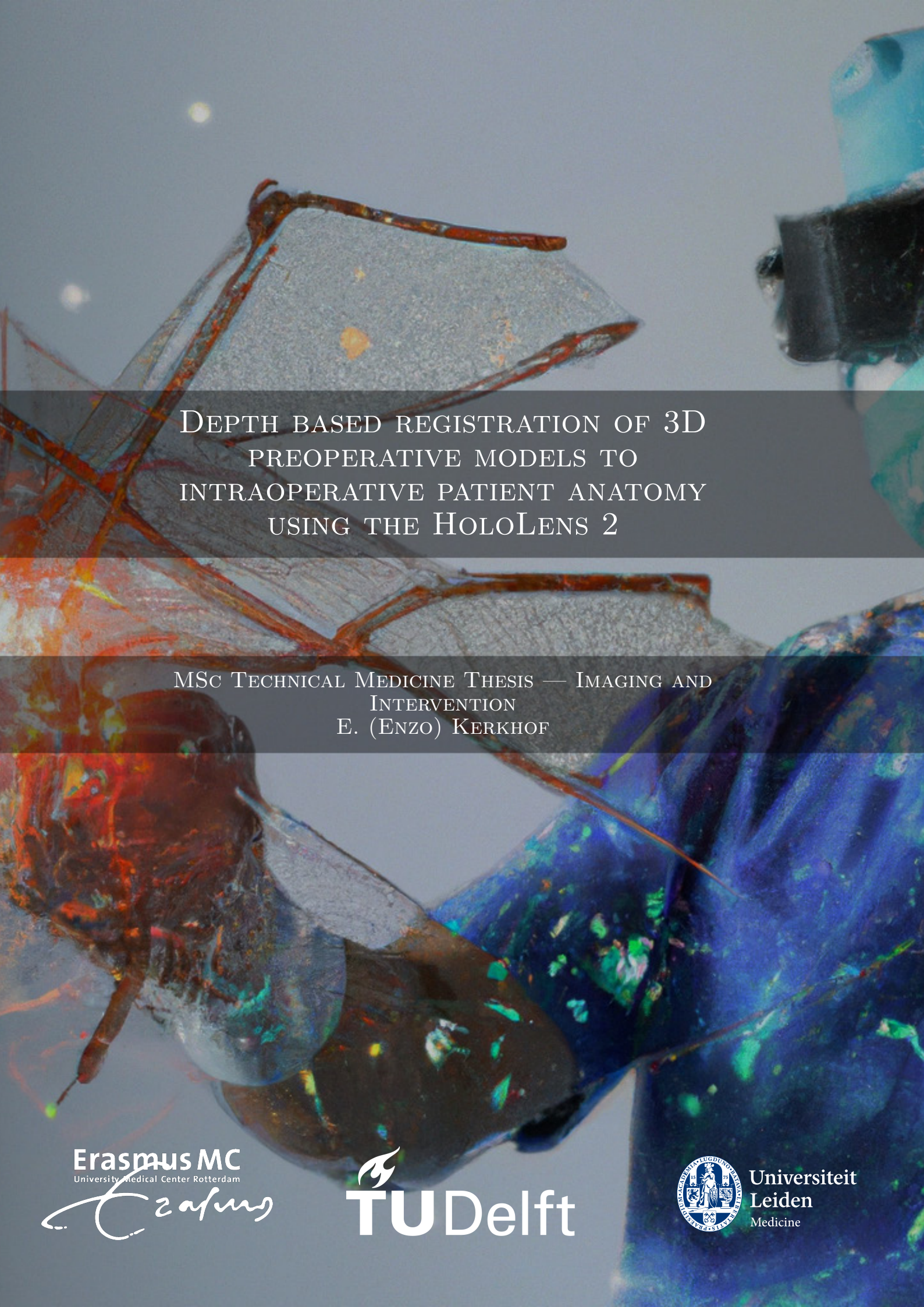


Depth based registration of 3D preoperative models to intraoperative patient anatomy using the HoloLens 2

MSc Technical Medicine Thesis — Imaging and Intervention

E. (Enzo) Kerkhof

Erasmus MC
University Medical Center Rotterdam

TU Delft

Universiteit Leiden
Medicine

Depth based registration of 3D preoperative models to intraoperative patient anatomy using the HoloLens 2

by

Enzo Kerkhof

Student number: 4448855

July 2, 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University — Delft University of Technology — Erasmus University Rotterdam

Master thesis project (TM30004, 35 ECTS)
Erasmus MC
Department of Surgery and Radiology & Nuclear Medicine
September 2022 - June 2023

Chair and technical supervisor:	Dr. Theo van Walsum	Erasmus MC
Medical supervisor:	Dr. Tessa van Ginhoven	Erasmus MC
Independent committee member:	Dr. Denise Hilling	LUMC
Optional committee member:	MSc. Abdullah Thabit	Erasmus MC

TU Delft

Erasmus MC
Universitair Medisch Centrum Rotterdam

LUMC Leids Universitair Medisch Centrum

Preface

This thesis concludes my time in the educational programme of Technical Medicine. The last 8 years have been a great adventure of following a new study programme paired with lots of uncertainties. My interest and passion for the field have (mostly) provided the counterbalance to take these uncertainties for granted and see them as potential opportunities. Hospitals and the treatment of patients as a team is something I've always found compelling, and I honestly think that technical physicians can make a difference there. I'd like to make a difference. With great excitement I look up to the accomplishments of my fellow students and colleagues. These examples motivate me. And thank you Mom, Dad, friends and others that provided the rest of the counterbalance by means of listening ears and emotional support.

I'd like to express my gratitude to all my supervisors for their personal investment in my graduation process. Tessa for giving me the unique opportunity to develop my clinical side within the OGC surgical ward and join the diagnostic team. Theo, Abdullah, Mohamed, and Pierre for their technical expertise and didactic skills for patiently supporting my learning process. And of course, Eveline and Robert for their openhearted coaching moments of confidence. These investments have led to my personal development as a technical physician and the promising results presented in this thesis.

Enzo Kerkhof
Rotterdam, June 2023

Depth based registration of 3D preoperative models to intraoperative patient anatomy using the HoloLens 2

Enzo Kerkhof[1,2,3*], Abdullah Thabit[2], Mohamed Benmahdjoub[4], Tessa van Ginhoven[3], Theo van Walsum[2]

[1*]Educational program Technical Medicine, Leiden University Medical Center, Delft University of Technology & Erasmus MC, Rotterdam, The Netherlands.
[2]Department of Radiology & Nuclear Medicine, Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, The Netherlands.
[3]Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus MC Cancer Institute, Rotterdam, The Netherlands.
[4]Department of Oral and Maxillofacial Surgery, Erasmus MC Cancer Institute, Rotterdam, The Netherlands.

*Corresponding author(s). E-mail(s): e.kerkhof@erasmusmc.nl;

Abstract

Purpose: Image-guided surgery can decrease perioperative complication incidences. Augmented Reality (AR) with head-mounted displays (HMDs) provides an accessible method of visualizing 3D preoperative models intraoperatively. This could provide the surgeon with an easy to use intraoperative image-guided surgery system.[1] Image-to-patient registration, the other key step of image-guided surgery, can still be intricate and time-consuming with traditional systems.[2, 3] This research explores the feasibility of using the depth sensors of the HoloLens 2, a state-of-the-art AR HMD, for depth-based image-to-patient registration. This research contributes to the advancement of less complex and more efficient image-guided surgical techniques.

Methods: To achieve these objectives, three experiments were conducted using a pilot system based on the HoloLens 2's depth sensors. The first experiment evaluated the accuracy of the depth sensors quantitatively. The second experiment compared four registration initialization methods, including manual and automated approaches. The accuracy and success rate of alignment were assessed using a multi-modal ground truth. Finally, a qualitative assessment of the pilot

system was performed on various objects and materials. This experiment aimed to evaluate the system's performance and usability in real-world scenarios.
Results: The depth accuracy experiment showed that both the AHAT and LT sensors had mean overestimation errors of 5.7 and 9.0 mm, respectively. In the registration experiment, the two manual initialization methods consistently achieved successful registration (100%), while the two automatic methods had varying success rates (23.3% and 50%). Three out of four depth registration methods completed the registration within 5 seconds. The mean translation errors ranged from 12.6 to 14.7 mm, and rotation errors ranged from 1.5 to 1.8 degrees. The minimum observed translation and rotation errors were 6.9 mm and 0.5 degrees, respectively, while the maximum errors were 18.8 mm and 3.2 degrees, respectively.
Conclusion: The study's results suggest the potential for achieving sub-10 mm registration accuracy within 5 seconds with depth-based image-to-patient registration. This offers a fast and convenient alternative to other tracking systems that require invasive fiducial markers and time-consuming calibration steps. However, the current accuracy level of the system poses some limitations. Nonetheless, the developed system holds promise for a wide range of surgical procedures that currently do not utilize image guidance due to its complexity. By enabling faster and more accessible image guidance, depth-based registration has the potential to enhance surgical outcomes, such as improving tumor resection margins and avoidance of vulnerable tissues, making it highly beneficial for various procedures.

Keywords: Image-guided surgery, Augmented Reality, Head-mounted displays, HoloLens 2, Depth-based registration, Image-to-patient registration.

1 Introduction

Surgeons require extensive comprehension of patient's three-dimensional (3D) anatomy when performing operations. Navigating between anatomical structures and performing surgery with the minimal amount of collateral damage takes years of training. To decrease the amount of perioperative complications and be able to perform high complex surgeries, more medical-technological development is aimed at supporting surgeons performing these procedures.

One approach to assist surgeons is by minimizing the cognitive effort required for mentally reconstructing the 3D geometry of patients' anatomy. Medical tomographic imaging has enabled detailed 3D preoperative insight and planning, but often the 3D anatomy still needs to be cognitively mapped from the 2D slices. Fortunately, it is currently possible to segment the anatomy of interest and reconstruct a 3D model of the patient.[4] Rendering these 3D models on a PC-monitor removes the task to cognitively visualise the third dimension for surgeons and enables 3D preoperative planning.

To extend this 3D visualization beyond the preoperative phase and enable intraoperative image-guided navigation, the preoperative 3D models need to be registered and visualised intraoperatively. Intraoperative 3D imaging modalities such as ultrasound (US) or cone beam CT (CBCT) could be used as well. However, 3D intraoperative

imaging modalities tend to perform less compared to preoperative modalities in terms of contrast and resolution. This is due to intraoperative limitations such as time constraint and radiation exposure. Intraoperative imaging does not reach the detailed level that is necessary for complex surgical procedures. Also, intraoperative imaging does not provide access to preoperative planning information. Therefore, methods to register and visualise the high-quality 3D preoperative models in the operative field are of significant relevance for intraoperative navigation. To use preoperative imaging for intraoperative support, the accuracy of registration and visualisation is key to a reliable system.

The use of Augmented Reality (AR) in surgical applications is an emerging research topic with the number of publications increasing significantly over the past decade.[3, 5, 6] AR head mountable displays (HMD) show the user an augmented version of their reality by projecting 3D holograms or overlays into the operative field. This is different from virtual reality (VR) HMD that let the user experience a complete virtual environment. AR therefore has several potential advantages for intraoperative applications: i) it can visualise preoperative models in 3D and ii) the physician experiences no "attention switch" by presenting information in the operative field.

Registration is the process of aligning two or more images or 3D models of the same object or patient to create a single, unified representation. The goal of registration is to achieve the most optimal alignment between the two images or models. In the case of intraoperative navigation this is called an image-to-patient registration, where the source 3D model comes from preoperative medical imaging and the intraoperative patient's anatomy is the target 3D model. Registration of 3D models can be achieved by transforming one moving or *source* model to a fixed or *target* model. A registration algorithm searches for the optimal alignment between two models, based on a set of predefined parameters such as the similarity of image intensities or specific corresponding 3D points or surfaces.

Accurate registration methods are already available, especially for operating on rigid anatomical structures such as in neurosurgery or orthopedic surgery. For example, neurosurgical navigation systems with registration errors of sub millimeter accuracy are the current golden standard.[7] These navigation systems use optical or electromagnetic (EM) tracked markers to acquire tool positions and perform a landmark based or surface based registration mainly for rigid surgical regions of interest (ROI). However, these systems in general have some drawbacks and limitations that need to be considered: 1) Complexity: existing image-guided navigation systems can be complex and require specialized equipment, including optical or electromagnetic tracking devices, sensors, and dedicated software. This complexity adds to the setup time and may require trained personnel to operate and maintain the system effectively. 2) Cost: implementing and maintaining image-guided navigation systems can be costly, requiring investments in equipment, software licenses, and staff training. 3) Limited application: due to the complexity and cost associated with existing navigation systems, they are typically implemented for specific surgeries or in specialized surgical centers. 4) Line-of-Sight limitations: optical tracking systems, which rely on line-of-sight between the tracking cameras and the tracked markers, can be susceptible to occlusion and line-of-sight limitations. This can impact the accuracy and reliability

of the navigation system, particularly in complex surgical environments. Therefore, while existing image-guided navigation systems have made significant advancements in surgical guidance, these drawbacks highlight the need for further research and development to address the limitations and improve the accessibility and usability of such systems in a wider range of surgical procedures.

With the development of AR HMDs the accessibility of visualising 3D preoperative models intraoperatively is increasing.[1, 6] However, to navigate with AR HMDs the 3D models need to be registered to intraoperative patient space[2, 3], and using an EM or optical image-guided navigation system for these registration problems adds the previously mentioned drawbacks to the system. AR HMDs image-guided surgery has potential for both replacing complex systems and for surgical applications that currently do not use any traditional image-guided navigation system due to the added complexity.[1] An example application are procedures with a localization challenge, such as sentinel lymph node excisions, which currently rely on directional-guided navigation systems using invasive markers.[8] With accessible image-guidance navigation, these procedures could be performed more efficiently by providing more than only directional information. This and other application fields would accordingly benefit from less complex methods for image registration, especially now 3D preoperative model visualization has become more accessible.

The Microsoft HoloLens 2, released in 2019 as the successor to HoloLens 1 (2016), is one of the most recently developed AR HMDs. The potential of the HoloLens for image-guided surgery has had considerable research interest since its first release.[2, 3, 5, 6] Gsaxner et al. (2023)[1] conducted a comprehensive systematic review that examined the broader applications of the HoloLens in medicine but also provided many specific examples to image-guided interventions. In this surgical context different tracking methods and image registration methods using the HoloLens without external system have also been subject of investigation.[2, 3, 9] Often, the cameras of the HoloLens are utilized to detect and track fiducial or optical markers.[10–13] However, the use of markers attached to a patient is often invasive, having markers in the surgical field is cumbersome, and the necessary manual calibration steps can add time and complexity to the procedures.

One promising avenue for addressing the registration challenge lies in leveraging the depth sensors of the HoloLens 2. These depth sensors have the potential to provide the necessary 3D data for registering preoperative models to the intraoperative patient space, eliminating the need for tracked instruments and user input. By utilizing depth-based methods, it could become feasible to achieve image-to-patient registration with greater simplicity and automation. Depth-based registration approaches have been proposed before by several studies.[14–25] Nonetheless, the current published research is still limited in numbers, often application specific and rarely translated to routine patient care.

The aim of this study is to evaluate the feasibility of using the depth sensors of the HoloLens 2 AR HMD for general depth-based image-to-patient registration. Specifically, the experiments focus on assessing the accuracy, robustness, speed and automation of different depth-based registration methods using the pilot system described in the methods section. By investigating these factors, this study contributes

to the broader goal of achieving precise and automated image-to-patient registration, thereby enhancing the efficiency and accessibility of image-guided surgery.

2 Methods

In AR-based surgical navigation the preoperative imaging and planning data needs to be aligned with the patient. Several steps are involved in performing the alignment based on the intraoperative patient anatomy. This section provides a detailed description of these steps (see Figure 1), beginning with the preparation of the imaging data and acquisition of the preoperative models. It then covers the process of intraoperative imaging and the direct overlay of the imaging data onto the patient. Additionally, the proposed HoloLens 2 depth-based registration system is introduced, along with the hardware and software utilized in the system.

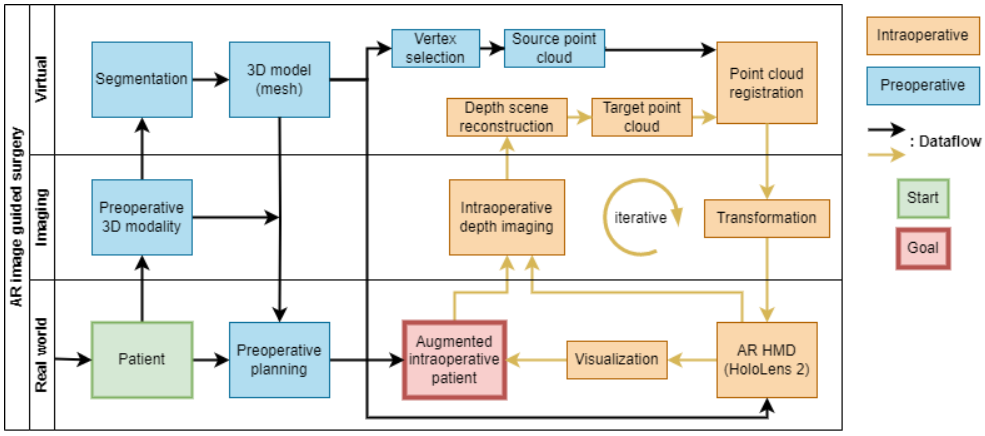


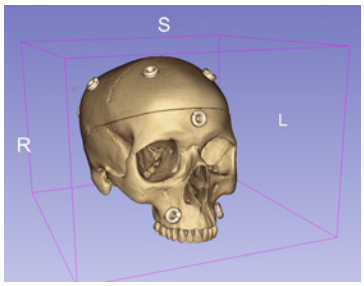
Fig. 1: Flowchart of AR HMD image guided surgery with depth-based registration

2.1 Preoperative imaging & planning

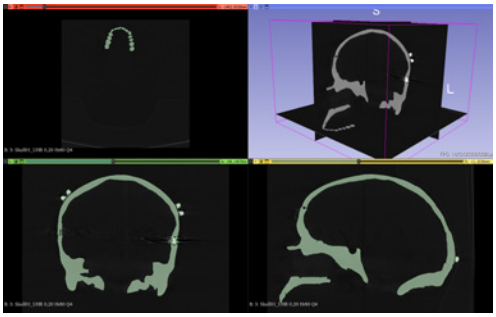
Preoperative imaging, such as CT or MRI scanning, is commonly performed to assist surgeons in planning surgical procedures. However, for the imaging data to be directly superimposed on the patient during surgery, a simpler representation of the preoperative image is needed. This can be done by modeling the target structure(s) from the 3D imaging data. The left side of the flowchart in Figure 1 shows these steps. The process involves segmenting (see Figure 2b) the desired anatomical structures of interest from the scans and generating a 3D representation in the form of a surface mesh model. A surface mesh consists of vertices, edges, and faces. The vertices represent the points in 3D space, while the edges and faces connect the vertices to form the surface of the model.

Different representations, such as the surface mesh model and point cloud, serve different purposes throughout the surgical navigation process. Figure 2 shows several

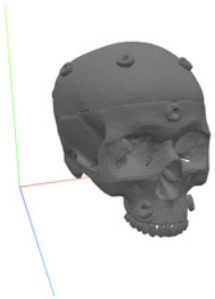
representations of a phantom skull. The surface mesh model, seen in 2c, is particularly useful for visualization purposes. Its faces provide a smooth and continuous surface that closely resembles the anatomical structure. This allows surgeons to gain a clear understanding of the geometry and spatial relationships of the structures they will be working with. On the other hand, by retaining only the vertices of the mesh a point cloud is obtained (see Figure 2e). This information can be utilized in various registration algorithms and spatial mapping processes to align the preoperative models with the patient anatomy.



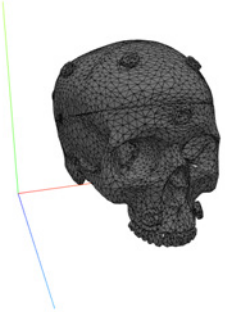
(a) CT volume ray casting rendering with shading



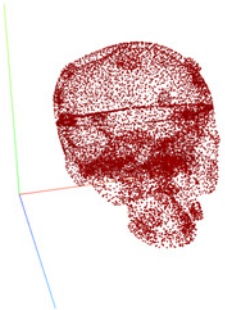
(b) CT slices with the segmented skull



(c) 3D mesh model



(d) Wireframe of edges visualized on surface model



(e) Point cloud of vertices

Fig. 2: Several digital representations of a phantom skull

2.2 Intraoperative Imaging

To align the preoperative image with the patient anatomy during surgery, an intraoperative representation is required. This can be achieved by capturing the surface of the patient's anatomy, as the surface is also represented in the preoperative scan. Depth sensors, such as those integrated into the Microsoft HoloLens 2, can therefore be utilized for intraoperative data acquisition. These sensors capture depth information of the scene, enabling the reconstruction of the 3D environment. The Microsoft HoloLens

2 utilizes two types of depth cameras: the Articulated Hand Tracking (AHAT) camera, which operates at a high frequency of 45 frames per second (FPS), and the Long-throw (LT) camera, which operates at a lower frequency of 1-5 FPS. Both cameras utilize Time-of-Flight (ToF) infrared technology to measure the distance to objects in the environment.[26]

The HoloLens 2 is furthermore equipped with other cameras and sensors to enable Simultaneous Localization and Mapping (SLAM). With SLAM the HoloLens tracks its own location in the world space, allowing the projection of holograms at fixed locations. The SLAM functionality is achieved through four grayscale cameras and an inertial measurement unit (IMU). The grayscale cameras detect points of interest in the environment, while the IMU provides information about force, angular rate, and headset orientation. The device also features a Photo Video (PV) camera capable of capturing 8 MP 1080P30 video and allows spatial input such as eye gaze tracking and voice recognition.

During intraoperative image acquisition (see the right side of Figure 1), the depth frame captured by the HoloLens 2 containing estimated depth values for each pixel is converted into a point cloud representation using a calibration mapping function. The points in this point cloud are still in the depth sensor space. Device specific calibration transforms for every HoloLens are available to transform from any sensor space to a common rigNode space.[26] Therefore, with the SLAM-based localization of the HoloLens and the thereby provided depth camera poses in world space, the points in the point cloud can be transformed from depth sensor space to the HoloLens world space coordinate system.

Subsequently, postprocessing techniques can be applied to refine the initial point cloud data. One well-established method for depth scene reconstruction is the Truncated Signed Distance Function (TSDF) volume integration.[26–29] TSDF volume integration involves voxel-based representation of the 3D space and updating the occupancy and distance values based on a stream of incoming depth data. This allows for the creation of a more detailed and accurate 3D representation of the intraoperative scene, which can then be used for alignment during the surgical procedure.

2.3 Image to patient alignment

Once the preoperative data and models are processed, and the intraoperative target data is acquired, the alignment between the two needs to be established (see top right Figure 1). In the case of rigid anatomical structures, a rigid transformation is sufficient for alignment. A rigid transformation includes translation and rotation of the source model, while preserving its original geometry. To achieve this alignment, point cloud representations and point cloud registration algorithms are commonly used. One frequently employed algorithm is the Iterative Closest Point (ICP) algorithm. The ICP algorithm iteratively aligns the preoperative point cloud with the intraoperative point cloud by minimizing the distance between corresponding points. This iterative process allows for the estimation of the rigid translation and rotation necessary to accurately align the two point clouds.

There are other registration algorithms available, each with different characteristics in terms of speed, accuracy, robustness to noise, and other factors. Some

examples of registration algorithms are Coherent Point Drift (CPD)[30], Globally Optimal ICP (Go-ICP)[31], 2-Point-Normal Sets[32] and Support Vector Registration (SVR)[33]. These algorithms offer alternative approaches for point cloud registration. Python libraries such as Open3D[28] provide implementations of various point cloud registration algorithms, including different variants of the ICP algorithm and other registration methods like RANSAC global registration[34] and fast global registration[35]. Open3D also includes feature extraction methods like Fast Point Feature Histograms (FPFH)[36] to assist in global registration. Global registration algorithms use feature extraction methods to extract distinctive features from point clouds, which are then used to find initial alignments between the preoperative and intraoperative point clouds. This helps to avoid falling into local minima and therefore reduces the failure rate of alignment attempts.

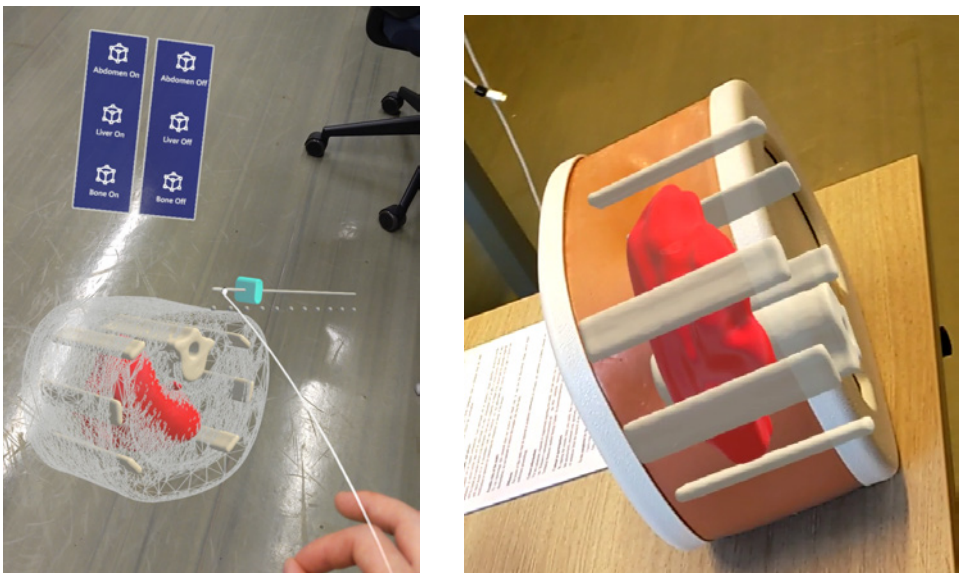
In current practices, a global registration algorithm is often used to establish an approximate alignment and is followed by a refinement algorithm such as ICP for fine-tuning and achieving higher accuracy. Once the transformation is determined, the AR hologram can be updated and visualized in the correct pose, augmenting the target anatomy during the surgical procedure.

2.4 AR Visualization

Traditional navigation systems can display augmentations such as planning annotations, tracked instruments, and anatomical segmentations superimposed on the preoperative data. However, Augmented Reality (AR) takes this a step further by overlaying these augmentations onto the user's reality. We focus on the Microsoft HoloLens 2 because the device can function as a stand-alone navigation system.

It is important to note that a different more detailed model may be used for the registration process compared to the one used for visualization. In the case of AR visualization with the Microsoft HoloLens 2, the models used for visualization are specifically tailored to the device's computational power, additive display characteristics, transparency, and other relevant factors. For example, to optimize the computational performance, models with a large number of vertices and faces may be remeshed to sparser and simpler models, reducing the rendering complexity. Additionally, user interface (UI) options can be integrated into the AR application. These options can include sliders to adjust the transparency of the models and buttons to activate or deactivate the rendering of specific models based on user preferences (see Figure 3).

An example workflow in AR visualization involves initially visualizing the surface model to assist the user in assessing the registration result. Once the user is satisfied with the alignment, the surface model can be deactivated to focus on the models projected inside the patient as seen in Figure 3b. These internal models can represent various relevant tissues such as lymph nodes, organs, tumors, bones, nerves, or vessels, providing guidance to the user during the surgical procedure. Furthermore, preoperatively planned incisions, drill trajectories, or cutting planes can be visualized on demand, enhancing the precision and accuracy of the surgical intervention.



(a) User interaction with transparency slider

(b) Wireframe surface model deactivated after alignment

Fig. 3: HoloLens app for abdominal phantom with transparency slider and on/off buttons

2.5 Pilot HoloLens 2 system

The pilot system consists of a HoloLens 2 device and a PC connected to the same network. In summary, the HoloLens captures depth sensor frames and transmits them to the PC. The PC processes the data to determine the intraoperative position of the preoperative model and sends this information back to the HoloLens. The HoloLens then visualizes an aligned hologram based on the received location. Figure 4 provides an overview of the pilot system components, illustrating their interconnections and data flow.

More specifically, the system employs a Unity engine application, developed for the HoloLens 2 using Mixed Reality Toolkit (MRTK), which offers a user interface and enables interaction with preloaded models.[37] The Unity application incorporates a server that allows the PC client to establish a connection with the HoloLens device over the network. To enable this client-server communication with data streaming functionality, the pilot system utilizes a Unity plugin called HoloLens 2 Sensor Streaming (HL2SS), developed and published by Diben et al. in 2022 [38]. This plugin, which is open-source, leverages the capabilities of the Microsoft HoloLens 2 Research Mode [26] to access raw sensor data. By integrating the HL2SS plugin into a Unity application, it becomes possible to establish TCP-based streams of sensor data from the HoloLens 2 to the PC client. The HL2SS server not only facilitates the streaming of sensor data but also allows clients to send a message back to the server. Noteworthy

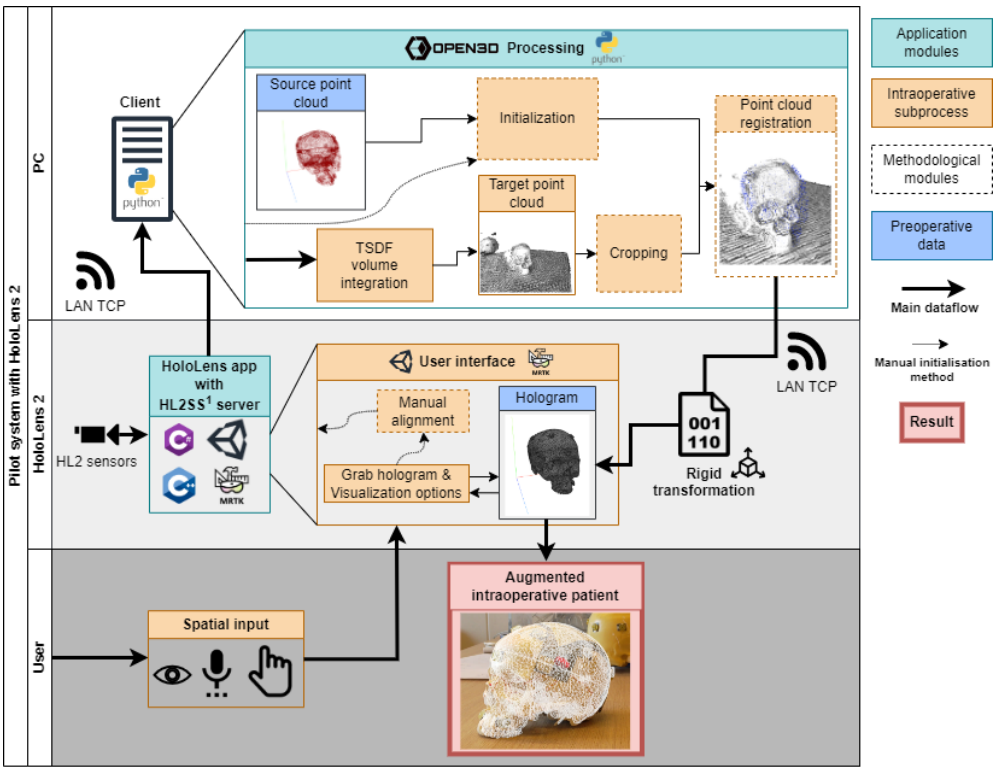


Fig. 4: Data flow and processing options of depth-based HoloLens 2 registration system

[1]HoloLens 2 Sensor Streaming Unity plugin (dll) GitHub Repository

is that we made modifications to the HL2SS source code to enhance its functionality for our specific requirements. The original plugin did not support remote interaction with existing Unity GameObjects. Therefore, we adapted to code to enable the Unity application to receive a world space transformation matrix for a predefined GameObject, specifically the preoperative model. Another modification was made to enable the PC client to request the world space transformation matrix of a preloaded preoperative model GameObject. This capability proved useful as it could be used as an initialization position and rotation for the registration algorithms. Consequently, the pilot system can stream frames from the depth sensors of the HoloLens to a Python application running on the PC, which, in turn, can send a registration matrix back to the Unity application on the HoloLens.

On the client side (a python script running on the PC) we use the Open3D Python library for processing due to its active development and the extensive range of implemented features. The Open3D based processing handles all incoming HoloLens data and performs various operations on them. These operations can include point cloud registration, feature extraction, surface reconstruction, or any other processing steps for aligning the preoperative models with the intraoperative data. For our experiments

in this study, we compared several methods that could be implemented in the processing module, with a specific focus on the steps denoted by the dotted borders in Figure 4, such as initialization and cropping of the region of interest in the target point cloud. We limited our investigations to the state-of-the-art implementations available in the Open3D library. In particular, for the reconstruction of the target point cloud, we chose to utilize TSDF (Truncated Signed Distance Function) volume integration. TSDF volume integration is considered the state of the art in dense scene reconstruction. While there are numerous options for optimization in dense scene reconstruction, we focused on the feasibility of the system rather than extensive optimization for our experiments.

3 Experiments & Results

We performed three experiments aimed at assessing the feasibility of utilizing the depth sensors of the HoloLens 2 AR HMD for general depth-based image-to-patient registration. The first experiment focused on evaluating the accuracy of the depth sensors, as these results would provide insights into the expected performance of the image-to-patient registrations. In the second experiment, a quantitative assessment using the pilot system described in the method section 2.5 was conducted. This experiment aimed to compare various registration approaches of the pilot system. Lastly, a qualitative experiment was conducted using different target objects to assess the system's performance with varying material types and geometries. These selected experiments were conducted to gain valuable insights and determine the potential of the system in real-world scenarios.

3.1 HoloLens depth sensors accuracy

This first experiment aims to evaluate the accuracy of the depth frames captured by the Articulated HAnd Tracking (AHAT) and Long-throw (LT) cameras of the HoloLens 2 under various distances and lighting conditions. Accurate depth measurements are crucial for ensuring the reliability of the image-to-patient registration process, as the depth data serves as the intraoperative target for registration. Our main focus was to assess the accuracy behavior of these sensors in a surgical environment.

3.1.1 Experimental setup

Similar to the work of Gu et al. (2021)[22], this experiment aims to compare distances estimated by the HoloLens sensors with ground-truth known distances. The experiment involves observing objects placed on a table under specific conditions while recording the depth data from the HoloLens sensors. To establish a ground-truth, the dimensions of the objects are known and the distance from the HoloLens to the table is measured using a different method. The different measurement method of the distance is further explained in section 3.1.2.

For the experiment setup, as seen in Figure 5a, three objects with a flat surface and of known dimensions were placed on a flat table, with the table itself functioning as a fourth surface. The objects used in the experiment consisted of a 3D printed

rectangular house-shaped model, a Lego cube, and a box, all covered with regular matte white printer paper to ensure similar surface material characteristics. Table 1 reports the heights of these objects.

Table 1: Height of observed objects

Object	**Height [mm]**
Table	0
House Model	60
LEGO Cube	95
Box	247

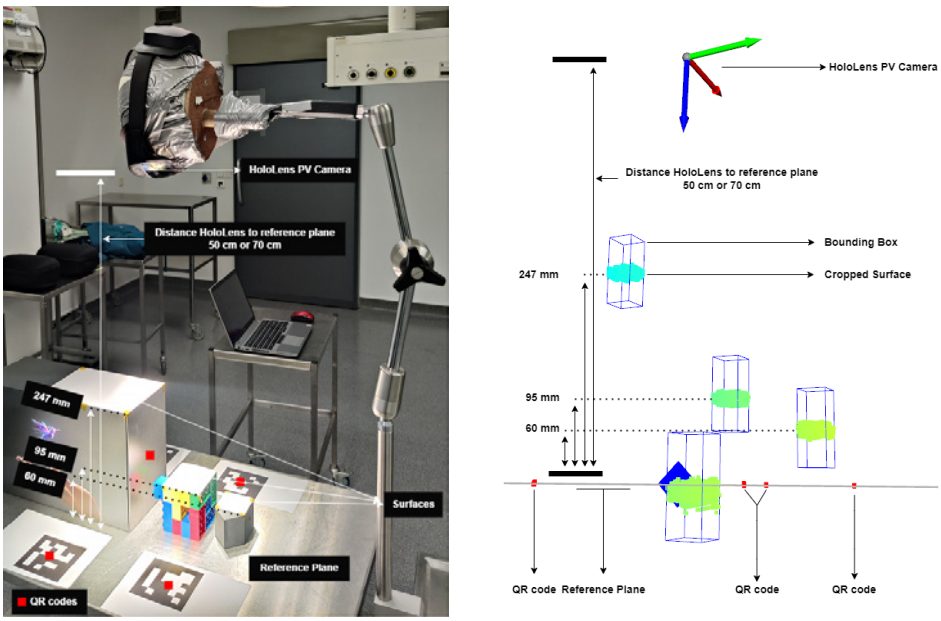
Table 2: Acquisition Variables

Hololens Device	**OR Lighting**	**Distance to table [cm]**
A / B	On / Off	50 / 70

The experiment involved two HoloLens devices and specific independent environmental variables to ensure a relevant evaluation of the sensors (Table 2). To achieve this, the depth values were analyzed within a range that is relevant to surgical procedures. The angle and distance between the HoloLens and the table were perpendicular and fixed at 50 cm or 70 cm (see Figure 5). Additionally, we investigated the impact of operating room (OR) lights on the depth measurements, providing an assessment of the sensors' performance under realistic conditions encountered during surgical procedures. In our setup, four surfaces and two setup distances resulted in eight different surface distances to be observed. Subsequently due to using two lighting conditions and the two depth sensors, 8 acquisitions needed to be done per HoloLens device. Resulting in 16 acquisitions that ran for 30 seconds.

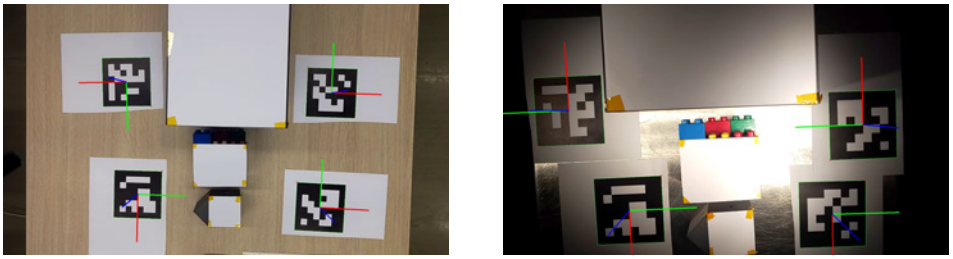
3.1.2 Distance evaluation metrics

To establish a ground truth, a reference plane was created using four ArUco[39, 40] QR codes with a width of 10 cm. The purpose of establishing the reference plane is to provide an external reference for the estimated depths. By creating a reference plane for each acquisition, the point cloud points can be projected onto this plane, enabling the measurement of the distance between the point cloud points and the reference plane. For each observed surface, the expected distance to the reference plane is known and should be either 0 mm, 60 mm, 95 mm, or 247 mm, corresponding to the specific objects used in the experiment. The length between the estimated distance (D_{est}) and the reference distance (D_{ref}) subtracted from the known object height (D_{real}) gives the error in depth estimation (D_{error}). This error indicates whether the depth is



(a) OR light setup (b) Virtual side view of setup

Fig. 5: Experiment setup of depth camera accuracy evaluation



(a) Normal lighting conditions (b) OR lighting conditions

Fig. 6: RGB frames of HoloLens PV camera with detected QR-codes

overestimated or underestimated:

$$D_{\text{error}} = D_{\text{real}} - |\overrightarrow{D_{\text{est}} D_{\text{ref}}}|$$

Figure 5 shows the experimental setup as well as the side view of the setup in the virtual world coordinate space.

The QR codes were positioned around the target objects on the table. The detection and pose estimation of the QR codes in the HoloLens PV camera space were performed using OpenCV and ArUco libraries.[39–41] The PV camera was configured

at a resolution of 1080p and 30 fps. The size of the marker and the highest resolution of the HoloLens 2 PV camera have been chosen based on the publication of Thabit et al. (2022)[9], where they evaluated the accuracy of ArUco marker tracking using the HoloLens 2 PV camera and reported a mean translational error of 1.8 ± 0.6 mm for 10 cm marker size at 50 cm distance, using the same resolution of 1080p. For this experiment we assume this to be accurate enough as a ground truth.

The HoloLens provides device specific transforms that allow for the transformation from any camera space to a common device rigNode space.[26] This capability was utilized to transform the pose estimations of the QR codes in the PV space and the point cloud points in the AHAT or LT space to a shared world space coordinate system. For every acquisition all RGB frames and their respective focal and principal points were used for QR code pose estimation. For a single frame example see Figure 6. The mean coordinates of the four QR codes were used to fit a plane using the PyVista python library implementation of the singular value decomposition (SVD) algorithm.[42]

Before the start of every acquisition the angle between the reference plane (based on a single RGB frame) and the forward vector of the PV camera was set to 180 ± 5 degrees to additionally ensure that the HoloLens was perpendicular to the surfaces. Likewise the distance between the HoloLens and the reference plane was set at 50 cm or 70 cm, with a tolerance of 1 cm. For each individual acquisition, the depth frames were converted into point clouds in world space, and all points were combined to form a single point cloud representing the entire 30-second acquisition. From this comprehensive point cloud, the flat surfaces of the objects and the table were manually selected. Using bounding boxes with dimensions of $3 \times 3 \times 8$ cm, the surfaces were cropped to ensure a consistent surface area for calculating the depth estimation error.

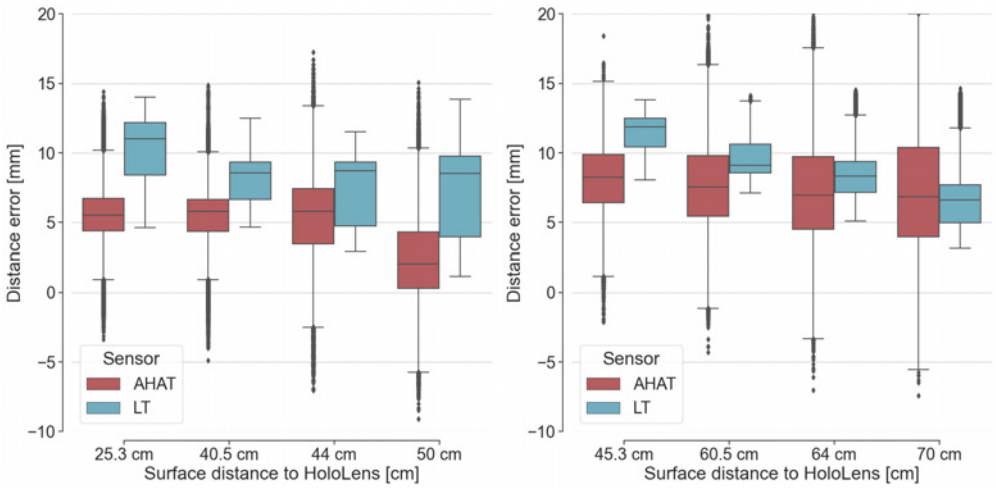
3.1.3 Results

For the 50 cm distance and the OR light on, the QR codes had to be re-positioned to the center of the frame to ensure detection. (Figure 6) For all 16 acquisitions, the standard deviation of the ArUco pose estimation for the QR codes in world coordinate space was determined. These are based on the PV camera that reports a pose per RGB frame based on the HoloLens SLAM. Since the HoloLens is fixed in position during acquisition, the SLAM drift can be removed from the pose estimation by only using the PV SLAM based pose of the first acquired frame. Table 3 shows the standard deviations of translation for these three factors.

Table 3: Standard deviations of SLAM and estimated QR-code positions

Translation standard deviation (in world coordinate space)	**x [mm] mean/min/max**	**y [mm] mean/min/max**	**z [mm] mean/min/max**
QR code positions with SLAM poses	0.18/0.05/0.57	0.36/0.11/0.75	0.19/0.06/0.73
SLAM PV camera poses	0.17/0.04/0.59	0.17/0.05/0.67	0.21/0.03/0.53
QR code positions with single PV pose	0.09/0.05/0.14	0.27/0.09/0.53	0.08/0.04/0.14

For all 16 acquisitions the maximum reported standard deviation of all SLAM based PV camera poses was sub-millimeter translation. Using these poses for QR code position estimation the maximum standard deviation of 15 acquisitions was still found to be sub-millimeter, but slightly higher. This implies a stable reference plane, ensuring precise error calculations. However, one acquisition conducted under OR lighting conditions exhibited significant axial QR code pose estimation standard deviations of 10 to 20 millimeters. This large variance in QR code position detection potentially leads to an unreliable reference plane. Therefore, this particular acquisition was excluded from further analysis. Estimating the QR code positions in world coordinate space based on a single PV pose to remove the SLAM drift, in these cases translation standard deviation decreased. Appendix Table A1 provides an overview of the acquired frames and QR code poses for each acquisition.



(a) 50 cm distance setup

(b) 70 cm distance setup

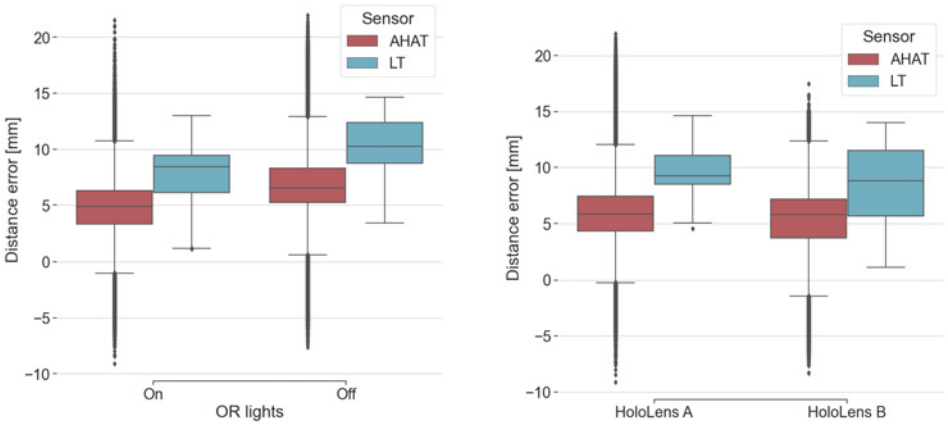
Fig. 7: Depth estimation errors at different distances to the HoloLens for the AHAT and LT sensor

Figure 7 presents the distance errors (with respect to the depth measured via the ArUco markers) per sensor and per observed surface. A positive distance error indicates depth overestimation. The following observations can be made:

- Both sensors (AHAT and LT) exhibit a mean overestimation error for all surfaces.
- The LT sensor generally shows a higher mean overestimation, although this difference decreases for surfaces located further than 50 cm.
- The standard deviation of the LT sensor is lower for surfaces positioned beyond 50 cm.
- The standard deviation of the AHAT sensor increases with the distance to the surface.

- The AHAT sensor exhibits a higher number of outliers compared to the LT sensor.

Furthermore, Figure 8 displays the distance errors for all acquisitions categorized by lighting conditions and HoloLens device. The difference observed between the two HoloLens devices is small: difference of mean error for AHAT and LT sensor of 0.74 mm and 1.34 mm. A larger difference in the mean error appears to be present for acquisitions conducted under other OR lighting conditions: a decrease of mean error for AHAT and LT sensor of 1.95 mm and 2.40 mm with the OR lights on.



(a) OR lighting conditions (b) Two different HoloLens 2 devices

Fig. 8: Depth estimation errors for the AHAT and LT sensor

Upon visual inspection, the point clouds generated by the AHAT and LT sensors exhibited various artifacts. Both sensors displayed a distinctive pattern of spherical layers when examining the cross-section of a surface, originating from the direction of the HoloLens (see Figure 9). The LT sensor, in particular, failed to provide depth values for areas in close proximity to the HoloLens. This issue was most prominent when capturing the surface of the tallest object, the box, during the acquisition setup at a distance of 50 cm from the reference plane. On the other hand, the AHAT sensor demonstrated more artifacts in the air close to the sensor and at the edges of the depth frame. In contrast, the point clouds generated by the LT sensor did not exhibit these artifacts. Additionally, both sensors showed irregularities in the pattern of the black area of the QR codes on the table. Examples of these artifacts can be found in Appendix A. Considering these observations, the LT sensor is preferred for use in the registration methods of the next experiment due to its relatively fewer artifacts and lower standard deviations for the depth error in the feasible operating range.

3.2 Depth registration methods

This experiment aims to evaluate the accuracy, robustness, and speed of depth registration with the HoloLens. Four different depth-based registration methods were assessed

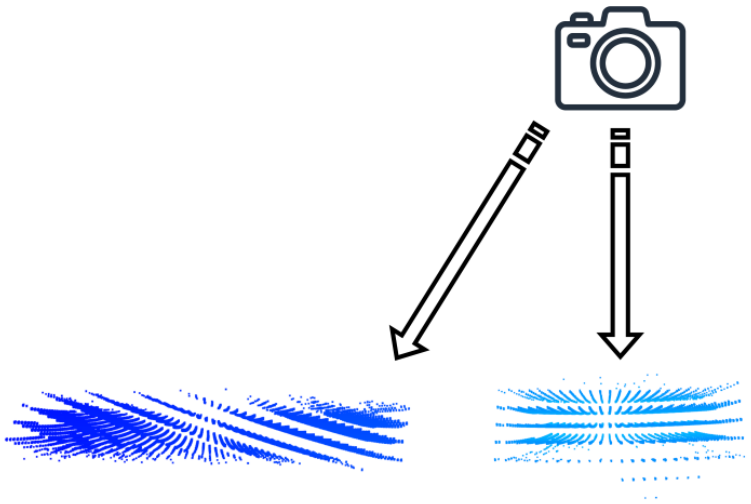


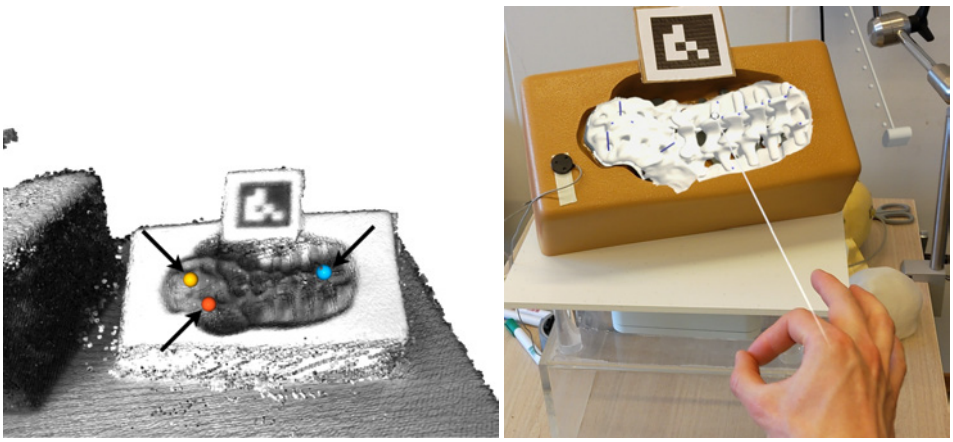
Fig. 9: Surface crops of complete 30 seconds AHAT acquisition point cloud with annotated point of view of the HoloLens

and compared using the pilot system described in Section 2.5. The objective was to compare a limited set of methods to assess the core features of the system and determine the potential of using the HoloLens 2 LT sensor for depth-based image-to-patient registration in a surgical user environment. This experiment serves as a starting point to assess the feasibility and potential of the system, and the results will guide future improvements and optimizations to enhance its accuracy and performance.

3.2.1 Experimental setup

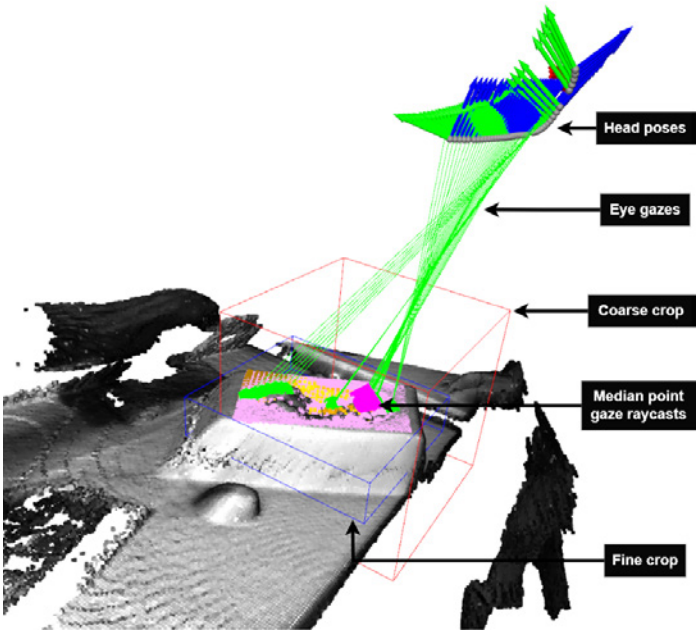
The pilot system allows for the implementation of various point cloud registration methods, depicted as methodological modules in Figure 4. As mentioned in the methods section ICP is an established point cloud registration method, but it does require a good initialization to converge. To determine what initialization is required to solve the image-to-patient registration problem with the ICP algorithm as a succeeding step for fine registration, this experiment focuses on four specific registration initialization methods.

The rationale to compare these four different methods is to know their speed, success rate, and their influence on the resulting registration accuracy. The first method is a manual corresponding point picker initialization method, which is also provided as an example function in Open3D. However, since the user has to manually select these points on the client PC, the process is cumbersome and may not suitable for a clinical workflow. Therefore, a second method was developed to transfer the manual initial alignment task to the HoloLens app, making it easier to use in a clinical setting. The other two methods are automatic, aiming to further enhance user convenience. These



(a) 3 point picking on target point cloud

(b) Manual coarse alignment



(c) Eye gaze raycast on target point cloud

Fig. 10: Different registration initialisation methods

automatic methods utilize the eye gaze tracking capability of the HoloLens 2 for determining the region of interest (surgical field) to initialize the registration automatically. The four methods are described in more detail below.

The target phantom used for registration was a Sawbones spine phantom, as shown in Figure 11. In this experiment, the HoloLens app was build with the preoperative model of this target phantom, acquired from a CT scan as mentioned in section 2.1.

The app included five visualizations of the model differing in colour, each corresponding to a specific registration method and one assigned for the ground truth reference. These visualizations update based on the received registration results from the PC. On the PC, the four methods received the same source point cloud based on a preoperative model. Depth data from the HoloLens LT sensor was used for the target point cloud, since the previous experiment showed the LT sensor can provide higher quality point clouds compared to the AHAT sensor. For every registration task the four different methods were also provided with the same target point cloud. The target point cloud was generated using the truncated signed distance function (TSDF) volume integration implementation of Open3D.[27, 28] TSDF volume integration parameters were set to a voxel size of 1.3 mm and the truncation value for the SDF was set to 0.002. These parameters provided a decent balance between computational time, detailed feature reconstruction and noise filtering.

A registration task proceeded as follows: 1) The model coupled with the "Manual" method was grabbed and placed roughly aligned on the target phantom by the user. 2) The target phantom was observed by the user from several sides while making sure to also look at the target phantom. 3) The user gave the "align" voice-command. 4) The world position of the ground truth model was saved. 5) The last 50 frames and poses of the LT sensor and the left front grey-scale sensor stream are used as input for the TSDF volume integration. (Since the LT sensor is 5 FPS, this was approximately the last 10 seconds of data.) 6) The user went to the client PC to pick 3 points on the reconstructed target point cloud and then 3 point on the source point cloud. 7) All registration results were sent to their corresponding models in the HoloLens app to visualize the registration results. The reason the target phantom was observed by the user from several sides was to provide the TSDF volume integration more heterogeneous camera poses and depth frames and therefore potentially capture more features of the target phantom.

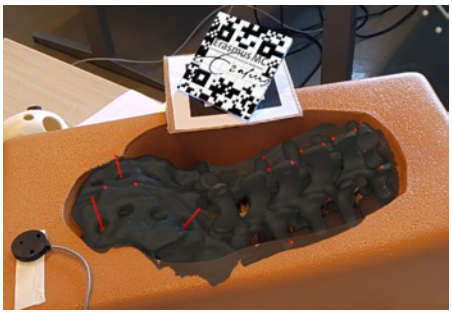
1. **Method 1 - "3P" (3 Point Picking)**: This method involved user interaction with the client PC monitor, where both the target point cloud and the source point cloud were displayed consecutively. The user had to manually select three corresponding points between the point clouds to initialize a ICP. To facilitate the selection of corresponding points, the user was initially presented with the reconstructed target point cloud. This approach was adopted due to the uncertainty regarding which features would be reconstructed most prominently, thus enabling easier identification and selection of corresponding points. Figure 10a shows a target point cloud with 3 picked points visualized as coloured spheres.
2. **Method 2 - "Manual"**: The world position of the hologram model coupled with this method was retrieved from the HoloLens app. This was used to crop the target point cloud and as the initialization pose for ICP. The user was therefore required to place the hologram of the preoperative model coarsely aligned with the phantom before initiating the registration process. The hologram placement was performed from a single viewing angle and was limited to approximately 5 seconds. Figure 10b shows the user aligning the model with blue evaluation markers with the target phantom. The user had the capability to set the white mesh of the hologram to (semi-)transparent.

3. **Method 3 - "Eye"**: The eye gaze spatial input of the HoloLens was utilized in this method. Therefore the user had to look at the target phantom while observing several sides up until they gave the "align" voice-command. After the voice-command the last 100 spatial input data points, captured at a frequency of 30 Hz, were processed for this method. The eye gaze ray was cast onto the target point cloud, and the median of these raycasted points was calculated. This median was used to crop the target point cloud based on a bounding box of 1.2x the size of the phantom. The translation of the source point cloud to this median point was used as ICP initialisation. Figure 10c shows the target point cloud, coordinate frames that represent poses of the head tracking, green lines that represent the eye rays originating from these tracked head poses, and the red bounding box used to crop the target point cloud based on the median raycasted eye point displayed in purple.
4. **Method 4 - "Eye RANSAC"**: Similar to Method 3, this method utilized the eye gaze spatial input. During development of method 3 the success rate appeared to be low. The hypothesis for that was that the translational initialization method 3 provides is not sufficient enough. Therefore we developed method 4. Method 4 includes additional steps to potentially overcome the lack of rotational initialization: First, a fast global registration was applied with the coarse cropped target point cloud surrounding the median eye gaze point. The output of the fast global registration was then used to initiate a RANSAC global registration. The resulting pose from the RANSAC registration was used to crop the target point cloud a second time. This more finely cropped target point cloud, along with the RANSAC registration pose, eventually served as input for ICP. Figure 10c additionally shows a blue bounding box, which is the bounding box used to crop the target point cloud once more after the RANSAC global registration.

3.2.2 Registration evaluation metrics

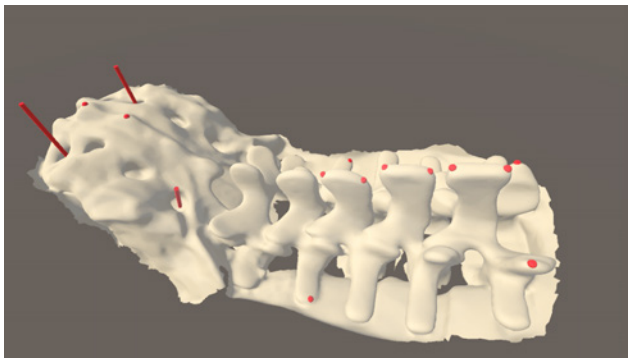
To assess the performance of the registration methods, a ground truth was established using the multi-modal method described in Benmahdjoub et al. (2022).[43] For detailed technical specifications and implementation details, please refer to the full text of Benmahdjoub et al. (2022).[43] The key aspects of this method for clarity and context are: 1) landmark-based registration of the preoperative model to the spine phantom using the NDI Aurora system and an attached EM tracker 2) a Vuforia Augmented Reality SDK image target attached and calibrated to an EM tracker 3) a Unity server streaming the NDI Aurora positions relative to the Vuforia image target to the Unity app in the HoloLens. By building the Unity app on the HoloLens with the Vuforia engine as plugin, the HoloLens can track the Vuforia image target. This results in the HoloLens being able to display all holograms superimposed on the objects tracked by the NDI system, as can be seen in Figure 11.

This established ground truth, known to have achieved a target registration error (TRE) of 2 mm in previous studies [44, 45], provided a benchmark for evaluating the performance of our point cloud based registrations. The registration error matrix R_{error} was used to calculate the rotation error. Of the rotation matrix from R_{error} the magnitude was calculated to represent the error of rotation as a single angle. The



(a) Ground truth registration

(b) Ground truth registration, with only red evaluation markers visualized



(c) Preoperative model of spine with embedded evaluation markers

Fig. 11: Ground truth multi-modal registration with superimposed holograms on the Sawbones spine phantom.

translation error was calculated by the distance d between two points.

$$R_{\text{error}} = R_{\text{ndi}} \cdot R_{\text{depth}}^{-1}$$

$$d = |R_{\text{ndi}} CT_{\text{center}} - R_{\text{depth}} CT_{\text{center}}|$$

The two points were defined by transforming the geometric center of the preoperative model from CT space to world space (${}^{\text{CT}}T_{\text{WCS}}$) twice. Once using the ground truth registration matrix R_{ndi} and once using the depth-based registration matrix R_{depth}. If the rotation error (RE) exceeded 20 degrees, the registration was marked and reported as failed. Additionally, the registration time of each method was measured between the "align" voice-command and visualization of the result.

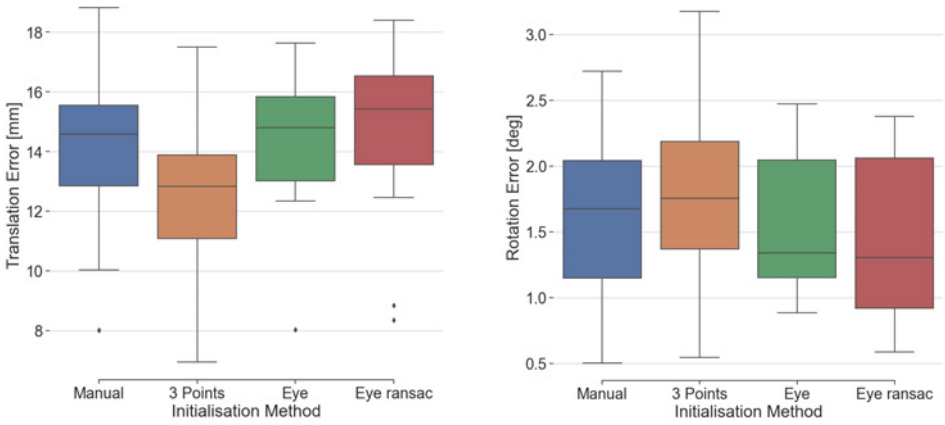
3.2.3 Results

We executed 30 registrations with the phantom model in different positions. The quantitative results for each registration method are summarized in Table 4. The mean translation errors of the four initialization methods ranged between 12.6 and 14.7 mm, rotation errors between 1.5 and 1.8 degrees. The minimum translation and rotation error observed was 6.9 mm and 0.5 degrees, respectively, while the maximum translation and rotation error was 18.8 mm and 3.2 degrees, respectively. Methods "3P" and "Manual" consistently provided successful registration (100%). The "Eye" and "Eye RANSAC" initialization methods did not always succeed, with "Eye" 23.3% and "Eye RANSAC" 50% of times. Method "3P" was the slowest method with a mean time of 34.5 seconds. The other three methods all achieved registration within 5 seconds. Box plots depicting the translation error (TE) and rotation error (RE) for each method can be seen in Figure 12. Additionally, Figure 13a presents box plots illustrating the registration time for each method.

Figure 13b illustrates the relationship between the magnitude of ground truth rotation and the success or failure of registrations for each method. It can be observed that the "Eye" method had a restricted number of successful registrations, which were achieved only when the orientation initialization fell within a specific range (130-180°). Similarly, the "Eye RANSAC" method also had a limited number of successful registrations. However, Figure 13b demonstrates that "Eye RANSAC" achieved 8 additional successful registrations across a wider range of orientation initializations.

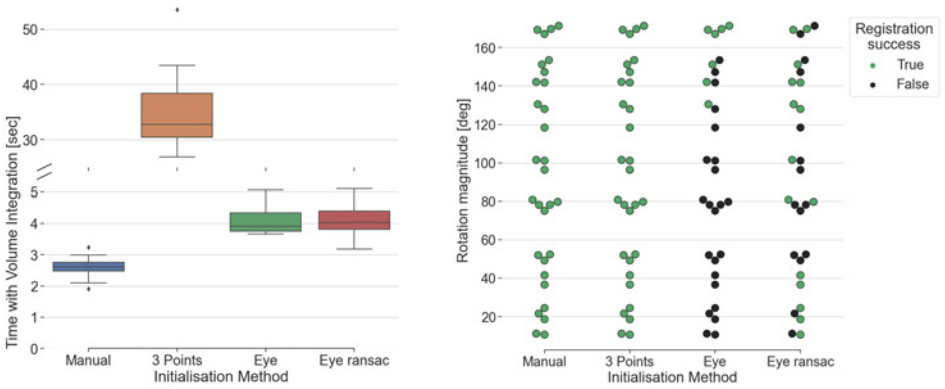
Table 4: Results for 30 registrations

Method	**Fail**	**TE [mm]**	**RE [deg]**	**Time [sec]**
Manual	0	14.06 ± 2.54	1.64 ± 0.6	2.60 ± 0.3
3P	0	12.59 ± 2.52	1.80 ± 0.7	34.55 ± 6.0
Eye	23	14.02 ± 3.13	1.58 ± 0.6	4.12 ± 0.5
Eye RANSAC	15	14.69 ± 3.01	1.45 ± 0.6	4.08 ± 0.5



(a) Translation error (b) Rotation error

Fig. 12: Registration errors of depth-based methods vs multi-modal ground truth



(a) Time per initialization method (b) Success vs ground truth rotation

Fig. 13: Registration time and success per initialization method

Qualitatively the registrations were assessed in renderings using Open3D and in mixed-reality (MR) using the MR capture of the HoloLens. As can be seen in Figure 11c the preoperative model was extended with several markers to help assist qualitative evaluation in MR. Three cylinders were added that pierced some of the dorsal sacral foramina, and multiple spheres were added to the tips of spinous processes. By not having to visualize the entire preoperative model, which causes overlapping and occlusion problems, these evaluation markers aided visualization of small translational and rotational differences between the ground truth and depth-based registration methods. In order to be distinctive, the four methods and the ground truth were assigned a

different colour for visualization. Table 5 shows the colour code for each method and the ground truth.

Table 5: Colour code for models and evaluation markers

Method	**Colour**
Ground truth	Red
Manual	Blue
3P	Green
Eye	Yellow
Eye RANSAC	Orange

The next figures provide insights into the qualitative registration results obtained using different depth-based methods. Figure 14 and Figure 15 show examples of registration results in MR, while Figure 16 and Figure 17 display registration results in the rendering.

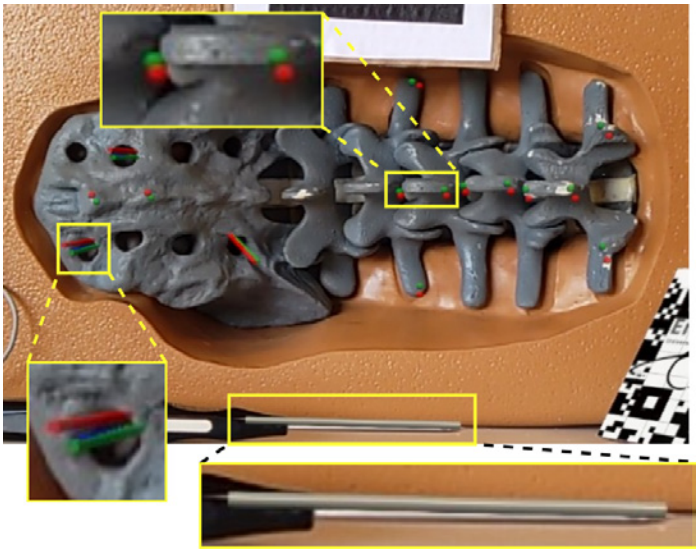
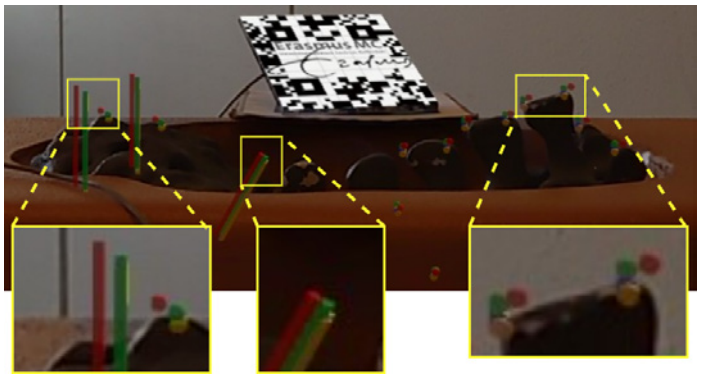


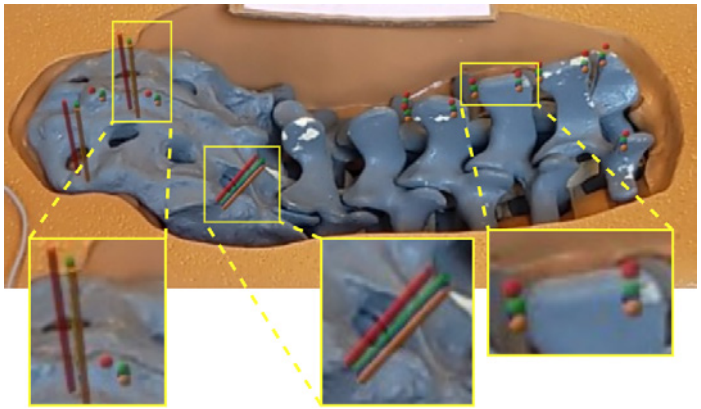
Fig. 14: Mixed reality photo of registration result. The tracked NDI EM pointer (grey cylinder) is linked to the ground truth.

In Figure 14, registration results from depth-based methods are shown with the ground truth. The zoomed-in areas display the hardly visible blue markers (method "manual"), green markers (method "3P"), and red markers (ground truth). It can be observed that the blue and green markers are partially overlapping, though there is an observable shift comparing them to the red ground truth. However, it is not

immediately clear if the observed ground truth in Figure 14 represents a qualitatively better registration compared to the depth-based methods. It is important to note that the EM-tracked NDI pointer is associated with the ground truth visualization of the multi-modal method. The shift between the superimposed grey cylinder of the NDI pointer and the real NDI pointer indicates that the ground truth itself may not always be perfect, as evidenced by the misalignment between the two.



(a) Side view of a result visualized with evaluation markers

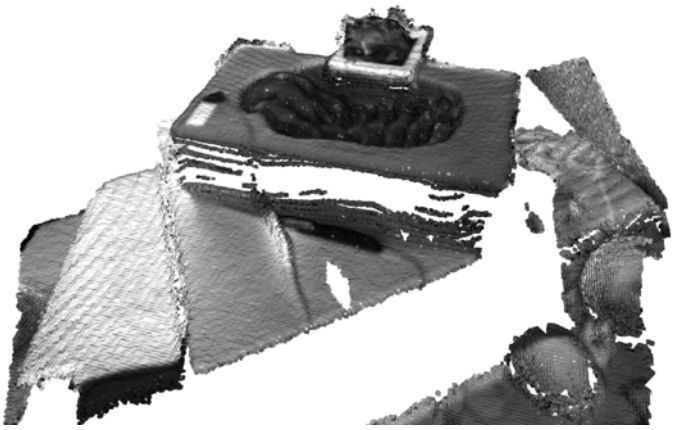


(b) Angle view of a result visualized with evaluation markers

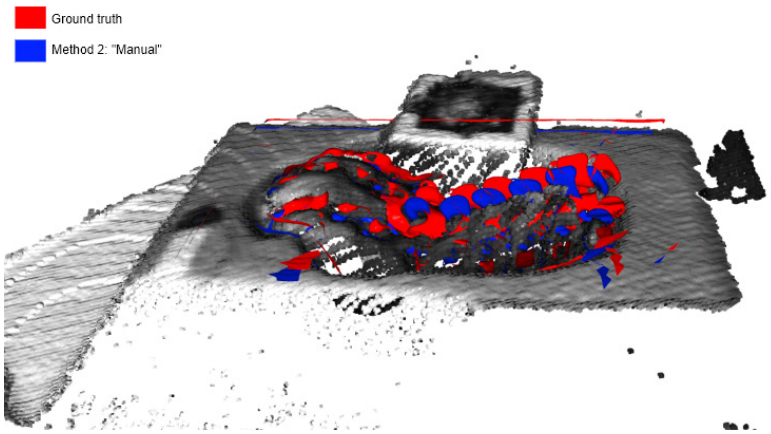
Fig. 15: Mixed reality photos of registration results with evaluation markers.

Figures 15a and 15b provide further comparisons between the registration results obtained from depth-based methods and the ground truth. In addition to the green and blue markers representing the "3P" and "Manual" methods, Figure 15a and 15b also include the orange markers representing the "Eye RANSAC" method. Observing these figures, it can be noted that the orange markers of the "Eye RANSAC" method are slightly more shifted compared to the partially overlapping green and blue markers. However, all successful depth-based registrations remain closer to each other than to

the ground truth. Specifically, when examining the piercing cylinders in Figure 15b and the spheres on the spinous process in Figure 15a, the ground truth appears to be shifted upward but qualitatively slightly better than the depth-based registrations.



(a) Example of a reconstructed target point cloud

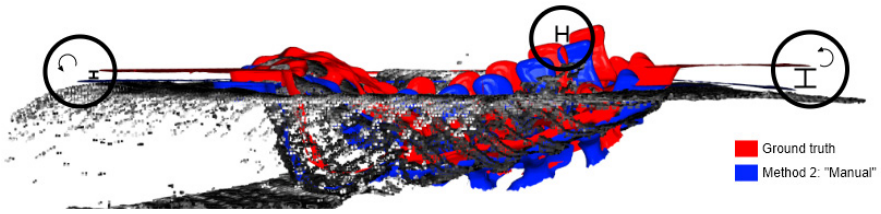


(b) Rendering of ground truth (red) and manual method (blue) result

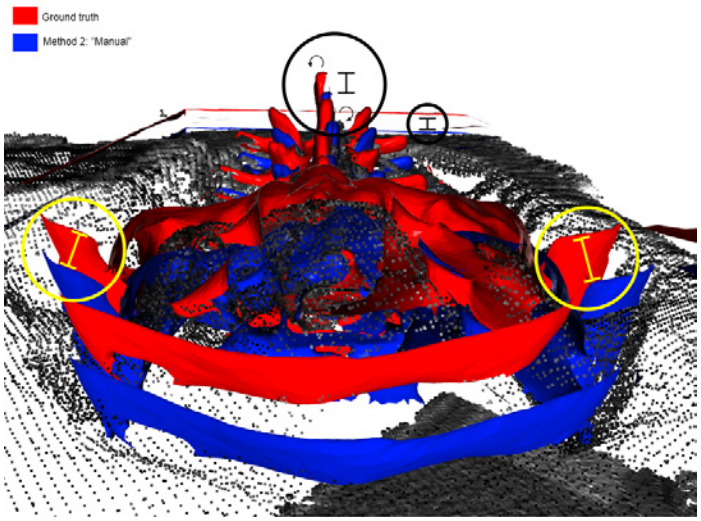
Fig. 16: Overview of the virtual scene and registration

Figure 16a displays the reconstructed target point cloud after TSDF volume integration, with the grey-scale values of the left front camera mapped onto it. The target phantom is clearly visible, along with its geometric features such as the spinous processes and the sacrum. Figure 16b provides an overview of two preoperative models placed on the target point cloud, based on the saved world poses of the ground truth (red) and the "Manual" method (blue). In Figure 17, the same scene is shown with

annotated details. Examining Figure 17a, it can be observed that the blue model (representing the "Manual" method) is accurately aligned with the target point cloud. In contrast, the ground truth model appears to be slightly rotated and mostly translated upwards. Figure 17b presents a different perspective and highlights the annotated translation and rotation errors. While these figures represent a single example, analyzing all registration results revealed that the observed translation errors, characterized by a shift downwards compared to the ground truth, were systematic in nature.



(a) Side view showing rotation and translation error



(b) Zoomed-in result showing translation and rotation error

Fig. 17: Specific visual errors in the rendered results

3.3 Qualitative assessment

Based on the results of the registration method experiment, the "Manual" method and the "Eye RANSAC" method were chosen for further qualitative analysis in this experiment. The "Manual" method was selected due to its robustness in providing successful registrations and its user-friendly workflow. Likewise, the "Eye RANSAC" method was chosen because it was the most successful automatic method.

The objective of this experiment was to evaluate the performance of the "Manual" and "Eye RANSAC" methods on different objects with diverse geometric features and material properties. Preoperative models of a set of objects and phantoms were obtained and loaded into the HoloLens app. The same implementation details as in the registration methods experiment were used to test the "Manual" and "Eye RANSAC" methods on these objects.

Since there was no quantitative ground truth available for comparison, the analysis in this experiment relied on qualitative assessments of the mixed-reality photos and renderings. The focus was on evaluating the visual outcomes and assessing the effectiveness of the registration methods in overlaying the preoperative models onto the real-world objects and phantoms, providing insights into the usability and effectiveness of the methods in various clinical scenarios.

3.3.1 Objects & Models

The objects selected for the qualitative phantom assessment experiment were chosen based on their diverse geometric features and material properties. These objects included:

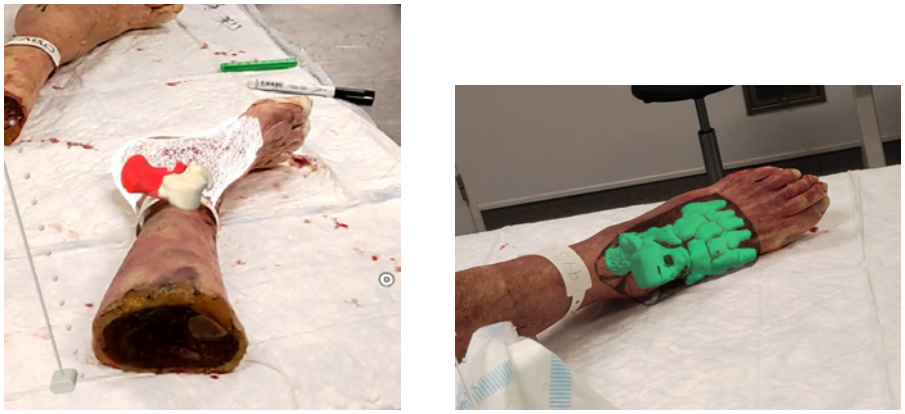
- Lower leg cadaver: A lower leg specimen obtained from a cadaver.
- Triple Modality 3D Abdominal Phantom: CIRS model 057A
- Skull phantom: A 3B Scientific skull phantom
- Lego cube: A simple geometric object, a cube made of Lego bricks.
- Mannequin: A mannequin of the upper body made of black reflective plastic.

CT scans were used for the lower leg cadaver, abdominal phantom, skull phantom, and Lego cube. The CT scan data allowed for the creation of detailed preoperative models, including internal structures in the case of the cadaver foot and abdominal phantom. In the case of the skull phantom, an evaluation marker and a tumor-shaped model were added to the preoperative model for illustrative purposes. For the mannequin, a detailed surface mesh was obtained using a 3D scanner (3dMD).

Fig. 18: Objects used for the qualitative assessment (left to right: lower leg cadaver, abdominal phantom, skull phantom, Lego cube, Mannequin)

3.3.2 Alignment Results

In Figure 19, two lower leg cadaver cadavers with superimposed holograms can be observed after a successful registration using the "Eye RANSAC" method. The registration process required multiple attempts before achieving a satisfactory alignment, but the desired result was obtained within approximately thirty seconds. Upon closer inspection of Figures 19a and 19b, a noticeable shift in alignment can be observed, this shift was particularly present in the direction in which the feet were observed during acquisition.



(a) Superimposed hologram with wireframe skin, red calcaneus and white talus model

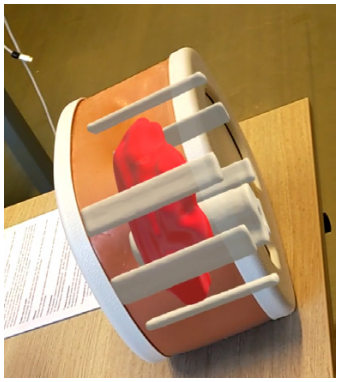
(b) Superimposed hologram with wireframe skin and green bone models

Fig. 19: Cadaver feet with aligned preoperative models.

In Figure 20, the abdominal phantom is depicted along with its aligned preoperative model. Both the "Manual" and "Eye RANSAC" methods were tested on this phantom, and both methods achieved successful alignment results in under 10 seconds. In Figure 20a a small shift can be observed on the left side of the phantom. Figure 20b shows the perspective of the user once the wireframe mesh model is deactivated and only the internal structures of the phantom are visualized.

Figure 21 shows a visually satisfactory alignment of the preoperative skull model with the skull phantom. This alignment was achieved using the "Manual" method, as the "Eye RANSAC" method did not yield consistent successful results. However, even with the "Manual" method, it was necessary to execute the alignment process multiple times before achieving an accurate alignment. Additionally, particular attention needed to be paid to the anterior features of the skull during data acquisition. The limitations of the "Eye RANSAC" method in achieving consistent results highlight the complexities involved in aligning objects with intricate geometric features, such as the skull phantom.

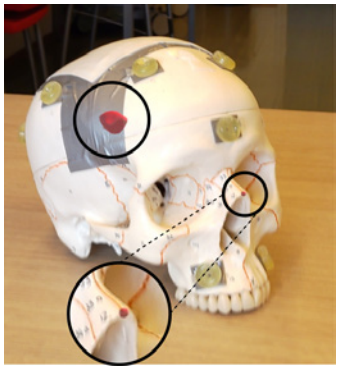
(a) Superimposed hologram with wireframe skin model active



(b) Superimposed holograms of liver and bone model

Fig. 20: Triple modality abdominal phantom with aligned preoperative models.

(a) Adequate alignment of the wireframe preoperative model

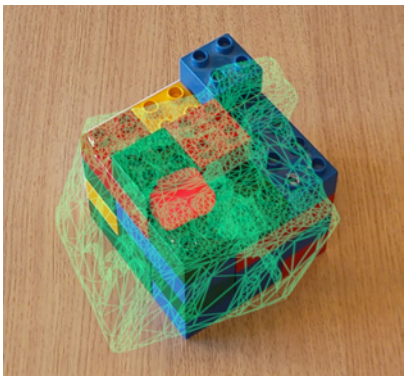


(b) Red evaluation marker on os nasale and illustrative tumour hologram

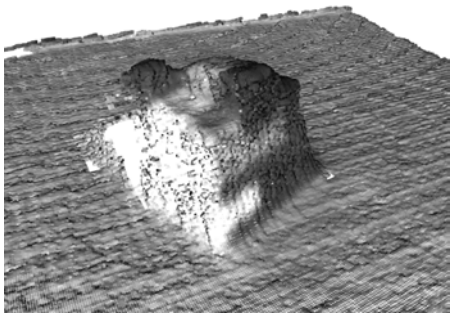
Fig. 21: Skull phantom results

In the test involving the Lego cube, no successful alignments were achieved. Figure 22a illustrates a failed result obtained using the "Manual" method. Further investigation of the reconstructed target point cloud, as shown in Figure 22b, revealed potential limitations contributing to the unsuccessful alignments. The surfaces of the Lego cube, which are composed of Acrylonitrile Butadiene Styrene (ABS) with a reflective finish, exhibited a lower density of points in the target point cloud, particularly in the reflective areas. Furthermore, the sharp edges and intricate geometric features of the Lego cube were not accurately represented in the reconstructed target point cloud.

In the case of the upper body mannequin, the "Eye RANSAC" method was tested to assess the influence of the larger surface area available for registration. The mannequin had similar material properties to the Lego cube. The "Eye RANSAC" method



(a) Failed alignment



(b) Sparse reconstructed target point cloud with rounded edges

Fig. 22: Lego cube results

often failed to provide an alignment. Only under specific lighting conditions that minimized reflections, a result similar to Figure 23a was obtained. Using the "Manual" method, all registration attempts failed. Upon visually inspecting the reconstructed target point clouds and analyzing the raw depth frames, it was evident that very few depth values could be estimated from the surface of the mannequin (Figure 23b). In cases where alignment was achieved using the "Eye RANSAC" method, the sparse points that were estimated were also found to be inaccurate. This is evident from the significant shift of the preoperative model inside the mannequin, as shown in Figure 23a.

(a) Poorly aligned mannequin hologram

(b) Sparse target point cloud with registered source point cloud

Fig. 23: Registrations of mannequin chest model with dark reflective material

4 Discussion

In our evaluation of the AHAT sensor of the HoloLens 2, which has similar specifications as the AHAT sensor of the HoloLens 1, we observed a trend of overestimated depth points, consistent with the findings of Gu et al. (2021) [22]. Our distance error values for the same material type and HoloLens-to-surface distance range corresponded with theirs, with a mean error of 7.18 ± 4.19 mm at 70 cm. These values were slightly higher than the mean error reported by Gu et al. (5.68 ± 2.21 mm at 70 cm). There could be several potential reasons for this discrepancy. Firstly, we evaluated a significantly larger number of frames (around 1300 frames) compared to Gu et al.'s evaluation of 50 frames. This larger sample size may have contributed to a higher variation in our results. Additionally, we observed that the mean depth estimation error improved by approximately 2 mm under bright operating room lighting conditions, indicating that lighting conditions play a significant role in the accuracy of depth estimation.

We also evaluated the LT sensor of the HoloLens 2 and found that it exhibited higher mean errors of overestimation compared to the AHAT sensor. However, we noticed that this difference in mean errors decreased when the HoloLens-to-surface distance was in the range of 60 to 70 cm. Interestingly, the LT sensor also showed significantly fewer outliers compared to the AHAT sensor. This suggests that the systematic error of overestimation exhibited by the LT sensor might be easier to calibrate compared to the AHAT sensor.

In our experiment, we used the ArUco marker as a reference plane, which may have resulted in an underestimation of the distance between the HoloLens and the surfaces. Thabit et al. (2022).[9] describes that the ArUco tracking error increases with distance, lower resolution and smaller markers sizes. Kalaitzakis et al. (2021)[46] describes a similar trend of decreasing accuracy with lower resolution and larger distances. They found that as the tracking error increases, there is a tendency for the distance from the camera to the ArUco marker to be underestimated. Therefore, to accurately quantify the exact amount of overestimation by the AHAT and LT sensors, a more accurate ground truth measurement would be required. Nonetheless, even considering the most unfavorable scenario of underestimation of the ArUco marker and overestimation of the point clouds, it can still be argued that both the AHAT and LT sensors tend to overestimate depth.

As mentioned in section 3.1.3 the ArUco QR code detection sometimes failed during the setup in OR lighting conditions. The large difference in brightness from the center to the side in the RGB frames looked like the cause of this, as seen in Figure 6b. This is likely to have been the cause of the large standard deviation of the QR code pose estimation during one of the acquisitions.

Our experimentation also observed the significant influence of material type on the accuracy and validity of depth estimation, which aligns with the findings reported by Gu et al. (2021)[22]. Quantifying the overestimation and performing calibration to mitigate the systematic errors associated with material types will be a challenging task. Therefore, it is crucial to conduct quantitative assessments of depth estimation accuracy and validity for different material types. Furthermore, exploring alternative methods of point cloud processing or scene reconstruction may contribute to reducing

the overestimation error. If the depth estimation error is predominantly in the direction that the sensor is observing, a potential approach could be to observe the scene from multiple angles and employ a scene reconstruction method that compensates for the overestimation, thereby reducing the directional error component.

In the registration method experiment, we performed 30 registrations with a spine phantom model in various positions. The mean translation errors of the four initialization methods ranged from 12.6 to 14.7 mm, while rotation errors ranged from 1.5 to 1.8 degrees. In comparing our registration methods with previous studies, such as Gsanxer et al. (2019) and Haxthauser et al. (2021), who also developed markerless depth-based methods with the HoloLens for image-to-patient registration, we found that our mean accuracy of 12.6 to 14.7 mm is in the same order of magnitude. These results are promising, considering that our system still has room for optimization. For example, if the depth overestimation could be calibrated for a systematic error the performance could increase. This all suggests that our methods are on par with existing approaches and have the potential for further improvement.

The successful registration results obtained for all 30 cases using the two manual registration methods, "3P" and "Manual," indicate their robustness. However, there was a quantitative difference in translation accuracy between the two methods, with "3P" performing approximately 1.5 mm better ($p=0.028$). This difference could be attributed to the advantage of users being able to pick corresponding points on the point cloud at the PC monitor, as it allows for higher precision in all three axes. And thus contributing to the accurate initialization of the ICP algorithm. On the other hand, the "Manual" method relied on a rough alignment from a single viewing perspective, which may have limited the axial precision of the ICP initialization. This limitation could explain the better performance of the "3P" method in terms of translation accuracy. While the "3P" method offers better quantitative results, it also has drawbacks such as the additional time required for users to walk back to the PC and the need for consistent selection of corresponding points. In contrast, the "Manual" method provides a more user-friendly workflow by allowing rough alignment from a single perspective with the HoloLens user interface. This would also be more favorable in a sterile environment.

The automatic registration methods, "Eye" and "Eye RANSAC," faced challenges in achieving accurate registrations. The "Eye" method had a limited success rate of only 23%, and successful alignments were achieved only when the orientation initialization fell within a specific range. Figure 13b illustrates that the successful alignments of the "Eye" method were concentrated in the orientation initialization range of 130-180°. To address these limitations, additional steps were introduced in the "Eye RANSAC" method to improve rotational initialization. This resulted in a significantly higher success rate of 50% compared to the "Eye" method. Figure 13b demonstrates that the successful alignments of the "Eye RANSAC" method were more distributed but still concentrated in the ranges of 130-180° and 0-40°. This indicates that accurately initializing rotation remains a challenge for the automatic methods, even though translation also plays a role in initialization accuracy. Further assessment could explore the range of translation and rotation initialization that different subsequent ICP algorithms can

handle, although the geometry of the target object would significantly influence the outcomes and make general conclusions difficult.

Despite these challenges, when the automatic methods were successful, both qualitatively and quantitatively, the results were comparable to those of the manual methods. This suggests that with a suitable initialization, the automatic methods can reach an accurate optimum for the registration task. However, it is evident that there is room for improvement in the automatic initialization methods, especially when it comes to addressing the challenge of providing appropriate orientation initialization for symmetrical target geometries. These findings emphasize the importance of refining the automatic initialization methods to enhance their reliability in achieving successful registrations for a wider range of applications. Further research and development efforts should focus on improving the rotational initialization process to overcome the limitations observed in the current methods.

The depth-based methods investigated in the depth registration experiment have shown advantages over point-based marker registration in terms of speed and potential accuracy. By eliminating the need for manual landmark selection, depth-based methods can reduce user errors and minimize sensitivity to outliers. Among the four depth-based methods tested, three of them were able to perform registration in less than 5 seconds, with most of the time being spent on the TSDF volume integration step. It's worth noting that the TSDF volume integration was implemented using Open3D on the CPU. However, running the code on a GPU with CUDA could significantly reduce computation time, further improving the overall speed of the depth-based registration process. Moreover, depth-based methods have the potential to achieve higher accuracy, provided that the target point cloud is reliable and occlusion challenges are overcome. This emphasizes the importance of optimizing the target point cloud reconstruction process to ensure better alignment between the preoperative model and the physical object.

Although marker-less systems have some clear advantages, our system relies on the HoloLens' SLAM after registration. SLAM drift of the HoloLens 1 has been described before and can be of the centimeter scale.[47] The HoloLens 2 is assumed to have improved on the world origin drift and is presumably millimeter scale. Especially with smaller ranges of motion from the headset and shorter usage time. However, even at this scale, SLAM drift remains a significant challenge that needs to be addressed. To mitigate the impact of drift in our system, one possible approach is to explore the implementation of periodic re-registration, leveraging the speed of the registration process and the high-quality initialization provided by the previous registration. Additionally, ongoing research is actively investigating alternative methods to correct for SLAM drifting, and incorporating these advancements into future systems will be essential for addressing this challenge effectively.[48]

Despite the potential shown in our results, there are several limitations and areas for future work that need to be addressed. One limitation is the assumption of the multi-modal system using NDI EM tracking and Vuforia SDK as the ground truth for registration. It is possible that the distance underestimation for the 3D pose of 2D fiducial markers could contribute more significantly to our translation and rotation errors than initially anticipated. From all three experiments our findings suggest the

target point clouds from the HoloLens have a systematic depth overestimation. Translation errors could be further amplified by the underestimation of the 2D Vuforia image target resulting in higher translation errors. Therefore it could be that the depth-based registration is performing better then the translation errors reported using this multi-modal ground truth.

In the registration experiment, we maintained a fixed reconstruction of the target point cloud by sending the last 50 LT depth frames and using the same TSDF parameters. However, it is important to note that this approach may not consistently yield optimal results, necessitating further optimization. The qualitative assessment revealed that the quality and density of the target point cloud significantly influence the registration process. Objects with sharp edges and small topography features, such as the Lego cube, suffered from a sparse and suboptimal target point cloud, leading to decreased registration success. Insufficiently reconstructed target point clouds fail to capture the intricate features of these objects, resulting in misalignments between the preoperative model and the physical object. Similarly, the surface properties of the mannequin, characterized by glossiness and low infrared reflectivity, posed challenges by producing sparse and inaccurate depth measurements, also hindering the registration process and causing misalignments. Consequently, the success of the registration heavily relies on the accuracy and density of the depth measurements obtained from the target object's surface. To address these issues, future work should focus on improving the quality and density of the target point cloud. This could involve exploring alternative depth sensing techniques, optimizing the reconstruction algorithms, or employing additional sensors to enhance the accuracy of the depth measurements.

A notable limitation that applies to the general use of AR HMDs in surgery is the challenge of accurately assessing the quality of alignment. In our qualitative testing, we employed conventional mixed reality visualization methods, such as transparency or wireframe rendering, to assist with alignment assessment. A study conducted by Gu et al. (2022)[49] reported a user preference for wireframe rendering in alignment assessment. However, their findings also indicated that participants were not consistently effective in identifying misalignments using these visualization methods. This raises concerns that if an image-to-patient registration method were to yield a slightly misaligned result with an AR HMD, users may not perceive the misalignment and continue with the procedure, potentially impacting the overall outcome. Therefore, further research is warranted to develop more reliable and intuitive visualization techniques that enable users to assess alignment with higher accuracy and confidence.

Future work should include assessing the accuracy of the system at the user endpoint, first validating in phantom and cadaver studies and then assessing outcomes in surgical procedures. While our experiments focused on evaluating the accuracy of the registration methods and depth measurements, it is important to consider the impact of the system on the surgical outcomes. Surgeons may be able to achieve higher accuracy and precision using this system compared to their conventional methods. Therefore, conducting clinical studies and evaluating the system's performance in actual surgical scenarios is essential to determine its true effectiveness and potential benefits in improving surgical outcomes. This would provide valuable insights into

the practical utility of the system and guide further refinements and optimizations for specific procedures.

5 Conclusion

In conclusion, this study presented a comprehensive evaluation of a depth-based image-to-patient registration system using the HoloLens 2. Through three experiments, we obtained valuable insights into the accuracy, success rates, and limitations of various registration methods. The quantitative assessment of depth accuracy revealed millimeter-scale mean overestimation errors for both the AHAT and LT sensors in surgically relevant scenarios. Various factors had positive or negative impact on this, including variations in operating room lighting conditions, distances between the sensor and the target, and the material type. This highlights the need for further investigation into refining the depth measurements of the HoloLens' sensors. The comparison of registration initialization methods using the spine phantom demonstrated the potential of manual approaches to consistently achieve successful alignments, while the automatic methods showed promise but require enhancements, particularly in accurate initialization for subsequent ICP steps. The qualitative assessment of the system's performance in real-world scenarios showcased its potential for improving surgical procedures, albeit with limitations observed for small-sized and symmetrical objects.

Future work should prioritize the refinement of both manual and automatic initialization methods to enhance their reliability, particularly in complex surgical scenarios. Additionally, improving the overall accuracy of the system is crucial, which can be achieved through calibrating for systematic errors based on quantitative experiments. It is imperative to validate the surgical accuracy on the user end by performing phantom surgical tasks with and without the system. This can be coupled with exploring techniques to provide users with more intuitive and reliable visualization methods for assessing alignment quality.

The study's results indicate the potential for achieving sub-10 mm registration accuracy within 5 seconds using depth-based image-to-patient registration. This is particularly promising for surgical applications that do not require high-level registration accuracy, and offers a fast and convenient alternative to other tracking systems that require invasive fiducial markers and time-consuming calibration steps. Conclusively, the goal of our research is to enable enhanced surgical outcomes, including improved tumor resection margins and avoidance of vulnerable tissues. By providing faster and more accessible image guidance, depth-based registration systems have the potential to revolutionize a field of surgical procedures that currently don't use image guidance and contribute to better patient outcomes.

Supplementary information.

- Mixed-reality video capture from the HoloLens 2

Demo using the depth-based registration method with manual initialization for the Sawbones spine phantom, and the automatic eye gaze method for the multi-modal abdominal phantom CIRS 57A.

Appendix A Depth accuracy supplement

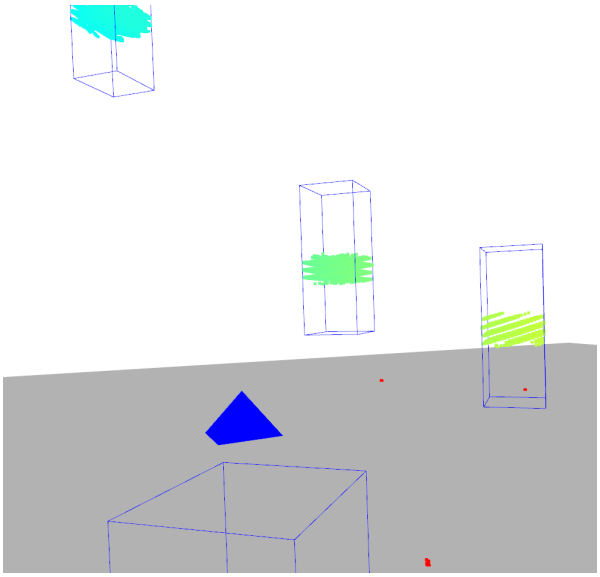


Fig. A1: Side view of point cloud with cropped surfaces of AHAT acquisition: Spherical layering artifact visible.

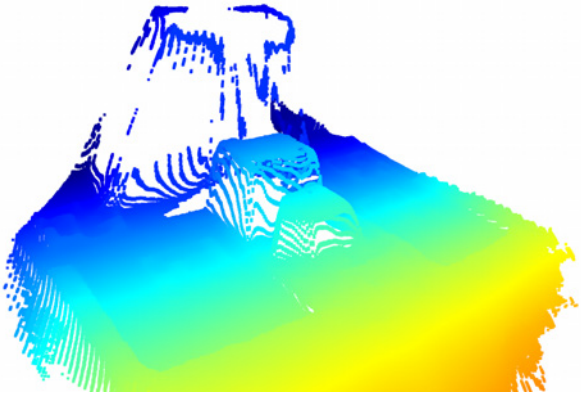


Fig. A2: Point cloud of LT 50 cm acquisition: No depth values for a large surface on the box object, some "waterfall" effects around the table and objects.

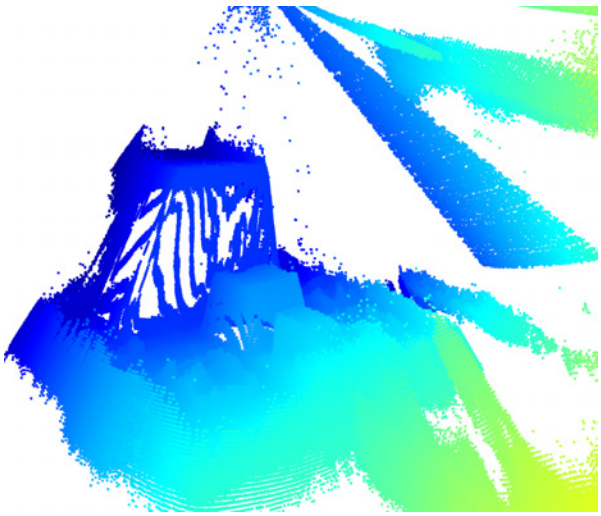


Fig. A3: Point cloud of AHAT 50 cm acquisition: Large "waterfall" effects around the table and objects, additional noise in the air where no objects were present.

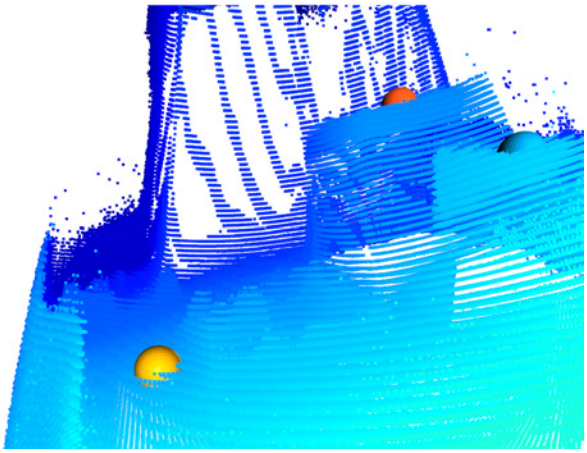


Fig. A4: Point cloud of AHAT acquisition: Right of the yellow picked point the table the surface is less smooth and the interruption is the same shape of the QR code placed there.

Table A1: Overview of 30 seconds depth acquisitions

HoloLens	Sensor	Distance [cm]	OR light	Depth frames	RGB frames	QR-code poses	QR max std X [mm]	QR max std Y [mm]	QR max std Z [mm]
1	AHAT	50	on	1335	842	3368	0.181	0.739	0.156
1	AHAT	70	on	1344	840	3360	0.142	0.403	0.113
1	AHAT	50	off	1362	842	3368	0.046	0.183	0.057
1	AHAT	70	off	1340	844	3376	0.574	0.431	0.655
1	LT	50	on	147	858	3432	0.322	0.746	0.442
1	LT	70	on	147	832	3328	0.127	0.560	0.125
1	LT	50	off	148	771	3084	0.139	0.245	0.062
1	LT	70	off	146	637	2548	0.565	0.486	0.727
2	AHAT	50	on	1340	738	2952	0.121	0.370	0.100
2	AHAT	50	off	1343	765	3060	0.066	0.111	0.084
2	AHAT	70	off	1341	847	3388	0.067	0.161	0.066
2	LT	50	on	147	597	2388	0.080	0.207	0.065
2	LT	70	on	147	788	3152	0.177	0.420	0.093
2	LT	50	off	148	672	2688	0.062	0.127	0.058
2	LT	70	off	147	754	3016	0.076	0.196	0.071

Appendix B Registration errors without pass/fail filtering

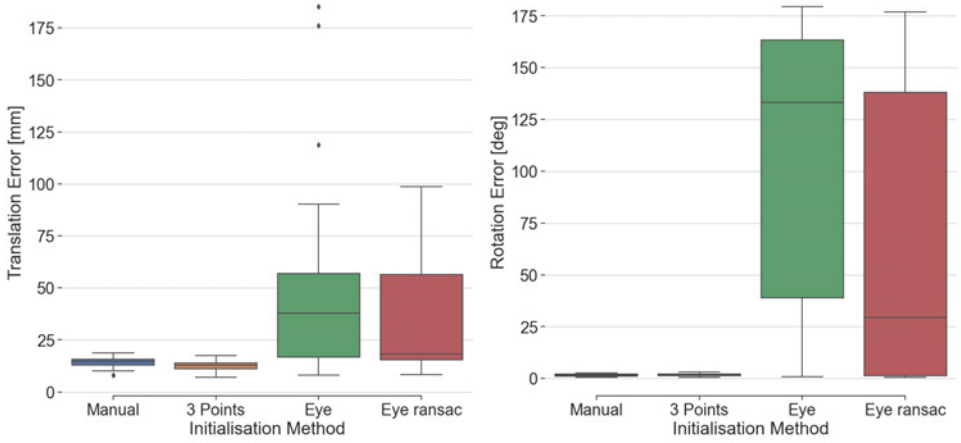


Fig. B5: Translation and rotation error of depth registration methods vs multi-modal ground truth

References

[1] Gsaxner C, Li J, Pepe A, Jin Y, Kleesiek J, Schmalstieg D, et al. The HoloLens in medicine: A systematic review and taxonomy [Journal Article]. Medical Image Analysis. 2023;85:102757. https://doi.org/https://doi.org/10.1016/j.media.2023.102757.

[2] Andrews CM, Henry AB, Soriano IM, Southworth MK, Silva JR. Registration Techniques for Clinical Applications of Three-Dimensional Augmented Reality Devices [Journal Article]. IEEE J Transl Eng Health Med-JTEHM. 2021;9.

[3] Perez-Pachon L, Poyade M, Lowe T, Groning F. Image Overlay Surgery Based on Augmented Reality: A Systematic Review [Journal Article]. Adv Exp Med Biol. 2020;1260:175–195. https://doi.org/10.1007/978-3-030-47483-6_10.

[4] Mauro MA, Murphy KPJ, Thomson KR, Venbrux AC, Morgan RA. Image-guided interventions e-book: expert radiology series. Elsevier Health Sciences; 2020.

[5] Casari FA, Navab N, Hruby LA, Kriechling P, Nakamura R, Tori R, et al. Augmented Reality in Orthopedic Surgery Is Emerging from Proof of Concept Towards Clinical Studies: a Literature Review Explaining the Technology and Current State of the Art [Journal Article]. Curr Rev Musculoskelet Med. 2021;14(2):192–203.

[6] Barcali E, Iadanza E, Manetti L, Francia P, Nardi C, Bocchi L. Augmented Reality in Surgery: A Scoping Review [Journal Article]. Applied Sciences. 2022;12(14):6890.

[7] Gumprecht HK, Widenka DC, Lumenta CB. BrainLab VectorVision Neuronavigation System: technology and clinical experiences in 131 cases [Journal Article]. Neurosurgery. 1999;44(1):97–104; discussion 104–5. https://doi.org/10.1097/00006123-199901000-00056.

[8] Foster R, Bathla S, Harris O, Jois R, Kiernan T, Ray A, et al. Feasibility of breast lesion localisation using Sirius Pintuition; A new magnetic seed localisation system - Report of the first dozen cases in the UK [Journal Article]. European Journal of Surgical Oncology. 2022;48(5):e228. https://doi.org/https://doi.org/10.1016/j.ejso.2022.03.167.

[9] Thabit A, Niessen WJ, Wolvius EB, van Walsum T, Linte CA, Siewerdsen JH. Evaluation of marker tracking using mono and stereo vision in Microsoft HoloLens for surgical navigation [Journal Article]. Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling. 2022;https://doi.org/10.1117/12.2607262.

[10] Hoch A, Liebmann F, Carrillo F, Farshad M, Rahm S, Zingg PO, et al. Augmented reality based surgical navigation of the periacetabular osteotomy of ganz – a pilot cadaveric study [Journal Article]. Mechanisms and Machine Science. 2021;93:192–201. https://doi.org/10.1007/978-3-030-58104-6_22.

[11] Sebastian A, Alex Johnson MD, Mathias U, Alexander FW, Kevin Y, Javad F, et al. On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial [Journal Article]. Journal of Medical Imaging. 2018;5(2):021209. https://doi.org/10.1117/1.JMI.5.2.021209.

[12] Song T, Yang C, Dianat O, Azimi E. Endodontic guided treatment using augmented reality on a head-mounted display system [Journal Article]. Healthcare Technology Letters. 2018;5(5):201–207. https://doi.org/https://doi.org/10.1049/htl.2018.5062.

[13] Andong C, Ali D, Jianbo S, Terence PG, Brian JP. Image-based marker tracking and registration for intraoperative 3D image-guided interventions using augmented reality [Journal Article]. ProcSPIE. 2020;11318:1131802. https://doi.org/10.1117/12.2550415.

[14] Haxthausen FV, Chen Y, Ernst F. Superimposing holograms on real world objects using HoloLens 2 and its depth camera [Journal Article]. Current Directions in Biomedical Engineering. 2021;7(1). https://doi.org/10.1515/cdbme-2021-1024.

[15] Gsaxner C, Pepe A, Li J, Ibrahimpasic U, Wallner J, Schmalstieg D, et al. Augmented Reality for Head and Neck Carcinoma Imaging: Description and

Feasibility of an Instant Calibration, Markerless Approach [Journal Article]. Comput Methods Programs Biomed. 2021;200. https://doi.org/10.1016/j.cmpb.2020.105854.

[16] Sta S, Ogor J, Letissier H, Stindel E, Hamitouche C, Dardenne G. Towards markerless computer assisted surgery: Application to total knee arthroplasty [Journal Article]. Int J Med Robot. 2021;17(5):e2296. https://doi.org/10.1002/rcs.2296.

[17] Hu X, Rodriguez y Baena F, Cutolo F. Head-mounted augmented reality platform for markerless orthopaedic navigation [Journal Article]. IEEE J Biomedical Health Informat. 2021;https://doi.org/10.1109/jbhi.2021.3088442.

[18] Hu X, Nguyen A, Baena FRY. Occlusion-robust Visual Markerless Bone Tracking for Computer-Assisted Orthopaedic Surgery [Journal Article]. IEEE Transactions on Instrumentation and Measurement. 2021;https://doi.org/10.1109/TIM.2021.3134764.

[19] Liebmann F, Stutz D, Suter D, Jecklin S, Snedeker JG, Farshad M, et al. SpineDepth: A Multi-Modal Data Collection Approach for Automatic Labelling and Intraoperative Spinal Shape Reconstruction Based on RGB-D Data [Journal Article]. Journal of Imaging. 2021;https://doi.org/10.3390/jimaging7090164.

[20] Li J, Deng Z, Shen N, He Z, Feng L, Li Y, et al. A fully automatic surgical registration method for percutaneous abdominal puncture surgical navigation [Journal Article]. Comput Biol Med. 2021;136:104663. https://doi.org/10.1016/j.compbiomed.2021.104663.

[21] Kalfas IH. Machine Vision Navigation in Spine Surgery [Journal Article]. Front surg. 2021;8:640554. https://doi.org/10.3389/fsurg.2021.640554.

[22] Gu W, Shah K, Knopf J, Navab N, Unberath M. Feasibility of image-based augmented reality guidance of total shoulder arthroplasty using microsoft HoloLens 1 [Journal Article]. Comput Methods Biomech Biomed Eng Imaging and Visualization. 2021;9(3):261–270. https://doi.org/10.1080/21681163.2020.1835556.

[23] Félix I, Raposo C, Antunes M, Rodrigues P, Barreto JP. Towards markerless computer-aided surgery combining deep segmentation and geometric pose estimation: application in total knee arthroplasty [Journal Article]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization. 2021;9(3):271–278. https://doi.org/10.1080/21681163.2020.1835554.

[24] Fan X, Duke RB, Ji S, Mirza SK, Paulsen KD. Stereovision surface stitching for image updating in open spine surgery [Journal Article]. Progr Biomed Opt Imaging Proc SPIE. 2021;11598. https://doi.org/10.1117/12.2582111.

[25] Gsaxner C, Pepe A, Wallner J, Schmalstieg D, Egger J. Markerless Image-to-Face Registration for Untethered Augmented Reality in Head and Neck Surgery

[Journal Article]. MICCAI 2019. 2019;p. 236–244.

[26] Ungureanu D, Bogo F, Galliani S, Sama P, Duan X, Meekhof C, et al. Hololens 2 research mode as a tool for computer vision research [Journal Article]. arXiv preprint arXiv:11239. 2020;.

[27] Zhou QY, Koltun V. Dense scene reconstruction with points of interest [Journal Article]. ACM Transactions on Graphics. 2013;32(4):1–8.

[28] Zhou QY, Park J, Koltun V. Open3D: A modern library for 3D data processing [Journal Article]. arXiv preprint arXiv:09847. 2018;.

[29] Zollhöfer M, Stotko P, Görlitz A, Theobalt C, Nießner M, Klein R, et al. State of the Art on 3D Reconstruction with RGB-D Cameras [Journal Article]. Computer Graphics Forum. 2018;37(2):625–652. https://doi.org/https://doi.org/10.1111/cgf.13386.

[30] Myronenko A, Song X. Point set registration: Coherent point drift [Journal Article]. IEEE transactions on pattern analysis and machine intelligence. 2010;32(12):2262–2275.

[31] Yang J, Li H, Campbell D, Jia Y. Go-ICP: A globally optimal solution to 3D ICP point-set registration [Journal Article]. J IEEE transactions on pattern analysis machine intelligence. 2015;38(11):2241–2254.

[32] Raposo C, Barreto JP. Using 2 point+normal sets for fast registration of point clouds with small overlap [Journal Article]. 2017 IEEE International Conference on Robotics and Automation (ICRA). 2017;p. 5652–5658. https://doi.org/10.1109/ICRA.2017.7989664.

[33] Dylan Campbell LP. An Adaptive Data Representation for Robust Point-Set Registration and Merging [Journal Article]. 2015 IEEE International Conference on Computer Vision (ICCV). 2015;https://doi.org/10.1109/iccv.2015.488.

[34] Choi S, Zhou QY, Koltun V. Robust Reconstruction of Indoor Scenes [Journal Article]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015;p. 5556–5565. https://doi.org/10.1109/CVPR.2015.7299195.

[35] Zhou QY, Park J, Koltun V. Fast global registration [Journal Article]. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. 2016;p. 766–782.

[36] Rusu RB, Blodow N, Beetz M. Fast Point Feature Histograms (FPFH) for 3D registration [Journal Article]. 2009 IEEE International Conference on Robotics and Automation. 2009;p. 3212–3217. https://doi.org/10.1109/ROBOT.2009.5152473.

[37] Haas JK. A history of the unity game engine [Journal Article]. Diss Worcester Polytechnic Institute. 2014;483(2014):484.

[38] Dibene JC, Dunn E. HoloLens 2 Sensor Streaming [Journal Article]. arXiv preprint arXiv:221102648. 2022;.

[39] Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas F, Medina-Carnicer R. Generation of fiducial marker dictionaries using Mixed Integer Linear Programming [Journal Article]. Pattern Recognition. 2015;51. https://doi.org/10.1016/j.patcog.2015.09.023.

[40] Romero-Ramirez F, Muñoz-Salinas R, Medina-Carnicer R. Speeded Up Detection of Squared Fiducial Markers [Journal Article]. Image and Vision Computing. 2018;76. https://doi.org/10.1016/j.imavis.2018.05.004.

[41] Bradski G. The OpenCV Library [Journal Article]. Dr Dobb's Journal of Software Tools. 2000;.

[42] Bane Sullivan AK. PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK) [Journal Article]. Journal of Open Source Software. 2019;4(37):1450. https://doi.org/10.21105/joss.01450.

[43] Benmahdjoub M, Niessen WJ, Wolvius EB, Walsum Tv. Multimodal markers for technology-independent integration of augmented reality devices and surgical navigation systems [Journal Article]. Virtual Reality. 2022;26(4):1637–1650. https://doi.org/10.1007/s10055-022-00653-3.

[44] Thabit A, Benmahdjoub M, van Veelen MLC, Niessen WJ, Wolvius EB, van Walsum T. Augmented reality navigation for minimally invasive craniosynostosis surgery: a phantom study [Journal Article]. International Journal of Computer Assisted Radiology and Surgery. 2022;17(8):1453–1460. https://doi.org/10.1007/s11548-022-02634-y.

[45] Benmahdjoub M, Thabit A, Veelen MLCv, Niessen WJ, Wolvius EB, Walsum Tv. Evaluation of AR visualization approaches for catheter insertion into the ventricle cavity [Journal Article]. IEEE Transactions on Visualization and Computer Graphics. 2023;29(5):2434–2445. https://doi.org/10.1109/TVCG.2023.3247042.

[46] Kalaitzakis M, Cain B, Carroll S, Ambrosi A, Whitehead C, Vitzilaios N. Fiducial Markers for Pose Estimation [Journal Article]. Journal of Intelligent and Robotic Systems. 2021;101(4):71. https://doi.org/10.1007/s10846-020-01307-9.

[47] Hübner P, Clintworth K, Liu Q, Weinmann M, Wursthorn S. Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications [Journal Article]. Sensors. 2020;20(4):1021.

[48] Gu W, Knopf J, Cast J, Higgins LD, Knopf D, Unberath M. Nail it! vision-based drift correction for accurate mixed reality surgical guidance [Journal Article]. International Journal of Computer Assisted Radiology and Surgery. 2023;https://doi.org/10.1007/s11548-023-02950-x.

[49] Gu W, Martin-Gomez A, Cho SM, Osgood G, Bracke B, Josewski C, et al. The impact of visualization paradigms on the detectability of spatial misalignment in mixed reality surgical guidance [Journal Article]. Int J Comput Assist Radiol Surg. 2022;17(5):921–927. https://doi.org/10.1007/s11548-022-02602-6.