

Stereoscopic Depth Imaging for Improved Point Cloud Generation of Skeletal Surfaces in IR-guided Orthopaedic Surgery

Master's thesis in computer science and engineering

MATS CEDERVALL
OSCAR DAHLQVIST

MASTER'S THESIS 2024

Stereoscopic Depth Imaging for Improved Point Cloud Generation of Skeletal Surfaces in IR-guided Orthopaedic Surgery

MATS CEDERVALL
OSCAR DAHLQVIST



Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Stereoscopic Depth Imaging for Improved Point Cloud Generation of Skeletal Surfaces in IR-guided Orthopaedic Surgery

MATS CEDERVALL
OSCAR DAHLQVIST

© MATS CEDERVALL, 2024.
© OSCAR DAHLQVIST, 2024.

Advisor: Hassan Nemati, Ortoma AB
Supervisor: Ulf Assarsson, Department of Computer Science and Engineering, Chalmers
Examiner: Erik Sintorn, Department of Computer Science and Engineering, Chalmers

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: The image depicts three components: a surgical navigation system, a depth camera, and a 3D reconstruction of intraoperative knee geometry.

Typeset in L^AT_EX
Gothenburg, Sweden 2024

Stereoscopic Depth Imaging for Improved Point Cloud Generation of Skeletal Surfaces in IR-guided Orthopaedic Surgery
MATS CEDERVALL
OSCAR DAHLQVIST
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis proposes a flexible approach of combining traditional optical navigation systems with a visible light-based depth camera for aligning preoperative CT models with the corresponding bone exposed during active surgery, providing real-time feedback of implant positioning to the surgeon based on a preoperative plan.

By leveraging the Intel[®] RealSense[™] D405 depth camera, the study investigates the benefits and challenges of using stereoscopic depth imaging for 3D reconstruction of musculoskeletal surfaces. The primary goal being to evaluate how this technology can be used to reduce surgery duration in orthopaedic total knee arthroplasty (TKA) surgery.

The proposed method transforms the depth images captured using the depth camera into a navigation system's frame of reference, accomplished through a series of preoperative calibration steps. The reconstructed intraoperative 3D model is then aligned with the CT model using iterative closest point (ICP) algorithms. This thesis also includes an investigation into using light polarisation filters, analysis of hyperparameter tuning, and accuracy evaluation of the Intel[®] RealSense[™] D405 camera's depth estimation capabilities.

Experimental validation includes a mock surgery on pig cadaver parts to simulate intraoperative conditions. Results demonstrate that the proposed approach achieves good alignment accuracy at around 0.5 mm, though it exhibits higher variability compared to competing methods.

In conclusion, the integration of this class of depth cameras with optical navigation systems is a viable solution for improving the speed of knee registration in orthopaedic surgery, but future work is encouraged to address depth estimation inaccuracies.

Keywords: surgical navigation tools, image-guided orthopaedic surgery, total knee arthroplasty, stereoscopic RGBD camera, Intel[®] RealSense[™] D405

Acknowledgements

We would like to express our sincere thanks to our mentor, Hassan Nemati, for his unwavering support and insightful guidance throughout the project. Our appreciation also extends to the exceptional team at Ortoma AB, who generously shared their expertise and warmly welcomed us during our time at their office.

We are also grateful to our supervisor Ulf Assarsson at Chalmers, for his support and contributions to the project. Lastly, we acknowledge everyone else who contributed to the project in various ways, helping us finalise this thesis.

Mats Cedervall & Oscar Dahlqvist, Gothenburg, 2024-06-24

Contents

Glossary	xi
1 Introduction	1
1.1 Problem Description	2
1.2 Purpose of the Study	4
1.2.1 Research Questions	4
1.3 Delimitations	4
1.4 Ethical Considerations and Risks	5
2 Background	7
2.1 Traditional IGOS Method	7
2.2 Total Knee Arthroplasty	7
2.2.1 Anatomy of the Knee	8
2.3 Computed Tomography (CT)	8
2.4 3D Representations	9
2.4.1 Point Clouds and Meshes	9
2.4.2 Homogeneous Coordinates and Transformations	9
2.5 Medical Navigation Systems	10
2.5.1 Patient Space	11
2.5.2 Navigation System Tool Calibration	11
2.5.3 Using Navigation Systems To Find Bone	11
2.6 Cameras	12
2.6.1 Intrinsics	12
2.6.2 Extrinsics	12
2.6.3 Camera Space	12
2.6.4 Perspective-n-Point (PnP)	13
2.6.5 Stereoscopic Depth Imaging	13
2.7 Registration Algorithms	14
2.7.1 Least Squares Point Alignment	14
2.7.2 Iterative Closest Point (ICP)	15
2.7.2.1 Local Registration	15
2.7.2.2 Global Registration	15
2.8 RANSAC	15
2.9 Polarised Light and Reflective Surfaces	16
3 Related work	17
3.1 Summary	19
4 Methodology	21
4.1 Experiments	22

4.1.1	Equipment	22
4.2	Camera Tracking	23
4.2.1	Camera Space	24
4.2.2	Camera Space to Navigation Space	24
4.2.2.1	Shared Geometry Calibration	25
4.2.2.2	Multi Frame Calibration	26
4.2.3	Camera-Navigation Time Synchronisation	27
4.3	CT Alignment	27
4.3.1	Filtering	28
4.3.2	Frame Averaging	28
4.3.3	ICP Registration	29
4.4	Hyperparameters	29
4.5	Polarisation	30
4.6	Data Evaluation	31
4.6.1	Baseline Evaluation	31
4.6.2	Alignment Evaluation	32
5	Results	33
5.1	Depth Estimation	33
5.2	Camera Marker Calibration	35
5.3	Frame Synchronisation	36
5.4	Manual Pointer Tip Positions	36
5.5	Alignment	36
5.5.1	Hyperparameters	36
5.5.2	Calibration Sensitivity	38
5.5.3	Comparison to Traditional Method	39
5.5.4	Depth Accuracy	41
6	Analysis & Discussion	43
6.1	Calibration	43
6.2	Depth Estimation	43
6.3	Hyperparameters	44
6.4	Depth Accuracy	45
6.5	Comparison to Traditional Alignment	46
6.6	Time	46
6.7	Future work	46
6.8	Summary	47
7	Conclusion	49
	Bibliography	51

Glossary

- Affine** A geometric transformation that maintains lines and parallelism. [9](#), [14](#), [24](#), [26](#)
- Cartesian Product** The set of all ordered pairs that can be obtained by combining each element of one set with each element of another set. [27](#)
- CT** Computed tomography, uses X-rays to create detailed images and 3D models of the internal structure of the human body. Detailed in Section [2.3](#). [1–3](#), [7](#), [8](#), [14](#), [17](#), [18](#), [21–23](#), [27–29](#), [32](#), [36](#), [40](#), [45](#), [46](#), [49](#)
- Distal femur** The lower end of the femur, part of the knee. Illustrated in Figure [2.2.1](#). [8](#), [18](#)
- Fiducials** Easily recognisable shapes, such as dots, crosses, or patterned targets, designed to be clearly identifiable for computers. [3](#), [11](#), [23–26](#)
- Fill rate** The percentage of a scene’s depth information successfully captured, affected by occlusions and surface properties. Detailed in Section [2.6.5](#). [3](#), [14](#), [18](#), [33](#), [34](#), [43–45](#)
- Frobenius norm** The square root of the sum of the squared elements of a matrix. [26](#)
- ICP** Iterative closest point. An algorithm to compute the affine transformation between two similar structures. Described further in Section [2.8](#). [4](#), [15](#), [17–19](#), [28–30](#), [32](#), [44](#), [46](#), [47](#)
- IGOS** image-guided orthopaedic surgery. [1](#), [2](#), [7](#), [14](#), [17](#), [19](#), [21](#)
- IR** Infrared. [3](#), [10](#), [11](#), [18](#), [22](#), [23](#), [47](#)
- Least Squares Point Alignment** An algorithm to compute the affine transformation between two differently oriented structures. Described further in Section [2.7.1](#). [14](#), [15](#), [25](#)
- Marker** An uniquely shaped object equipped with IR reflective surfaces, tracked by IR cameras. Illustrated in Figure [2.5.1a](#). [3](#), [7](#), [11](#), [22–26](#), [31](#), [43](#), [46](#)
- Patellar surface** A region at the lower end of the femur, part of the knee. Illustrated in Figure [2.2.1](#). [8](#), [28](#), [33](#)
- Patient space** A frame of reference relative to the patient during surgery. Detailed in Section [2.5.1](#). [3](#), [24](#)
- PnP** Perspective-n-Point. An algorithm that computes the pose of a camera in some world space. Described further in Section [2.6.4](#). [13](#), [25](#)

- RANSAC** Random sample consensus. An algorithm for handling data sets with outliers. Described further in Section 2.8. 15, 30, 44
- RGBD** An image that also contains depth information (red, green, blue, depth). 4, 25, 47
- RGBD camera** A camera that in addition to capturing ordinary colour data also estimates depth information for each pixel. Produces RGBD images instead of typical RGB images. (red, green, blue, depth). 3, 4, 21, 22, 33, 44, 45, 47
- ROI** Region of interest. 1, 2
- Sobel filter** A type of commonly used edge-detection filter. 27
- Temporal noise** A metric quantifying fluctuations in-depth data over time, defined as the standard deviation of repeated depth captures, and detailed in Section 2.6.5. 3, 13, 18, 31, 33, 34, 43, 44, 46
- TKA** total knee arthroplasty, also known as joint replacement surgery. 1, 2, 4, 7, 8, 18, 19, 43, 46
- Voxel** A voxel is a 3D unit of space, representing a value in a grid, used in 3D modelling and imaging. 8, 29, 44

1

Introduction

Modern surgical navigational tools have been developed over the past 30 years and have become increasingly important since their introduction [1]. In addition to enabling significantly safer, more accurate, and less invasive procedures, the technology has allowed for new sophisticated techniques to be integrated into the surgical routine [1]–[3]. While surgical navigational tools consist of a broad span of technologies, this paper focuses on a limited part of image-guided orthopaedic surgery (IGOS) in the context of total knee arthroplasty (TKA), also known as total knee replacement [4]. Specifically, it addresses the task of determining the pose of the target bone during ongoing surgery.

The central idea of orthopaedic surgery is to restore the physiological function of the musculoskeletal system and includes tasks such as reconstruction of torn ligaments, treatment for arthritis, dislocations, or joint replacement [3], [5]. TKA is but one example of joint replacement surgery; in which the damaged cartilage and bone in the joint is removed and replaced with artificial parts [4], [6], see Figure 1.0.1.

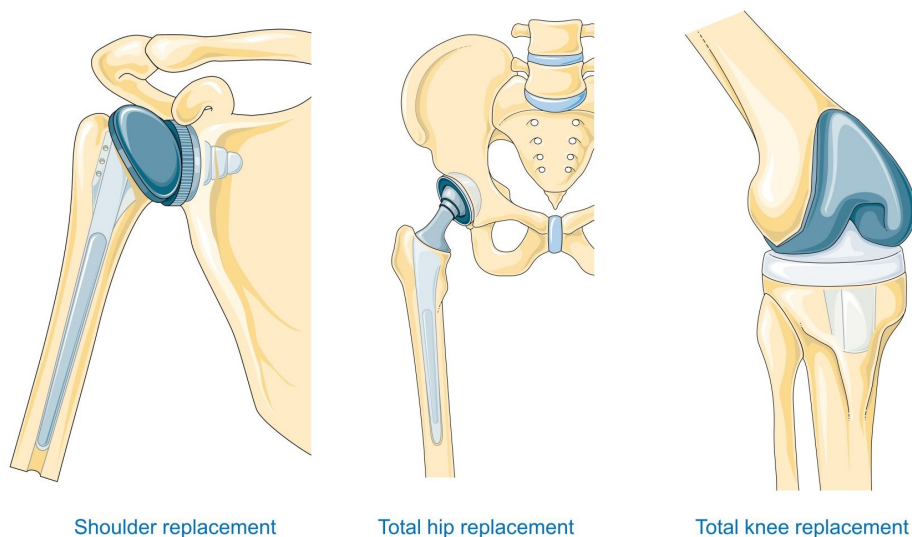


Figure 1.0.1: Illustrates the results of common orthopaedic joint replacement surgeries. From [6] CC-BY 4.0 / Cropped from original.

Whereas the typical surgical routine can be split into three stages: pre-, intra-, and post-operative stage, IGOS is primarily only concerned with the first two. The preoperative stage involves preparation and planning for the surgery, and in IGOS this typically includes the collection of 3D anatomic skeletal data of the patient using, for example, computed tomography (CT) [3]. The 3D model of the skeletal region of interest (ROI) obtained from CT scans will hereinafter be referred to as the *CT model*.

An orthopaedic surgical plan may typically involve a number of procedures performed on the skeletal ROI, that are mapped out in varying detail using the CT model. In IGOS systems, the intraoperative stage heavily relies on this CT model in order to provide accurate and real-time feedback to the surgeon [3]. The specific type of feedback provided by IGOS systems can vary, but may include visualisation of tool positioning in relation to pre-planned incision points, or instantaneous assessment of implant placement compared to the preoperative plan [7].

The intent of TKA surgery is to alleviate patient pain and to improve their range of motion, such that they may resume daily activities without being hindered by significant pain [4], [8]. However, with increasing numbers of younger and more active individuals opting for TKA surgery, stricter requirements are being placed on implant longevity and natural movement [7]. Thus, in pursuit of improved surgical outcomes, personalised alignment strategies are becoming more widely adopted, elevating the importance of precise customised implant positioning and the techniques for achieving it [9], [10].

In this paper, we propose a practical approach of combining traditional optical navigation systems with a visible light-based depth camera, presenting a fast and versatile solution for aligning a preoperative CT model of a patient’s knee with the patient’s physical knee during surgery. Furthermore, we investigate the effects of leveraging polarised light filters to improve the depth estimation quality of the depth camera for reflective musculoskeletal surfaces, combined with a detailed analysis of the impact of several key parameters in the aforementioned alignment procedure.

1.1 Problem Description

A key element of IGOS systems is the image registration process, which involves the aligning of preoperative 3D models with the patient’s corresponding surgical area during the operation [3]; see Figure 1.1.1 for an illustrative example. The preciseness of the registration is of paramount importance for the preoperative plan to be of proper use, since any misalignment introduces erroneous offsets between the procedures planned out using the CT model and the actual physical positioning of the patient during surgery [11].

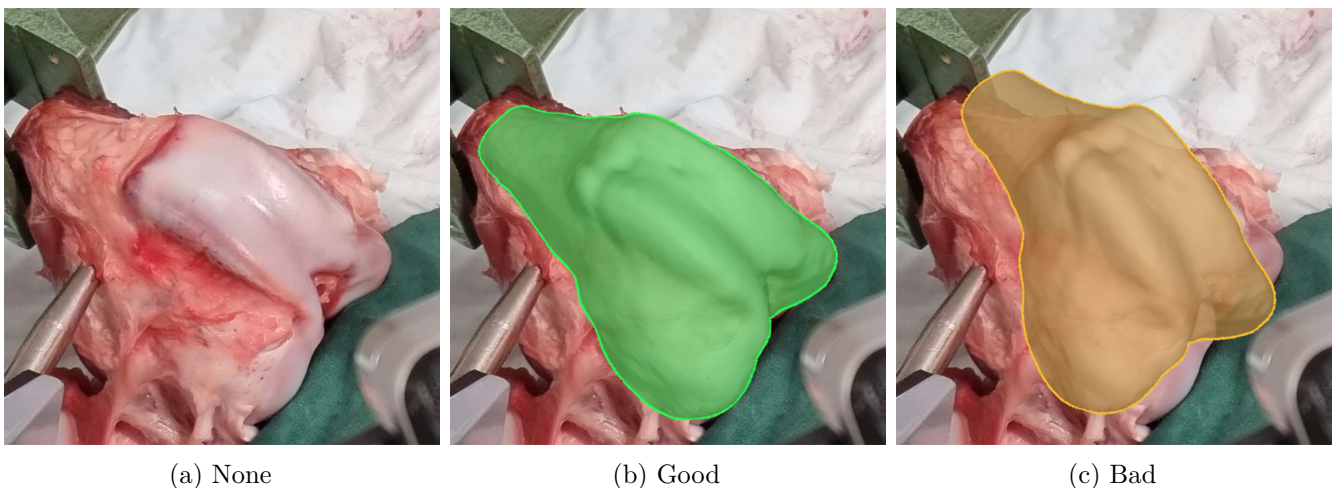


Figure 1.1.1: Illustrative example alignment between a pig knee and its CT-scanned counterpart.

While modern TKA encompasses multiple schools of thought regarding the optimal approach for performing the procedure, the base requirement of achieving a good implant positioning remains a common goal [9], [12]. The preciseness of the angle for the implant is regularly expressed in the range of ± 3 to 5° , motivating the need for high precision feedback systems [10], [12].

The first part of leveraging the CT-scanned anatomical 3D model is collecting intraoperative data to match it with. In this work, we employ the Intel[®] RealSense[™] D405 stereoscopic depth camera (also referred to as an [RGBD camera](#)) to 3D capture the musculoskeletal surfaces of the surgical incision site. Using a stereoscopic depth camera for 3D object reconstruction is, however, not without its own host of problems. Among the issues is the quality of the depth estimation, [temporal noise](#), [fill rate](#), frame aggregation, and the computation time for generating the final 3D reconstructed model. Please refer to Section [2.6.5](#) for detailed explanations of these terms.

Following the 3D surface reconstruction procedure, the subsequent step of aligning the captured 3D anatomical surfaces with the preoperative CT model also presents several non-trivial challenges. The source of which most notably originate from the fact that the CT model and the 3D captured surfaces during surgery are always dissimilar, and these differences can occasionally be substantial. The asymmetry stems from the differences in 3D reconstruction methodology; whereas the CT model is derived from the application of computed tomography (CT), the 3D model during surgery is obtained using a stereoscopic depth camera.

The key differences between the CT model and the intraoperative 3D model are the types of tissues captured and the context in which they are. While CT uses X-rays to capture anatomical data of the insides of a patient, see Section [2.3](#), it struggles to accurately capture cartilage. Which results in a fundamental difference between the preoperative CT model and what the [RGBD camera](#) sees during surgery. Additionally, the view of the [RGBD camera](#) is strictly limited to the musculoskeletal surfaces exposed through the surgical incision, meaning only part of the bone will be represented in the produced point cloud from the [RGBD camera](#).

Importantly, the [RGBD camera](#) is also limited in terms of depth accuracy and the estimated depth will fluctuate due to noise even with a stationary camera and scene. This variation is called [temporal noise](#) and introduces errors between frames that need to be accounted for when computing a 3D model from the produced point clouds [[13](#)], [[14](#)]. Furthermore, the [RGBD camera](#) is highly sensitive to surface textures, for reasons that will be expanded upon in Section [2.6.5](#), and will have a hard time capturing larger regions with homogeneous surfaces, such as smooth sections of bone and cartilage; making it harder to accurately capture the anatomical regions we are looking for. Together, these discrepancies between the CT model and the 3D model reconstruction from the [RGBD camera](#) make for a non-trivial 3D model-to-model alignment procedure.

Further, the alignment between the CT model and the camera-captured surfaces is to be performed in a reference space that is relative to the patient, such that if the patient is moved during surgery, the overlaid CT model should automatically be moved the corresponding amount. This is generally achieved using a surgical navigation system; in this paper, we employ a medical navigation system, which uses infrared (IR) cameras and IR reflective objects, generally referred to as [fiducials](#), attached in a unique shape to form a [marker](#) that can be tracked in the 3D space in view of the IR cameras. This system and its [markers](#) allow tools and objects in the scene to have their position and orientation to be tracked in real-time, and by attaching such a [marker](#) to the patient we can ensure objects are tracked in relation to the patient. We called this reference frame the [patient space](#).

To leverage this navigational system for tracking the surgical incision site relative to the patient, we have attached one of the aforementioned [markers](#) to the [RGBD camera](#). Through a series of calibration steps, the captured 3D point clouds from the [RGBD camera](#) are transformed to be expressed in [patient space](#), allowing the CT model to be aligned relative to the patient and be robust against limb movements during surgery.

This study focuses on evaluating the applicability of using the Intel[®] RealSense[™] D405 depth camera as a point cloud source for the CT model alignment procedure, in particular in combination with a medical navigational system for external tracking of the patient and the handheld [RGBD camera](#).

1.2 Purpose of the Study

This work aims to evaluate the efficacy of employing an affordable stereoscopic [RGBD camera](#) for 3D reconstruction of musculoskeletal surfaces exposed intraoperatively in orthopaedic surgery. Specifically, the study uses the Intel® RealSense™ D405 depth camera in the context of total knee arthroplasty (TKA) surgery. Additionally, the effects of leveraging polarised light filters will be evaluated, applied in order to remove surface specular highlights and thereby improving the 3D reconstruction quality of the target surfaces. This was done due to a concern of the reflective nature of musculoskeletal surfaces.

The primary intent is to propose a faster methodology for intrasurgical CT model alignment compared to a handheld manual pointer sampling based approach. A faster alignment procedure makes the routine more appealing for the surgical personnel, and reduces the inherent risks and inconvenience associated with longer surgery durations. Moreover, due to reduced user interaction during data acquisition compared to handheld pointer sampling, this approach is more appropriate for a sterile environment. That is, the proposed methodology carries benefits for both the surgeon and the patient.

The gap in literature identified in the study is the lack of 3D reconstruction quality evaluation for depth cameras applied to capture reflective and homogeneous musculoskeletal surfaces during orthopaedic surgery. This study seeks to address this gap by providing comprehensive data on the performance and potential advantages of using an [RGBD camera](#) in this context.

1.2.1 Research Questions

Below is a list of the primary research questions for this study, condensed from the discussion in the previous Section 1.2.

- **RQ1:** Is it viable to replace a few high quality sampled points with a large number of low quality points using the Intel® RealSense™ D405 depth camera ...
 - **RQ1.1:** ... in terms of point cloud to CT model alignment accuracy?
 - **RQ1.2:** ... while keeping total compute time from last [RGBD](#) frame acquisition to completed alignment under 60 seconds?
- **RQ2:** How does the application of polarised light filtering effect the Intel® RealSense™ D405 depth camera's point cloud acquisition for musculoskeletal surfaces, and in turn impact the point cloud to CT model alignment accuracy?
- **RQ3:** What is the depth estimation accuracy of the Intel® RealSense™ D405 depth camera for musculoskeletal surfaces?

1.3 Delimitations

This work will not be improving on existing registration methods, but rather utilise off-the-shelf iterative closest point (ICP) algorithms provided by the open-source library Open3D [15], combined in specific configurations to fit the problem at hand. In addition, the proposed methodology will mainly be compared to the traditional method of manual handheld pointer sampling technique. The collected points used for traditional method are manually aligned visually by us while assisted by an experienced professional at Ortoma AB. It should be noted that this introduces certain limitations regarding the generalisability and reproducibility of some of our findings. Furthermore, the study is confined to a single set of pig cadaver parts with corresponding CT models, mainly due to time and cost constraints.

1.4 Ethical Considerations and Risks

This research focuses on designing a solution meant for use during complex human surgery, hence it is of great importance that the solution considers reasonable diversity and differences between patients. The solution must be robust to diverse body compositions in terms of muscle mass, fat, and most importantly, differences in bone and cartilage in the target hip and knee regions. Without proper care, a solution could avoidably fail to properly handle the diverse conditions expected in real surgery and risk poor prosthetic alignment for the patient. The proposed methodology has therefore been designed under the requirement of being generalisable from the start.

Moreover, the project involves a limited amount of analysis of images and videos captured during real-patient orthopaedic surgeries, which naturally demands great responsibility for the handling of the material, to preserve and respect patient privacy and integrity. Precautions include offline data processing, a strict no-share policy, and thorough removal of the related data from utilised devices upon project completion.

Additionally, the work includes tests performed on pig parts sourced from a local butcher, that would otherwise be used for food. The ethical considerations are thus greatly limited, as the animal has already been slaughtered for another purpose.

Lastly, any work targeting the domain of human surgery typically involves similar risks to those discussed above, making them an inherent aspect of these types of works. We believe modern surgery should continue to develop and employ digital technologies to further improve surgery outcomes and decrease patient suffering, hence justifying the need for this paper.

2

Background

This chapter delves into the technicalities of what image-guided orthopaedic surgery (**IGOS**) entails, with a specific focus on its application in total knee arthroplasty (**TKA**). It aims to provide an appropriate level of understanding of the domain and equip the reader with the necessary background to comprehend the subsequent chapters of the thesis.

2.1 Traditional IGOS Method

The commonly used method for **IGOS** has remained largely unchanged since it was pioneered by Merloz et al. in 1995 [2]. This technique is used in a similar manner regardless of the specific bones being treated or the complexity of the skeletal intervention, and its general steps are outlined below:

1. Some time before the operation, the patient undergoes a **CT** scan. This scan is utilised to plan the upcoming procedure, identifying what is damaged, determining the necessary implants, and deciding their placement.
2. The operation room is fitted with a navigation system that tracks the position and orientation of purpose-calibrated surgical tools equipped with **markers**. Navigation systems and their **markers** will be further detailed in Section 2.5.
3. A reference navigation system **marker** is fitted to the bone that is to be tracked, creating a frame of reference relative to the patient. This is useful for when the patient is moved during surgery.
4. A navigation system pointer, shown in Figure 2.5.1b, is repeatedly positioned on the bone to create a collection of points that then correspond to the bone's surface.
5. A computer derives a transformation matrix that transforms the **CT** scan into the navigation system's frame of reference.
6. Some kind of mounting tool is used to position implants at the pre-planned locations and orientations. The correct position and orientation are generally indicated in real-time by a screen attached to the navigation system.

2.2 Total Knee Arthroplasty

TKA is a type of joint replacement surgery, in which damaged cartilage and bone is removed and replaced with artificial parts with the intent of relieving pain and restoring joint functionality [4], [6]. In Sweden, approximately 17,000 **TKA** surgeries were performed in 2022, with arthritis being the predominant reason for the majority of cases [16]. **TKA** is, in fact, one of the most widely performed orthopaedic surgeries in the world, hence unsurprisingly representing a notable portion of all orthopaedic procedures done in Sweden 2022 [5], [12].

The procedure is performed through a series of steps beginning with a detailed preoperative plan, often involving imaging techniques such as [CT](#) to create a 3D model of the patient’s knee [1], [12]. This model is used for precise planning of bone cuts and implant placement to achieve optimal performance of the implant and restoration of joint functionality [10]. Ensuring the accuracy of implant positioning is critical for reducing the risk of postoperative complications and enhancing the longevity of the implant [9], [11]. To achieve this, intraoperative navigation systems are employed, providing real-time feedback to the surgeon on how closely the implant aligns with its pre-planned location and rotation [1], [3].

2.2.1 Anatomy of the Knee

For the purposes of this study, only a very limited understanding of the anatomy of the knee is necessary. Specifically, the [distal femur](#) region which is exposed during [TKA](#) surgery, and illustrated in [Figure 2.2.1](#). The two highlighted regions in the figure corresponds to the [patellar surface](#) and the tibial surfaces respectively.

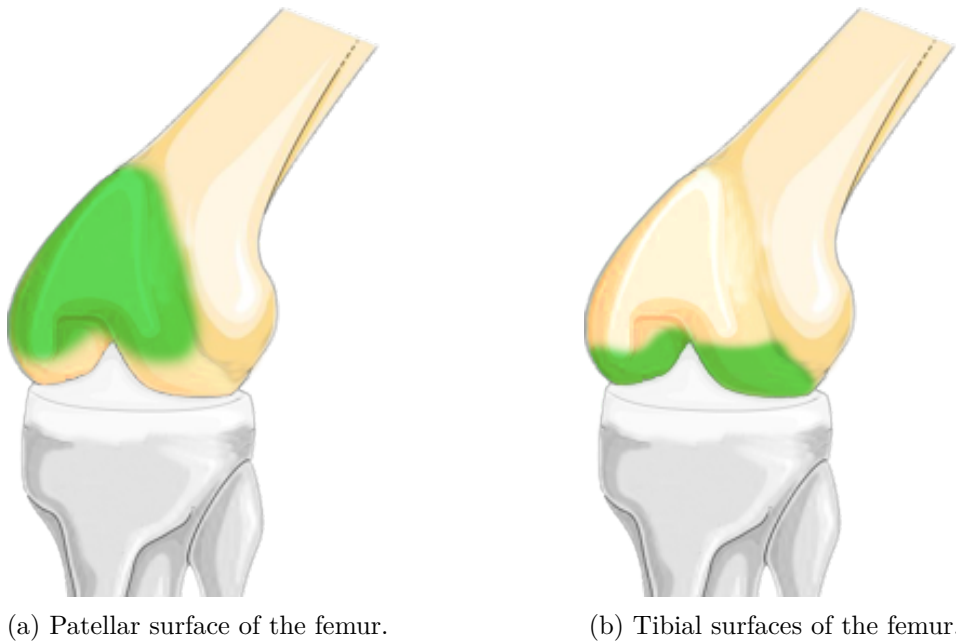


Figure 2.2.1: Illustrates the key regions of the [distal femur](#) for this study [17]. From [6] CC-BY 4.0 / Edited from original.

2.3 Computed Tomography (CT)

In the context of medical imaging, computed tomography ([CT](#)) uses X-rays to create detailed images of the internal structure of the human body. It involves the capturing of a series of 2D slices (called radiographs) from multiple angles that are computed upon using a reconstruction algorithm to form a single 3D grid of data, often referred to as a tomogram [18]. Each [voxel](#) (i.e. sub-volume) in the 3D spatial data grid (the tomogram) is assigned a, so-called, CT number that relates to the level of attenuation of the X-ray passing through that location [18], [19]. Attenuation is the intensity reduction of an X-ray passing through an object due to the absorption or scattering of photons, it roughly corresponds to the energy of the photons and the properties of the material being passed through, especially its density and atomic numbers [18], [19]. The CT number of each [voxel](#) is then compared to threshold values to categorise the corresponding tissue type, enabling the segmentation of for instance bone and soft tissues [18].

2.4 3D Representations

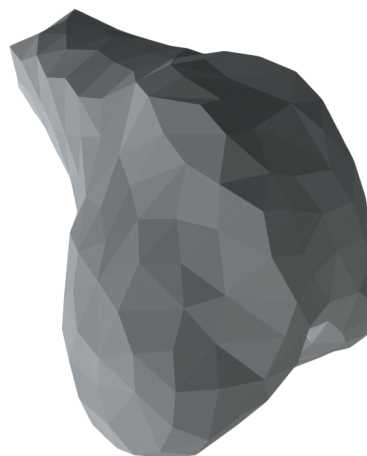
This section provides an overview of 3D data presentation and the use of homogeneous coordinates to transform 3D data between different frames of reference.

2.4.1 Point Clouds and Meshes

Point clouds are collections of 3D points, each defined by (x, y, z) coordinates relative to an origin, and often also include some associated colour data. Figure 2.4.1a illustrates an example of a point cloud. Meshes, on the other hand, consist of 3D points that are interconnected to form triangles, which define the surfaces of a 3D object. Figure 2.4.1b illustrates an example of a 3D mesh.



(a) Example of a 3D point cloud. Points are visualised as individual squares.



(b) Example of a 3D mesh.

Figure 2.4.1: Examples of the two types of 3D geometries relevant for this project.

2.4.2 Homogeneous Coordinates and Transformations

Representing transformations in 3D space is straightforward when using matrix multiplication on homogeneous coordinates. Homogeneous 3D coordinates are represented as a 4D vector, with the constant 1 appended to the $[x, y, z]$ positions, making it an element of $\mathbb{R}_{4 \times 1}$.

$$P = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Using homogeneous coordinates simplifies rotation and translation operations, as they can be performed easily using ordinary matrix multiplication. The 4×4 [affine](#) transformation matrices \mathbf{T} used for this purpose are matrices that when multiplied by a point in homogeneous coordinate form transform the point into \mathbf{T} 's frame of reference. \mathbf{T} is composed of three key parts: a 3×3 rotation matrix that handles rotation, a 3×1 translation vector that manages translation, and a padding row to enable the use of matrix multiplication. By multiplying a point with a transformation matrix it rotates and translates the point at the same time.

$$T = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$$

All coordinates are expressed in a particular frame of reference, denoted as ${}^A P$, where A is the frame of reference and P is the point. To convert a point to a different frame of reference, you multiply it by the transformation matrix of the target frame.

$${}^A P = \mathbf{T}_A \cdot P = \mathbf{T}_A \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

The previous equations apply not only to points but also to transformations. Transforming a transformation can be conceptualised as transforming a point with an attached direction vector.

Frames of reference can be converted between each other. The transformation matrix that converts positions from frame A to frame B is denoted as $\mathbf{T}_{A \rightarrow B}$. For all frames of reference and all points, the following equation holds¹:

$${}^B P = \mathbf{T}_{A \rightarrow B} \cdot {}^A P$$

Finally, matrix algebra can be performed as normal and produce meaningful results. Here is an example that showcases that by knowing the transformation matrix of A and B to some shared space X it is possible to transform points between A to B without ever knowing the point's positions in X .

$$\begin{aligned} {}^X P &= \mathbf{T}_{A \rightarrow X} \cdot {}^A P \\ {}^X P &= \mathbf{T}_{B \rightarrow X} \cdot {}^B P \\ \mathbf{T}_{B \rightarrow X} \cdot {}^B P &= \mathbf{T}_{A \rightarrow X} \cdot {}^A P \\ {}^B P &= \mathbf{T}_{B \rightarrow X}^{-1} \cdot \mathbf{T}_{A \rightarrow X} \cdot {}^A P \end{aligned}$$

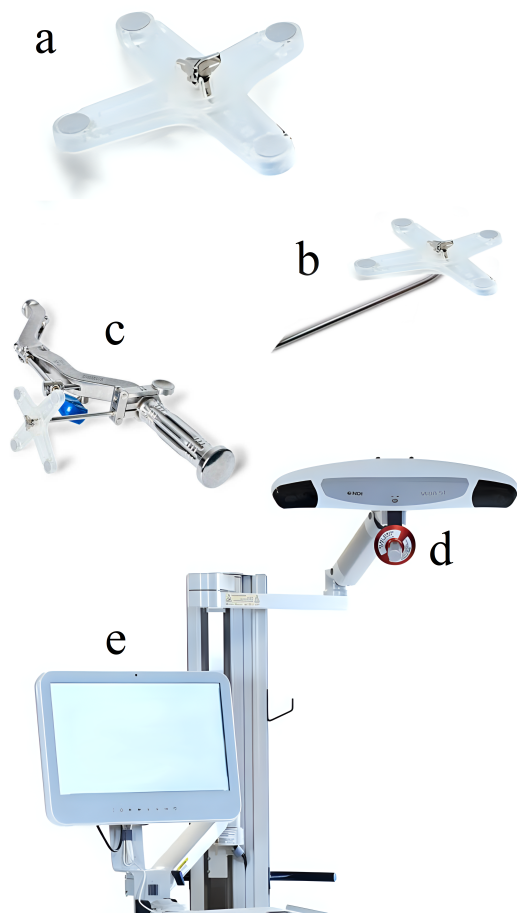
2.5 Medical Navigation Systems

Navigation systems used in intraoperative orthopaedic imaging consist of hardware and software that allow tracking of surgical equipment and assisting in delicate procedures during surgery. At its most basic, a navigation system uses some kind of imaging technique to see where surgical equipment is located with high precision. The choice of imaging technique varies, it can be infrared, visible light-based, electromagnetic, or X-rays [3]. The reason for tracking surgical equipment is to ensure that incisions and procedures are performed to a level of accuracy and precision that surpasses traditional handheld methods in categories such as cutting, drilling, and implant installation [3].

For this study, we focus on a particular class of navigation systems: a system comprising an **IR** camera that observes the surgical scene by tracking small **IR**-reflecting surfaces attached to specific shapes. Figure 2.5.1 illustrates the components of this type of navigation system.

¹This equation is a reformulation of the change-of-basis equation which is true for all matrices, not just 3D homogeneous coordinate transformations.

Figure 2.5.1: Components of a typical navigation system used for orthopaedic surgery. Images courtesy of Ortomax AB, used with explicit permission. Not to scale.



a Marker

A shape equipped with **fiducials** that can be tracked by the navigation system (d). These **fiducials** come in various forms; including optical patterns, **IR** reflective surfaces, light sources, or other features detectable in some arbitrary spectra.

b Pointer

A lightweight and nimble tool used to sample points on the surface of the bone.

c Mounting Tool

Used to prepare for and insert the prosthetic, varies greatly depending on the type of surgery and method. A navigation system can include multiple other tools, such as drills and cutting devices.

d Navigation System Camera

A specialised stereoscopic camera that detects the location of **markers** and provides real-time information on the position of the tools.

e Navigation System Screen

Displays real-time information regarding the position of tools and surgical steps, often connected to a standard PC running associated software.

2.5.1 Patient Space

Typically, during surgery, patient movement necessitates attaching a second **marker** to the patient [9]. This approach ensures consistent navigation by incorporating the frame of reference of the reference **marker**, which moves with the patient, rather than relying solely on the assumption that the bone remains fixed in the navigation system's frame of reference.

2.5.2 Navigation System Tool Calibration

A crucial feature of navigation systems is their ability to track the geometry of surgical instruments equipped with **markers**, rather than just the **markers** themselves. This requires identifying which specific tool is associated with each **marker** and understanding the tool's geometry. By knowing the **marker's** position and the tool's geometry, the system can accurately determine the location of specific parts, such as the tip of a scalpel or the end of a probe, relative to the navigation system [20].

2.5.3 Using Navigation Systems To Find Bone

The first step in finding the location of the intraoperative position of the bone is by calibrating a pointer, as depicted in Figure 2.5.1b. By putting the pointer's tip directly onto the bone during surgery it is possible to calculate where exactly that point of the bone is in the navigation systems frame of reference [21]. By taking multiple points it becomes possible to get an accurate understanding of both the position and rotation of the intraoperative bone.

2.6 Cameras

This subsection introduces the camera concepts and algorithms required to convert images into a meaningful 3D space.

2.6.1 Intrinsics

Camera intrinsics refer to the internal parameters of a camera that define how it projects a 3D scene onto a 2D image [22]. These parameters are typically combined into a 3×3 matrix and include:

Focal Length: f_x, f_y

The horizontal and vertical distances between the camera's lens and its image sensor, determining how much of the scene is captured and the size of objects in the image.

Principal Point c_x, c_y

The coordinates where the camera's optical axis intersects the image plane, representing the centre of the image.

Skew s

Skew defines the aspect ratio of pixels in the image. If the pixels are square, the skew is 1.

2.6.2 Extrinsics

Camera extrinsics define a camera's pose within a coordinate space, often referred to as world space. These parameters describe the camera's location and its orientation relative to the scene it captures [23]. Typically, extrinsics are represented as a 4×4 affine transformation matrix $\mathbf{T}_{W \rightarrow C}$, which transforms coordinates from world space to camera space.

To transform a point $P = [x, y, z]$ from world space to camera space, you first convert the point into homogeneous coordinates by appending a 1 to the vector. Then, you multiply this homogeneous vector by the extrinsics matrix (2A). Alternatively, to transform points from camera space back to world space, you multiply by the inverse of the extrinsics matrix (2B), (2C) [23].

$${}^C P = \mathbf{T}_{W \rightarrow C} \cdot {}^W P \quad (2A)$$

$${}^W P = \mathbf{T}_{W \rightarrow C}^{-1} \cdot {}^C P \quad (2B)$$

$${}^W P = \mathbf{T}_{C \rightarrow W} \cdot {}^C P \quad (2C)$$

${}^C P \in \mathbb{R}_{4 \times 1}$ *The position of some point in camera space.*

${}^W P \in \mathbb{R}_{4 \times 1}$ *The position of some point in world space.*

It is important to note that, unlike camera intrinsics, camera extrinsics are not fixed properties of the camera. Instead, camera extrinsics represent the relative pose of the camera at the moment the image is captured within a chosen world space. The world space serves as a frame of reference, which can be defined in relation to a point on a table, another camera, a room, or even the entire world, as long as the coordinate system is Euclidean.

2.6.3 Camera Space

Camera space refers to the three-dimensional frame of reference of the camera, where x , y , and z represent the horizontal, vertical, and depth offsets from the camera's origin. Consequently, moving

the camera also shifts the camera space relative to the surrounding environment.

If depth can be estimated using, for instance, a depth camera, a flat 2D image can be transformed into a 3D point cloud. This conversion is achieved by combining the depth image with the camera's intrinsic parameters, enabling the accurate mapping of the point cloud within the camera's frame of reference. Assuming there is no distortion or skew in the camera, Equation (2D) can be applied for this conversion [15].

$$P_{x,y} = \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = \begin{bmatrix} (x - c_x) \cdot D_{x,y}/f_x \\ (y - c_y) \cdot D_{x,y}/f_y \\ D_{x,y} \\ 1 \end{bmatrix} \quad (2D)$$

Focal length (part of camera intrinsics): $f_x, f_y \in \mathbb{R}$

Principal point (part of camera intrinsics): $c_x, c_y \in \mathbb{R}$

Depth Image: $D \in \mathbb{R}_{width \times height}$

Position in Camera Space: $P_{x,y} \in \mathbb{R}_{4 \times 1}$

2.6.4 Perspective-n-Point (PnP)

Perspective-n-Point (PnP) is an algorithm that computes the extrinsics matrix $\mathbf{T}_{W \rightarrow C}$ given 2D points on a flat camera image and corresponding 3D points in world space [24]. PnP determines the extrinsics matrix that minimises the error between the positions of features on the 2D camera image and their corresponding 3D ground truth as projected onto the camera plane [24]. As with any projection method, knowledge of the camera intrinsics is required in this process. Equation (2E) shows the arguments of the PnP algorithm.

$$\mathbf{T}_{W \rightarrow C} = solvePnP(I, \{(^{2D}p_i, ^W P_1), (^{2D}p_2, ^W P_2), \dots\}) \quad (2E)$$

$^{2D}p_i \in \mathbb{R}_{2 \times 1}$ The 2D coordinate of a point on the camera image.

$^W P_1 \in \mathbb{R}_{3 \times 1}$ The 3D coordinate of a point in world space.

$\mathbf{T}_{W \rightarrow C} \in \mathbb{R}_{4 \times 4}$ The camera extrinsics matrix.

$I \in \mathbb{R}_{3 \times 3}$ The camera intrinsics matrix.

2.6.5 Stereoscopic Depth Imaging

Stereoscopic depth imaging, often referred to as stereo vision, is a technique used to infer depth information from two slightly different perspectives, mimicking human vision [25]. This method involves capturing two images from two slightly offset cameras (left and right) and comparing them to determine depth. Achieving high-quality stereoscopic imaging is challenging because even minor and subtle differences are critical. Small, sometimes imperceptible, changes in the environment can lead to significant variations in the calculated depth perception.

No stereoscopic depth camera or algorithm is perfect, and common metrics used to evaluate their performance include:

Temporal Noise

A metric that quantifies fluctuation in depth data over time. It is defined as the standard deviation of the depth data captured across a series of captures [13]. High temporal noise

indicates that the depth measurements are inconsistent, which can affect the reliability of the depth information.

Fill rate

This metric measures the proportion of the scene that the camera successfully captures depth information for [13]. A low **fill rate** can occur if parts of the scene are not visible to both cameras or if the reconstruction algorithm fails to generate accurate depth data for certain regions. This can be due to occlusion, low-texture areas, or reflective surfaces.

By analysing these metrics, one can assess the performance and reliability of a stereoscopic depth camera under different conditions.

The effectiveness of stereoscopic imaging is greatly dependent on the ability to find correspondences between images [25]. Richly textured surfaces facilitate this process by providing distinct features that can be matched between the left and right images, enabling accurate depth calculations. In contrast, plain, monochromatic surfaces present challenges because they lack distinct features for establishing correspondences. In such cases, depth reconstruction must rely on interpolating data from surrounding better-textured areas, which can result in significant inaccuracies.

To address this weakness, a common solution is to introduce more texture in poorly textured areas [25]. For instance, by placing distinctly textured markers, coating the surfaces of the object with some material with more texture, or by projecting light patterns onto the surfaces. Naturally, the sterile nature of surgeries imposes constraints on what can be utilised in orthopaedic surgery, making for example coating-based methods impracticable.

Apart from stereoscopic depth imaging, there are numerous other methods for 3D reconstruction. For instance, structured-light-based techniques, that involves projecting a known light pattern onto the surface of an object, where the pattern deforms according to the object's geometry. These deformations are captured by a scanner's sensors and used to reconstruct a detailed 3D model of the object [25]. Another common method for depth estimation is time-of-flight, including techniques such as SONAR and RADAR. Which, in short, uses a transmitter and receiver pair that determines distance to an object by measuring the time from sending a signal and receiving it again as the result of it reflecting against the object [25].

2.7 Registration Algorithms

A registration algorithm is used to align points from two different frames of reference by determining an **affine** transformation between them [11]. This process ensures that features in one frame align with the corresponding regions in the other. Enabling, for instance, the merging of different data sources or seamless combination of multiple 3D scans of a common scene into a single cohesive scene. In image-guided orthopaedic surgery (**IGOS**), registration algorithms are essential for accurately mapping preoperative **CT** scans into the navigation system space, enabling precise localisation and guidance during surgery.

2.7.1 Least Squares Point Alignment

Least squares point alignment is a simple form of alignment algorithm that relies on having pairs of corresponding points where each item in the pair describes the same point from two different frames of reference. Using this collection of pairs, it finds a transformation matrix $\mathbf{T}_{A \rightarrow B}$ that transforms the points ${}^A P$ to ${}^B P$ with as little error as possible. Specifically, the objective is to find the $\mathbf{T}_{A \rightarrow B}$ that minimises the sum of the squared distances between coordinates in pairs [26].

$$\mathbf{T}_{A \rightarrow B} = \text{LeastSquares}(\{(^A P_1, ^B P_1), (^A P_2, ^B P_2), \dots\}) \quad (2F)$$

2.7.2 Iterative Closest Point (ICP)

ICP is a very popular and widely used point cloud registration method, and the original paper by Besl and McKay [27] has been cited over 12,000 times. **ICP** is an iterative, robust and stable algorithm [28] that unlike **least squares point alignment**, does not require points to have predefined correspondences at the start, instead finding these on its own.

The general **ICP** algorithm starts with two point clouds and an initial guess for the relative transformation matrix for positioning the source point cloud onto the target point cloud [29]. Note, the initial guess can be the identity matrix, but a better guess is most often preferred since the algorithm converges to the closest local minima [27]. Next, a correspondence set of point pairs between the two point clouds is created, by finding the closest point in the target point cloud for each point in the source point cloud. An improved guess of the relative transformation matrix is then computed using the set of correspondences and least squares, minimising the summed distance of all correspondence pairs [27]. The algorithm reiterates these two steps of populating a correspondence set and minimising the distances between corresponding points until the summed distance falls below a target threshold, or a maximum number of iterations is reached.

2.7.2.1 Local Registration

General **ICP** is not without problems; it only works if the initial alignment is reasonably correct, otherwise, **ICP** ends up finding a local minima that is likely to be considerably different than the global minima [28]. However, if the initial alignment is rather good then a better registration can commonly be achieved by fine-tuning it using general **ICP**. This type of fine-tuning using **ICP** is sometimes called local **ICP**.

2.7.2.2 Global Registration

An important class of **ICP** variants is those meant for global registration. These are used to provide an initial rough alignment guess for the local registration to start out with. A popular version is the one developed by Rasu et al. [30], which instead of 3D coordinates uses a higher dimensional coordinate space. These coordinates are called feature vectors and describe the local geometric characteristics of a 3D point. Open3D [15] implements a feature-matching-registration adaptation of **ICP**. It is largely the same as a normal feature vector but with some extra outlier checks using **RANSAC**.

2.8 RANSAC

RANdom **SAM**ple **CON**sensus (**RANSAC**) is an algorithm that estimates characteristics of datasets containing many outliers [31]. Although it was originally devised by Fischler and Bolles [32] for the purposes of image detection, **RANSAC** has proven very useful for large variety of tasks. The algorithm works by iteratively selecting random subsets of the data points and fitting a model to these subsets, evaluating how many data points fit the model within a predefined tolerance to identify inliers. By repeating this process and selecting the model with the most inliers, **RANSAC** effectively isolates the best-fit model even in the presence of significant noise and outliers. The model can be described by any number of objective functions and it is widely applicable in fields such as computer vision, 3D reconstruction, and various other forms of data fitting.

2.9 Polarised Light and Reflective Surfaces

When light reflects off many types of surfaces, it becomes partially polarised in the direction parallel to the surface, effectively turning the surface into a polariser [33]. Some non-reflective surfaces do the opposite, when polarised light interacts with them, it is partly absorbed and then re-emitted with a random polarisation [33].

Reflections are important because reflective materials cause specular highlights in images, which is generally undesirable since the regions corresponding with the specular highlights become overexposed, making their details hard to discern. By placing a linear polarisation filter in front of the camera lens, or lenses in the case of stereoscopic cameras, it is possible to remove some of these highlights by ensuring the direction of the polarisation filter is orthogonal to the polarisation direction of the highlights [33]. Notably, because the polarisation of the reflected light depends on the angle of the surface, complex scenes will contain surface reflections with varying polarisation; making complete exclusion of specular highlights infeasible for practical applications.

Specular highlights are of particular interest in this study, because feature-based 3D scanning techniques, such as those utilised in stereoscopic depth cameras, have been shown to be susceptible to holes and artefacts in the generated 3D point-clouds, due to the reflective and homogeneous nature of the skeletal surfaces [34].

3

Related work

This chapter provides an overview of relevant previous work conducted in the domain of image-guided orthopaedic surgery (IGOS), by summarising a number of earlier papers that are particularly significant for framing the context of this study.

He et al. [21] addressed the time-consuming process of intraoperative patient bone to CT registration in orthopaedic surgery by developing a novel approach utilising a compact 3D structured light surface scanner, a technique briefly described in Section 2.6.5. They employed the EinScan™ SP desktop scanner from Shining 3D Tech Co., Ltd., a device priced around \$2,000 at the time of reading. The scanner produced accurate high-density point clouds of the target bone geometry in about 30 seconds without using ionising radiation. However, this duration only includes the capturing of the bone's 3D surface, not the registration process itself. The study evaluated this method on 24 CT-scanned femur sawbones, which are synthetic anatomical models commonly used in medical training [35]. Accuracy was measured by comparing known points at pre-drilled holes on the sawbones to their corresponding positions on the registered CT model, achieving an accuracy of 0.44 ± 0.22 mm, surpassing the traditional pointer sampling technique's 0.68 ± 0.14 mm. This new method proved faster and more precise than conventional techniques. However, the authors caution that this technology should be tested in scenarios more realistic than sawbone models before being used clinically.

Other recent studies have, in addition to structured light scanners, employed other 3D reconstruction devices. Wein et al. [36] used a handheld ultrasound scanner combined with a real-time bone detection solution based on a novel bone-specific feature descriptor using filtering and confidence maps computed using a random walk framework [37]. The detected bone surfaces were then globally aligned using a controlled random search (CRS) algorithm with local mutation, followed by a fine adjustment using soft tissue-aware intensity-based registration. Evaluation was performed on the femur, tibia, and fibula of a human cadaver leg with implanted electromagnetic bone markers and calibrated tracking sensors for use as ground truth information. Median system error was reported at 3.7 mm across 11 datasets. The authors attributed most of this error to the two electromagnetic sensors used for ground truth, while noting that the visual alignment appeared excellent.

Similarly, Ciganovic et al. [38] used a handheld ultrasound scanner for 3D capturing of bone surfaces in orthopaedic procedures targeting the forearm. Employing phase symmetry to locate bone surfaces and a factor graph-based outlier removal strategy. The study did not involve intraoperative scanning and instead used the ultrasound scanner to capture the bones inside the forearm without involving surgery. Registration of the non-exposed patient bone was carried out against a 3D model sourced from magnetic resonance imaging (MRI). In total, three registration methods were tested: coherent point drift (CPD), point-to-plane ICP, and a Gaussian mixture models-based algorithm. Of these, only point-to-plane ICP achieved good registration accuracy within a reasonable amount of time. Although not directly comparable to knee anatomy, the accuracy in forearm registration was 0.57 ± 0.08 mm in about 15 seconds of compute time.

Hu et al. [39] recently published a proof of concept study that combined the use of an Intel®

RealSense™ D415 stereoscopic depth camera, priced around \$270 at the time of writing, with an IR navigation system for TKA surgery. Using the depth camera to capture the 3D surface of the distal femur exposed during surgery. This was combined with machine learning (ML) based bone segmentation and an improved ICP algorithm, called bounded iterative closest point (BICP), for the intraoperative patient bone to CT registration. Evaluated on a synthetic model of a human knee the system achieved an average drilling error of 3.64 ± 1.49 mm in position and $2.13 \pm 0.81^\circ$ in orientation. The authors conclude that future work should incorporate a higher accuracy depth camera and retraining of the bone segmentation models. Furthermore, it is important to note that the automatic bone segmentation ML model was tested solely on a synthetic knee. Consequently, real surgical conditions are anticipated to introduce new challenges, both in collecting training data and in achieving the robustness required to handle the more varied conditions expected in actual TKA surgeries.

In a comparison study by Burger et al. [13], the low-cost Intel® RealSense™ D415 stereoscopic depth camera, also used in the aforementioned [39] study, was evaluated against two other similarly priced stereoscopic depth cameras: Intel® RealSense™ D405, and Stereolabs® ZED-Mini. These tests were designed to evaluate their performance in surgical simulation environments. The Intel® RealSense™ D405 depth camera outperformed the other two devices in four out of five metrics, including depth accuracy and fill rate. However, it did not excel in temporal noise, where the RealSense™ D415 performed better. The study also noted poor performance for thin structures and reflective surfaces, such as surgical tools. The authors suggest further tests using light field cameras, which may be more robust to reflections and occlusions than the stereoscopic depth cameras used in the study.

In addition to utilising different imaging techniques and devices, significant efforts have been made to improve the registration procedure by enhancing the ICP algorithm. An example is the study by Hu et al. [39], as mentioned earlier, which used bounded iterative closest point (BICP) instead of the general ICP algorithm, outlined in Section 2.7.2. BICP enhances traditional ICP in femur registration tasks by using the rotational centre of the femur-pelvis joint to constrain its convergence [11]. The method requires four inputs: (1) 3D point samples on the distal femur during surgery, (2) reasonably corresponding 3D points in the CT model, (3) the 3D point of the hip centre of rotation during surgery, and (4) the corresponding estimated hip centre in the CT model [39]. The primary addition compared to traditional ICP is the hip centre of rotation, which can be estimated by attaching a tracker to the femur, rotating the leg, and fitting a sphere to the collected 3D points. The sphere's centre then provides a reasonable estimate of the rotational centre for the patient's leg during surgery [39]. BICP first aligns the corresponding hip centres of rotation and then incorporates additional steps in the iterative alignment process. A key step involves rotation around the hip centre to align the axis between the rotational centre and points on the distal femur before essentially performing the typical alignment of the corresponding distal femur points. Hu et al. [39] reported significant improvements in rotational accuracy with BICP compared to the traditional ICP algorithm, achieving a 4-8x reduction in rotational error and slightly shorter computation times. However, while improved robustness in both positional and rotational accuracy was also observed, only minor improvements were found for positional accuracy. Rodriguez y Baena et al. [11] noted that BICP is also suited for other long bone orthopaedic surgeries, as long as there is an estimable point on the axis to be aligned that is sufficiently distant from the other sampled points.

In their paper, Zhang et al. [40] introduce a hybrid feature-based registration method named CR-RAMSICP. This approach incorporates a two-step process: a coarse global registration using the singular value decomposition (SVD) algorithm, followed by a fine local registration with an improved ICP algorithm. The objective was to develop a straightforward yet effective intraoperative patient bone to CT registration technique for robot-assisted orthopaedic surgeries. The method specifically tackles the issue of local minima convergence, which arises from the asymmetry between dense point clouds from CT scans and sparse point clouds collected from patients using traditional pointer

sampling techniques. This challenge involves aligning a densely populated point cloud with one that has significantly fewer points. The experimental setup was nearly identical to the traditional **IGOS** method described in Section 2.1, with the addition of an orthopaedic robot system. The tool of the robot system was calibrated to enable its position to be expressed in the frame of reference of the surgical navigation system. The **ICP** algorithm was improved by primarily replacing the fixed distance threshold for valid correspondence pairs with an adaptable threshold that accounts for the local distribution around each source point. This was achieved by calculating a separate distance threshold for each source point based on the distance to its nearest neighbour in the target point cloud, thus making the algorithm more robust against outliers and poor coarse registrations. Alignment experiments conducted on a pig cadaver pelvis demonstrated an alignment accuracy of 0.56 ± 0.08 mm, compared to 0.89 ± 0.14 mm for standard **ICP**, with a computation time of approximately 10 seconds. The authors conclude that this method is suitable for other types of orthopaedic surgery, with the next step being evaluation in human phantom experiments, i.e., using models mimicking the human anatomy.

3.1 Summary

The previous work in image-guided orthopaedic surgery (**IGOS**) has explored various imaging techniques, devices, and enhancements to registration algorithms to improve the applicability and accuracy of these types of systems in real surgery. Studies have demonstrated the utility of structured light 3D scanners [21] and handheld ultrasound devices [36], [38] for accurate bone detection and alignment, though each comes with trade-offs such as capture speed and complexity of integration into surgical workflows. Research comparing registration methods [11], [38]–[40], has highlighted the potential of improved algorithms, such as bounded iterative closest point (**BICP**) [39], to enhance accuracy and robustness over traditional **ICP**, especially in rotational alignment.

The cost of imaging devices naturally influences depth estimation accuracy, with higher-end devices typically providing superior performance [21]. However, the strategic use of advanced registration algorithms can offset these disparities, making high-precision outcomes achievable even with more affordable equipment [39]. While standard **ICP** is widely used, customised improvements can provide substantial benefits for specific surgical applications like total knee arthroplasty (**TKA**) [39]. Evaluations of various depth cameras [13], [21], [39] have shown differing performance levels, emphasising the importance of selecting appropriate imaging technology for optimal results. These studies collectively highlight the necessity for advanced registration techniques and the integration of high-performance imaging devices with surgical navigation systems to enhance the precision and efficiency of orthopaedic procedures.

4

Methodology

This study intends to evaluate the performance of a specific depth camera, the Intel® RealSense™ D405, in CT alignment as part of an IGOS system, compared to the traditional manual point-taking method. The traditional point-taking technique, outlined in Section 2.1, will be modified by replacing steps 4 and 5 with a process using the aforementioned RGBD camera. A flowchart of our proposed method is depicted in Figure 4.0.1.

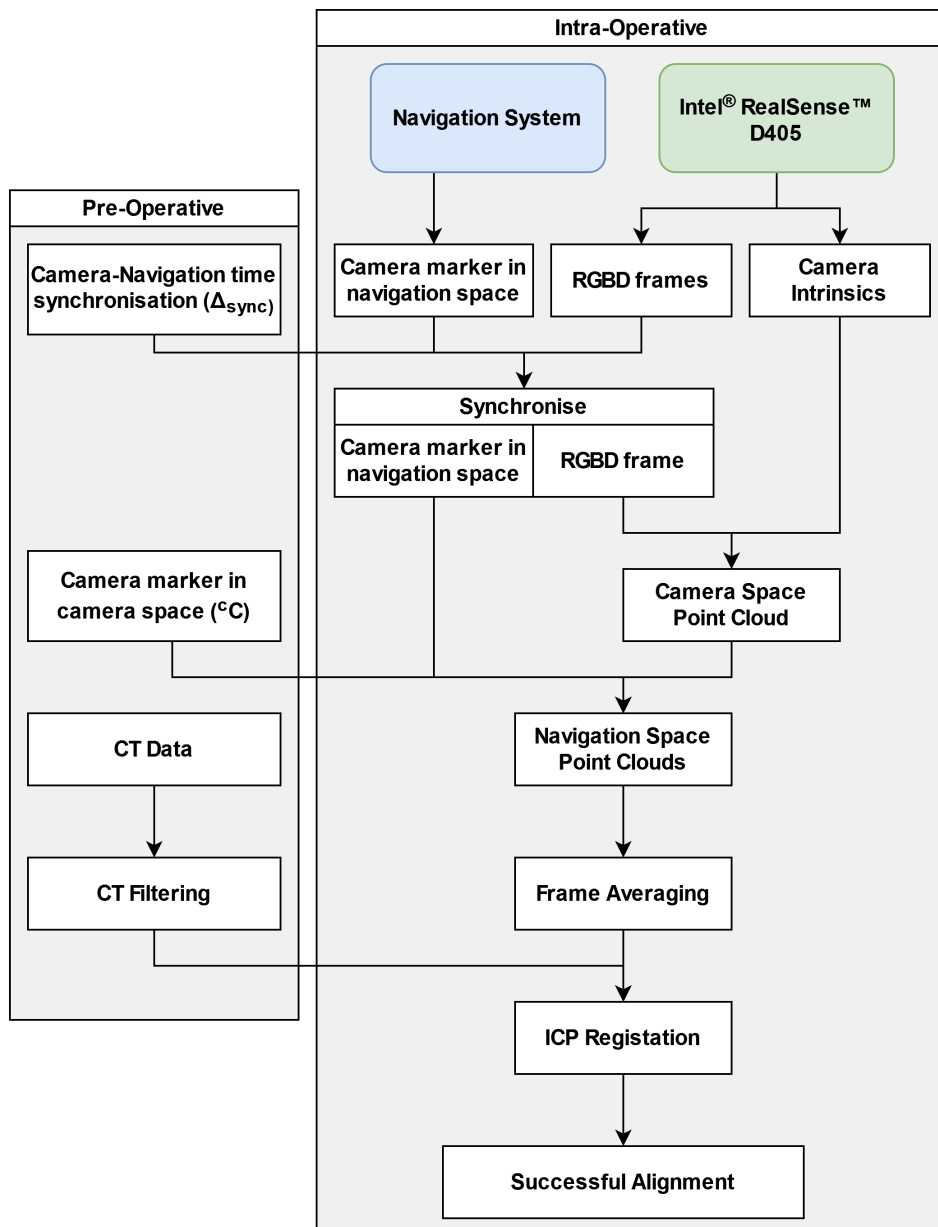


Figure 4.0.1: A flowchart of how our method obtains an alignment of CT data in navigation space.

The method can be divided into two parts: preoperative and intraoperative. The preoperative process consists of the following [CT](#) data preparation and system calibration steps:

1. Filter the [CT](#) data to isolate regions important for alignment. Described in [Section 4.3.1](#)
2. Calibrate the camera-space position of the navigation system [marker](#) that is attached to the camera ${}^C C$. Described in [Section 4.2.2.1](#) and [4.2.2.2](#).
3. Synchronise the system clocks between the navigation system and the [RGBD camera](#). Described in [Section 4.2.3](#).

The intraoperative parts of our proposed method are more complex. The goal is to utilise the [CT](#) scan, calibration data, and to combine the simultaneous real-time streams from both the navigation system and [RGBD camera](#) to produce an accurate [CT](#) alignment. The high-level steps are as follows:

1. Synchronise the streams of the navigation system and depth camera system.
2. Transform camera images into point clouds in navigation-space. Described in [section 4.2](#)
3. Filter out noise and view-dependant effects. Described in [section 4.3.2](#)
4. Align [CT](#) model in navigation space [4.3.3](#).

4.1 Experiments

To develop and evaluate our method, we required recordings of scenarios that closely mimic intraoperative conditions for knee surgery. These recordings were generated through two distinct approaches.

Firstly, we conducted continuous testing in an office environment equipped with a medical navigation system and 3D-printed bone models. This setup facilitated ongoing experimentation and iterative refinement of our method in a controlled setting.

Secondly, a thorough testing session was carried out at Sahlgrenska University Hospital, where a surgeon performed a mock surgery on the knee region of a pig. The primary goal of this mock surgery was to expose the knee bones to similar extent as in a real surgical procedure, without involving any steps related to attaching or preparing for a prosthetic. Once the bones were appropriately positioned, we conducted multiple passes with the camera to record depth camera frames and navigation system communication. Additionally, we performed several repetitions of the traditional method to serve as a benchmark for comparison.

4.1.1 Equipment

The following equipment was used during the thorough testing at Sahlgrenska University Hospital in addition to the preliminary testing in an office setting.

Intel® RealSense™ D405

The Intel® RealSense™ D405 is a compact [RGBD camera](#) that uses two visible-light cameras to estimate depth through stereoscopic imaging. We selected this depth camera because it is affordable, referenced in previous work, and supports a close working range of 7 cm to 100 cm, making it an excellent fit for our needs [\[41\]](#). As of the time of writing, it is priced at approximately \$270 [\[42\]](#).

Medical navigation system

We used an optical [IR](#)-based medical navigation system comprising of a pair of stereoscopic [IR](#)

cameras. It operates at 60 Hz, providing continuous updates on the location of [markers](#) with [IR](#) reflective [fiducials](#).

Navigation system marker

A [marker](#) tracked by the navigation system, shown alongside the Intel® RealSense™ D405 depth camera in Figure 4.1.1a.

Navigation pointer tool

A navigation system tool with a pointed tip, used for calibration and traditional method comparison, similar to the one shown in Figure 2.5.1b.

A medium power laptop

A consumer-grade laptop equipped with an Intel® Core™ i7-13700H CPU and 32GB RAM.

Camera Polarisation Filter

A camera attachable 3D-printed ring, fitted with a polarisation filter and depicted in Figure 4.1.1b.

Light Source Polarisation Filter

A box with a hole fitted with polarisation filters, depicted in Figure 4.1.1c.



(a) The Intel® RealSense™ D405 fitted with a custom 3D printed mount along with a navigation system [marker](#).



(b) Custom 3D printed polarisation filter, attachable to the depth camera.



(c) A box with a hole fitted with polarisation filters.

Figure 4.1.1: Selection of equipment used in the project.

4.2 Camera Tracking

This section describes the process of converting a series of depth camera images into point clouds within the navigation system's frame of reference.

The objective of this project is to replace manual point-taking with automated methods. Therefore, it is essential that the data captured by the camera can be transformed into the navigation system's frame of reference. This is crucial for aligning the [CT](#)-scanned bone with the patient, allowing the mounting tool (as described in Figure 2.5.1c) to be accurately positioned according to the preoperative plan.

For our study, we define the world space using the navigation system, which we will refer to as "Navigation Space" or "Navspace." Since the navigation system camera and the bone remain stationary during surgery, this provides a stable frame of reference.

For the sake of conciseness and clarity, we will not discuss conversions to [patient space](#). Our experiments involve stationary, non-living materials, so converting to [patient space](#) has no practical implications. However, adapting our method to use [patient space](#) would only require an additional matrix multiplication.

4.2.1 Camera Space

The Intel[®] RealSense[™] D405 camera produces two 2D images: a standard 8-bit RGB image matrix and a depth image matrix. The depth image matrix is half the resolution of the colour image and contains 16-bit integer scalars, where each step represents a tenth of a millimetre. Since the Intel[®] RealSense[™] D405 camera comes pre-calibrated with intrinsics and has negligible amounts of distortions, we can use Equation (2D) to convert these images into a 3D point cloud in camera space.

4.2.2 Camera Space to Navigation Space

To compute the transformation between two different frames of reference, we can use one of the following methods:

- A) **Using a Shared World Space:** Establish the relationship of each coordinate space relative to a common world space.
- B) **Identifying Common Points:** Determine the positions of multiple (linearly independent) points in both frames of reference.
- C) **Using an Affine Transformation:** Identify the [affine](#) transformation of a shared object in both frames of reference.

For our purposes, the third option is the most practical.

Using a shared world space (option A) would require either an additional navigation system or the placement of [fiducials](#) within the operating scene that are visible to both the camera and the navigation system. This approach is complicated by the different wavelengths used by the camera and navigation system and the challenge of placing [fiducials](#) in the limited field of view of the Intel[®] RealSense[™] D405 camera. The need to position the camera close to the incision for optimal resolution and parallax, combined with the restricted visibility beyond the patient's body, further complicates this method.

Identifying common points (option B) faces similar visibility challenges. However, both this method and using an [affine](#) transformation (option C) allow [fiducials](#) to be attached directly to the camera, making their transformations constant in camera space. Given that the navigation system directly provides the transformation of [markers](#), the third option is more efficient as it simplifies the process compared to identifying common points.

To implement the [affine](#) transformation method, we attach a navigation system [marker](#) to the camera. The transformation of this [marker](#) in camera space (${}^C C$) remains constant regardless of the camera's position in navigation space. Retrieving the transformation in navigation space (${}^N C$) is straightforward, as the navigation system provides this information directly.

By knowing the transformation matrix of the [marker](#) in both camera space and navigation space, we can compute the transformation between these spaces using the equations in (4A).

$$\begin{aligned}
{}^N C^{-1} \cdot {}^N P &= {}^C C^{-1} \cdot {}^C P \\
{}^N P &= {}^N C \cdot {}^C C^{-1} \cdot {}^C P \\
{}^N P &= \mathbf{T}_{C \rightarrow N} \cdot {}^C P
\end{aligned} \tag{4A}$$

${}^N C \in \mathbb{R}_{4 \times 4}$ Transformation of the [marker](#) attached to the camera in navigation space.

${}^C C \in \mathbb{R}_{4 \times 4}$ Transformation of the [marker](#) attached to the camera in camera space.

$\mathbf{T}_{C \rightarrow N} \in \mathbb{R}_{4 \times 4}$ Transformation to convert camera space to navigation space.

${}^C P \in \mathbb{R}_{4 \times 1}$ Point in camera space.

${}^N P \in \mathbb{R}_{4 \times 1}$ Point in navigation space.

Thus, with ${}^N C$ and ${}^C C$, we can easily convert the camera space point cloud to navigation space. We obtain ${}^N C$ directly from the navigation system, while ${}^C C$ requires some calculation. The crucial point is that ${}^C C$ is constant because the [marker](#) is fixed to the camera. Although we cannot measure ${}^C C$ directly, we have developed two methods to compute it. The following subsections will describe these methods in detail.

4.2.2.1 Shared Geometry Calibration

This idea is largely an implementation of method B as described in Section 4.2.2, by identifying the location of points in both frames of reference. In essence, this method uses a machined calibration block with distinct features that are easily visible in the camera’s [RGBD](#) image as well as measurable in navigation space. By using the block’s manufacturing specifications as a ground truth, we can determine transformation matrices where the block’s frame of reference is the world space.

This was accomplished using a metal block with machine-drilled divots that are clearly visible by the camera and have well-defined geometry that can be sampled by a navigation system pointer. The block’s manufacturing specifications define the block’s frame of reference, where one of the divots is set as the origin at $[0, 0, 0]$, with the remaining divots having coordinates as millimetre offsets from the origin. For clarity, we refer to this frame of reference as block space.

The following steps describe the process of (1) finding the transformation of navigation space relative to block space, (2) obtaining features in camera space and determining the relationship between camera space and block space, and (3) calculating ${}^C C$, the transformation of the camera marker in camera space.

1) Navigation Space to Shared Geometry

The navigation [markers](#) were too large to fit inside the calibration divots. To address this, we placed the tip of a navigation system pointer in the divot positions and extracted the tip positions in navigation space. By comparing the navigation-space and block-space coordinates of the divots, we computed the transformation matrix $\mathbf{T}_{N \rightarrow Block}$ between the two spaces using [least squares point alignment](#), described in Section 2.7.1.

2) Camera Space to Shared Geometry

Finding the transformation between camera space and block space $\mathbf{T}_{C \rightarrow Block}$ is more challenging because we do not know any inherent extrinsics. Instead, we relied on analysing the camera image and computing the extrinsics that could have produced the image of the block. This was solved using [PnP](#), described in Section 2.6.4. To run [PnP](#), we needed to locate the machined block’s divots on the 2D RGB image. Unlike traditional [fiducials](#), such as grids or QR codes, there is no pre-made solution for finding divots on this calibration block. Therefore, we performed a series of manually adjusted steps, including edge detection, box filters, convolutions, circle

finders, and image alignment, to locate the holes.

3) Camera Marker in Camera Space

With the transformation matrices from camera-space and navigation-space to block-space determined, we used Equation (4B) to solve for the ${}^C C$ matrix.

$$\forall M \in \mathbb{R}_A. \quad \begin{cases} {}^{Block} M &= \mathbf{T}_{N \rightarrow Block} \cdot {}^N M \\ {}^{Block} M &= \mathbf{T}_{C \rightarrow Block} \cdot {}^C M \end{cases} \\ \implies {}^C C = \mathbf{T}_{C \rightarrow Block}^{-1} \cdot \mathbf{T}_{N \rightarrow Block} \cdot {}^N C \quad (4B)$$

$\mathbb{R}_A \subset \mathbb{R}_{4 \times 4}$ *The set of all [affine](#) transformations.*

${}^C C \in \mathbb{R}_{4 \times 4}$ *The transformations of the [marker](#) attached to the camera in camera space.*

${}^C N \in \mathbb{R}_{4 \times 4}$ *The transformations of the [marker](#) attached to the camera in navigation space.*

$\mathbf{T}_{C \rightarrow Block} \in \mathbb{R}_{4 \times 4}$ *Transformation to convert camera space to block space.*

$\mathbf{T}_{N \rightarrow Block} \in \mathbb{R}_{4 \times 4}$ *Transformation to convert navigation space to block space.*

4.2.2.2 Multi Frame Calibration

This method provides a shortcut that bypasses the need to know the location of points in navigation space.

We record the camera images and navigation positions while viewing well-defined [fiducials](#), in this case, a chess grid. These patterns are ideal for determining the camera's extrinsics relative to the grid as the world space. The advantage of this method is that it eliminates the need to measure a world space using the navigation system. The only assumption made is that the chess grid remains stationary and fixed in navigation space.

Using the navigation space transformation formula (4A), we can set up a system of equations where ${}^C C$ is the only unknown variable. For each camera pose, the unknown navigation transform should result in the grid points being transformed to the same navigation space positions. This constraint is described in Equation (4C).

$$\left\{ \forall a, b \in [1 \dots (\#Frames)]^2. \quad {}^N C_a \cdot {}^C C^{-1} \cdot G_a = {}^N C_b \cdot {}^C C^{-1} \cdot G_b \right\} \quad (4C)$$

$G_i \in \mathbb{R}_{4 \times n}$ *Camera space positions of chess grid corners for frame i in homogeneous coordinates.*

${}^N C_i \in \mathbb{R}_{4 \times 4}$ *Camera [marker](#) in navigation space for frame i .*

${}^C C \in \mathbb{R}_{4 \times 4}$ *The unknown camera-marker-in-camera-space matrix.*

With a set of three linearly independent frames, it is possible to find an exact solution for ${}^C C$. However, this solution is susceptible to noise and variance in the camera-to-grid extrinsics, leading to slight differences depending on which three frames are selected.

To improve calibration consistency, we investigate solving an optimisation problem in an over-determined system that includes significantly more than three frames and a greater variance in angles. This approach aims to reduce the effects of noise, inaccuracies in extrinsics, and potential view-dependent effects. We use the open-source convex optimisation solver CVXPY [43] [44], with the goal of minimising the [Frobenius norm](#) of the difference between each pair. Instead of adding

optimisation criteria for the full [Cartesian product](#) of pairs, we examine the results for combinations, permutations, and consecutive pairs.

4.2.3 Camera-Navigation Time Synchronisation

A final aspect of transforming the camera’s view into navigation space is making sure that the two systems’ data streams are synchronised. This involves ensuring that recorded frames from both systems are paired up properly, so the transformation uses frame pairs taken at the same time. Both the navigation system and the depth camera have internal clocks and add timestamps to their respective frames. By determining the difference between these clocks, we can match the corresponding camera and navigation system frames by caching the most recent frames. This clock difference is referred to as Δ_{sync} .

Calibrating the clock discrepancy is achieved by observing a shared event, which in our case is moving the camera. Detecting this event in the navigation system is straightforward; we simply examine the length of the translation vector of the tracked depth camera between frames. For the depth camera, event detection is accomplished by comparing consecutive frames. This comparison involves applying a [Sobel filter](#) and then evaluating the average absolute difference between the frames. The [Sobel filter](#) is used to amplify small differences, as demonstrated in [Figure 4.2.1](#).

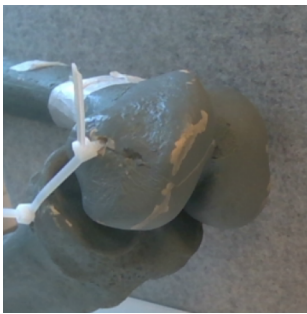


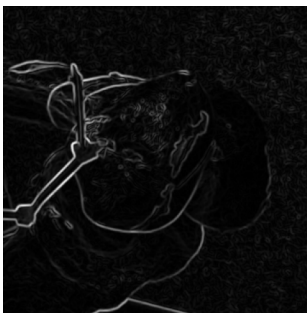
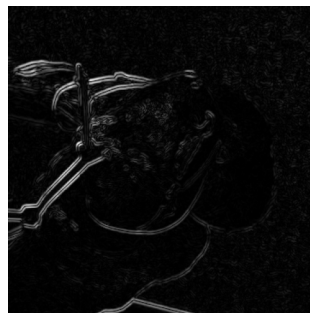
Image 1



Absolute difference of images



Image 2

Image 1 with [Sobel filter](#)

Absolute difference of Sobel-filtered images

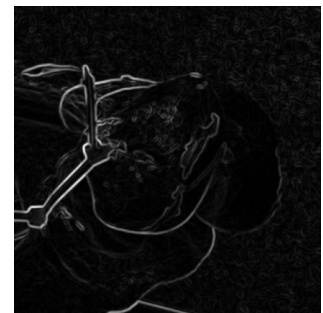
Image 1 with [Sobel filter](#)

Figure 4.2.1: Two slightly different images of a synthetic knee model to showcase [Sobel filtering](#) and how it amplifies differences.

4.3 CT Alignment

The overarching goal of our project is to find an accurate transformation matrix that maps the preoperative [CT](#) data onto the patient during surgery, specifically the $\mathbf{T}_{CT \rightarrow N}$ matrix. The process to achieve this involves the following steps:

1: Filtering

Filter the **CT** scan to primarily only include areas that should be visible and accurate in the camera view.

2: Frame Averaging

Combine multiple images to produce a more consistent and reliable representation than an individual frame.

3: Registration

Utilise **ICP** for precise point-to-point alignment and registration.

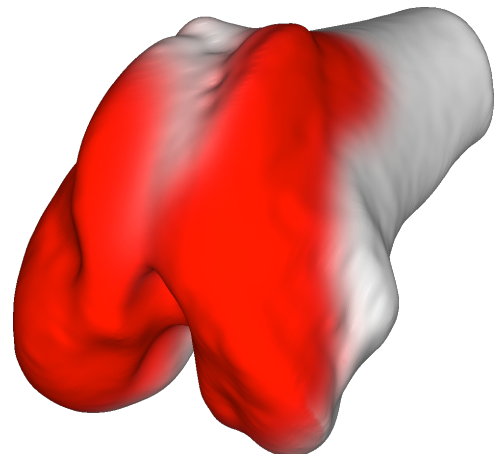
The next few sections expand on these steps.

4.3.1 Filtering

To enhance alignment efficiency and reduce false solutions, it is necessary to exclude certain regions of the **CT** scan that are not visible from the camera’s perspective during **ICP** computation. Furthermore, even visible regions may still be different in the **CT** model due to the aforementioned differences in the 3D reconstruction method. Figure 4.3.1 provides a preview of the regions considered important for inclusion in the alignment procedure. Notably, the upper centre of the **patellar surface** of the femur is excluded due to its thicker cartilage layer.



(a) Picture of an in-progress intraoperative 3D capture, taken during the pig animal lab at Sahlgrenska Universitetssjukhuset in Mölndal.



(b) Corresponding **CT** model with important regions marked in red.

Figure 4.3.1: Comparison of the intraoperative view during a 3D capture and its **CT** model counterpart, where only the red regions are used in the alignment procedure.

4.3.2 Frame Averaging

To address noise and view-dependent effects, we combine multiple frames of the navigation space point clouds of the bone. However, simply merging these point clouds results in poor **ICP** performance because the algorithm is sensitive to noisy data [45]. Misaligned aggregated point clouds lead to numerous incorrect point-to-point correspondences, making it difficult for the **ICP** algorithm to converge to an accurate solution.

Instead, we use a density-based averaging method, which conceptually identifies a curve through the point cloud where the point density is the highest. This approach, however, tends to smooth out

sharp edges and fine features because the density is typically higher inside corners than outside, as illustrated in Figure 4.3.2.

We chose Poisson surface reconstruction for its robustness against noisy data [46]. This technique allows us to create a smooth, continuous surface from the noisy point cloud data. After reconstruction, the Poisson surface is uniformly down-sampled to match the point cloud density of the CT scan. With the frame averaging complete, the Poisson mesh is now ready for CT alignment.

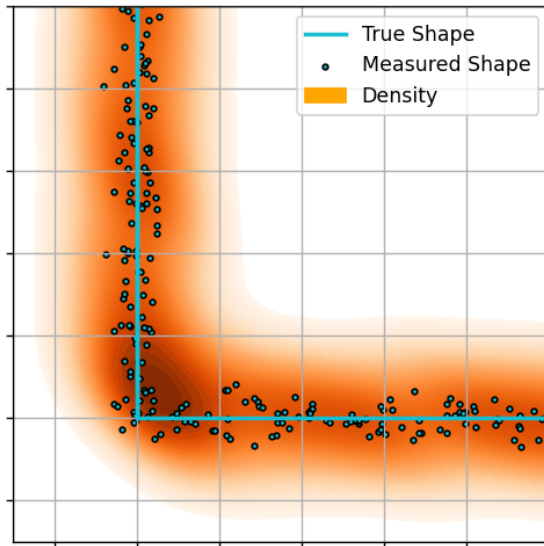


Figure 4.3.2: Depiction of how density estimation of a sharp corner does not match the real shape, the inside of the corner ends up with a higher density than the outside. Resulting in a smoothing of the corner.

4.3.3 ICP Registration

The point-cloud-to-point-cloud registration between the CT model and Poisson surface is performed in two stages using ICP. The first stage involves a global coarse alignment to ensure the CT scan is approximately positioned and oriented relative to the bone, achieving accuracy within the order of a centimetre. The second stage utilises traditional ICP for optimal fit refinement. Both global and local ICP algorithms are constrained to prevent scaling, as the camera and navigation system scales are known to be identical. Any ICP solution introducing scaling in $\mathbf{T}_{CT \rightarrow N}$ would thus be invalid.

The ICP registration is implemented using Open3D, beginning with global registration as described in Section 2.7.2.2, followed by local registration as outlined in Section 2.7.2.1. For global registration, we do not use the full 33-dimensional vectors; instead, we exclude attributes related to point orientation, as this information is not inherently available from the depth images.

4.4 Hyperparameters

Our approach depends on various hyperparameters. To achieve optimal results, we analyse the impact of these hyperparameters and select the most effective ones. These include both algorithmic parameters and those that vary based on the recording. Below is a comprehensive list of all relevant hyperparameters, organised into categories.

Global Registration

Feature Vector Voxel Size

To ensure that the feature vectors capture large-scale geometric shapes, we apply substantial, uniform voxel down-sampling before calculating them. Without this step, the feature vectors would fail to capture the macro-scale changes we are interested in. This parameter sets the millimetre spacing between each voxel.

Feature Vector Feature Radius

Feature vectors are computed based on local geometric characteristics and the feature radius determines the extent of what is "local".

RANSAC Confidence

The confidence threshold of RANSAC in the feature vector ICP, represents the desired probability that the inlier subset is representative. Higher values increase computation time, reduce the inlier-subset's size, and increase accuracy.

RANSAC Max Iteration

This parameter sets the maximum number of iterations allowed for the ICP algorithm before it stops, even if the desired RANSAC confidence level has not been reached.

Local Registration

Distance Threshold

The maximum distance allowed between a source and target pair of points for them to be considered as corresponding in the ICP algorithm.

Other Computation

Aggregation Random Down-Sampling Percent

To avoid creating excessively large point clouds that slow down computation, we apply random down-sampling to each frame before running Poisson averaging.

Poisson Depth

The depth parameter in the Poisson surface reconstruction algorithm controls the level of detail captured. Lower depth settings result in greater smoothing, which can eliminate fine details and sharp corners. Higher depth settings, on the other hand, can misinterpret noise as detail while also significantly increasing computation time.

Recording Attributes

Number of frames

The number of frames included in Poisson averaging, uniformly distributed over a recording.

Polarisation

Indicates whether polarisation filters are used. Details on the methodology for applying and testing polarisation can be found in Section 4.5.

4.5 Polarisation

We also explore the effects of polarising both the light in the scene and the light received by the depth camera. This investigation is inspired by an Intel white paper by Sweetser and Grunnet-Jepsen [47], which suggests that polarising filters on their RealSense[®] devices can enhance depth estimation accuracy, especially for reflective surfaces. Given the reflective nature of organic tissue and bone, we consider it valuable to examine the potential benefits of adding polarisation filters.

To utilise polarisation on the camera, we designed a jig that attaches a rotating filter to the camera lenses. This setup, illustrated in Figures 4.5.1a and 4.5.1b, allows the filter to be adjusted on the fly, effectively reducing reflective highlights in the scene.

Applying polarisation to the scene lighting is relatively straightforward in a surgical environment, as they are typically illuminated by a primary light source on a movable arm. By polarising this main light source, the majority of the light on the bone will be polarised. Figures 4.5.1c and 4.5.1d depict our crude polarisation filter fitted onto a surgical light.

For completeness, we also investigate the effect of polarising only the light source, without attaching a filter to the camera, to identify any effects introduced by the decreased light intensity caused by using linear polarisation filters.



(a) Custom 3D printed polarisation filter, attachable to the depth camera. (b) The Intel® RealSense™ D405 depth camera fitted with a custom 3D printed mount along with a navigation system marker. (c) A box with a hole fitted with polarisation filters, used to filter the surgical spotlight. (d) Light source polarisation filter mounted to the surgical spotlight.

Figure 4.5.1: Equipment and experiment setup during polarisation light experiment.

4.6 Data Evaluation

Our main goal is to achieve accurate intraoperative pose alignment and compare this with the traditional method detailed in Section 2.1. The following subsection will describe the metrics used for evaluation and their significance.

4.6.1 Baseline Evaluation

Baseline metrics, while not directly used in alignment, provide valuable insights into the reliability of our technology and calibration methods, demonstrating the consistency and accuracy of our equipment and procedures.

Temporal Noise

Evaluates the consistency of the Intel® RealSense™ D405 camera’s depth measurements over time in an intraoperative scenario. It identifies regions with varying levels of uncertainty and assesses the impact of polarisation.

Sync Calibration

Measures the variability of synchronisation calibration between the camera and the navigation system.

Camera Marker Calibration

Assesses the variability of ${}^C C$ and evaluates the time required for different calibration methods to execute and compute.

Traditional Method Tip Accuracy

Examines the variability of sampled points using our equipment to ensure the reliability of our implementation of the traditional method.

4.6.2 Alignment Evaluation

A robust metric for evaluating alignment is essential. Alignment measures how accurately our solution calculates the transformation of a **CT** scan into navigation space, specifically evaluating the correctness of the $T_{CT \rightarrow N}$ matrix in transforming the **CT** scan.

To compare transformations, we use a scalar metric that calculates the average point-to-point distances between pairs of feature-filtered **CT** point clouds transformed into navigation space. This metric, referred to as TF_{score} , is formally defined in Equation (4D). It specifies the average distance that each point in a point cloud ends up transformed in comparison to another transformation.

$$TF_{score}({}^C P, \mathbf{T}_{CT \rightarrow N}, \mathbf{U}_{CT \rightarrow N}) = \|\mathbf{T}_{CT \rightarrow N} \cdot {}^C P - \mathbf{U}_{CT \rightarrow N} \cdot {}^C P\|_{Frobenius} \quad (4D)$$

${}^C P \in \mathbb{R}_{4 \times m}$ *The feature filtered point cloud as homogeneous coordinates in **CT** Space where m is the number of points in the cloud.*

$\mathbf{T}_{CT \rightarrow N} \in \mathbb{R}_{4 \times 4}$ *First **CT** to navigation space transformation.*

$\mathbf{U}_{CT \rightarrow N} \in \mathbb{R}_{4 \times 4}$ *Second **CT** to navigation space transformation.*

Instead of only comparing pairs of CT-Navigation transforms, TF_{score} is also used to evaluate the consistency of an alignment. Ideally, the same alignment should be achieved given the same recording, even if the recording is cropped or down-sampled. In practice, however, ICP-based registration is significantly affected by randomness.

The consistency of alignment is evaluated by comparing the TF_{score} of each transformation with the transformation that produces the median point cloud for all configurations and recordings. This approach helps us understand the distribution and variability of the alignment results.

With this alignment metric established, we address the following questions:

How do hyperparameters effect alignment

Algorithm parameters will be evaluated through individual parameter sweeps. The ranges were chosen as either linear, geometric, or hard-coded, based on initial small-scale tests. Polarisation and the number of frames will be given special handling and tested against all parameter sweeps instead of on their own.

How does calibration effect alignment

Without external measurements to verify the accuracy of Δ_{sync} and ${}^C C$ calibration, we investigate the impact of intentionally incorrect calibration. If a small additional calibration error worsens alignment, we could conclude that our calibration method is insufficient.

How does polarisation effect alignment

As described in Section 4.5, we will study how alignment is influenced by the polarisation of the camera and the light source.

5

Results

The following chapter presents the findings of the study, organised in a series of subsections starting with the evaluation of the [RGBD camera's](#) depth estimation capabilities, followed by an investigation of the sensitiveness of the camera calibration procedure, the results of hyperparameter tuning, and finally concludes with the presentation of achieved alignment quality between the CT model and the intraoperative 3D model captured using the Intel[®] RealSense[™] D405 camera.

5.1 Depth Estimation

The estimated depth data computed by the Intel[®] RealSense[™] D405 camera is prone to a considerable amount of [temporal noise](#) and coverage gaps in difficult-to-capture regions. The results of the experiments conducted are presented in Figure [5.1.1](#). The results show tangible improvements in [fill rate](#) when properly using polarised filters as well as higher levels of [temporal noise](#) in regions with reflective surfaces and homogeneous texture, especially apparent on the [patellar surface](#) of the femur.

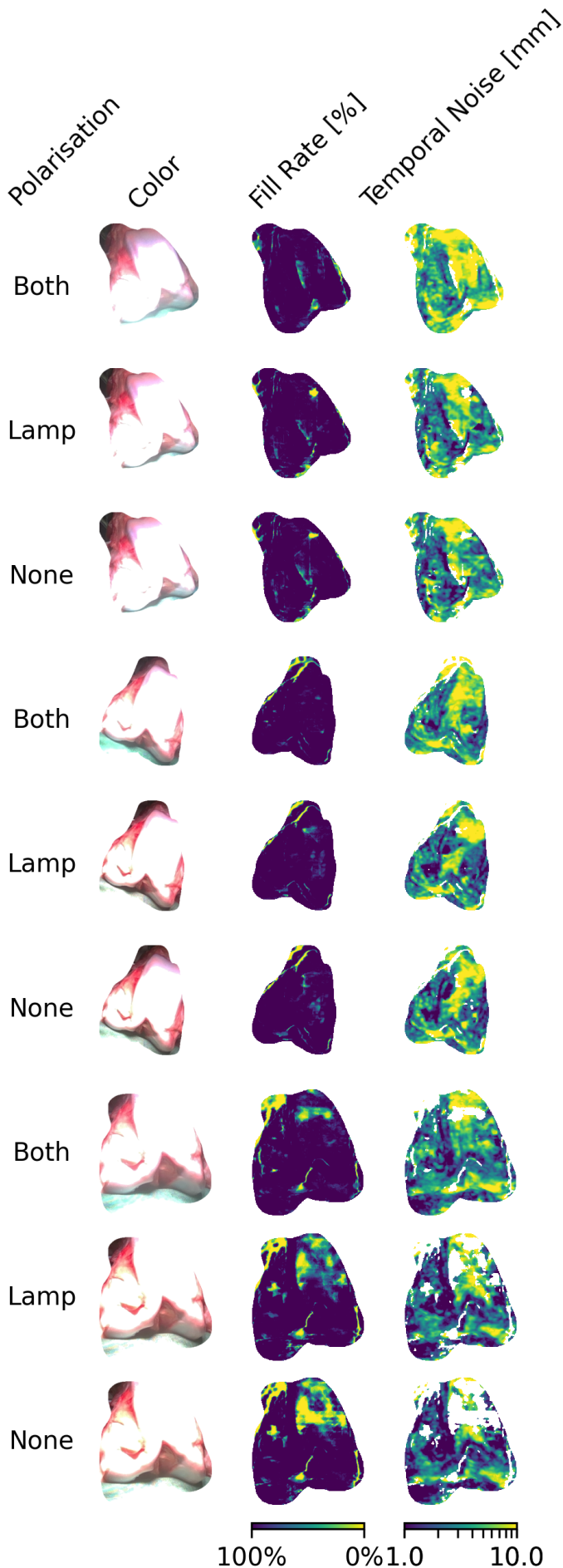


Figure 5.1.1: Analysis of the depth uncertainty of the knee region, based on three distinct angles.

Polarisation

The type of polarisation setup that was used. 'Both' indicates that polarisation filters were applied to both the surgical light and the camera; 'Lamp' signifies that only the surgical light was equipped with a polarisation filter; 'None' denotes that no polarisation filters were used.

Depictions of the equipment used to test polarisation can be found in Figure 4.5.1.

Fill rate

Illustrates how often the camera manages to capture depth information for each pixel. Higher is better. Yellow is bad.

Temporal Noise

Illustrates the per-pixel camera depth confidence. For any given pixel it is defined as the standard deviation of that position's depth value. Pixels with less than 50% fill rate are excluded. Lower is better. Yellow is bad.

Summary

Using polarisation filters on both the camera and the light source reduces data loss but increases **temporal noise**. This conclusion is supported by the observed decrease of yellow regions in the **fill rate** column and a corresponding increase of yellow regions in the **temporal noise** column when polarised filters are used. In contrast, polarising only the light source has a negligible effect.

5.2 Camera Marker Calibration

This section presents our calibration results from determining the transformation for the navigation space marker attached to the camera in camera space, referred to as ${}^C C$.

Metric	Calibration Method				
	Shared Geometry	Multi Frame			
		Set of 3	Optimisation Solver		
			Combinations	Permutations	Pairs
Standard Deviation of Translation [mm]	1.8	4.3	1.7	1.7	2.3
Time to Compute [s]	~10	0	3.2	8.7	0.18
Time to Perform [s]	~100	~10	~10	~10	~10

Table 5.2.1: How different camera-marker-in-camera-space (${}^C C$) calibration methods affect the consistency of the transformation matrix and how long they take to compute and perform.

Shared Geometry

Implementation of the method described in Section 4.2.2.1. The results were derived from three repetitions of the shared geometry calibration using five different angles performed during the animal test at Sahlgrenska.

Multiframe

Implementation of the method outlined in section 4.2.2.2. Results are compiled from 20 repetitions of calibrations performed in the office setting. This calibration was not used for any results other than this table.

Standard Deviation of Translation

The standard deviation of the length of the translation vector making up ${}^C C$

Time to Compute

The time spent by our mid-powered laptop to solve the method

Time to Perform

The time a person spent performing the calibration steps, excluding the computation time.

5.3 Frame Synchronisation

This section presents our calibration results from determining the time difference between the system clocks of the navigation system and the Intel® RealSense™ D405, referred to as Δ_{sync} .

	Standard Deviation of Δ_{sync} [s]
Repetition 1	0.010
Repetition 2	0.011
Average	0.011

Table 5.3.1: Showcases how consistent our sync calibration is. Δ_{sync} is the difference between the internal clocks of the navigation system and the depth camera. Results were derived from two repetitions of sync calibration performed during the animal test at Sahlgrenska. Each repetition consisted of 10 instances of Δ_{sync} calibration. Each repetition took ~5 seconds.

Standard Deviation of Δ_{sync}

The standard deviation of the measured Δ_{sync} value.

5.4 Manual Pointer Tip Positions

This section presents the consistency achieved during the point sampling using the traditional method. Points were collected by placing the tip of a pointer on bone. The handheld pointer was kept stationary for around three seconds for each of the ~30 sampled points.

	Standard Deviation of Translation [mm]	Time to perform alignment
Repetition 1	0.049	~5 minutes
Repetition 2	0.038	~5 minutes
Repetition 3	0.057	~5 minutes
Average	0.048	~5 minutes

Table 5.4.1: Showcases how consistent our point sampling was for our equipment using the traditional method as described in 2.1. Data is collected from three repetitions of alignment using the traditional method during the animal test at Sahlgrenska. The tip position for each sample was recorded over ~5 seconds.

Standard Deviation of Translation

The average standard deviation of the length of the tip-position-vectors.

5.5 Alignment

This is the most important result section, as it describes how well our solution calculated the transformation of a CT scan into navigation space. More specifically we evaluate how correct the $\mathbf{T}_{CT \rightarrow N}$ matrix is. We evaluate this by comparing the consistency of our TF_{score} to the median transform of all alignments achieved by default hyperparameters.

5.5.1 Hyperparameters

This section presents the impact of hyperparameters on alignment and computation time. The results highlight the alignment distribution achieved with various hyperparameter values.

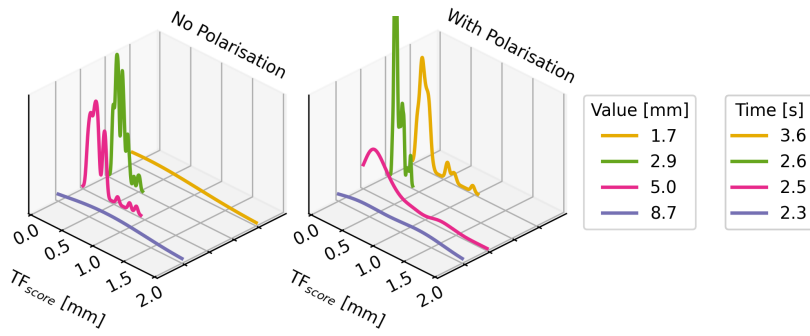


Figure 5.5.1: Parameter sweep of **Feature Vector Voxel Size**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.

TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

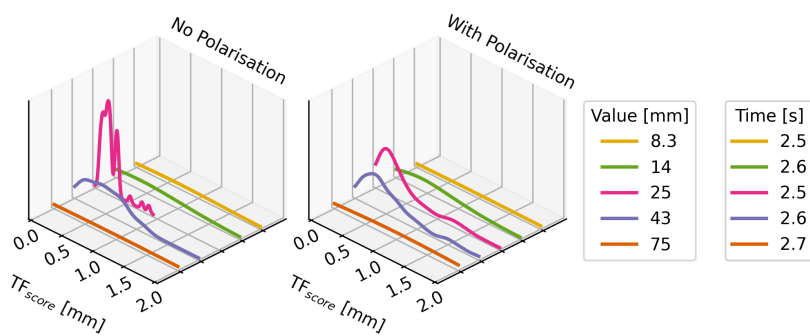


Figure 5.5.2: Parameter sweep of **Feature Vector Feature Radius**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.

TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

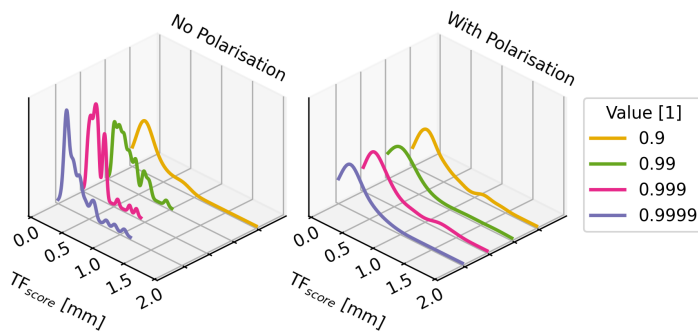


Figure 5.5.3: Parameter sweep of **RANSAC Confidence**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.

TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

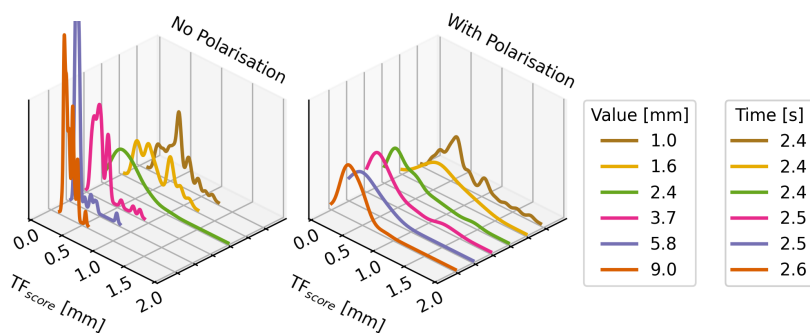


Figure 5.5.4: Parameter sweep of **Distance Threshold**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.

TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

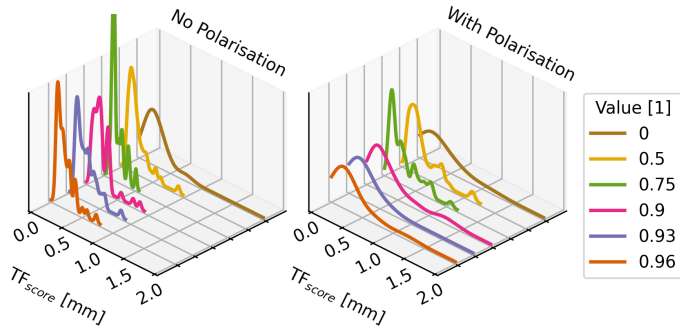


Figure 5.5.5: Parameter sweep of **Aggregation Random Down-Sampling Percent**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.
 TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

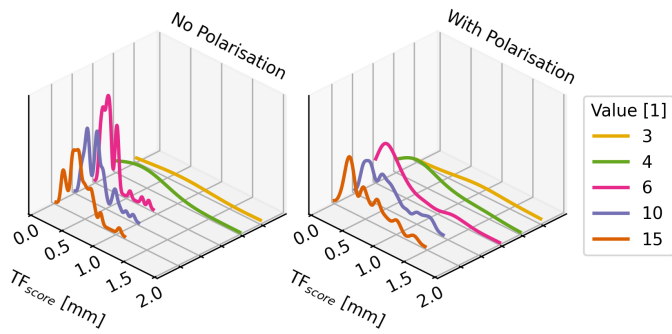


Figure 5.5.6: Parameter sweep of **Poisson Depth**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.
 TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

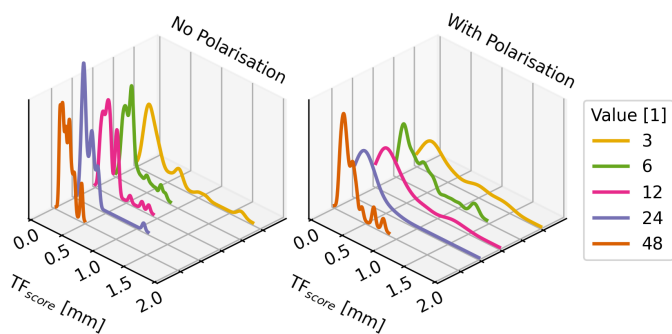


Figure 5.5.7: Parameter sweep of **Number of Frames**

Each curve consists of 54 repetitions.
Value The value of the parameter being swept.
Time The total execution time to finish the alignment.
 TF_{score} The average point-to-point distance to the median alignment. Lower values & lower variance is better.

5.5.2 Calibration Sensitivity

Figures 5.5.8 and 5.5.9 illustrate the variations in alignment quality resulting from the introduction of artificial errors during the camera calibration steps. As expected, larger artificial errors lead to a decrease in alignment quality. The sensitivity of alignment accuracy to these introduced errors highlights how crucial accurate camera calibration is for achieving good alignments using our suggested approach.

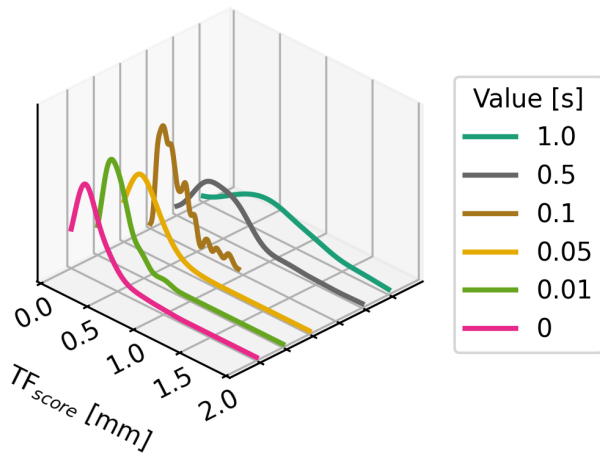


Figure 5.5.8: Effects of adding artificial error to Δ_{sync}

Each curve represents the distribution of 120 samples. *Value* The size of the added error introduced to the clock synchronisation, in seconds.

TF_{score} The average point-to-point distance to the median alignment, as specified in Section 5.5. A lower value and variance is better.

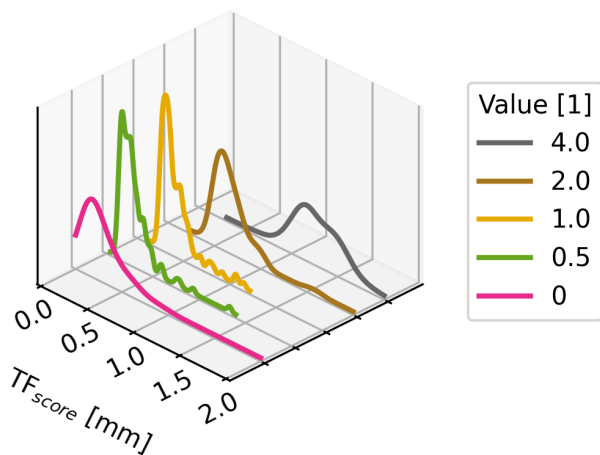


Figure 5.5.9: Effects of adding artificial error to the ${}^C C$ transformation matrix

Each curve represents the distribution of 120 samples. *Value* The size of the error introduced to the camera calibration, in standard deviations from the median calibration repetition. One standard deviation is 1.8 mm.

TF_{score} The average point-to-point distance to the median alignment, as specified in Section 5.5. A lower value and variance is better.

5.5.3 Comparison to Traditional Method

This section compares the alignment achieved by using our method compared to the traditional method, outlined in Section 2.1.

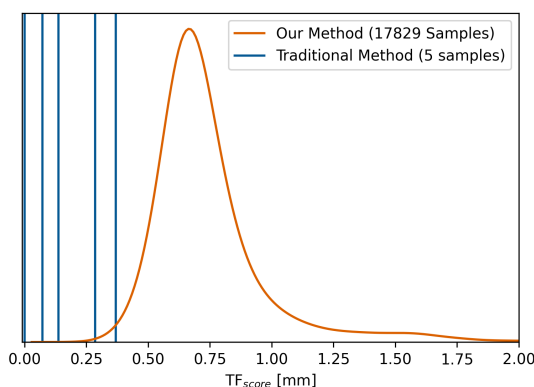


Figure 5.5.10: TF_{score} comparison between the alignment achieved using the traditional versus our method.

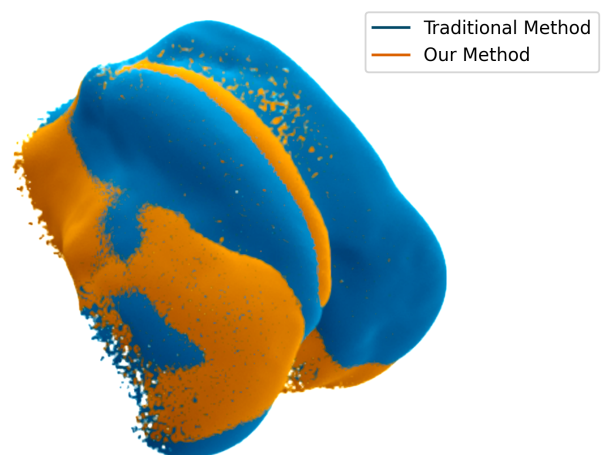


Figure 5.5.11: Visual comparison between the median alignment using the traditional method versus using our median solution. Median is defined in terms of TF_{score} .

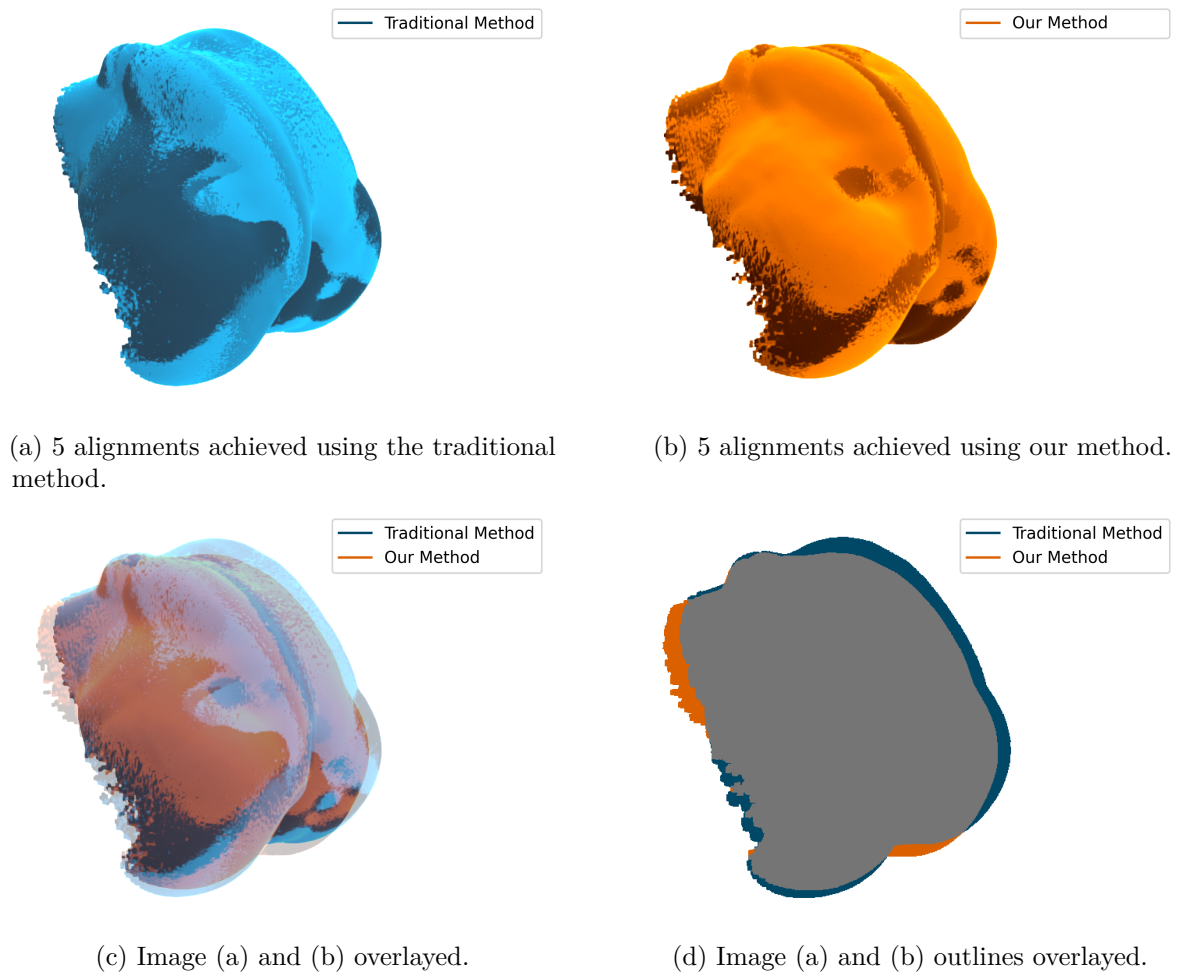


Figure 5.5.12: This visual comparison presents our aligned CT scan alongside those obtained using the traditional method. The display includes all five traditional alignments and five randomly selected alignments from our method. Each set of five is shown simultaneously, with variations in brightness for clarity.

5.5.4 Depth Accuracy

This section highlights the differences between the camera's view and the aligned CT model. The figures illustrate the same data but through various types of visualisation.

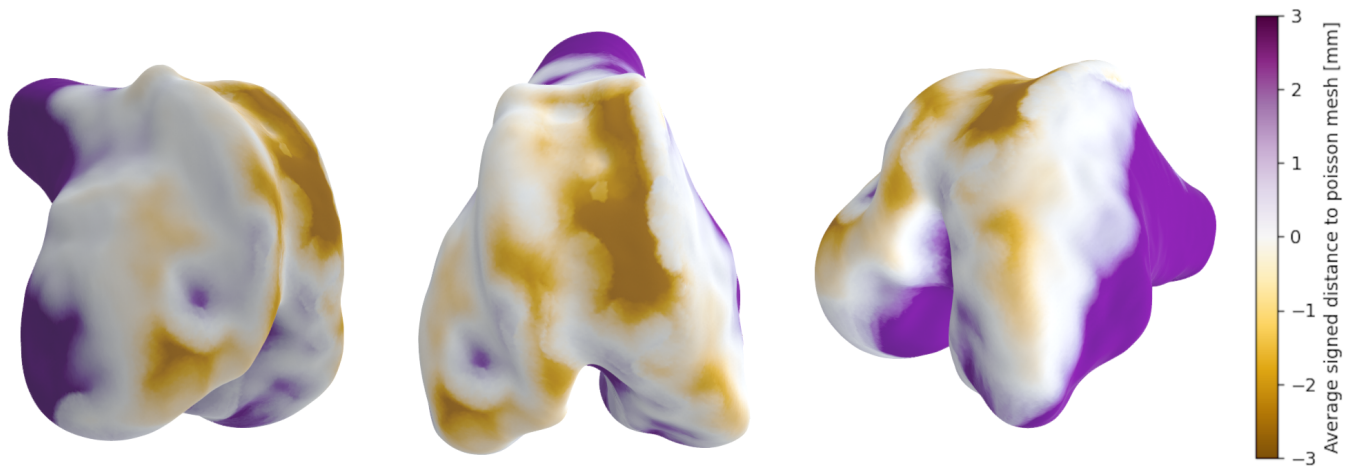


Figure 5.5.13: Illustration of the distance between our median aligned CT model and its corresponding Poisson mesh, as seen by the camera. Negative values indicate that the Poisson mesh is below the surface of the aligned CT model, while positive values indicate it is above.

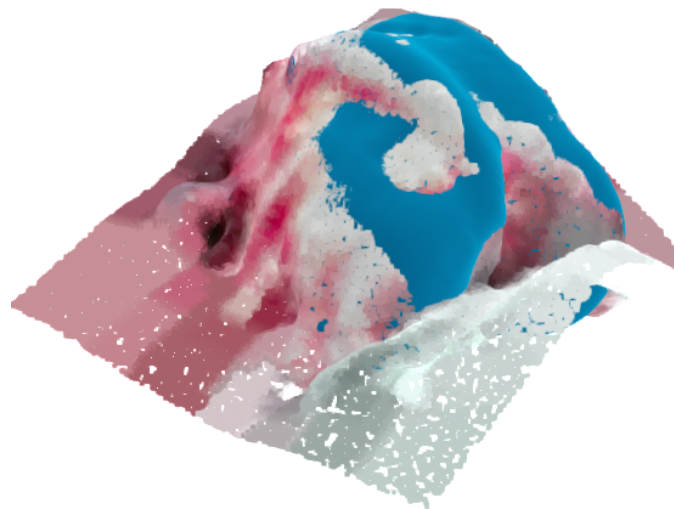


Figure 5.5.14: Illustration of how our median aligned CT model looks like together with its corresponding Poisson cloud. Blue is the CT model with alignment applied. Pink is the Poisson averaged camera view.

6

Analysis & Discussion

This chapter presents a thorough discussion of the results gathered in the study. By examining both the elements that were found to enhance alignment and those that impeded it, we aim to offer insights into the applicability of leveraging stereoscopic depth cameras in total knee arthroplasty (TKA) surgery.

6.1 Calibration

Our proposed approach involves two calibratable values. The first is ${}^C C$, the transformation matrix of the [marker](#) attached to the camera in camera space. The second parameter, Δ_{sync} , is the time difference between the internal clocks of the camera and the navigation system, ensuring synchronised frames from both devices.

Repeated calibrations of ${}^C C$ show a standard deviation of approximately 2 mm for the translation, which is less consistent than desired. Ideally, this value should be as low as possible, as a 1 mm error in ${}^C C$ would result in a similar magnitude error in the camera point cloud’s position in navigation space. Despite this, we found that introducing an incorrect ${}^C C$ did not significantly affect the results unless the error exceeded two standard deviations.

On the other hand, repeated calibrations of Δ_{sync} are very consistent, with a standard deviation of 0.01 seconds, which is less than one frame at the 60 frames per second rate of our navigation system. We also tested the impact of deliberately adding errors to Δ_{sync} . Our method proved robust, with noticeable issues only arising after a substantial 0.5-second offset, equivalent to a 30-frame deviation from the properly calibrated value.

These tests indicate that our method is robust against the demonstrated variability in the calibration procedure, requiring errors of several standard deviations to noticeably impact alignment quality. Given that purposeful miscalibration does not have a visible effect unless it is substantial, we conclude that poor calibration is unlikely to be the cause of any alignment issues experienced.

6.2 Depth Estimation

The purpose of evaluating depth estimation uncertainty was to determine whether the depth measurements from the Intel® RealSense™ D405 camera were reliable, considering both depth noise and the use of light polarisation filters. Two metrics were used: [temporal noise](#), which is the standard deviation of the measured depth for each point, and [fill rate](#), which is the ratio at which the camera could compute depth information at any given location.

Regions with low [fill rates](#) exhibited relatively higher [temporal noise](#), with values exceeding 10 mm being common, particularly in reflective areas. In contrast, non-reflective regions generally performed better, maintaining consistent [fill rates](#) and often demonstrating [temporal noise](#) of less than 1 mm.

However, this consistency was not universal; some non-reflective areas displayed noise levels ranging from 1 mm to 5 mm in small patches. As expected, textured regions consistently showed lower [temporal noise](#).

Using polarisation filters increased the overall [fill rate](#) but also worsened [temporal noise](#). The increase in [fill rate](#) was anticipated since polarisation removes specular highlights, which the [RGBD camera](#) perceives as saturated white values, making subtle parallax comparisons for estimating depth impossible in those regions. However, the increase in [temporal noise](#) contradicted previous findings, especially since regions without significant reflections also showed increased noise. Polarising only the surgical light, and not the camera, did not exhibit this pattern. Post-experiment inspection revealed that the polarisation filter attached to the camera had sustained scratches during experimentation, likely contributing to the observed increase in [temporal noise](#). For more accurate testing, it is important to prevent damage to the polarisation filters and consider using more durable filters if possible.

6.3 Hyperparameters

This section analyses the impact of hyperparameters on alignment and computation time, with the discussion organised by parameter.

Feature Vector Voxel Size

The results show that only a narrow range of [voxels](#) size values is effective for alignment, with values outside this range causing alignment to fail. This parameter is crucial for global registration; if global registration fails, fine alignment will almost certainly fail as well. Very small [voxels](#) slowed down computation time.

Feature Vector Feature Radius

Similar to the [voxel](#) size, the feature radius parameter affects global registration and has a narrow band of viable values. Outside this range, alignment fails. This parameter did not significantly impact execution time.

RANSAC Confidence

This parameter relates to global [ICP](#). Increasing the [RANSAC](#) confidence above 0.999 did not affect alignment but did increase computation time.

RANSAC Max Iteration

This parameter, also related to global [ICP](#), did not affect alignment or computation time, as the alignment iterations typically complete before reaching the maximum iteration limit.

Distance Threshold

Increasing the local [ICP](#) distance threshold improved alignment by making the [ICP](#) correspondence stricter. However, this effect was less noticeable in polarised data due to its higher noise levels. Strict [ICP](#) slightly increased computation time, as expected.

Aggregation Random Down-Sampling Percent

Surprisingly, low or no random down-sampling worsened alignment, suggesting quirks in the frame-averaging implementation when handling large amounts of data. Minimal or no down-sampling increased computation time due to the higher memory usage and the need for more averaging points, though the effect on [ICP](#) was minimal.

Poisson Depth

The Poisson depth parameter significantly affects alignment. Too low a Poisson depth results in an overly smooth averaged point cloud, worsening alignment, while too high a Poisson depth

interprets noise as fine detail, also worsening alignment. Higher Poisson depth significantly increased computation time.

Number of frames

A small number of frames results in poor alignment due to the camera's noise, making it hard to distinguish noise from high-frequency features. More frames improved alignment but with diminishing returns. Increasing the number of frames increased computation time almost linearly.

Polarisation

Contrary to expectations, polarisation filters worsened depth accuracy and alignment, except for [fill rate](#). Scratches on the polarisation filter, discovered during post-experiment inspection, likely contributed to the poor results. As a result, findings regarding the use of polarisation filters on the Intel® RealSense™ D405 should be considered unreliable unless further evidence is provided.

Despite these challenges, valuable insights were gained: high depth uncertainty worsens alignment. The [RGBD camera's](#) noise and depth inaccuracies significantly impact alignment quality. Using a more accurate depth camera would likely yield better and more consistent results.

6.4 Depth Accuracy

As shown in Figures [5.5.13](#) and [5.5.14](#), the Poisson-averaged cloud and the aligned CT model do not overlap perfectly. Approximately half of the areas used for alignment are within 1 mm of accuracy, while other regions, including those not visible to the camera, deviate significantly more. However, even in the areas with the greatest discrepancies, as long as they are visible to the camera, the error remains below 5 mm. Regions not visible to the camera, such as those underneath, show significantly worse alignment, but this is not so important as we filter out most of these regions.

There are several reasons why the Poisson mesh and the aligned CT model may differ in shape, and we will discuss five key factors below.

CT scan does not match the intraoperative bone

An inherent issue is that the CT scan and the intraoperative bone are not identical. CT imaging fundamentally differs in how it determines surfaces and what qualifies as bone. Unlike a CT scan, the [RGBD camera](#) cannot distinguish bone from cartilage.

The averaged meshes are not correctly overlapped

For effective averaging, all camera space clouds must be converted to a shared frame of reference. Misalignment leads to the removal of fine details. Despite calibration, some misalignment is expected because the camera and navigation system captures are not simultaneous. This effect is noticeable when the camera is moved quickly.

Hypothetically, moving the camera at 1 meter per second with a navigation system operating at 60 frames per second, even with perfect sync calibration, would still result in a discrepancy of up to half a frame between the camera capture and the navigation system recording, leading to an 8 mm displacement. However, our experiments did not involve such high-speed movements.

Poisson averaging is a bad approach to use as an average

As with any density-based method, Poisson averaging tends to smooth out sharp features, causing inaccuracies in areas like the edges of the knee. However, we do not consider this a major issue, as the expected characteristic effects are not present. Excessive edge smoothing would typically cause outer edges of the CT to appear more underneath the Poisson mesh

and inner edges to appear more on top, which is only observed at the reflective centre of the bone—an area deemed untrustworthy through [temporal noise](#) analysis.

ICP finds the incorrect solution given the Poisson cloud

Figures [5.5.13](#) and [5.5.14](#) show that a perfect [CT](#) alignment on the Poisson cloud does not exist, with large regions not corresponding. Because of this, even if there were a theoretically better fit given the [ICP](#) constraints, it would not guarantee that the alignment is more correct.

Camera depth information is not correct

A significant factor is the inaccuracy of the depth camera information. Comparing the depth accuracy plot in Figure [5.5.13](#) with the [temporal noise](#) in Figure [5.1.1](#), it is evident that areas with greater distance from the aligned [CT](#) generally correspond to areas with high [temporal noise](#).

6.5 Comparison to Traditional Alignment

The alignment achieved by our method is approximately normally distributed around a transformation that is about 0.5 mm different from the manual alignments, as shown in Figures [5.5.10](#) and [5.5.11](#). Although the manual alignment intersects with our distribution, our average alignment converges to a different point compared to the traditional method.

One possible explanation is that our implementation of the traditional method may be inaccurate. While a professional assisted with the final alignments, the rest of the process was not externally validated. We are confident in the consistency of the manually taken points, as demonstrated in Table [5.4.1](#). However, errors could have been introduced during the choice of points, world transformations, and tool calibration, such as discrepancies between the calibrated tip position and the actual physical position.

6.6 Time

Achieving the desired algorithm speed was not an issue, with most configurations taking approximately 2.5 seconds. Increasing the strictness of parts of the algorithm and allowing more iterations did not improve alignment, except for the ICP Distance Threshold, which had a minimal impact on computation time.

To get a truthful understanding of the overall saved time, it is necessary to account for the time required for calibration as well. Synchronisation calibration must be redone if either the camera or the navigation system is power cycled, due to the reset of internal clocks and potential drift between the systems. Fortunately, one iteration of sync calibration takes only about 5 seconds, and we believe a more efficient approach could reduce this to less than a second. On the other hand, $^C C$ calibration does not need to be repeated before each surgery, if the [marker](#) attachment is robustly constructed. However, multi-frame calibration could be performed without making the timeline unreasonable.

6.7 Future work

Building on the findings of this study, some potential improvements and areas for further research have been identified that could enhance the use of stereoscopic depth cameras in intraoperative imaging for [TKA](#) surgeries.

A straightforward improvement in depth estimation quality can be achieved by employing artificial texture projection. This method involves projecting patterns onto the surfaces to be captured by the

RGBD camera, adding texture to otherwise homogeneous areas and thereby enhancing the camera’s ability to capture detailed and accurate 3D reconstructions. It is important to note that the Intel® RealSense™ D405 camera used in this study is equipped with an **IR** cut filter, which naturally needs to be removed if this approach is combined with an **IR** projector. For further details on suitable projectors for this application, the white paper by Grunnet-Jepsen et al. [48] at Intel® may serve as a good starting reference.

Enhancing the frame aggregation process is another key area for improving alignment quality. Currently, the averaging of point clouds from multiple **RGBD** frames taken from various angles results in the loss of high-frequency geometric details, which diminishes finer anatomical features and reduces the overall quality of the 3D reconstruction. This is primarily due to the noisy depth data obtained from the Intel® RealSense™ D405 camera, but could also arise from poor calibration of the ${}^C C$ transformation matrix. Future research could address these problems by utilising higher quality depth cameras with less noise or by pursuing alternative algorithms that better preserve high-frequency details during frame aggregation.

Another area for future work involves focusing on the fundamental differences between the CT model and the intraoperative 3D model. Although improvements in this area may not be as prominent as those achieved through the employment of artificial texture or the enhancement of frame aggregation, they may still prove significant. For instance, more detailed cartilage modelling could improve the accuracy of alignment by allowing the **ICP** algorithm to prioritise more reliable regions and deprioritise regions likely to differ between the two models.

6.8 Summary

This section summaries our discussion according to our research questions.

- **RQ1:** *Is it viable to replace a few high quality sampled points with a large number of low quality points using the Intel® RealSense™ D405 depth camera ...*
 - **RQ1.1:** *... in terms of point cloud to CT model alignment accuracy?*
Yes, we achieve a discrepancy of approximately 0.5 mm compared to manual alignment. However, we are concerned about discrepancies between the averaged surface and the CT scan. More problematic discrepancies could occur in other environments, potentially affecting alignment quality detrimentally.
 - **RQ1.2:** *... while keeping total compute time from last **RGBD** frame acquisition to completed alignment under 60 seconds?*
Yes, it is feasible. Computation time is approximately 2.5 seconds, allowing sufficient room to include calibration and still remain under the 1-minute limit.
- **RQ2:** *How does the application of polarised light filtering effect the Intel® RealSense™ D405 depth camera’s point cloud acquisition for musculoskeletal surfaces, and in turn impact the point cloud to CT model alignment accuracy?*
We found that using polarisation filters had an overall negative impact. However, we believe these findings are unreliable due to scratches on the polarisation filters.
- **RQ3:** *What is the depth estimation accuracy of the Intel® RealSense™ D405 depth camera for musculoskeletal surfaces?*

While many regions achieve very good accuracy of less than 1 mm, other regions, particularly texture-less ones, exhibit deviations of more than 3 mm. These are not the only problematic areas; some richly textured regions also show poor accuracy, though no consistent pattern explaining this was identified.

7

Conclusion

This thesis introduces a method that combines optical medical navigation systems with an affordable visible light-based depth camera to achieve quicker intraoperative CT-to-patient bone alignment in total knee arthroplasty.

Our approach achieves a good comparative alignment with an average error of around 0.5 mm from the alignment of the traditional method, though it has a larger standard deviation than earlier works. Additionally, the proposed method shows great potential for improving knee registration speed and convenience, with computation times around 2.5 seconds, allowing for calibration within the 1-minute limit.

Experiments using polarisation unexpectedly resulted in worse depth data estimation and alignments. This negative impact is hypothesised to be due to scratches on the polarisation filters incurred during testing.

Significant discrepancies between the CT and the averaged point cloud were observed. Some regions achieved a disparity of less than 1 mm, while others, particularly texture-less areas, showed deviations exceeding 3 mm. Regions with large discrepancies partially correlated with those having high temporal noise and low fill rates. Due to the ambiguous nature of these discrepancies, other environments may lead to less favourable alignments.

Future work should address depth estimation issues and test in more diverse environments. Leveraging other registration algorithms and high-accuracy imaging devices should further enhance the precision and efficiency of intraoperative CT-to-patient bone alignment.

In summary, while there are areas for improvement, the Intel® RealSense™ D405 depth camera presents a promising option for faster and more convenient knee registration.

Bibliography

- [1] U. Mezger, C. Jendrewski, and M. Bartels, “Navigation in surgery,” *Langenbeck’s Archives of Surgery*, vol. 398, no. 4, pp. 501–514, 2013, ISSN: 1435-2443. DOI: [10.1007/s00423-013-1059-4](https://doi.org/10.1007/s00423-013-1059-4). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627858/> (visited on 02/08/2024).
- [2] P. Merloz, J. Tonetti, L. Pittet, *et al.*, “Computer-Assisted Spine Surgery,” en, *Computer Aided Surgery*, vol. 3, no. 6, pp. 297–305, Jan. 1998, ISSN: 1092-9088, 1097-0150. DOI: [10.3109/10929089809148150](https://doi.org/10.3109/10929089809148150). [Online]. Available: <http://www.tandfonline.com/doi/full/10.3109/10929089809148150> (visited on 05/07/2024).
- [3] X. Fan, Q. Zhu, P. Tu, L. Joskowicz, and X. Chen, “A review of advances in image-guided orthopedic surgery,” en, *Physics in Medicine & Biology*, vol. 68, no. 2, 02TR01, Jan. 2023, ISSN: 0031-9155, 1361-6560. DOI: [10.1088/1361-6560/aca9e9](https://doi.org/10.1088/1361-6560/aca9e9). [Online]. Available: <https://iopscience.iop.org/article/10.1088/1361-6560/aca9e9> (visited on 01/23/2024).
- [4] E. Tokgöz, “Surgical Approaches Used for Total Hip Arthroplasty,” en, in *Total Hip Arthroplasty*, Cham: Springer International Publishing, 2023, pp. 1–23, ISBN: 978-3-031-08927-5. DOI: [10.1007/978-3-031-08927-5_1](https://doi.org/10.1007/978-3-031-08927-5_1). [Online]. Available: https://link.springer.com/10.1007/978-3-031-08927-5_1 (visited on 05/15/2024).
- [5] Socialstyrelsen, *Statistikområden, operationer på rörelseapparaten*. [Online]. Available: <https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikdatabasen/> (visited on 05/08/2024).
- [6] O. R. for Nursing (Open RN), *Nursing Skills*, First. Wisconsin: WI Technical Colleges Open Press, 2021, This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>, ISBN: 9781734914122. [Online]. Available: <https://openwa.pressbooks.pub/nursingskills/> (visited on 05/22/2024).
- [7] C. Rivière and P.-A. Vendittoli, Eds., *Personalized Hip and Knee Joint Replacement*, en. Cham: Springer International Publishing, 2020, ISBN: 978-3-030-24243-5. DOI: [10.1007/978-3-030-24243-5](https://doi.org/10.1007/978-3-030-24243-5). [Online]. Available: <https://link.springer.com/10.1007/978-3-030-24243-5> (visited on 01/23/2024).
- [8] E. Tokgöz, “Preexisting Conditions Leading to Total Hip Arthroplasty,” en, in *Total Hip Arthroplasty*, Cham: Springer International Publishing, 2023, pp. 25–43, ISBN: 978-3-031-08927-5. DOI: [10.1007/978-3-031-08927-5_2](https://doi.org/10.1007/978-3-031-08927-5_2). [Online]. Available: https://link.springer.com/10.1007/978-3-031-08927-5_2 (visited on 05/15/2024).
- [9] T. Karasavvidis, C. A. Pagan Moldenhauer, F. S. Haddad, M. T. Hirschmann, M. W. Pagnano, and J. M. Vigdorchik, “Current Concepts in Alignment in Total Knee Arthroplasty,” en, *The Journal of Arthroplasty*, vol. 38, no. 7, S29–S37, Jul. 2023, ISSN: 08835403. DOI: [10.1016/j.arth.2023.01.060](https://doi.org/10.1016/j.arth.2023.01.060). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0883540323000803> (visited on 05/31/2024).
- [10] Y. Yang, L. Jiang, X. Zhou, X. Zhou, H. Chen, and Z. Chen, “Robotic-assisted total knee arthroplasty improves implant position and early functional recovery for the knee with severe varus/valgus deformity,” en, *BMC Musculoskeletal Disorders*, vol. 25, no. 1, p. 92, Jan.

- 2024, ISSN: 1471-2474. DOI: [10.1186/s12891-024-07203-9](https://doi.org/10.1186/s12891-024-07203-9). [Online]. Available: <https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-024-07203-9> (visited on 05/31/2024).
- [11] F. Rodriguez Y Baena, T. Hawke, and M. Jakopec, “A bounded iterative closest point method for minimally invasive registration of the femur,” en, *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 227, no. 10, pp. 1135–1144, Oct. 2013, ISSN: 0954-4119, 2041-3033. DOI: [10.1177/0954411913500948](https://doi.org/10.1177/0954411913500948). [Online]. Available: <http://journals.sagepub.com/doi/10.1177/0954411913500948> (visited on 05/30/2024).
- [12] J. O. Aflatooni, A. E. Wininger, K. J. Park, and S. J. Incavo, “Alignment options and robotics in total knee arthroplasty,” *Frontiers in Surgery*, vol. 10, p. 1106608, Feb. 2023, ISSN: 2296-875X. DOI: [10.3389/fsurg.2023.1106608](https://doi.org/10.3389/fsurg.2023.1106608). [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fsurg.2023.1106608/full> (visited on 05/31/2024).
- [13] L. Burger, L. Sharan, R. Karl, *et al.*, “Comparative evaluation of three commercially available markerless depth sensors for close-range use in surgical simulation,” en, *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 6, pp. 1109–1118, May 2023, ISSN: 1861-6429. DOI: [10.1007/s11548-023-02887-1](https://doi.org/10.1007/s11548-023-02887-1). [Online]. Available: <https://link.springer.com/10.1007/s11548-023-02887-1> (visited on 01/30/2024).
- [14] A. Grunnet-Jepsen, J. N. Sweetser, and J. Woodfill, *Best-Known-Methods for Tuning Intel® RealSense™ Depth Cameras D400 series for Best Performance*, en, Rev 2.0, Feb. 2023. [Online]. Available: <https://dev.intelrealsense.com/docs/tuning-depth-cameras-for-best-performance> (visited on 02/12/2024).
- [15] Q.-Y. Zhou, J. Park, and V. Koltun, “Open3D: A modern library for 3D data processing,” *arXiv:1801.09847*, 2018.
- [16] A. W-Dahl, J. Kärrholm, C. Rogmark, *et al.*, “Årsrapport 2023 Svenska Ledprotesregistret,” en, Svenska ledprotesregistret, Tech. Rep., 2023, Artwork Size: 25557537 byte Medium: application/pdf Version Number: 1.0, 25557537 byte. DOI: [10.18158/R1HHFZIAH](https://doi.org/10.18158/R1HHFZIAH). [Online]. Available: <http://refdocs.registercentrum.se/10.18158/r1HHfziah> (visited on 02/05/2024).
- [17] F. R. Noyes and S. D. Barber-Westin, Eds., *Noyes’ knee disorders: surgery, rehabilitation, clinical outcomes*, eng, Second edition. Philadelphia, PA: Elsevier, 2017, OCLC: 945377883, ISBN: 978-0-323-42855-2.
- [18] P. J. Withers, C. Bouman, S. Carmignato, *et al.*, “X-ray computed tomography,” en, *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 18, Feb. 2021, ISSN: 2662-8449. DOI: [10.1038/s43586-021-00015-4](https://doi.org/10.1038/s43586-021-00015-4). [Online]. Available: <https://www.nature.com/articles/s43586-021-00015-4> (visited on 06/04/2024).
- [19] E. Seeram, *Computed tomography: physical principles, clinical applications, and quality control*, eng, Fourth edition. St. Louis, Missouri: Elsevier, 2016, OCLC: 910279905, ISBN: 978-0-323-31288-2.
- [20] J. L. M. De La Barrera, A. Sarvestini, S. L. G. Kg, and Co, “Method and system for calibrating a surgical tool and adapter thereof,” en, U.S. Patent 7166114B2, Sep. 2002. [Online]. Available: <https://patents.google.com/patent/US7166114B2/en>.
- [21] G. He, V. M. Mustahsan, M. R. Bielski, I. Kao, and F. A. Khan, “Report on a novel bone registration method: A rapid, accurate, and radiation-free technique for computer- and robotic-assisted orthopedic surgeries,” *Journal of Orthopaedics*, vol. 23, pp. 227–232, Jan. 2021, ISSN: 0972-978X. DOI: [10.1016/j.jor.2021.01.010](https://doi.org/10.1016/j.jor.2021.01.010). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0972978X21000106> (visited on 06/05/2024).
- [22] K. Simek, *Dissecting the camera matrix, part 3: The intrinsic matrix*, 2013. [Online]. Available: <https://ksimek.github.io/2013/08/13/intrinsic/>.
- [23] K. Simek, *Dissecting the camera matrix, part 2: The extrinsics matrix*, 2012. [Online]. Available: <https://ksimek.github.io/2012/08/22/extrinsic/>.

- [24] *Pose computation overview*. [Online]. Available: https://docs.opencv.org/4.x/d5/d1f/calib3d_solvePnP.html.
- [25] S. Giancola, R. Sala, and M. Valenti, *A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies* (SpringerBriefs in Computer Science), 1st ed. 2018. Cham: Springer International Publishing : Imprint: Springer, 2018, ISBN: 978-3-319-91761-0. DOI: [10.1007/978-3-319-91761-0](https://doi.org/10.1007/978-3-319-91761-0).
- [26] J. Lambert, *Iterative closest point (icp) algorithm*, Accessed: 2024-06-20, Nov. 2018. [Online]. Available: <https://johnwlambert.github.io/icp/>.
- [27] P. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb. 1992, ISSN: 0162-8828, 2160-9292. DOI: [10.1109/34.121791](https://doi.org/10.1109/34.121791). [Online]. Available: <http://ieeexplore.ieee.org/document/121791/> (visited on 06/05/2024).
- [28] P. Li, R. Wang, Y. Wang, and W. Tao, "Evaluation of the icp algorithm in 3d point cloud registration," *IEEE Access*, vol. 8, pp. 68 030–68 048, 2020. DOI: [10.1109/ACCESS.2020.2986470](https://doi.org/10.1109/ACCESS.2020.2986470).
- [29] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, Quebec City, Que., Canada: IEEE Comput. Soc, 2001, pp. 145–152, ISBN: 978-0-7695-0984-6. DOI: [10.1109/IM.2001.924423](https://doi.org/10.1109/IM.2001.924423). [Online]. Available: <http://ieeexplore.ieee.org/document/924423/> (visited on 06/05/2024).
- [30] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D registration," in *2009 IEEE International Conference on Robotics and Automation*, Kobe: IEEE, May 2009, pp. 3212–3217, ISBN: 978-1-4244-2788-8. DOI: [10.1109/ROBOT.2009.5152473](https://doi.org/10.1109/ROBOT.2009.5152473). [Online]. Available: <http://ieeexplore.ieee.org/document/5152473/> (visited on 03/25/2024).
- [31] M. Zuliani, "Ransac for dummies," *Vision Research Lab, University of California, Santa Barbara*, 2009.
- [32] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, ISSN: 0001-0782. DOI: [10.1145/358669.358692](https://doi.org/10.1145/358669.358692). [Online]. Available: <https://dl.acm.org/doi/10.1145/358669.358692> (visited on 06/07/2024).
- [33] en. [Online]. Available: <https://dev.intelrealsense.com/docs/optical-filters-for-intel-realsense-depth-cameras-d400>.
- [34] O. Kahmen, N. Haase, and T. Luhmann, "ORIENTATION OF POINT CLOUDS FOR COMPLEX SURFACES IN MEDICAL SURGERY USING TRINOCULAR VISUAL ODOMETRY AND STEREO ORB-SLAM2," en, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B2-2020, pp. 35–42, Aug. 2020, ISSN: 2194-9034. DOI: [10.5194/isprs-archives-XLIII-B2-2020-35-2020](https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-35-2020). [Online]. Available: <https://isprs-archives.copernicus.org/articles/XLIII-B2-2020/35/2020/> (visited on 01/23/2024).
- [35] B. M. Hetaimish, "Sawbones laboratory in orthopedic surgical training," en, *Saudi Medical Journal*, vol. 37, no. 4, pp. 348–353, Apr. 2016, ISSN: 0379-5284, 1658-3175. DOI: [10.15537/smj.2016.4.13575](https://doi.org/10.15537/smj.2016.4.13575). [Online]. Available: <https://smj.org.sa/lookup/doi/10.15537/smj.2016.4.13575> (visited on 06/19/2024).
- [36] W. Wein, A. Karamalis, A. Baumgartner, and N. Navab, "Automatic bone detection and soft tissue aware ultrasound–CT registration for computer-aided orthopedic surgery," en, *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 971–979, Jun. 2015, ISSN: 1861-6410, 1861-6429. DOI: [10.1007/s11548-015-1208-z](https://doi.org/10.1007/s11548-015-1208-z). [Online]. Available: <http://link.springer.com/10.1007/s11548-015-1208-z> (visited on 05/13/2024).
- [37] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," en, *Medical Image Analysis*, vol. 16, no. 6, pp. 1101–1112, Aug. 2012, ISSN: 13618415.

- DOI: [10.1016/j.media.2012.07.005](https://doi.org/10.1016/j.media.2012.07.005). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1361841512000977> (visited on 06/21/2024).
- [38] M. Ciganovic, F. Ozdemir, F. Pean, P. Fuernstahl, C. Tanner, and O. Goksel, “Registration of 3D freehand ultrasound to a bone model for orthopedic procedures of the forearm,” en, *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 6, pp. 827–836, Jun. 2018, ISSN: 1861-6429. DOI: [10.1007/s11548-018-1756-0](https://doi.org/10.1007/s11548-018-1756-0). [Online]. Available: <https://doi.org/10.1007/s11548-018-1756-0> (visited on 06/05/2024).
- [39] X. Hu, H. Liu, and F. R. Y. Baena, “Markerless Navigation System for Orthopaedic Knee Surgery: A Proof of Concept Study,” en, *IEEE Access*, vol. 9, pp. 64 708–64 718, 2021, ISSN: 2169-3536. DOI: [10.1109/ACCESS.2021.3075628](https://doi.org/10.1109/ACCESS.2021.3075628). [Online]. Available: <https://ieeexplore.ieee.org/document/9416444/> (visited on 01/23/2024).
- [40] C. Zhang, Y. Liu, Y. Zhang, and H. Li, “A hybrid feature-based patient-to-image registration method for robot-assisted long bone osteotomy,” en, *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 9, pp. 1507–1516, Sep. 2021, ISSN: 1861-6410, 1861-6429. DOI: [10.1007/s11548-021-02439-5](https://doi.org/10.1007/s11548-021-02439-5). [Online]. Available: <https://link.springer.com/10.1007/s11548-021-02439-5> (visited on 05/07/2024).
- [41] Intel Corporation, “Intel RealSense D400 Series Product Family Datasheet,” Tech. Rep. 337029-017, Sep. 2023. [Online]. Available: <https://dev.intelrealsense.com/docs/intel-realsense-d400-series-product-family-datasheet> (visited on 02/12/2024).
- [42] I. R. Store, *Intel® realsensetm depth camera d405*, en. [Online]. Available: <https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d405.html> (visited on 06/03/2024).
- [43] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.
- [44] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, “A rewriting system for convex optimization problems,” *Journal of Control and Decision*, vol. 5, no. 1, pp. 42–60, 2018.
- [45] D. Rodriguez-Losada and J. Minguez, “Improved Data Association for ICP-based Scan Matching in Noisy and Dynamic Environments,” en, in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, ISSN: 1050-4729, Rome, Italy: IEEE, Apr. 2007, pp. 3161–3166, ISBN: 978-1-4244-0601-2. DOI: [10.1109/ROBOT.2007.363960](https://doi.org/10.1109/ROBOT.2007.363960). [Online]. Available: <http://ieeexplore.ieee.org/document/4209578/> (visited on 05/07/2024).
- [46] M. Berger, A. Tagliasacchi, L. M. Seversky, *et al.*, “A Survey of Surface Reconstruction from Point Clouds,” en, *Computer Graphics Forum*, vol. 36, no. 1, pp. 301–329, Jan. 2017, ISSN: 0167-7055, 1467-8659. DOI: [10.1111/cgf.12802](https://doi.org/10.1111/cgf.12802). [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1111/cgf.12802> (visited on 03/21/2024).
- [47] J. Sweetser and A. Grunnet-Jepsen, *Optical Filters for Intel® RealSense™ Depth Cameras D4xx*, en, Rev 1.0, Feb. 2023. [Online]. Available: <https://dev.intelrealsense.com/docs/optical-filters-for-intel-realsense-depth-cameras-d400> (visited on 02/12/2024).
- [48] A. Grunnet-Jepsen, J. N. Sweetser, P. Winer, A. Takagi, and J. Woodfill, *Projectors for Intel® RealSense™ Depth Cameras D4xx*, en, Rev 1.0, Feb. 2023. [Online]. Available: <https://dev.intelrealsense.com/docs/projectors> (visited on 02/12/2024).