Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery

Liebmann, Florentin ; von Atzigen, Marco ; Stütz, Dominik ; Wolf, Julian ; Zingg, Lukas ; Suter, Daniel ; Cavalcanti, Nicola A ; Leoty, Laura ; Esfandiari, Hooman ; Snedeker, Jess G ; Oswald, Martin R ; Pollefeys, Marc ; Farshad, Mazda ; Fürnstahl, Philipp

Abstract: Established surgical navigation systems for pedicle screw placement have been proven to be accurate, but still reveal limitations in registration or surgical guidance. Registration of preoperative data to the intraoperative anatomy remains a time-consuming, error-prone task that includes exposure to harmful radiation. Surgical guidance through conventional displays has well-known drawbacks, as information cannot be presented in-situ and from the surgeon's perspective. Consequently, radiation-free and more automatic registration methods with subsequent surgeon-centric navigation feedback are desirable. In this work, we present a marker-less approach that automatically solves the registration problem for lumbar spinal fusion surgery in a radiation-free manner. A deep neural network was trained to segment the lumbar spine and simultaneously predict its orientation, yielding an initial pose for preoperative models, which then is refined for each vertebra individually and updated in real-time with GPU acceleration while handling surgeon occlusions. An intuitive surgical guidance is provided thanks to the integration into an augmented reality based navigation system. The registration method was verified on a public dataset with a median of 100% successful registrations, a median target registration error of 2.7 mm, a median screw trajectory error of 1.6°and a median screw entry point error of 2.3 mm. Additionally, the whole pipeline was validated in an ex-vivo surgery, yielding a 100% screw accuracy and a median target registration error of 1.0 mm. Our results meet clinical demands and emphasize the potential of RGB-D data for fully automatic registration approaches in combination with augmented reality guidance.

DOI: https://doi.org/10.1016/j.media.2023.103027

Medical Image Analysis 91 (2024) 103027

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery

Florentin Liebmann [a,b,*], Marco von Atzigen [a,b], Dominik Stütz [c], Julian Wolf [d], Lukas Zingg [e], Daniel Suter [e], Nicola A. Cavalcanti [a,e], Laura Leoty [a], Hooman Esfandiari [a], Jess G. Snedeker [b], Martin R. Oswald [c,f], Marc Pollefeys [c,g], Mazda Farshad [e], Philipp Fürnstahl [a]

[a] Research in Orthopedic Computer Science, Balgrist University Hospital, University of Zurich, Zurich, Switzerland
[b] Laboratory for Orthopaedic Biomechanics, ETH Zurich, Zurich, Switzerland
[c] Computer Vision and Geometry Group, ETH Zurich, Zurich, Switzerland
[d] Product Development Group, ETH Zurich, Zurich, Switzerland
[e] Department of Orthopedics, Balgrist University Hospital, University of Zurich, Zurich, Switzerland
[f] Computer Vision Lab, University of Amsterdam, Amsterdam, Netherlands
[g] Microsoft Mixed Reality and AI Zurich Lab, Zurich, Switzerland

ARTICLE INFO

MSC:
41A05
41A10
65D05
65D17

Keywords:
Registration
RGB-D
Augmented reality
Pedicle screw

ABSTRACT

Established surgical navigation systems for pedicle screw placement have been proven to be accurate, but still reveal limitations in registration or surgical guidance. Registration of preoperative data to the intraoperative anatomy remains a time-consuming, error-prone task that includes exposure to harmful radiation. Surgical guidance through conventional displays has well-known drawbacks, as information cannot be presented in-situ and from the surgeon's perspective. Consequently, radiation-free and more automatic registration methods with subsequent surgeon-centric navigation feedback are desirable. In this work, we present a marker-less approach that automatically solves the registration problem for lumbar spinal fusion surgery in a radiation-free manner. A deep neural network was trained to segment the lumbar spine and simultaneously predict its orientation, yielding an initial pose for preoperative models, which then is refined for each vertebra individually and updated in real-time with GPU acceleration while handling surgeon occlusions. An intuitive surgical guidance is provided thanks to the integration into an augmented reality based navigation system. The registration method was verified on a public dataset with a median of 100% successful registrations, a median target registration error of 2.7 mm, a median screw trajectory error of 1.6° and a median screw entry point error of 2.3 mm. Additionally, the whole pipeline was validated in an ex-vivo surgery, yielding a 100% screw accuracy and a median target registration error of 1.0 mm. Our results meet clinical demands and emphasize the potential of RGB-D data for fully automatic registration approaches in combination with augmented reality guidance.

1. Introduction

Complex orthopedic procedures, such as pedicle screw placement, can benefit from computer assistance in regards to safety and accuracy (Gelalis et al., 2012; Perdomo-Pantoja et al., 2019). Nevertheless, computer-assisted orthopedic surgery (CAOS) only accounts for an estimated 5% of all orthopedic surgeries performed in North America, Europe and Asia (Joskowicz and Hazan, 2016). Many state-of-the-art navigation systems for pedicle screw placement comprise three main components: planning, registration and navigation. The latter two strongly contribute to the low clinical adoption (Härtl et al., 2013; Nadeau et al., 2015; Joskowicz and Hazan, 2016). Existing registration approaches tend to be time-consuming, cumbersome and often involve radiation. This hinders real-time application for surgical navigation, which itself suffers from limitations caused by conventional visualization techniques.

State-of-the-art CAOS systems commonly require intraoperative imaging or manual anatomy digitization by the surgeon for registration (Markelj et al., 2012). Both are time-consuming processes. Innovative approaches, such as the co-calibration of fluoroscopy and RGB or RGB-D video sources exist, allowing for "virtual fluoroscopy" (Navab et al., 2009; Lee et al., 2017). In addition, machine learning based methods have the potential to solve typical challenges in 2D/3D registration (Unberath et al., 2021). However, except for ultrasound, intraoperative

* Corresponding author at: Laboratory for Orthopaedic Biomechanics, ETH Zurich, Zurich, Switzerland.
E-mail address: florentin.liebmann@balgrist.ch (F. Liebmann).

https://doi.org/10.1016/j.media.2023.103027
Received 2 March 2023; Received in revised form 29 October 2023; Accepted 9 November 2023
Available online 10 November 2023

imaging, always comes along with bulky equipment and radiation exposure for the patient as well as the OR personnel. Despite a multitude of techniques for 2D/3D registration (Sundar et al., 2006; Esfandiari et al., 2019; Miao et al., 2016), which reduces radiation, registration failures are an existing problem even when reference markers are used (Zhang et al., 2019).

Besides the registration issue, most state-of-the-art navigation systems provide visualizations on 2D monitors in the OR periphery, which can cause attention shift and may increase the cognitive load for the surgeons, e.g. hand-eye coordination (Brendle et al., 2020; Qian et al., 2017; Léger et al., 2017). Given the recent advent of augmented reality (AR) solutions and their potential in the realm of medicine, their application in intraoperative settings to provide surgical guidance should be considered in hopes of alleviating the aforementioned limitations (Eckert et al., 2019; Birlo et al., 2022). The use of AR for spine surgery, and pedicle screw placement in particular, has been investigated thoroughly in the past few years, showing the benefits that the technology could bring to the field (Ma et al., 2017; Elmi-Terander et al., 2019; Gibby et al., 2019; Liebmann et al., 2019; Molina et al., 2019; Farshad et al., 2021b,a; Liu et al., 2021; Uddin et al., 2021; von Atzigen et al., 2022; Chytas et al., 2019; Ma et al., 2018). While there are mitigation strategies tailored to the visualization challenges of the existing CAOS solutions (e.g. Wolf et al. (2023)), the question about an ideal registration remains.

As a potential remedy for the aforementioned registration difficulties, one approach towards a more automatic and radiation-free registration is the intraoperative 3D reconstruction of the target anatomy using depth sensing hardware and the associated computer vision software. Ji et al. (2015) used two co-calibrated RGB cameras mounted to a surgical microscope for registration of preoperative spine CT images in a clinical trial. In their work, the authors pursued a semi-automatic segmentation approach to localize the anatomy of interest in the 2D RGB images based on manual surgeon annotations and region growing followed by a 3D reconstruction module. They reached a registration accuracy of 1.43 mm, but the manual annotations make real-time use in a surgical setting cumbersome. The path of stereo feature matching in open spine surgery was further demonstrated to be promising in recent work by Manni et al. (2020), who achieved a better than 0.5 mm 3D triangulation error on grayscale images as evaluated on data from 23 patients. Besides methods that have been reported and evaluated in academic settings, a commercial navigation system for surface-reconstructing, radiation-free spine surgery navigation is available that operates based on a structured light sensor integrated into OR lamps (7D Surgical Inc., Toronto, ON, Canada, Faraji-Dana et al. (2020). The authors have reported the registration time to be less than 20 s, and re-registration (in case of perceived registration inaccuracies) to be even faster. However, the system's footprint is big, as it comes along with an entire OR lamp. Furthermore, registration starts with manual point sampling and has to be performed for each vertebral level individually. Motion detection and compensation relies on markers clamped to the anatomy, which is the standard approach. While such techniques can diminish some of the concerns related to radiation exposure, time and cost of a common registration pipeline, they still require a residual registration process between the reconstructed anatomy and the patient, making them susceptible to issues such as sub-optimal manual input, poor initialization and small capture range. More recent algorithms have looked into achieving a higher level of registration autonomy through the utilization of artificial intelligence (AI) concepts. In Félix et al. (2021), an RGB-D sensor was used for automatic registration of preoperative femural and tibial 3D models to intraoperative cadaveric anatomy. The RGB images were segmented with a neural network allowing for the corresponding 3D segmentation of the reconstructed anatomy. The reconstructed 3D models were then automatically registered to the anatomy using a RANSAC-based method. Through the analysis of the results, the authors have reported that a considerable part of the error was attributed to the infrared-based RGB-D sensor, an observation that was affirmed in other studies (e.g., Gu et al. (2021b)). Hu et al. (2022) investigated a femur registration and tracking approach using point cloud data that finds a global alignment between a peroperative 3D reference model and intraoperative depth camera data based on RANSAC, followed by an iterative closest point (ICP) refinement. A PointNet-based network was proposed in this study to restore the surface of the unmodified bone captured by the depth camera before using it for registration, coping intraoperative bone surface modification. The employment of the network reduced the registration RMSE from 2.40 mm to 2.07 mm, but the improvement did not reflect significantly on the pose error when compared to a ground truth tracking. An early prototype of a complete AR-based navigation approach using an optical see-through head-mounted display (HMD) for total shoulder arthroplasty evaluated on synthetic bone models was presented in Gu et al. (2021a). The 3D model counterpart of the synthetic bone model is first aligned manually as a movable AR rendering, followed by an ICP refinement. Thereby, the intraoperative data was a point cloud, originating from a co-calibrated external RGB-D sensor. The point cloud was computed from a disparity map which was generated by a transformer-based network with the two RGB images of the depth sensor as input. After registration, an additional fiducial marker clamped to the anatomy is responsible for motion compensation. An average pin placement accuracy of 4.66° and 3.8 mm was achieved. These recent research contributions show that AI-based algorithms using surface data, potentially in combination with RGB, could advance registration approaches in CAOS. What stands out is that there remains a dependency on a coarse initial alignment as well as reference markers for motion detection and/or compensation.

Apart from registration, accurate instrument tracking is a prerequisite for CAOS systems. The aforementioned AR based navigation solutions rely on a broad variety of approaches: external high-end infrared tracking systems (Ma et al., 2017), infrared tracking on a custom-made HMD (Molina et al., 2019; Liu et al., 2021), as well as stereo or monocular tracking of single fiducial markers on off-the-shelf HMD (Liebmann et al., 2019; Farshad et al., 2021b,a). Recent work shows that submillimeter tracking accuracy based on monocular infrared is also possible on an off-the-shelf HMD (Martin-Gomez et al., 2023). When external RGB cameras are in place, similar accuracies can be achieved using stereo based methods (Liu et al., 2013; Ma and Zhao, 2017).

The goal of this study was to tackle the aforementioned drawbacks of the current navigation systems by developing an efficient, radiation-free and real-time approach for automatic registration of intraoperative RGB-D data to the underlying patient with the potential downstream goal of providing a more accurate and faster pedicle screw placement alternative under AR guidance. Our method comprises of a registration module for marker-less, automatic piecewise registration of preoperative lumbar spine 3D models to intraoperative RGB-D data with pose updates during surgical interaction as well as a navigation module for AR-guided pedicle screw placement. The registration module was developed and evaluated on the public SpineDepth dataset (Liebmann et al., 2021) of pose-annotated cadaveric surgery RGB-D recordings. Data collected from simulated pedicle screw placement interventions was used to evaluate the registration success. Finally, the full pipeline (registration + navigation) was validated in an *ex-vivo* setup, where a surgeon placed ten pedicle screws in a cadaveric lumbar spine under AR guidance.

2. Material and methods

The proposed hardware setup consists of the following components: an RGB-D sensor, a GPU-enabled workstation and a HMD. In our case, a ZED Mini (Stereolabs Inc., San Francisco, CA, USA), a HP Z2 (HP Inc., Palo Alto, CA, USA) with an Nvidia GeForce RTX 2080 SUPER (Nvidia Corporation, Santa Clara, CA, USA) and a Microsoft HoloLens

2 (Microsoft Corporation, Redmond, WA, USA) were used (Fig. 1). The RGB-D sensor has a very small footprint (124.5×30.5×26.5 mm, 60 g) and is therefore versatile and easy to integrate. It offers stereo RGB-based depth reconstruction with aligned color and depth frames.

The main contribution of this work lies in the registration and navigation method. It can be subdivided into two modules: the registration module and the navigation module (Fig. 1). The RGB-D sensor observes the surgical site from the top and serves as input for both modules. The registration module (Section 2.1) is responsible for the automatic segmentation and registration of lumbar vertebrae L1–L5 and outputs five rigid 3D transformation estimations, one for each vertebra $V_i, i \in \{1, 2, \dots, 5\}$ and incoming RGB-D frame f: $\hat{\mathbf{T}}_{V_i}(f)$, $\hat{\mathbf{T}}$ denoting that the transformation is estimated. The navigation module (Section 2.2) is responsible for tracking a surgical drill sleeve (rigidly attached to the drill) and finding its 3D transformation in each frame: $\mathbf{T}_D(f)$. For each frame, the navigation module streams the aforementioned six transformations ($\hat{\mathbf{T}}_{V_i}(f), i \in \{1, \dots, 5\}$ and $\mathbf{T}_D(f)$) to the HMD via a User Datagram Protocol (UDP) connection. Note that no other data is sent to the HMD and that the HMD is not involved in any part of the registration module. The second part of the navigation module consists of AR guidance for pedicle screw placement on the HMD, based on the received vertebra and drill sleeve poses. Finding the relative transformation ${}^{HMD}\mathbf{T}_S$ between the coordinate frame of the RGB-D sensor S and the HMD device is required upon re-positioning of the RGB-D sensor and is based on standard chessboard detection (Bradski, 2000).

The registration module and the first part of the navigation module were implemented as a real-time capable C++ application with an OpenGL (Woo et al., 1999) window for live visualization and controlling purposes, referred to as the server app. Libraries and implementation details are provided throughout the following sections. The second part of the navigation module, the AR guidance for pedicle screw placement, was implemented in Unity (2019.4.39f1, Unity Technologies, San Francisco, CA, USA) and will hereafter be referred to as the client app. The source code is available upon request.

2.1. Registration module

The registration module has two goals: first, given an unoccluded RGB-D frame, referred to as the *initial frame*, it estimates the 3D pose of each lumbar vertebra L1–L5 with respect to the sensor coordinate frame. Second, given subsequent frames of the same viewpoint with surgeon interaction, referred to as the *interaction frames*, it updates the poses of L1–L5, if visible. In our workflow, the surgeon positions the RGB-D sensor above the incision without any occlusion by personnel or instruments and initiates the process.

The registration method is illustrated in Fig. 2. It is divided into segmentation/pose initialization and pose refinement. The first stage relies on a deep neural network that combines the concepts of 2D U-Net (Ronneberger et al., 2015) and regression-based orientation prediction (Mahendran et al., 2017). For a given *initial frame* and during the inference time, the network outputs a 2D binary segmentation mask for the lumbar spine (segmentation path) and an estimate of the spine's rotation in the RGB-D sensor's coordinate system (orientation path) represented in form of a quaternion (**R** in Fig. 2). The segmentation output is used to mask the corresponding depth image, leading to a segmented point cloud of the lumbar spine. The preoperative 3D models are transformed with the predicted orientation as the rotation and the center of mass of the segmented point cloud as the translation (**T** in Fig. 2). In the second stage, an en-bloc registration of the combined preoperative 3D models is performed using ICP (Besl and McKay, 1992) registration (general alignment), followed by a piecewise ICP registration of each vertebra (piecewise refinement). Using the accurate pose determined from an *initial frame*, efficient motion compensation in subsequent *interaction frames* is achieved by iterative application of segmentation and piecewise refinement steps, indicated as dotted arrows in Fig. 2.

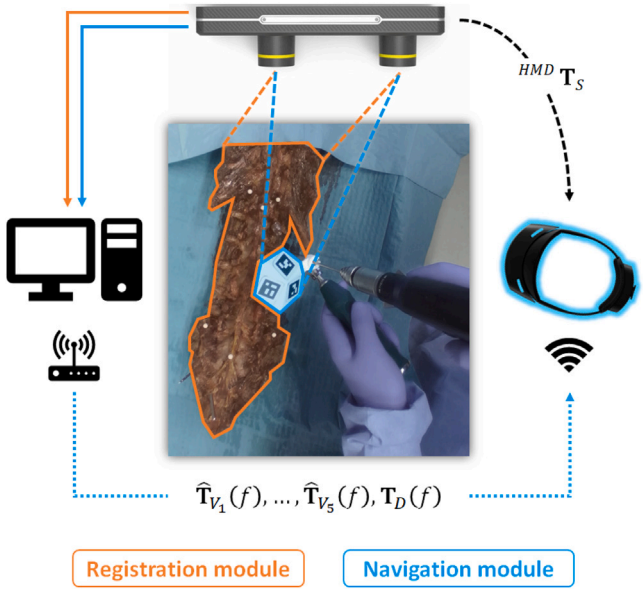

Fig. 1. Setup overview. Solid lines denote a wire, dotted lines denote a wireless connection and dashed lines denote a transformation. The RGB-D sensor observes the surgical site from the top and serves as input for the registration and navigation module. The former outputs five poses, one for each vertebra V_i and frame f: $\hat{\mathbf{T}}_{V_1}(f), \dots, \hat{\mathbf{T}}_{V_5}(f)$. The latter tracks the surgical drill sleeve and outputs its pose $\mathbf{T}_D$ for each frame: $\mathbf{T}_D(f)$. All poses are streamed to the HMD wirelessly. The second part of the navigation module comprises AR guidance for pedicle screw placement, based on the received poses. ${}^{HMD}\mathbf{T}_S$ denotes the transformation between the RGB-D sensor's coordinate frame and the one of the HMD. It is found using standard chessboard detection after the RGB-D sensor has been positioned.

2.1.1. Data preparation

The SpineDepth dataset that was created in our preceding publication (Liebmann et al. (2021), available under: rocs.balgrist.ch/en/open-access/spinedepth) provides pose-annotated RGB-D recordings of mockup spine surgeries performed on ten cadaveric human specimens. The extent of anatomical exposure in Specimen 1 was significantly less than in the other nine. It was therefore excluded from this study. Furthermore, we excluded Specimen 10 as its anatomy is extremely different from the remaining eight, i.e., it is much smaller. When the dataset was created, in each surgery, ten pedicle screws were placed bilaterally into vertebrae L1–L5. The placement of each screw was divided into four surgical steps, each captured in a separate recording by two downward facing RGB-D sensors simultaneously. After each screw placement, the sensors were repositioned to capture the surgical site from a new perspective. This resulted in a total of 80 recordings within the SpineDepth dataset from 20 different viewpoints per specimen. Within the same dataset and for each specimen, preoperative 3D models of vertebrae L1–L5, referred to as PreOp models, are available that are spatially aligned to the actual position of the anatomy. In other words, the SpineDepth dataset includes the aforementioned transformation $\hat{\mathbf{T}}_{V_i}$ between each vertebra and the RGB-D frame observing it. However, in the dataset it is the ground truth transformation and therefore referred to as $\mathbf{T}_{V_i}$. Applying it on the PreOp 3D models transforms them to their ground truth location in the camera coordinate system of the respective RGB-D sensor.

In order to investigate the generalizability of our method to unseen anatomy, the data was prepared to support a leave-one-out cross-validation strategy with eight folds, one for each specimen. For each screw and surgical step three frames were selected, the first frame (*initial frame*) and two random frames (*interaction frames*), resulting in twelve frames per screw. The sensor viewpoints did not change within the four recordings. The *initial frame* never contained surgeon interaction, while the *interaction frames* had a chance thereof. The resulting

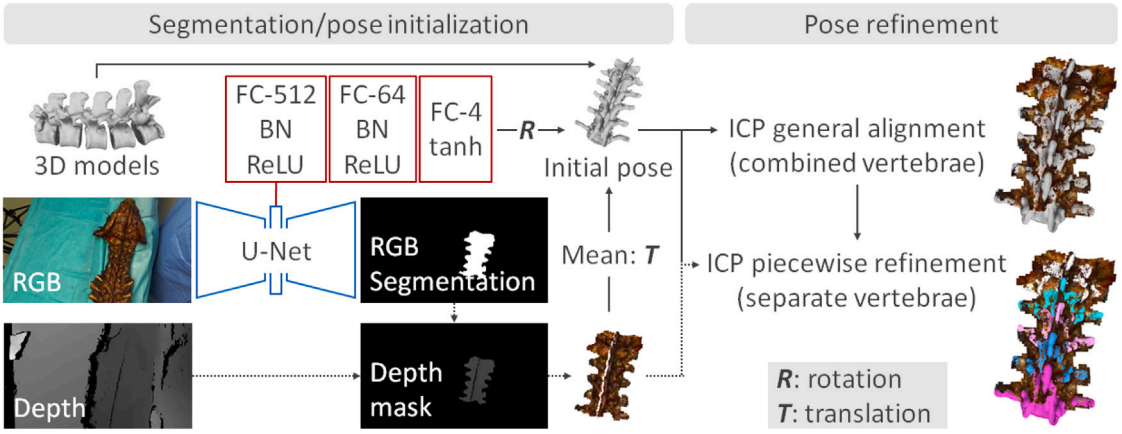

Fig. 2. For an *initial frame*, all arrows are executed once: the network outputs result in a segmented point cloud and an initial pose (R, T) for the preoperative 3D models, which are then registered to the segmented point cloud (general alignment), followed by an individual registration (piecewise refinement). For subsequent *interaction frames*, only the dotted arrows are executed, updating the poses of visible vertebrae individually.

240 frames (10 screws × 12 frames × 2 RGB-D sensors) per specimen were used for training the network. They are referred to as *training folds*. For later testing of the entire registration module, only the first two surgical steps, i.e., recordings, for each screw were considered, as they represent the relevant steps for navigation, i.e., entry point preparation and trajectory drilling with a surgical awl. These 40 recordings of a specimen were used in their full lengths with the first frame as *initial frame* and all subsequent frames as *interaction frames*. They are referred to as *testing folds*.

The pose annotations provided by the SpineDepth dataset cannot be used directly for training our network, as the segmentation and orientation path of our network require a binary segmentation mask and a quaternion as ground truth, respectively. The former was generated as follows. A depth image of vertebrae L1–L5, transformed according to their ground truth poses, was rendered, using the method of Guney and Geiger (2015) and Geiger and Wang (2015), and subtracted from the corresponding sensor depth image for all non-zero pixels in Matlab (R2021a, MathWorks, Portola Valley, CA, USA). The resulting mask consisted of all pixels with an absolute difference below 10 mm. This threshold correctly handled clear occlusions, while maintaining contiguous mask regions, despite measurement noise. Further smoothing of those regions was achieved by applying a 2D convolution with a kernel size of 15 and uniform weights of $1/225$, followed by a thresholding at 0.5, resulting in a binary mask again. The network orientation path only predicts the overall lumbar spine orientation (Fig. 2), and not the one of each vertebra. Therefore, the rotation of vertebra L3 was stored as a quaternion for each frame, representing the overall spine orientation. Note that in the experimental setup of the SpineDepth dataset all specimens had been mounted in a very similar same way with respect to the ground truth high-end optical tracking system, and the orientation of vertebra L3 was similar enough such that orientation normalization was given implicitly.

2.1.2. Network architecture and training

The network architecture is indicated in Fig. 2. The main structure is inspired by U-Net, taking downsampled RGB images of size 144 × 256 ($H \times W$) as input. It consists of four downsampling blocks, each of the form *Conv–BN–ReLU–Conv–BN–ReLU–MaxPooling*, with 64, 128, 256 and 512 filters of size 3 × 3. The bottleneck block is of the same form, with 512 filters, but without the *MaxPooling* layer. The upsampling blocks mirror the downsampling ones, but with a leading upconvolution layer instead of a *MaxPooling* layer at the end. Skip connections connect the down- and upsampling blocks. The output *Conv* layer has 1 filter and *sigmoid* activation, for which a dice loss L_D is minimized. An additional branch is appended to the bottleneck where the 9 × 16 × 512 feature representation is used for regression-based orientation prediction. After flattening, two blocks of *FC–BN–ReLU*, with 512 and 64 units, follow. Another *FC* layer with 4 units then predicts the spine orientation. The quaternion codomain of $[-1, 1]$ is accounted for with a *tanh* activation. For frame i, the geodesic loss L_{G_i} is computed between the ground truth quaternion q_{t_i} and its normalized predicted counterpart q_{p_i}, as given in Eq. (1). Quaternions should have magnitude 1 in order to represent valid rotations. As suggested in Langlois et al. (2018), a penalization term L_{N_i} helped the network predicting such.

$$L_{G_i} = 2\cos^{-1}|\langle q_{t_i}, q_{p_i}\rangle| \quad \text{and} \quad L_{N_i} = \left(1 - ||q_{p_i}||\right)^2 \tag{1}$$

Including L_D, the network's total loss for a batch of size B is

$$\mathcal{L} = \frac{1}{B}\sum_{i=1}^{B} L_{D_i} + \frac{1}{B}\sum_{i=1}^{B}\left(L_{G_i} + L_{N_i}\right). \tag{2}$$

Augmentations were employed to enrich the SpineDepth dataset that included a limited number of specimens and viewpoints. 2D image rotation of α radians was applied on the input images and the corresponding ground truth quaternion by multiplication with quaternion $[\cos\frac{\alpha}{2}, 0, 0, \sin\frac{\alpha}{2}]$, which represents a rotation around the camera's z-axis. Each frame was augmented by rotations of 30, 90, 150, 210, 270 and 330 degrees, resulting in 1680 frames ($240 + 6 \times 240$) per *training fold*. The 240 frames are: 10 screws × 12 frames × 2 RGB-D sensors (Section 2.1.1).

The resulting network with ~58M trainable parameters was implemented in Keras (2.7.0, Chollet et al. (2015)). For each specimen, it was trained from scratch on the remaining eight *training folds* during 30 epochs with batch size of 32 using the Adam optimizer (Kingma and Ba, 2014). The learning rate was set to $10^{-4-\lfloor\frac{epoch}{10}\rfloor}$. Training took roughly 30 min on a NVIDIA Quadro P6000. Three trainings were performed per specimen, the only difference being the weight initialization. For each specimen, the best results in terms of L_D are presented.

2.1.3. Registration and pose update

The registration for an *initial frame* consists of an initial pose, a general alignment and a piecewise refinement (Fig. 2). They are denoted as $\hat{\mathbf{T}}_{init}$, $\hat{\mathbf{T}}_{gen}$ and $\hat{\mathbf{T}}_{ref_{V_i}}$. The following sections elaborate on each part followed by an explanation of the pose update for the *interaction frames*. Note that only the points visible from an orthogonal posterior view were selected from the PreOp models (3-matic, Materialise NV, Leuven, Belgium) and used for registration and pose updates (Fig. 3).

Initial pose estimation. For an *initial frame's* RGB image, our network predicts a resized binary segmentation mask M_p (1080 × 1920 pixels) and a normalized quaternion q_p. Using M_p, the corresponding full point

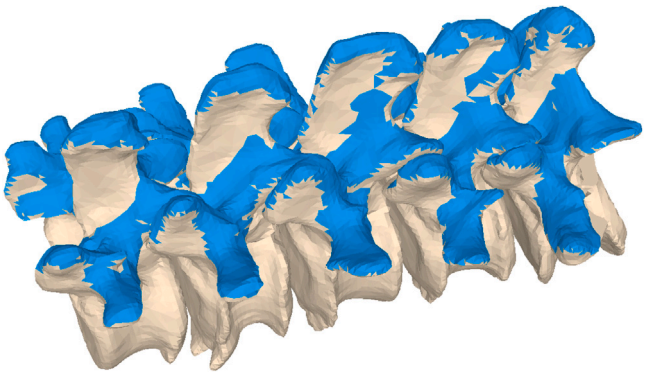

Fig. 3. Points used for registration and pose updates (blue).

cloud PC_f is masked to produce a segmented point cloud PC_s. The initial pose for the first frame, which is used as the *initial frame*, $\hat{\mathbf{T}}_{init}$ for the PreOp models is constructed from the center of mass of the largest connected component (Bradski, 2000) of PC_s of size N_{CC} as the translation and the inferred q_p as the rotation:

$$\hat{\mathbf{T}}_{init} = [R, T] = \left[q_p, \frac{1}{N_{CC}} \sum_{n=1}^{N_{CC}} PC_{s_n} \right] \tag{3}$$

The trained network was converted to Open Neural Network Exchange (ONNX, Foundation (2022)) format and integrated into the server app using TensorRT (8.0.3, NVIDIA (2021)). This way, the RGB-D sensor data, which is available on GPU memory, can be directly fed to the network. Furthermore, PC_f is masked by M_p on the GPU using CUDA (8.0.3.4, NVIDIA (2022)), resulting in PC_s. Note that PC_s contains all segmented points, not only the largest connected component. The latter was only employed for finding the translational part of the initial pose to mitigate the influence of outliers in M_p.

General alignment. The general alignment consists of point-to-point ICP between the combined PreOp models of L1–L5, transformed by $\hat{\mathbf{T}}_{init}$, and PC_s. It is denoted as $\mathbf{T}_{\text{ICP}_{combined}}$. The pose after general alignment is:

$$\hat{\mathbf{T}}_{gen} = \hat{\mathbf{T}}_{init} \mathbf{T}_{\text{ICP}_{combined}} \tag{4}$$

The Point Cloud Library (1.12.0, Rusu and Cousins (2011)) implementation with a maximum correspondence distance of 5 mm and stopping criteria of 50 iterations or a transformation epsilon of $< 10^{-8}$ was used.

Piecewise refinement. The piecewise refinement is based on the general alignment result. In essence, this is another point-to-point ICP alignment. However, given the piecewise nature of this alignment, PreOp models L1–L5 are aligned individually. First, the nearest neighbor points for each PreOp model are found in the masked point cloud PC_s. Point correspondences with a Euclidean distance below 2.0 mm were considered inliers. The Umeyama method (Umeyama, 1991; Guennebaud et al., 2010) method was then used to find the optimal transformation, which was applied to the respective PreOp model. The two steps are repeated 50 times, or until the root-mean-square error of point correspondences does not decrease anymore. The ICP result for vertebra i is denoted as $\mathbf{T}_{\text{ICP}_{piecewise_i}}$. The pose after piecewise refinement is:

$$\hat{\mathbf{T}}_{ref_{V_i}} = \hat{\mathbf{T}}_{init} \mathbf{T}_{\text{ICP}_{combined}} \mathbf{T}_{\text{ICP}_{piecewise_i}} \quad i \in \{1, \dots, 5\} \tag{5}$$

As the same functionality is employed for real-time pose updates during interaction frames (see next paragraph), the nearest neighbor search is performed in parallel on the GPU using CUDA. The EIGEN (Guennebaud et al., 2010) implementation of the Umeyama method is computationally inexpensive, and was therefore performed serially for each PreOp model on the CPU.

Pose update. After performing the registration based on the *initial frame*, the poses of PreOp models L1–L5 are updated individually once a new frame is available. After network inference on the new frame, the same technique as for the piecewise refinement is used. Given that the extent of change in the vertebra poses is minimal for a single patient during an intervention and in order to boost performance, only a single iteration was performed, denoted as $\hat{\mathbf{T}}_{update_i}(f)$ for vertebra i in frame f. Furthermore, due to surgeon interactions, the visibility of the anatomy during *interaction frames* may be obscured. Consequently, only PreOp models that have at least 90% of the number of inliers as after the last iteration during piecewise refinement are updated. The pose $\hat{\mathbf{T}}_{V_i}(f)$ (Fig. 1) in any *interaction frame* f after transformation of each vertebra by $\hat{\mathbf{T}}_{ref_{V_i}}$ is:

$$\hat{\mathbf{T}}_{V_i}(f) = \hat{\mathbf{T}}_{V_i}(f-1) \mathbf{T}_{upd_i}(f) \quad i \in \{1, \dots, 5\}, f \in \{2, 3, \dots\} \tag{6}$$

2.2. Navigation module

The navigation module comprises two parts: tracking of a surgical drill sleeve and AR guidance for pedicle screw placement on the HMD. Both parts are explained in the next sections.

2.2.1. Drill sleeve tracking

Literature findings show that stereo based tracking using RGB cameras can deliver high accuracy (Liu et al., 2013; Ma and Zhao, 2017). In addition, retrieving tracking results in the same coordinate system as the one used for registration allows using AR for pure visualization purposes. Therefore, tracking of the surgical drill sleeve was realized using the RGB-D sensor. It was based on a custom-made, 3D printed component attached to the drill sleeve (Fig. 4). Further development in the scope of an ongoing clinical study based on the publications of Farshad et al. (2021b,a) showed that multiple non-planar markers are beneficial in terms of accuracy (non-planar depth information) and trackability (number of covered viewpoints with at least two markers visible). Thus, the component was equipped with three nonplanar, sterile markers (Clear Guide Medical, Baltimore MD, USA) showing unique AprilTag (Olson, 2011; Wang and Olson, 2016) patterns. The tracking was integrated into our server app and was performed on a separate thread, which is initiated upon app startup. Whenever a new frame was available from the RGB-D sensor, the undistorted left and right grayscale images were made available to the tracking thread. In both images, the markers were detected by the ArUco (Muñoz-Salinas et al., 2018; Garrido-Jurado et al., 2014, 2016) library. If at least two corresponding markers were found in both images, the respective 2D corner coordinates are used for triangulation (Bradski, 2000), yielding eight or twelve 3D corner coordinates. The actual pose is found by applying the Umeyama method between the ground truth corner coordinates, which are known by design, and their estimated counterparts resulting from the triangulation. A Kalman filter (Kalman, 1960; Bradski, 2000) with a constant acceleration model is used for noise reduction on the final drill sleeve pose.

2.2.2. AR guidance

The goal of the AR guidance for pedicle screw placement is the accurate and fast navigation of the screw entry point and trajectory. Upon startup of the client app, the HMD establishes a UDP connection to the server app and continuously receives the poses of PreOp models L1–L5 as well as the drill sleeve pose. The surgeon positions the RGB-D sensor such that a reasonable initial pose, which is visualized on a monitor in the periphery of the OR, is estimated by the server app. A standard chessboard (Bradski, 2000) is used to co-calibrate, i.e., determine ${}^{HMD}\mathbf{T}_S$ (Fig. 1), the coordinate frame of the RGB-D sensor and the one of the HMD. The surgeon can trigger the co-calibration in the client app by speech command and by key press in the server app. In the client app, the chessboard pose is estimated from a single RGB image originating from the HoloLens photo/video camera using chessboard

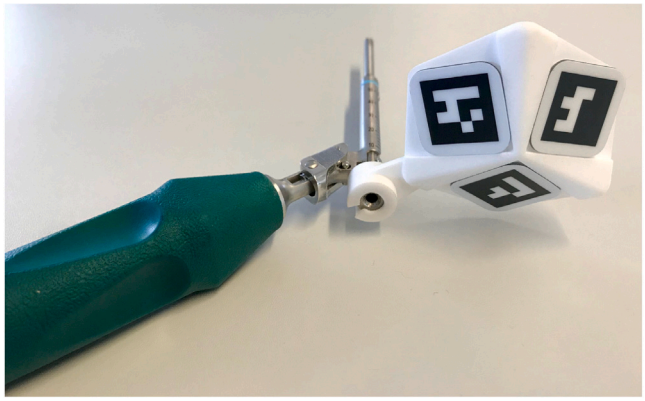

Fig. 4. Drill sleeve with three nonplanar, sterile markers for tracking.

detection and PnP. In the server app, the chessboard is detected in both the left and the right RGB image and the 3D chessboard corners are triangulated, from which the pose can be deducted. The two poses then serve as a common coordinate system, allowing the server app to send and the client app to visualize the estimated poses of vertebrae and drill sleeve relative thereto. Note that this is a one-time process that is only required upon repositioning of the RGB-D sensor. As soon as the chessboard is removed, the surgeon initiates the registration on an *initial frame*, which is followed by pose updates in subsequent *interaction frames* (Section 2.1.3).

In the client app, the surgeon is provided with three different visualization components. The most important component is the virtual twin presented in the work of Wolf et al. (2023), who investigated different user interfaces for AR-guided pedicle screw placement. Their virtual twin approach, where PreOp models and navigation information are not directly overlaid onto the anatomy, but rendered in an axis-aligned fashion with only a translational offset from the anatomy (Fig. 5), was integrated into the client app. Wolf et al. (2023) showed that this approach allows for accurate screw placement, while both ease of use and cognitive load were well rated by surgeons. On the virtual twin, the current drill sleeve pose was visualized with respect to the preoperatively planned screw entry point and trajectory. In addition, the angular 3D deviation between the drill sleeve and the screw trajectory was shown. Besides the virtual twin, a direct overlay of the entry point in form of an aiming cross could be shown/hidden by the surgeon. Lastly, the PreOp models could also be visualized on the anatomy upon request. This was particularly useful to qualitatively check the overall registration accuracy. For a more detailed verification of the registration, the surgeon was asked to touch certain anatomical landmarks using the drill sleeve on the anatomy and confirm their correspondence on the virtual twin before starting the actual navigation of a screw. If the registration was unsatisfactory, the RGB-D sensor was repositioned and the process was repeated from the co-calibration on. After successful navigation of a level, the surgeon selects the next level in the client app menu. Note that the frame rate of a HoloLens 2 application is 60 frames per second. The client app only receives poses and does arguably lightweight rendering, therefore the frame rate was not influenced in a significant way.

2.3. Evaluation

The method was evaluated in two stages. First, the registration module was evaluated separately on the SpineDepth dataset, referred to as the verification. The entire prototype (registration and navigation module) was then evaluated in a cadaveric experiment on an unseen lumbar spine anatomy, where ten real pedicle screws were placed under AR guidance, referred to as *ex-vivo* validation. The two evaluation stages and the respective outcome measures are described in the following sections.

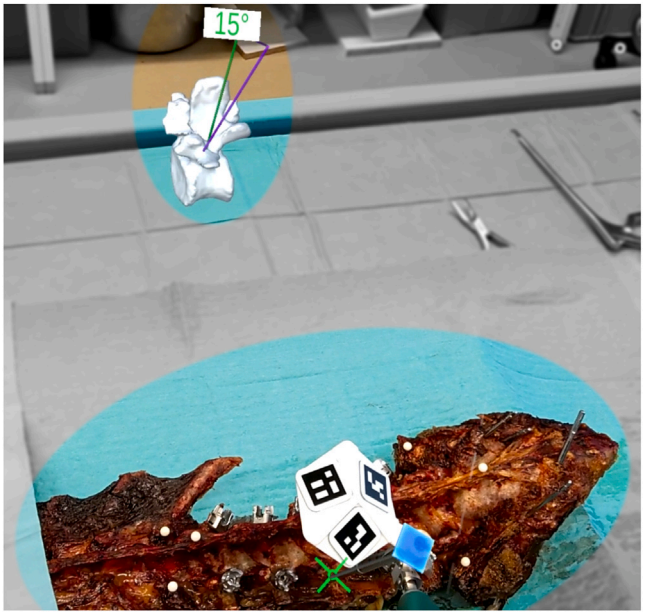

Fig. 5. Surgeon's view of AR navigation. The PreOp model is rendered as an axis-aligned virtual twin (top) and the current drill sleeve pose is visualized with respect to the preoperatively planned screw entry point and trajectory. In addition, the angular 3D deviation between the drill sleeve and the screw trajectory is shown. The direct overlay of the entry point (green cross) can be shown/hidden by the surgeon. The blue square denotes the origin of the navigation device. Note that white push-pins are only used for postoperative evaluation of target registration error.

2.3.1. Verification

For verification, the eight trained networks (Section 2.1.2) were employed. For each specimen, each of the 40 recordings (Section 2.1.1) in the respective *testing fold* was evaluated using the server app, with the first frame as the *initial frame* and all subsequent frames as *interaction frames*.

The recordings in the SpineDepth dataset were made from a broad variety of viewpoints, some of which providing a strongly inclined lateral view of the anatomy, thus potentially influencing the registration quality. Preliminary analysis had shown that the 3D angle between the RGB-D sensor's forward axis and the axis pointing in posterior direction correlate positively with the target registration error (TRE) after applying our proposed method. The correlations (Pearson correlation coefficient: PCC) are reported for each specimen in the *testing fold* (viewpoint-error correlation, VEC). It was defined that only recordings with a 3D angle below 30° are considered, which was assumed to be the range of angles allowing OR lamps to illuminate the surgical site without surgeon-induced occlusions. The number of recordings fulfilling this criterion is also part of the results (acceptable viewpoints, AcVp). Note that for generalization purposes, the viewpoints above 30° were not excluded from network training.

A single threshold for the accuracy of pedicle screw placement with respect to the optimally planned screw could not be defined, as the required accuracy is generally dependent on different anatomical and surgical factors such as the anatomical morphology and pathology of the patient, the underlying bone quality, or the utilized surgical approach. Furthermore, with an automated evaluation based on the given dataset, tapping landmarks on the anatomy to confirm the registration accuracy, as done during AR guidance (Section 2.2.2), is not possible. Therefore, a successful registration was measured by the established clinical criteria according to Modi et al. (2008), who define a screw perforation of less than 2 mm as safe. To this end, optimal pedicle screws (⌀: 5 mm) were planned bilaterally using an in-house developed preoperative planning software (CASPA, University Hospital Balgrist, Zurich, Switzerland).

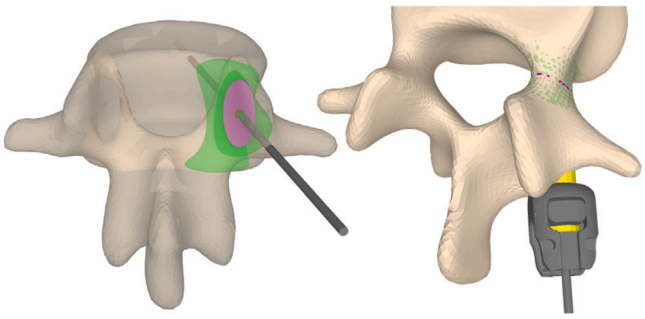

Fig. 6. Assessment of pedicle perforation. Left: calculation of intersection between screw axis (gray) and pedicle middle plane (violet). If the intersection point was outside the pedicle 3D model (green), the registration was not successful. Otherwise, it was verified whether any point of the pedicle 3D model was inside the cylinder (right, yellow, attached is the pedicle screw head in gray). If so, for any point within the cylinder, the distance to the cylinder mesh was calculated. The point having the largest distance to the was used to quantify the amount of perforation in mm.

The assessment of pedicle perforation is illustrated in Fig. 6. A 3D model of the pedicle was extracted from the PreOp model and imported into MATLAB. The screws were represented as cylinders (⌀ 5mm). For each frame of a recording, the screws were transformed according to the corresponding vertebra pose found by our method, while the pedicle 3D model was transformed according to the respective ground truth pose. The intersection between the screw axis and the pedicle middle plane was calculated. If the intersection point was outside of the pedicle 3D model, the screw, and therewith the registration, was considered as not successful. If the intersection point was inside the pedicle 3D model, it was verified whether any point of the pedicle 3D model was inside the cylinder. If so, the screw was said to perforate the pedicle. For any point within the cylinder, the distance to the cylinder mesh was calculated. The point having the largest distance was used to quantify the amount of perforation in mm.

Note that the registration success was defined on a per frame basis and for the target screw only (the screw that the surgeon works on in the respective recording), e.g. if the surgeon prepares the entry point of L2 left in the recording, the registration for a frame was considered successful when the previously described perforation assessment using the estimated pose for L2 ($\hat{\mathbf{T}}_{V_2}$) revealed that the perforation would have been below 2 mm. The success rate of a single recording was defined as the number of successful frames divided by the total number of frames. The success rate of an entire specimen equals the median success rate over all recordings in the *testing fold*. In the same way, the median 3D angular deviation between the optimal and estimated screw trajectory (trajectory error: $E_{E_T R}$) as well as the median 3D distance between the optimal and estimated screw entry point (entry point error: $E_{E_E P}$) are reported. As the TRE considers the registration for an entire vertebra, it can only be computed on a per vertebra level. The TRE for a recording of F frames, where L2 was targeted, with $K = 3$ landmarks (L_1: spinous process, L_2 and L_3: left and right transverse processes) and $\mathrm{d}(p_1, p_2)$ as the 3D Euclidean distance between two points p_1 and p_2, is defined in Eq. (7). The median over all 40 recordings is reported.

$$\mathrm{TRE} = \frac{1}{FK} \sum_{f=1}^{F} \sum_{k=1}^{K} \mathrm{d}(\hat{\mathbf{T}}_{V_2}(f) L_k, \mathbf{T}_{V_2}(f) L_k). \tag{7}$$

Preliminary analysis showed that the alignment of each PreOp model can stabilize during the first few *interaction frames*, presumably due to the slightly varying 3D reconstructions provided by the RGB-D sensor in the absence of surgeon interaction. Therefore, the TRE is reported as of frame 61 (~2 s).

As described in Section 2.1.3, the 3D points used for registration are a subset of the ground truth 3D model geometries in the dataset (Fig. 3). Consequently, a direct comparison between all 3D model points transformed according to our method and the corresponding ground truth 3D points is possible. This metric for pose estimation problems was defined in Hinterstoisser et al. (2013) and is referred to as average distance (ADD). The ADD for exemplary vertebra L2 in a recording of F frames is computed in the same way as the TRE (Eq. (7)), except L denoting 3D points instead of landmarks and K denoting the number of points in the 3D model. Analogously to the TRE, the ADD is computed on a per vertebra level and the median over all 40 recordings considering the respective target screws (the screw that the surgeon works on in the respective recording) is reported.

As an additional result, the percentage of updated poses (UpPo) for the vertebra of interest in *interaction frames*. For each recording, the number of frames where a pose update according to our method was performed (Section 2.1.3) divided by the number of possible frames was assessed. For each specimen, the median over all 40 recordings is reported.

The median runtimes of the registration step (initial pose estimation + general alignment + piecewise refinement) and the pose update step are reported over 40 recordings of an exemplary specimen. For the pose update step, the average over all frames in a recording was calculated, after which the median over all recordings was formed.

Besides the outcome measures related to registration and pose updates, the performance of the eight trained networks are reported. Segmentation accuracy was evaluated with the Dice similarity coefficient (DSC). As in Tulsiani and Malik (2015) and Mahendran et al. (2017), the orientation prediction was evaluated with the median geodesic angle error (MGAE), which equals the median loss defined in Eq. (1) over an entire fold, expressed in degrees. Note that these outcome measures are based on number of frames defined for the *training folds*, i.e., not full recordings but 240 frames per fold/specimen.

2.3.2. Ex-vivo validation

The goal of the ex-vivo validation was to place ten pedicle screws (L1–L5, left and right) under AR guidance using the herein presented method (registration and navigation module) on an unseen human lumbar spine. A fresh frozen specimen was used. Ethical approval was obtained from the ethical committee of Canton Zurich (Basec-Nr. 2017-00874). The specimen was CT scanned using a NAEOTOM Alpha© device (Siemens Healthineers, Erlangen, Germany) with a 0.8 mm slice thickness and a 0.41 × 0.41 mm in-plane resolution (x–y). 3D models of L1–L5 were extracted using the global thresholding, region growing and wrapping functionalities of the Mimics software (Materialise NV, Leuven, Belgium). Again, the points visible from an orthogonal posterior view were selected as described in Section 2.1.3 (Fig. 3). Optimal pedicle screws (⌀: 5 mm) were planned in CASPA. In preparation of the experiment, the specimen was thawed and dissected to have no soft tissues, e.g. paravertebral muscles, without damaging the intraspinous ligament, the ligamentum flavum as well as the facet joint capsule. The specimen was fixated to a wooden board with surgical pins through spinal levels T6/7 and S1.

The network was trained in the same way as described in Section 2.1.2, but using the *training folds* of all eight specimens in the SpineDepth dataset. As there was no ground truth available for the unseen specimen, the number of epochs was reduced to ten to mitigate overfitting to the experimental setup, e.g. the spine orientation w.r.t. the table, of the SpineDepth dataset. Due to the fact that the preoperative CT was taken from the frozen specimen, the inter-vertebral deformation between the pre- and intraoperative states was higher than in the SpineDepth dataset, where the CT was conducted in a fully thawed state. Therefore, the piecewise refinement (Section 2.1.3) was performed for 50 iterations, without a stopping criterion, to overcome local minima due to the 2 mm inlier threshold.

During the experiment, the RGB-D sensor was placed above the surgical site (Fig. 7) and the server and client apps were started. After that, the workflow was as described in Section 2.2.2: the sensor viewpoint was adjusted, such that the initial pose was reasonable,

Fig. 7. Setup during the ex-vivo validation. After inserting the five pedicle screws on the right side, the specimen was rotated by 180°. After re-registration, the surgeon could insert the five screws on the left side.

followed by the co-calibration of sensor and HMD. The registration and pose updates were initiated. For each vertebra, the surgeon checked the registration accuracy and inserted the respective pedicle screw (right side) according to the AR guidance. For screw insertion on the left side, the specimen was rotated 180° (the other side of the table was not optimal to stand for the surgeon), followed by a re-registration.

For the ex-vivo validation, TrEr, EpEr, TRE and ADD are reported. TrEr and EpEr were quantified following the same procedure as described in Liebmann et al. (2019). A postoperative CT of the specimen was acquired with the same imaging device and protocol as for the preoperative scans. 3D models of the bone anatomy and the screws were extracted. In the CASPA software, the PreOp models along with the planned screw trajectories were registered to the postoperative bone anatomy using point-to-plane ICP (Rusinkiewicz and Levoy, 2001). In the same fashion, generic cylindrical 3D models were aligned to the postoperative screw 3D models. The cylinders' main axes were compared to the planned screw trajectories, yielding the 3D angular deviation TrEr. The 3D Euclidean distance EpEr was determined by comparing the planned entry points to the intersection point of the cylinders' main axes with the registered preoperative 3D model.

In contrast to the SpineDepth verification where the ground truth vertebral poses were available, the data collected in the *ex-vivo* validation experiment lacked the registration ground truth; therefore, the experiment was captured as an RGB-D recording and the TRE was quantified retrospectively in a static manner. Six push-pins were inserted into the spinal levels T12 and S1 (three each) before the experiment (Fig. 5). The 3D positions of the push-pin head centers were determined in the postoperative CT using the Mimics software as well as in the left and right RGB images of the RGB-D *initial frame* using blob detection and triangulation techniques (Bradski, 2000). The best fit in a least-squares sense between the two point sets was found in the CASPA software and allowed for transforming the preoperative 3D models into the coordinate frame of the RGB-D sensor. The TRE is based on the same three landmarks per vertebra as for the verification and is also reported for the 61st frame (~2 s) after the *initial frame*. The ADD were computed analogously to the TRE.

As an additional outcome measure, the navigation time, defined according to Farshad et al. (2021b) as the time from picking up the drill sleeve until the drilling process was started, is reported.

Note that the reported median values for TrEr, EpEr, and navigation time were calculated based on all ten screws placed. TRE and ADD are based on the five vertebrae and all three registrations that were performed during the experiment.

2.3.3. Ablation study

To further understand the capabilities of the proposed registration method and the mechanisms leading to our results, an ablation study was conducted. For both verification stages, the server app was run three times with the following modifications (italic font denotes the name of the modification used hereinafter):

- *General:* Registration only included general alignment, no piecewise refinement, no pose updates
- *Refinement:* Registration included general alignment and piecewise refinement, but no pose updates
- *First-60:* Registration included general alignment, piecewise refinement and pose updates during the first 60 *interaction frames* of a recording

For the ablation study, only TRE and ADD are considered. Note that, for the ex-vivo validation, First-60 is equal to our primary results by definition and is therefore not reported.

3. Results

Table 1 summarizes the results. It comprises the verification and the ex-vivo validation. In the following, the five-number summary is given in the following format: median (minimum, first quartile, third quartile, maximum).

For the verification, the viewpoints and TRE correlated by 0.41 (0.12, 0.36, 0.69, 0.78). The number of acceptable viewpoints was 27 (22, 24, 32, 36), with 100% (0, 100, 100, 100) registrations being successful. The trajectory error was 1.6° (0.0, 1.0, 2.6, 37.8) and the entry point error was 2.3 mm (0.4, 1.6, 3.6, 39.3). The TRE was 2.7 mm (0.5, 1.7, 3.9, 41.1) and the ADD was 2.6 mm (0.5, 1.9, 3.8, 49.7). 9% (0, 0, 24, 100) of the poses were updated during *interaction frames*. The DSC was 0.74 (0.67, 0.72, 0.74, 0.76) and the MGAE was 14° (13, 13, 17, 21).

The runtimes were assessed on specimen 6. The registration step (initial pose estimation + general alignment + piecewise refinement) took 1475 ms (1344, 1428, 1595, 1755) and the pose update step took 20 ms (16, 18, 21, 27).

An exemplary case from the verification is illustrated and explained in Fig. 8. It shows segmentation and occlusion handling, initial pose, general alignment, piecewise refinement and pose updates as well as comparison of ground truth models and screws to their counterparts estimated by the proposed registration module.

During the ex-vivo validation, three registrations were necessary. After placing the first screw (L1, right), the client app crashed unexpectedly. Therefore a second registration became necessary. After

Table 1
Results overview. Specimen numbering is according to the SpineDepth dataset. Spec.: Specimen. perf.: performance. Med.: Median. Min.: Minimum. Q_1: first quartile. Q_3: third quartile. Max.: Maximum. VEC: viewpoint-error correlation. AcVp: acceptable viewpoints (out of 40). SuRe: successful registrations. TrEr: trajectory error. EpEr: entry point error. TRE: target registration error. UpPo: updated poses. DSC: Dice similarity coefficient. MGAE: median geodesic angle error. N/A: not applicable. * denotes that the five-number summary (median, minimum, first quartile, third quartile and maximum) was calculated from the 320 recordings (40 recordings per specimen) and not based on the median value per specimen given in this table. For the ex-vivo validation, TrEr and EpEr were calculated based on all ten screws placed, and TRE and ADD based on the five vertebrae and all three registrations that were performed during the experiment.

Spec.	Registration and pose updates								Network perf.	
	VEC	AcVp	SuRe* [%]	TrEr* [°]	EpEr* [mm]	TRE* [mm]	ADD* [mm]	UpPo* [%]	DSC	MGAE [°]
2	0.43	28	100	2.1	3.5	4.1	3.9	1	0.67	13
3	0.40	22	100	2.2	2.7	3.1	2.8	15	0.74	14
4	0.78	22	100	1.7	1.9	2.5	2.3	15	0.75	18
5	0.67	24	100	1.3	1.4	1.5	1.7	12	0.76	14
6	0.27	30	100	1.1	1.1	1.3	1.3	3	0.74	13
7	0.12	36	100	1.6	1.7	1.8	2.1	9	0.74	16
8	0.39	36	100	1.9	3.2	3.4	3.4	13	0.74	13
9	0.73	26	100	2.4	4.0	4.1	4.0	11	0.67	21
Med.	0.41	27	100	1.6	2.3	2.7	2.6	9	0.74	14
Min.	0.12	22	0	0.0	0.4	0.5	0.5	0	0.67	13
Q_1	0.36	24	100	1.0	1.6	1.7	1.9	0	0.72	13
Q_3	0.69	32	100	2.6	3.6	3.9	3.8	24	0.74	17
Max.	0.78	36	100	37.8	39.3	41.1	49.7	100	0.76	21
Ex-vivo validation										
Med.	N/A	N/A	N/A	2.4	2.2	1.0	1.0	N/A	N/A	N/A
Min.	N/A	N/A	N/A	0.7	0.5	0.5	0.5	N/A	N/A	N/A
Q_1	N/A	N/A	N/A	1.7	0.9	0.9	0.8	N/A	N/A	N/A
Q_3	N/A	N/A	N/A	2.6	2.6	1.3	1.5	N/A	N/A	N/A
Max.	N/A	N/A	N/A	6.8	3.7	2.6	2.8	N/A	N/A	N/A

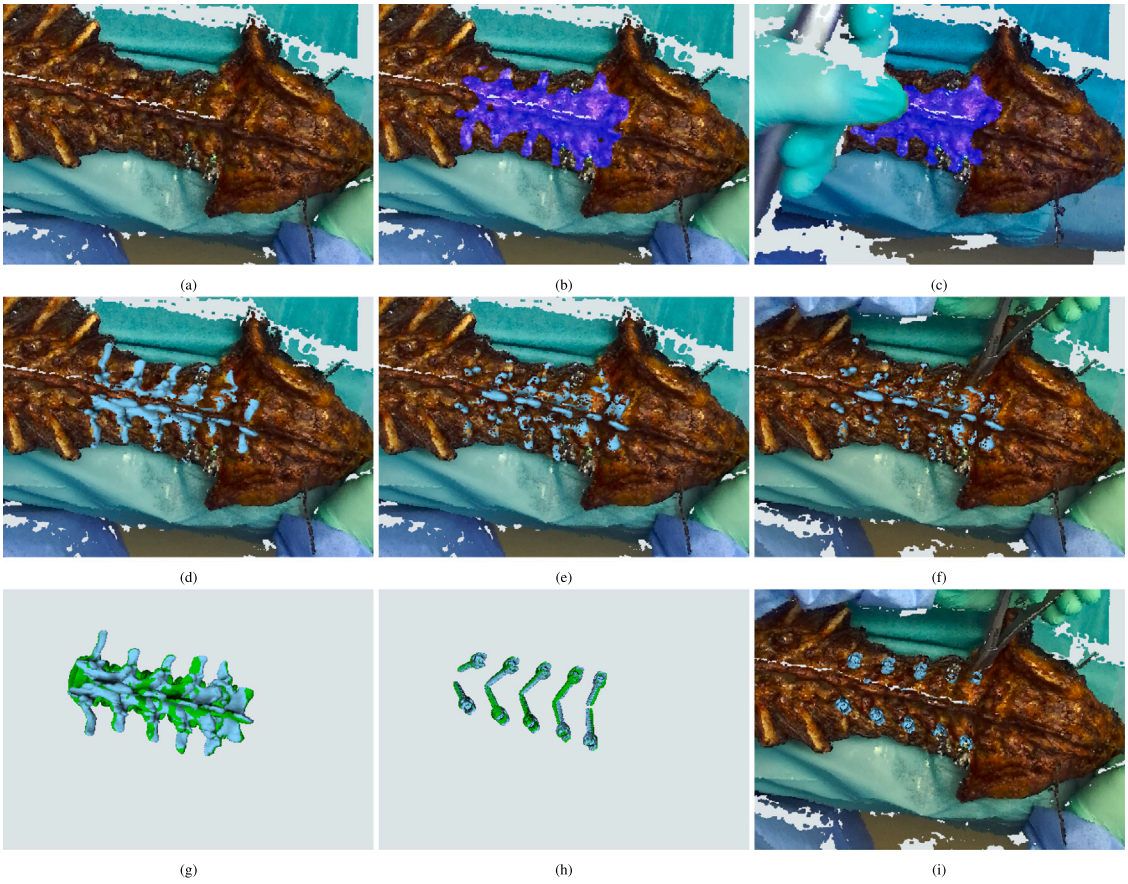

Fig. 8. Exemplary recording from verification. (a) Point cloud (b) Segmentation with screw occlusion handling (L4 left & right, L5 left). (c) Segmentation with surgeon occlusion handling. (d) Initial pose in *initial frame*. (e) General alignment in *initial frame*. (f) After piecewise refinement and 60 pose updates. (g) Estimated (blue) and ground truth (green) vertebra poses. (h) Estimated (blue) and ground truth (green) simulated screws. (i) Estimated simulated screws on point cloud.

placement of the remaining four screws on the right side, the specimen was rotated by 180°, followed by the third registration, such that the surgeon could operate on the left side. The trajectory error was 2.4° (0.7, 1.7, 2.6, 6.8), while the entry point error was 2.2 mm (0.5, 0.9, 2.6, 3.7). All screws were of grade 0, i.e., fully contained within the pedicle (Modi et al., 2008). The TRE was 1.0 mm (0.5, 0.6, 1.1, 1.1)

for the first, 0.9 mm (0.7, 0.9, 1.0, 1.2) for the second, and 1.8 mm (1.0, 1.4, 2.1, 2.6) for the third registration, respectively. The corresponding ADD were 0.9 mm (0.5, 0.8, 1.0, 1.2), 0.8 mm (0.7, 0.7, 1.2, 1.2), and 1.9 mm (0.9, 1.7, 2.1, 2.8). The navigation time per screw was 28 s (16, 22, 31, 49).

For the verification, the ablation study showed a TRE of 2.58 mm (0.4, 1.6, 4.0, 40.5) for *Refinement*, 2.64 mm (0.4, 1.7, 3.8, 40.5) for *First-60*, 2.672 mm (0.5, 1.7, 3.9, 41.1) for our primary results, and 2.674 mm (0.6, 1.9, 4.2, 40.2) for *General*, respectively. For the ADD, the order was the same with 2.47 mm (0.4, 1.8, 3.9, 49.1), 2.61 mm (0.5, 1.9, 3.8, 49.1), 2.62 mm (0.5, 1.9, 3.8, 49.7), and 2.67 mm (0.5, 1.8, 4.1, 48.5), respectively.

For the ex-vivo validation, our primary results were the most accurate in terms of TRE with 1.0 mm (0.5, 0.9, 1.3, 2.6), followed by *Refinement* with 1.1 mm (0.5, 0.7, 1.8, 3.7), and *General* with 1.7 mm (1.3, 1.6, 3.9, 5.3), respectively. For the ADD, *Refinement* was slightly more accurate than our primary results with 0.9 mm (0.4, 0.8, 1.6, 4.1), as opposed to 1.0 mm (0.5, 0.8, 1.5, 2.8). The least accurate was *General* with 1.6 mm (1.2, 1.5, 3.4, 4.9).

4. Discussion

Despite the fact that CAOS can increase accuracy as well as safety in complex orthopedic procedures, such as pedicle screw placement (Gelalis et al., 2012; Perdomo-Pantoja et al., 2019), the clinical adoption of such methods is arguably low (Joskowicz and Hazan, 2016; Härtl et al., 2013; Nadeau et al., 2015). Besides economic reasons, one major barrier along ubiquitous adaptation of the existing CAOS solutions is their interference with the standard surgical workflow. More specifically, main limiting factors associated with the current CAOS systems for surgical navigation can be noted as: cumbersome and time-consuming, ionizing radiation exposure, lengthy registration procedures and unintuitive visualization of spatial navigation information on 2D monitors in the OR periphery. In this work, we intended to tackle these drawbacks and presented a simplistic and radiation-free approach for automatic, accurate and fast pedicle screw placement in cadaveric lumbar spines under AR guidance.

The verification on the SpineDepth dataset showed a median registration success rate of 100%, meaning that the target screw would have been placed successfully within the clinical safe zone. In the study of Félix et al. (2021), who pursued a similar approach for femur and tibia, the success of a registration was defined based on the percentage of inliers, which had to be at least 80. They reached a success rate of 37.7% for the femur and 35.2% for tibia, respectively. The required registration accuracy, defined as 3° rotational error and 3–4 mm translational error, was only met in terms of translation. The surface-based femur registration and tracking approach of Hu et al. (2022) achieved a root-mean-square error of 2.40 mm on real-time captures of a bone phantom, which reduced to 2.07 mm when the bone phantom data was processed with the suggested PointNet-based restoration network. For spine surgery, a wide range of acceptable registration errors can be found in the literature, which depend on various factors. Rampersaud et al. (2001) defined that the maximum rotational and translational deviation for the lumbar spine reach from 2.1°/0.65 mm (L1) to 12°/3.8 mm (L5) for screws with a diameter of 6.5 mm. As the TRE reported in this work comprises the rotational and translational aspect, a comparison to TrEr and EpEr is more meaningful. While the TrEr (1.6°) is within the aforementioned limits in our case, the EpEr (2.3 mm) exceeds the limit for L1. Besides the targeted spine level, different methods for error calculation can affect the reported values (Holly and Foley, 2007). The TRE is a well-known measure to characterize the accuracy of navigation approaches (Ershad et al., 2014). Guha et al. (2019) investigated the error propagation of clinical-grade navigation systems w.r.t. a dynamic reference frame (DRF) attached to the anatomy, which is a common motion compensation technique, on four human cadavers. They compared intraoperative tip positions of a tracked awl (mimicking a bone screw) to typical pedicle screw entry points in a postoperative CT. An average 3D navigation (note that this error may differ from the registration error) error of 2.71 mm at DRF level was found. This error increased with a larger distance to the DRF level. Although the respective registration error must have been lower, the fact that the use of a DRF is the gold standard makes it eligible for comparison to our registration error, assuming tracking errors in current navigation systems are minimal: the EpEr (2.3 mm) of our verification was superior and the TRE (2.7) equal. When comparing to the TRE of 1.43 ± 0.35 mm in the semi-automatic microscopic RGB stereo method of Ji et al. (2015), three out of eight specimens in the verification can be considered within the range of their standard deviation. The required 2 mm maximum acceptable registration error for cranial and spine procedures (Faraji-Dana et al., 2020) is reached for three of our specimens. However, it should be considered that the dataset already comes with certain inaccuracies (ground truth TRE of 1.5 mm). The second (0.9 mm) registration in our ex-vivo validation showed sub-millimetric accuracy, which is equal or close to studies using navigation systems with manual point sampling for pedicle screw navigation (0.9 mm in Papadopoulos et al. (2005), 0.7 mm in Nottmeier and Crosby (2007)) or cutting-edge intraoperative CT device for cranial procedures (0.93 mm in Carl et al. (2018)). The screw accuracies with a TrEr of 2.4° and EpEr of 2.2 mm in the ex-vivo validation are in line with other studies investigating surgical navigation for pedicle screw placement. The AR system used in Felix et al. (2022) achieved a 3D accuracy of 2.5° and 1.9 mm for open surgery in cadavers. In van Dijk et al. (2015), the accuracy of 178 minimally invasive screws using a robotic system was assessed, resulting in a mean 2D in-plane error of 2.55° and a 3D entry point deviation of 2.0 mm. An even lower mean angular deviation of 1.53° can be found in the cadaveric study of Lamartina et al. (2015). However, again, the values originate from 2D in-plane measurements.

Besides showing similar accuracy, our registration method has two advantages over clinically established navigation systems. First, the registration is fully automated and is performed for all targeted levels simultaneously, while computation time required by our method was only approximately 1.5 s (and after that real-time with a median of 20 ms per frame) and therewith considerably lower compared to other clinical-grade systems based on surface data (less than 20 s in Faraji-Dana et al. (2020)) or manual point sampling (117 s in Nottmeier and Crosby (2007), 125 s in Farshad et al. (2021b)). Our ablation study shows that the piecewise refinement improves accuracy especially when the preoperative images were acquired in a different patient positioning. Second, anatomy displacement induced by surgical manipulation or respiration can be as high as 1.85 ± 1.48 mm and 1.09 ± 0.44 mm, respectively (Guha et al., 2019). Our ablation study could not show superior accuracy when applying real-time pose updates in *interaction frames* throughout entire recordings (primary results) as opposed to a registration based on an *initial frame* only (*Refinement*) or applying updates for the first 60 *interaction frames* (*First-60*). One of the main reasons could be the frame rate of the RGB-D sensor as well as motion blur in the images, leading to a insufficient 3D reconstruction. The surgical interactions in the SpineDepth dataset are of fast nature. However, slower motions, such as breathing, could be compensated with the method at hand (upon proper investigations in the future). For faster motion, different sensor types, not based on RGB or grayscale images, could further improve the performance of our method in this regard. Finally, the anatomical part that is moved the most is (partly) hidden by the surgeon, and therefore challenging to track. Nevertheless, we see our approach being a foundation for developing automatic level-wise motion compensation in real-time without needing a DRF clamped to the anatomy.

The median time for pre-drilling a screw trajectory with our method was 28 s per screw which can be considered as very fast. This is superior to navigation using C- or O-arm (248 s for C-arm and 134 s for O-arm in Liu et al. (2017)), as well as other studies using AR guidance in

cadaveric specimens (57.5 s in Müller et al. (2020), 67 s in Farshad et al. (2021b)) or a first in-human study (312 s in Elmi-Terander et al. (2019)).

In terms of network performance, the DSC of the segmentation path with a median of 0.74 on the SpineDepth dataset was comparable to Félix et al. (2021) (DSC for tibia: 0.73), although segmentation of the spinal anatomy might be considered more challenging. The accuracy of the orientation prediction (14°) is comparable to the ones reported in the two publications inspiring our method (16.63° in Mahendran et al. (2017), 13.59° in Tulsiani and Malik (2015)).

Further analysis revealed a PCC of −0.78 between TRE and DSC, suggesting that the segmentation quality plays a key role in finding an accurate registration. The TRE also correlates (PCC of 0.74) with the visible bone surface error (VBSE) reported in the SpineDepth publication (Liebmann et al., 2021), which essentially describes the reconstruction quality of the RGB-D sensor in use. While the dataset was recorded based on stereo calibrations created with a manufacturer-provided application, for our ex-vivo validation, standard OpenCV stereo calibration functionality (Bradski, 2000) was employed, leading to a much lower median TRE (1.0 mm). This potential of accurate spine 3D reconstruction was confirmed in the study of Manni et al. (2020), where features in stereo grayscale images were matched with a 3D triangulation error below 0.5 mm. Stereo calibration and reconstruction quality may not be the only factors influencing the accuracy of the proposed registration approach, but they can be seen as a key factor. Other such factors could be the presence of soft tissue and the missing facet joints/mammillary processes in the dataset specimens, not only regarding accuracy, but also for convergence during general alignment, as more soft tissue flattens important bony surfaces, as well as the presence of previously inserted screws. The latter is suspected to be the reason for the increase in error from the first and second registration in the ex-vivo validation to the third, for which the specimen was rotated by 180° and all screws on the right side had already been inserted. This imbalance is not accounted for, which is a limitation of our method. More importantly, the full anatomical exposure in the cadaveric specimens is unrealistic within a clinical setting, and the high visibility facilitates the registration as well as the navigation. Furthermore, our method did not generalize to all specimens in the verification: specimen 10 had to be excluded due to its much smaller size compared to the other eight considered specimens. Besides that, only a potential approach to real-time motion compensation could be shown. It needs to be found out whether the method would work as is for compensating slow movements, such as breathing, or fast movements when a different type of sensor is used.

For future work, the method needs to be evaluated on specimens with surgical approaches of varying sizes, i.e., less visibility of anatomical structures. Most importantly, the requirements for successful pose updates during surgeon interaction, i.e., sensor type, viewpoint and number of iterations needed should be investigated. Furthermore, transformer-based depth reconstruction, as proposed in Gu et al. (2021a), could be a promising way to increase registration accuracy, while feature-based tracking (Manni et al., 2020) should be investigated as a motion compensation strategy.

5. Conclusions

Our results suggest that fast, radiation-free, and fully automatic level-wise registration with real-time pose updates from RGB-D data for pedicle screw navigation under augmented reality guidance is feasible and meets clinical accuracy demands.

CRediT authorship contribution statement

Florentin Liebmann: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Marco von Atzigen:** Software, Validation, Investigation. **Dominik Stütz:** Methodology, Software, Validation, Formal analysis, Data curation. **Julian Wolf:** Software. **Lukas Zingg:** Validation, Investigation. **Daniel Suter:** Validation, Investigation. **Nicola A. Cavalcanti:** Validation, Investigation. **Laura Leoty:** Formal analysis. **Hooman Esfandiari:** Writing – review & editing. **Jess G. Snedeker:** Resources, Supervision. **Martin R. Oswald:** Methodology, Writing – review & editing. **Marc Pollefeys:** Resources, Supervision. **Mazda Farshad:** Validation, Investigation, Resources, Funding acquisition. **Philipp Fürnstahl:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mazda Farshad reports a relationship with Incremed AG that includes: board membership and equity or stocks.

Data availability

Public dataset is mentioned in the manuscript. Other data will be made available upon request.

Acknowledgments

This project is part of SURGENT under the umbrella of University Medicine Zurich/Hochschulmedizin Zürich.

References

Besl, P.J., McKay, N.D., 1992. Method for registration of 3-D shapes. In: Sensor Fusion IV: Control Paradigms and Data Structures, Vol. 1611. Spie, pp. 586–606.

Birlo, M., Edwards, P.E., Clarkson, M., Stoyanov, D., 2022. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. Med. Image Anal. 102361.

Bradski, G., 2000. The opencv library. Dr. Dobb's J. Softw. Tools.

Brendle, C., Schütz, L., Esteban, J., Krieg, S.M., Eck, U., Navab, N., 2020. Can a hand-held navigation device reduce cognitive load? A user-centered approach evaluated by 18 surgeons. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. Springer, pp. 399–408.

Carl, B., Bopp, M., Saß, B., Nimsky, C., 2018. Intraoperative computed tomography as reliable navigation registration device in 200 cranial procedures. Acta Neurochir. 160 (9), 1681–1689.

Chollet, F., et al., 2015. Keras. https://keras.io.

Chytas, D., Malahias, M.-A., Nikolaou, V.S., 2019. Augmented reality in orthopedics: current state and future directions. Front. Surg. 6, 38.

Eckert, M., Volmerg, J.S., Friedrich, C.M., et al., 2019. Augmented reality in medicine: systematic and bibliographic review. JMIR mHealth uHealth 7 (4), e10967.

Elmi-Terander, A., Burström, G., Nachabe, R., Skulason, H., Pedersen, K., Fagerlund, M., Ståhl, F., Charalampidis, A., Söderman, M., Holmin, S., et al., 2019. Pedicle screw placement using augmented reality surgical navigation with intraoperative 3D imaging: A first in-human prospective cohort study. Spine 44 (7), 517.

Ershad, M., Ahmadian, A., Dadashi Serej, N., Saberi, H., Amini Khoiy, K., 2014. Minimization of target registration error for vertebra in image-guided spine surgery. Int. J. Comput. Assist. Radiol. Surg. 9 (1), 29–38.

Esfandiari, H., Anglin, C., Guy, P., Street, J., Weidert, S., Hodgson, A.J., 2019. A comparative analysis of intensity-based 2D–3D registration for intraoperative use in pedicle screw insertion surgeries. Int. J. Comput. Assist. Radiol. Surg. 14, 1725–1739.

Faraji-Dana, Z., Mariampillai, A.L., Standish, B.A., Yang, V.X., Leung, M.K., 2020. Machine-vision image-guided surgery for spinal and cranial procedures. In: Handbook of Robotic and Image-Guided Surgery. Elsevier, pp. 551–574.

Farshad, M., Fürnstahl, P., Spirig, J.M., 2021a. First in man in-situ augmented reality pedicle screw navigation. North Am. Spine Soc. J. (NASSJ) 6, 100065.

Farshad, M., Spirig, J.M., Suter, D., Hoch, A., Burkhard, M.D., Liebmann, F., Farshad-Amacker, N.A., Fürnstahl, P., 2021b. Operator independent reliability of direct augmented reality navigated pedicle screw placement and rod bending. North Am. Spine Soc. J. (NASSJ) 8, 100084.

Felix, B., Kalatar, S.B., Moatz, B., Hofstetter, C., Karsy, M., Parr, R., Gibby, W., 2022. Augmented reality spine surgery navigation: increasing pedicle screw insertion accuracy for both open and minimally invasive spine surgeries. Spine 47 (12), 865–872.

Félix, I., Raposo, C., Antunes, M., Rodrigues, P., Barreto, J.P., 2021. Towards markerless computer-aided surgery combining deep segmentation and geometric pose estimation: application in total knee arthroplasty. Comput. Methods Biomech. Biomed. Eng.: Imaging Vis. 9 (3), 271–278.

Foundation, T.L., 2022. ONNX. https://onnx.ai/.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Marín-Jiménez, M., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognit. 47 (6), 2280–2292. http://dx.doi.org/10.1016/j.patcog.2014.01.005, URL http://www.sciencedirect.com/science/article/pii/S0031320314000235.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R., 2016. Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recognit. 51, 481–491. http://dx.doi.org/10.1016/j.patcog.2015.09.023, URL http://www.sciencedirect.com/science/article/pii/S0031320315003544.

Geiger, A., Wang, C., 2015. Joint 3d object and layout inference from a single rgb-d image. In: German Conference on Pattern Recognition. Springer, pp. 183–195.

Gelalis, I.D., Paschos, N.K., Pakos, E.E., Politis, A.N., Arnaoutoglou, C.M., Karageorgos, A.C., Ploumis, A., Xenakis, T.A., 2012. Accuracy of pedicle screw placement: A systematic review of prospective in vivo studies comparing free hand, fluoroscopy guidance and navigation techniques. Eur. Spine J. 21 (2), 247–255.

Gibby, J.T., Swenson, S.A., Cvetko, S., Rao, R., Javan, R., 2019. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. Int. J. Comput. Assist. Radiol. Surg. 14 (3), 525–535.

Gu, W., Shah, K., Knopf, J., Josewski, C., Unberath, M., 2021a. A calibration-free workflow for image-based mixed reality navigation of total shoulder arthroplasty. Comput. Methods Biomech. Biomed. Eng.: Imaging Vis. 1–9.

Gu, W., Shah, K., Knopf, J., Navab, N., Unberath, M., 2021b. Feasibility of image-based augmented reality guidance of total shoulder arthroplasty using microsoft HoloLens 1. Comput. Methods Biomech. Biomed. Eng.: Imaging Vis. 9 (3), 261–270.

Guennebaud, G., Jacob, B., et al., 2010. Eigen v3. http://eigen.tuxfamily.org.

Guha, D., Jakubovic, R., Gupta, S., Fehlings, M.G., Mainprize, T.G., Yee, A., Yang, V.X., 2019. Intraoperative error propagation in 3-dimensional spinal navigation from nonsegmental registration: A prospective cadaveric and clinical study. Glob. Spine J. 9 (5), 512–520.

Guney, F., Geiger, A., 2015. Displets: Resolving stereo ambiguities using object knowledge. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (CVPR), pp. 4165–4175.

Härtl, R., Lam, K.S., Wang, J., Korge, A., Kandziora, F., Audigé, L., 2013. Worldwide survey on the use of navigation in spine surgery. World Neurosurg. 79 (1), 162–172.

Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N., 2013. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In: Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11. Springer, pp. 548–562.

Holly, L.T., Foley, K.T., 2007. Image guidance in spine surgery. Orthop. Clin. North Am. 38 (3), 451–461.

Hu, X., et al., 2022. Automatic bone surface restoration for markerless computer-assisted orthopaedic surgery. Chin. J. Mech. Eng. 35 (1), 1–6.

Ji, S., Fan, X., Paulsen, K.D., Roberts, D.W., Mirza, S.K., Lollis, S.S., 2015. Patient registration using intraoperative stereovision in image-guided open spinal surgery. IEEE Trans. Biomed. Eng. 62 (9), 2177–2186.

Joskowicz, L., Hazan, E.J., 2016. Computer aided orthopaedic surgery: Incremental shift or paradigm change? Med. Image Anal. 100 (33), 84–90.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lamartina, C., Cecchinato, R., Fekete, Z., Lipari, A., Fiechter, M., Berjano, P., 2015. Pedicle screw placement accuracy in thoracic and lumbar spinal surgery with a patient-matched targeting guide: A cadaveric study. Eur. Spine J. 24 (7), 937–941.

Langlois, J., Mouchère, H., Normand, N., Viard-Gaudin, C., 2018. 3D orientation estimation of industrial parts from 2d images using neural networks. In: International Conference on Pattern Recognition Applications and Methods. pp. 409–416.

Lee, S.C., Fuerst, B., Tateno, K., Johnson, A., Fotouhi, J., Osgood, G., Tombari, F., Navab, N., 2017. Multi-modal imaging, model-based tracking, and mixed reality visualisation for orthopaedic surgery. Healthc. Technol. Lett. 4 (5), 168–173.

Léger, É., Drouin, S., Collins, D.L., Popa, T., Kersten-Oertel, M., 2017. Quantifying attention shifts in augmented reality image-guided neurosurgery. Healthc. Technol. Lett. 4 (5), 188–192.

Liebmann, F., Roner, S., von Atzigen, M., Scaramuzza, D., Sutter, R., Snedeker, J., Farshad, M., Fürnstahl, P., 2019. Pedicle screw navigation using surface digitization on the microsoft HoloLens. Int. J. Comput. Assist. Radiol. Surg. 14 (7), 1157–1165.

Liebmann, F., Stütz, D., Suter, D., Jecklin, S., Snedeker, J.G., Farshad, M., Fürnstahl, P., Esfandiari, H., 2021. SpineDepth: A multi-modal data collection approach for automatic labelling and intraoperative spinal shape reconstruction based on RGB-D data. J. Imaging 7 (9), 164.

Liu, A., Jin, Y., Cottrill, E., Khan, M., Westbroek, E., Ehresman, J., Pennington, Z., Sheng-fu, L.L., Sciubba, D.M., Molina, C.A., et al., 2021. Clinical accuracy and initial experience with augmented reality–assisted pedicle screw placement: the first 205 screws. J. Neurosurg.: Spine 1 (aop), 1–7.

Liu, W., Ren, H., Zhang, W., Song, S., 2013. Cognitive tracking of surgical instruments based on stereo vision and depth sensing. In: 2013 IEEE International Conference on Robotics and Biomimetics. (ROBIO), IEEE, pp. 316–321.

Liu, H., Wang, Y., Pi, B., Qian, Z., Zhu, X., Yang, H., 2017. Comparison of intraoperative O-arm-and conventional fluoroscopy (c-arm)-assisted insertion of pedicle screws in the treatment of fracture of thoracic vertebrae. J. Orthopaed. Surg. 25 (1), 2309499016684090.

Ma, L., Fan, Z., Ning, G., Zhang, X., Liao, H., 2018. 3D visualization and augmented reality for orthopedics. Intell. Orthopaed.: Artif. Intell. Smart Image-guided Technol. Orthop. 193–205.

Ma, S., Zhao, Z., 2017. A new method of surgical tracking system based on fiducial marker. In: Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21. Springer, pp. 886–896.

Ma, L., Zhao, Z., Chen, F., Zhang, B., Fu, L., Liao, H., 2017. Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: A pilot study. Int. J. Comput. Assist. Radiol. Surg. 12 (12), 2205–2215.

Mahendran, S., Ali, H., Vidal, R., 2017. 3D pose regression using convolutional neural networks. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2174–2182.

Manni, F., Elmi-Terander, A., Burström, G., Persson, O., Edström, E., Holthuizen, R., Shan, C., Zinger, S., van der Sommen, F., de With, P.H., 2020. Towards optical imaging for spine tracking without markers in navigated spine surgery. Sensors 20 (13), 3641.

Markelj, P., Tomaževič, D., Likar, B., Pernuš, F., 2012. A review of 3D/2D registration methods for image-guided interventions. Med. Image Anal. 16 (3), 642–661.

Martin-Gomez, A., Li, H., Song, T., Yang, S., Wang, G., Ding, H., Navab, N., Zhao, Z., Armand, M., 2023. STTAR: surgical tool tracking using off-the-shelf augmented reality head-mounted displays. IEEE Trans. Vis. Comput. Graphics.

Miao, S., Wang, Z.J., Liao, R., 2016. A CNN regression approach for real-time 2D/3D registration. IEEE Trans. Med. Imaging 35 (5), 1352–1363.

Modi, H.N., Suh, S.W., Fernandez, H., Yang, J.H., Song, H.-R., 2008. Accuracy and safety of pedicle screw placement in neuromuscular scoliosis with free-hand technique. Eur. Spine J. 17 (12), 1686–1696.

Molina, C.A., Theodore, N., Ahmed, A.K., Westbroek, E.M., Mirovsky, Y., Harel, R., Khan, M., Witham, T., Sciubba, D.M., et al., 2019. Augmented reality–assisted pedicle screw insertion: A cadaveric proof-of-concept study. J. Neurosurg.: Spine 31 (1), 139–146.

Müller, F., Roner, S., Liebmann, F., Spirig, J.M., Fürnstahl, P., Farshad, M., 2020. Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging. Spine J. 20 (4), 621–628.

Muñoz-Salinas, R., Marín-Jimenez, M.J., Yeguas-Bolivar, E., Medina-Carnicer, R., 2018. Mapping and localization from planar markers. Pattern Recognit. 73 (Supplement C), 158–171. http://dx.doi.org/10.1016/j.patcog.2017.08.010, URL http://www.sciencedirect.com/science/article/pii/S0031320317303151.

Nadeau, M., Batke, J., Fisher, C., Street, J., 2015. A qualitative web-based expert opinion analysis on the adoption of intraoperative CT and navigation systems in spine surgery. Glob. Spine J. 5 (1_suppl), s–0035.

Navab, N., Heining, S.-M., Traub, J., 2009. Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications. IEEE Trans. Med. Imaging 29 (7), 1412–1423.

Nottmeier, E.W., Crosby, T.L., 2007. Timing of paired points and surface matching registration in three-dimensional (3D) image-guided spinal surgery. Clin. Spine Surg. 20 (4), 268–270.

NVIDIA, 2021. Tensorrt. https://developer.nvidia.com/tensorrt.

NVIDIA, 2022. CUDA. https://developer.nvidia.com/cuda-toolkit.

Olson, E., 2011. AprilTag: A robust and flexible visual fiducial system. In: 2011 IEEE International Conference on Robotics and Automation. IEEE, pp. 3400–3407.

Papadopoulos, E.C., Girardi, F.P., Sama, A., Sandhu, H.S., Cammisa, Jr., F.P., 2005. Accuracy of single-time, multilevel registration in image-guided spinal surgery. Spine J. 5 (3), 263–267.

Perdomo-Pantoja, A., Ishida, W., Zygourakis, C., Holmes, C., Iyer, R.R., Cottrill, E., Theodore, N., Witham, T.F., Sheng-fu, L.L., 2019. Accuracy of current techniques for placement of pedicle screws in the spine: A comprehensive systematic review and meta-analysis of 51,161 screws. World Neurosurg. 126, 664–678.

Qian, L., Unberath, M., Yu, K., Fuerst, B., Johnson, A., Navab, N., Osgood, G., 2017. Towards virtual monitors for image guided interventions-real-time streaming to optical see-through head-mounted displays. arXiv preprint arXiv:1710.00808.

Rampersaud, Y.R., Simon, D.A., Foley, K.T., 2001. Accuracy requirements for image-guided spinal pedicle screw placement. Spine 26 (4), 352–359.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm. In: Proceedings Third International Conference on 3-D Digital Imaging and Modeling. IEEE, pp. 145–152.

Rusu, R.B., Cousins, S., 2011. 3D is here: Point cloud library (PCL). In: IEEE International Conference on Robotics and Automation. (ICRA), IEEE, Shanghai, China, pp. 1–4.

Sundar, H., Khamene, A., Xu, C., Sauer, F., Davatzikos, C., 2006. A novel 2D-3D registration algorithm for aligning fluoro images with 3D pre-op CT/MR images. In: Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display, Vol. 6141. SPIE, pp. 760–766.

Tulsiani, S., Malik, J., 2015. Viewpoints and keypoints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1510–1519.

Uddin, S.-A., Hanna, G., Ross, L., Molina, C., Urakov, T., Johnson, P., Kim, T., Drazin, D., 2021. Augmented reality in spinal surgery: highlights from augmented reality lectures at the emerging technologies annual meetings. Cureus 13 (10).

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13 (04), 376–380.

Unberath, M., Gao, C., Hu, Y., Judish, M., Taylor, R.H., Armand, M., Grupp, R., 2021. The impact of machine learning on 2d/3d registration for image-guided interventions: A systematic review and perspective. Front. Robotics AI 8, 716007.

van Dijk, J.D., van den Ende, R.P., Stramigioli, S., Köchling, M., Höss, N., 2015. Clinical pedicle screw accuracy and deviation from planning in robot-guided spine surgery: robot-guided pedicle screw accuracy. Spine 40 (17), E986–E991.

von Atzigen, M., Liebmann, F., Hoch, A., Spirig, J.M., Farshad, M., Snedeker, J., Fürnstahl, P., 2022. Marker-free surgical navigation of rod bending using a stereo neural network and augmented reality in spinal fusion. Med. Image Anal. 77, 102365.

Wang, J., Olson, E., 2016. AprilTag 2: Efficient and robust fiducial detection. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. (IROS), IEEE, pp. 4193–4198.

Wolf, J., Luchmann, D., Lohmeyer, Q., Farshad, M., Fürnstahl, P., Meboldt, M., 2023. How different augmented reality visualizations for drilling affect trajectory deviation, visual attention, and user experience. Int. J. Comput. Assist. Radiol. Surg. 1–9.

Woo, M., Neider, J., Davis, T., Shreiner, D., 1999. OpenGL Programming Guide: The Official Guide To Learning OpenGL, Version 1.2. Addison-Wesley Longman Publishing Co., Inc..

Zhang, J.N., Fan, Y., Hao, D.J., 2019. Risk factors for robot-assisted spinal pedicle screw malposition. Sci. Rep. 9 (1), 1–6.