans through acquisitions involving system motion [ARS+22]. Additionally, the system's data is employed to stitch together multiple volumes, enabling comprehensive full-body scans with the IRm.

To further investigate the potential of the system, a project with Friedrich-Alexander-University of Erlangen-Nürnberg exploring the capabilities for tool tracking in Brachytherapy scenarios is started. Another important open topic is the long-term and temperature stability of the system, which is especially important for unsupervised clinical use cases.

In conclusion, the proposed system, seamlessly integrated into a mobile CBCT device, is a unique and promising solution, demonstrating effectiveness across various use cases. It is important to acknowledge that while the system yields high accuracy for rigid phantoms, its performance could diminish in cases involving non-rigid patient motion. Addressing deformable motion would require more advanced methods, such as those based on deep learning, which are beyond the scope of this work and await future exploration.

List of Figures

List of Tables

Bibliography

[Alb22] J. Albrecht. Cone beam ct from complete data using saddle trajectories on a mobile robotic cbct scanner. Master's thesis, Ludwig Maximilian University of Munich, 2022.

[AR14] K. Abramovitch and D. D. Rice. Basic principles of cone beam computed tomography. *Dental Clinics*, 58(3):463–484, 2014.

[ARS+22] J Albrecht, S Rit, P Steininger, F Ginzinger, P Huber, I Messner, M Kraihamer, H Schmitz, S Corradini, C Belka, et al. Cone-beam ct from complete data using saddle trajectories on a mobile robotic cbct scanner. In *Medical Physics*, volume 49, pages E123–E123, 2022.

[Aya91] N. Ayache. *Artificial vision for mobile robots: stereo vision and multisensory perception*. Mit Press, 1991.

[BBO+09] F. Bergner, T. Berkus, M. Oelhafen, P. Kunz, T. Pan, and M. Kachelrieß. Autoadaptive phase-correlated (aapc) reconstruction for 4d cbct. *Medical Physics*, 36(12):5695–5706, 2009.

[BFS+14] M. Berger, C. Forman, C. Schwemmer, J. H. Choi, K. Müller, A. Maier, J. Hornegger, and R. Fahrig. Automatic removal of externally attached fiducial markers in cone beam c-arm ct. In *Bildverarbeitung für die Medizin 2014*, pages 318–323. Springer, 2014.

[BK08] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[BLH+11] B. M. Barney, R. J. Lee, D. Handrahan, K. T. Welsh, J. T. Cook, and W. T. Sause. Image-guided radiotherapy (igrt) for prostate cancer comparing kv imaging of fiducial markers with cone beam computed tomography (cbct). *International Journal of Radiation Oncology*Biology*Physics*, 80(1):301–305, 2011.

[BM92] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606, 1992.

[BMA+16] M. Berger, K. Müller, A. Aichert, M. Unberath, J. Thies, J.-H. Choi, R. Fahrig, and A. Maier. Marker-free motion correction in weight-bearing cone-beam ct of the knee joint. *Medical Physics*, 43(3):1235–1248, 2016.

[BNH+20] G. Burström, R. Nachabe, R. Homan, J. Hoppenbrouwers, R. Holthuizen, O. Persson, E. Edström, and A. Elmi-Terander. Frameless patient tracking with adhesive optical skin markers for augmented reality surgical navigation in spine surgery. *Spine*, 45(22):1598–1604, 2020.

[BRU+18] B. Bier, N. Ravikumar, M. Unberath, M. Levenston, G. Gold, R. Fahrig, and A. Maier. Range imaging for motion compensation in c-arm cone-beam ct of knees under weight-bearing conditions. *Journal of Imaging*, 4(1):13, 2018.

[BUG+17] B. Bier, M. Unberath, T. Geimer, N. Ravikumar, G. Gold, R. Fahrig, and A. Maier. Fusing motion estimates for cbct reconstruction. In *Proceedings of the 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (IEEE NSS/MIC 2017), Atlanta, GA, USA*, pages 21–28, 2017.

[CF98] T. A. Clarke and J. G. Fryer. The development of camera calibration methods and models. *The Photogrammetric Record*, 16(91):51–66, 1998.

[CKB18] S. Choi, J. Kim, and J. Baek. A hybrid approach to reduce cone-beam artifacts for a circular orbit cone-beam ct system. *IEEE Access*, 6:54595–54606, 2018.

[Cri89] F. Cristian. Probabilistic clock synchronization. *Distributed computing*, 3(3):146–158, 1989.

[CSS+21] S. Capostagno, A. Sisniega, J. W. Stayman, T. Ehtiati, C. R. Weiss, and J. H. Siewerdsen. Deformable motion compensation for interventional cone-beam CT. *Physics in Medicine & Biology*, 66(5):055010, feb 2021.

[FDK84] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612–619, Jun 1984.

[FFWN17] J. Fotouhi, B. Fuerst, W. Wein, and N. Navab. Can real-time rgbd enhance intraoperative cone-beam ct? *International journal of computer assisted radiology and surgery*, 12(7):1211–1219, 2017.

[FJSS21] R. Fahrig, D. A. Jaffray, I. Sechopoulos, and J. Stayman. Flat-panel conebeam ct in the clinic: history and current state. *Journal of Medical Imaging*, 8(5):052115, 2021.

[FSF+14] A. Fassi, J. Schaerer, M. Fernandes, M. Riboldi, D. Sarrut, and G. Baroni. Tumor tracking method based on a deformable 4d ct breathing motion model driven by an external surface surrogate. *International Journal of Radiation Oncology* Biology* Physics*, 88(1):182–188, 2014.

[GCO+19] M. Guo, G. Chee, D. O'Connell, J. Dhou, S.and Fu, K. Singhrao, D. Ionascu, D. Ruan, P. Lee, D. A. Low, et al. Reconstruction of a high-quality volumetric image and a respiratory motion model from patient cbct projections. *Medical Physics*, 46(8):3627–3639, 2019.

[Gen79] D. Gennery. Stereo-camera calibration. In *Proceedings ARPA IUS Workshop*, volume 1, pages 101–107, 1979.

[HBA+16] J. Hahn, H. Bruder, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelriess. Reduction of motion artifacts in cardiac ct based on partial angle reconstructions from short scan data. In *Medical Imaging 2016: Physics of Medical Imaging*, volume 9783, pages 321–329. SPIE, 2016.

[Hir07] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341, 2007.

[HS97] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[JKKR18] S. Jang, S. Kim, M. Kim, and J. Ra. Head motion correction based on filtered backprojection for x-ray ct imaging. *Medical Physics*, 45(2):589–604, 2018.

[JZG+19] C. Jiang, Q. Zhang, Y.i Ge, D. Liang, Y. Yang, X. Liu, H. Zheng, and Z. Hu. Wasserstein generative adversarial networks for motion artifact removal in dental ct imaging. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, pages 790–797, 2019.

[KBM+18] A. Klugmann, B. Bier, K. Müller, A. Maier, and M. Unberath. Deformable respiratory motion correction for hepatic rotational angiography. *Computerized Medical Imaging and Graphics*, 66:82–89, 2018.

[KMBS21] Y. Ko, S. Moon, J. Baek, and H. Shim. Rigid and non-rigid motion artifact reduction in x-ray ct using attention module. *Medical Image Analysis*, 67:101883, 2021.

[KNKF11] J. Kim, J. Nuyts, Z. Kuncic, and R. Fulton. A method of motion tracking during ct for motion correction. In *2011 IEEE Nuclear Science Symposium Conference Record*, pages 3816–3819, 2011.

[Kon98] K. Konolige. Small vision systems: Hardware and implementation. In *Robotics research*, pages 203–212. Springer, 1998.

[Kon10] K. Konolige. Projected texture stereo. In *2010 IEEE International Conference on Robotics and Automation*, pages 148–155, 2010.

[LNW+19] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass. Motion estimation and correction in cardiac ct angiography images using convolutional neural networks. *Computerized Medical Imaging and Graphics*, 76:101640, 2019.

[LSU16] W. Luo, A. G. Schwing, and R. Urtasun. Efficient deep learning for stereo matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5695–5703, 2016.

[MB12] J Moorees and E. Bezak. Four dimensional ct imaging: A review of current technologies and modalities. *Australasian physical engineering sciences in medicine / supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine*, 35:9–23, 02 2012.

[MBC+15] K. Müller, M. Berger, Jh. Choi, S. Datta, S. Gehrisch, T. Moore, M. P. Marks, A. K. Maier, and R. Fahrig. Fully automatic head motion correction for interventional c-arm systems using fiducial markers. *Proceedings of the 13th Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Newport, RI, USA*, pages 1–4, 2015.

[Mil85] D. Mills. Network time protocol (ntp). Technical report, 1985.

[MLS+20] R. Lee MacDonald, Y. Lee, J. Schasfoort, H. Soliman, A. Sahgal, and M. Ruschin. Real-time infrared motion tracking analysis for patients treated with gated frameless image guided stereotactic radiosurgery. *International Journal of Radiation Oncology* Biology* Physics*, 106(2):413–421, 2020.

[MNC+21] J. Maier, M. Nitschke, J.n Choi, G. Gold, R. Fahrig, B. Eskofier, and A. Maier. Rigid and non-rigid motion compensation in weight-bearing cbct of the knee using simulated inertial measurements. *IEEE Transactions on Biomedical Engineering*, pages 1–1, 2021.

[MNH+19] M. Dashtbani Moghari, P. Noonan, D. Lewis Henry, R. Fulton, N. Young, K. Moore, and A. Kyme. Characterization of the intel realsense d415 stereo depth camera for motion-corrected ct imaging. In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3, 2019.

[MPHG91] D. Marr, T. Poggio, E. C Hildreth, and W. L Grimson. *A computational theory of human stereo vision*. Springer, 1991.

[NST+19] S. Niebler, E. Schömer, H. Tjaden, U. Schwanecke, and R. Schulze. Projection-based improvement of 3d reconstructions from motion-impaired dental cone beam ct data. *Medical Physics*, 46(10):4470–4480, 2019.

[OJS+17] S. Ouadah, M. Jacobson, J. Webster Stayman, T. Ehtiati, C. Weiss, and J. H. Siewerdsen. Correction of patient motion in cone-beam ct using 3d–2d registration. *Physics in Medicine & Biology*, 62(23):8813, 2017.

[Par82] D. L. Parker. Optimal short scan convolution reconstruction for fan beam ct. *Medical physics*, 9(2):254–257, 1982.

[PAST15] R. Pauwels, K. Araki, JH Siewerdsen, and S. Thongvigitmanee. Technical aspects of dental cbct: state of the art. *Dentomaxillofacial Radiology*, 44(1):20140224, 2015.

[Pos80] J. Postel. User datagram protocol. Url: https://www.rfc-editor.org/rfc/rfc768 (visited 13.01.2024), 1980.

[RND04] T. Rodet, F. Noo, and M. Defrise. The cone-beam algorithm of feldkamp, davis, and kress preserves oblique line integrals. *Medical Physics*, 31(7):1972–1975, 2004.

[SBB+17] C. Syben, B. Bier, M. Berger, A. Aichert, R. Fahrig, G. Gold, M. Levenston, and A. Maier. Joint calibration and motion estimation in weight-bearing cone-beam ct of the knee joint using fiducial markers. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 494–497, 2017.

[SF08] W. Scarfe and A. Farman. What is cone-beam ct and how does it work? *Dental Clinics of North America*, 52(4):707–730, 2008.

[SHG+11] R. Schulze, U. Heil, D. Groβ, D. D. Bruellmann, E. Dranischnikow, U. Schwanecke, and E. Schoemer. Artefacts in cbct: a review. *Dentomaxillofacial Radiology*, 40(5):265–273, 2011.

[SHT09] A. Segal, D. Haehnel, and S. Thrun. Generalized-icp. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009.

[SHZ+21] A. Sisniega, H. Huang, W. Zbijewski, J. W. Stayman, C. R. Weiss, T. Ehtiati, and J. H. Siewerdsen. Deformable image-based motion compensation for interventional cone-beam CT with a learned autofocus metric. In Hilde Bosmans, Wei Zhao, and Lifeng Yu, editors, *Medical Imaging 2021: Physics of Medical Imaging*, volume 11595, pages 241 – 248. SPIE, 2021.

[SNCS+18] R. Spin-Neto, C. Costa, D. MRA Salgado, N. RM Zambrana, E. Gotfredsen, and A. Wenzel. Patient movement characteristics and the impact on cbct image quality and interpretability. *Dentomaxillofacial Radiology*, 47(1):20170216, 2018.

[SNKH+20] R Spin-Neto, C Kruse, L Hermann, L-L Kirkevang, and A Wenzel. Impact of motion artefacts and motion-artefact correction on diagnostic accuracy of apical periodontitis in cbct images: an ex vivo study in human cadavers. *International Endodontic Journal*, 53(9):1275–1288, 2020.

[SSC+16] A. Sisniega, J. W. Stayman, Q. Cao, J. Yorkston, J. H. Siewerdsen, and W. Zbijewski. Image-based motion compensation for high-resolution extremities cone-beam CT. In Despina Kontos and Thomas G. Flohr, editors, *Medical Imaging 2016: Physics of Medical Imaging*, volume 9783, pages 156 – 162. SPIE, 2016.

[SSY+17] A. Sisniega, J. W. Stayman, J. Yorkston, J. H. Siewerdsen, and W. Zbijewski. Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion. *Physics in Medicine and Biology*, 62(9):3712–3734, apr 2017.

[STS+19] A. Sisniega, G. Thawait, D. Shakoor, J. H. Siewerdsen, S. Demehri, and W. Zbijewski. Motion compensation in extremity cone-beam computed tomography. *Skeletal radiology*, 48(12):1999–2007, 2019.

[Tur01] H. Turbell. *Cone-beam reconstruction using filtered backprojection*. PhD thesis, Linköping University Electronic Press, 2001.

[WCH+92] J. Weng, P. Cohen, M. Herniou, et al. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence*, 14(10):965–980, 1992.

[WKKK13] J. Wicklein, Y. Kyriakou, W. A. Kalender, and H. Kunze. An online motion- and misalignment-correction method for medical flat-detector CT. In Robert M. Nishikawa, Bruce R. Whiting, and Christoph Hoeschen, editors, *Medical Imaging 2013: Physics of Medical Imaging*, volume 8668, pages 466 – 472. SPIE, 2013.

[WPS+19] N. S. West, M. J. Parkes, C. Snowden, J. Prentis, J. McKenna, M. Shahid Iqbal, J. Cashmore, and C. Walker. Mitigating respiratory motion in radiation therapy: Rapid, shallow, non-invasive mechanical ventilation for internal thoracic targets. *International Journal of Radiation Oncology*Biology*Physics*, 103(4):1004–1010, 2019.

[WYMF18] G. Wang, J. C. Ye, K. Mueller, and J. Fessler. Image reconstruction is a new frontier of machine learning. *IEEE transactions on medical imaging*, 37(6):1289–1296, 2018.

[XYZL18] S. Xie, C. Yang, Z. Zhang, and H. Li. Scatter artifacts removal using learning-based method for cbct in igrt system. *IEEE Access*, 6:78031–78037, 2018.

[Zha00] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[ZTS+20] X. Zhang, J.n Tang, G. C. Sharp, L. Xiao, S. Xu, and H. Lu. A new respiratory monitor system for four-dimensional computed tomography by measuring the pressure change on the back of body. *The British Journal of Radiology*, 93(1108):20190303, 2020. PMID: 31912746.