Inside-Out Instrument Tracking for Surgical Navigation in Augmented Reality

Christina Gsaxner
Antonio Pepe
Dieter Schmalstieg
Graz University of Technology
Graz, Austria

Jianning Li
Jan Egger
Essen University Hospital
Essen, Germany

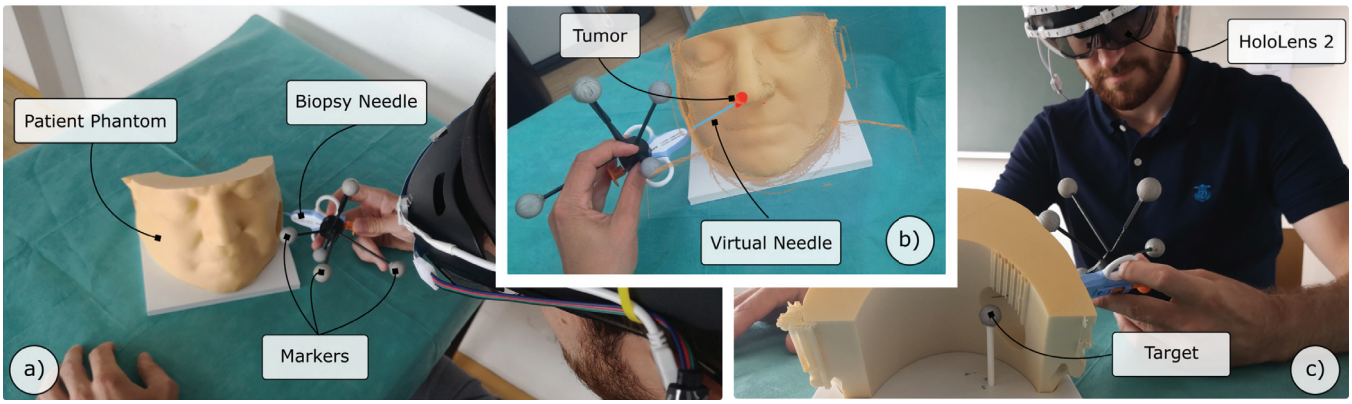

Figure 1: (a) Our inside-out tracking system enables the localization of tools fitted with reflective sphere markers purely through the HoloLens 2. This example shows a simulated core needle biopsy using our system. (b) The user aligns the virtual needle tip with the center of the tumor, as seen through the headset. (c) This allows the user to hit an otherwise occluded anatomical landmark.

ABSTRACT

Surgical navigation requires tracking of instruments with respect to the patient. Conventionally, tracking is done with stationary cameras, and the navigation information is displayed on a stationary display. In contrast, an augmented reality (AR) headset can superimpose surgical navigation information directly in the surgeon's view. However, AR needs to track the headset, the instruments and the patient, often by relying on stationary infrastructure. We show that 6DOF tracking can be obtained without any stationary, external system by purely utilizing the on-board stereo cameras of a HoloLens 2 to track the same retro-reflective marker spheres used by current optical navigation systems. Our implementation is based on two tracking pipelines complementing each other, one using conventional stereo vision techniques, the other relying on a single-constraint-at-a-time extended Kalman filter. In a technical evaluation of our tracking approach, we show that clinically relevant accuracy of 1.70 mm/1.11° and real-time performance is achievable. We further describe an example application of our system for untethered end-to-end surgical navigation.

VRST '21, Dec 07–10, 2021, Osaka, Japan

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

CCS CONCEPTS

• **Computing methodologies** → **Tracking**; *Mixed / augmented reality*; • **Applied computing** → Imaging.

KEYWORDS

Augmented Reality, Tracking, Surgical Navigation, HoloLens 2

ACM Reference Format:
Christina Gsaxner, Antonio Pepe, Dieter Schmalstieg, Jianning Li, and Jan Egger. 2021. Inside-Out Instrument Tracking for Surgical Navigation in Augmented Reality. In *VRST '21: ACM Symposium on Virtual Reality Software and Technology, Dec 07–10, 2021, Osaka, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Recent advances in medical imaging technology focus on integrating real-time information into clinical procedures, making interventions faster, safer and less invasive [49, 66]. Image guided interventions (IGI), such as ultrasound during core needle biopsies [52] or computed tomography (CT) during vascular procedures [58], are considered the gold standard nowadays. Surgical navigation systems (SNS) have also made their way into standard clinical care. Their main purpose is to determine the position of medical tools in relation to the patient's anatomy, thereby aiding the physician

in targeting (or avoiding) critical structures such as blood vessels, nerves or tumors. Most commercially available SNS use infrared (IR) emitting stereoscopic cameras [34], which can accurately track reflective marker spheres attached to surgical instruments and the patient. This allows the determination of the relative localization between instrument and anatomy in real-time. Nonetheless, current IGI systems still face limitations. Intra-operative X-ray or CT guidance burdens operators and patients with increased radiation exposure, while ultrasound guidance is only applicable for shallow structures and certain tissue types. Further, image guidance always increases the complexity of interventions. Conventionally, guidance information is displayed on an external 2D screen placed around the operating site. Dividing their attention between screen and patient disrupts the physician's focus, increases mental workload [46], and the misalignment between the operator's viewpoint and the navigation information deteriorates hand-eye coordination [27]. Therefore, image-guided procedures demand the highest levels of experience and concentration from physicians.

Augmented reality (AR) promises to overcome these limitations by displaying navigation data directly in the surgeon's view of the operating site. In particular, optical see-through head-mounted display (OST-HMD) can alleviate issues of focus switching and misinterpretation of information, while keeping the hands of the surgeon unoccupied. Moreover, compared to commercial SNS, C-Arm X-ray machines or similar devices, an OST-HMD is inexpensive, compact and flexible. Thus, an image guidance system built on top of an OST-HMD can not only enhance existing procedures with *in situ* AR visualization, but can enable navigation for procedures usually performed without image guidance, even at the bedside of the patient [2]. However, addressing the elevated requirements concerning accuracy and reliability of medical procedures with current AR hardware is challenging [66], and the AR scenario is complicated by the need to track the AR device in addition to instruments and patient. Consequently, existing AR-SNS prototypes often still rely on the integration of external tracking, using the AR device as a display only [15, 17, 21, 36, 42, 48, 63, 72]. Such a setup suffers from a bulky form factor and high cost.

In this contribution, we present an AR-SNS for a commercial OST-HMD, the HoloLens 2 (Microsoft, Redmond, WA, USA). We describe a method for the tracking of instruments fitted with reflective sphere markers, as commonly used by conventional SNS, using solely the HoloLens' built-in hardware. By implementing a robust and fast tracking algorithm, instruments can be localized in real-time and with unrestricted six degrees of freedom (6DOF). The HoloLens 2 is equipped with four environmental understanding cameras, mainly used by the built-in *simultaneous localization and mapping* (SLAM) algorithm of the device, which tracks the HoloLens itself with respect to a static environment. We re-use the streams of the two front-facing cameras to track sphere markers using robust stereo vision techniques, combining 3D triangulation and a single-constraint-at-a-time extended Kalman filter (SCAAT-EKF). We combine this inside-out instrument tracking with a markerless, untethered patient registration pipeline [23] to enable surgical navigation. Our AR system is the first drop-in replacement for an existing SNS: A surgeon wearing the HoloLens uses the familiar marked instruments, but without requiring external tracking.

2 RELATED WORK

Tracking for augmented environments. Most works about tracking in AR focus on the self-localization of the AR device within its environment and are therefore not directly applicable to our use case of high precision, real-time object tracking. Still, since we draw inspiration from work in this area, we provide a brief overview. In general, two tracking paradigms can be distinguished in AR: Outside-in tracking, on the one hand, relies on external sensors, placed around the user, observing movements from the outside. Inside-out tracking, on the other hand, directly uses sensors on the AR device itself, allowing it to work in unprepared environments. Although a large variety of tracking technologies exist (e.g., GPS, electromagnetic), vision-based technologies are most prevalent, due to their high flexibility, accuracy and update rates [33]. Early works extensively studied multi-view configurations for optical tracking of passively reflecting [13, 16, 59] or actively emitting [3, 69] markers, which are now well-established in commercial outside-in tracking systems and SNS. They typically require expensive, specialized hardware to ensure proper sensor synchronization and minimal lag. Algorithmic solutions can, to some extend, address this limitation, e.g., SCAAT-EKF can track with unsynchronized hardware [57, 68]. Still, stationary infrastructure restricts usage to delimited environments. In the past years, monoscopic inside-out tracking in the visible light range has gained considerable popularity due to the availability of cameras and increasing computational power of consumer devices. Marker-based inside-out approaches are facilitated by fiducial image tracking libraries like ARToolKit [32], ArUco [22] or PTC Vuforia [54]. While this enables AR on mobile devices, tracking error of such systems is typically in the range of several millimetres up to centimetres [1, 8, 10]. Marker-free inside-out systems mostly use SLAM [9, 33], i.e., they map a the static environment and localize the device against it. However, SLAM is not directly applicable for highly precise, dynamic object tracking [45]. Hybrid methods introduce additional sensors, such as inertial measurement units [53, 70], depth cameras [5] or combined sensor units [26, 67]. This adds complexity to the system, often at the expense of flexibility, and accurate calibration of all sensors is essential for precision [41]. The combined requirement of high accuracy, real-time object tracking and mobile self-contained operation rules out many established tracking techniques for our scenario. Therefore, in the following, we concentrate on related work in the same domain as ours: medical tool tracking using the HoloLens.

Outside-in approaches. Frequently, a stationary outside-in SNS is used to localize patients, instruments and HMD in a common coordinate system. In this scenario, the HMD only serves as a display. As IR-based optical tracking is most common in SNS; several systems combining it with the HoloLens have been proposed for a variety of surgical applications from orthopedics to general surgery [15, 17, 42, 48, 63, 72]. SNS based on electromagnetic tracking have also been explored together with the HoloLens [21, 36]. A slightly different course of action is taken by Liu et al. [44], who perform image-based catheter tracking based on intraoperative X-ray images. Outside-in AR systems rely on expensive, external hardware. Several, often bulky, components are needed, which makes such systems inflexible. While they may be able to fulfill the high demands in precision and reliability of medical tasks, they require

an accurate calibration between AR device and SNS, which is time-consuming and error-prone. Further, optical tracking requires a constant line of sight between the stationary camera and tracking targets, which restricts the freedom of movement of the physician, especially when multiple targets are used (e.g., when an additional AR device needs to be tracked). Electromagnetic tracking as an alternative to optical tracking is even less popular, as it suffers from interference with metal in the operating room [34].

Inside-out approaches. Inside-out tracking eliminates the dependence on additional hardware by relying on the AR device itself, thus, overcoming the problems mentioned above. The HoloLens features a variety of built-in sensors, which can be exploited for inside-out tracking. In medical scenarios, most commonly, the central color camera of the HoloLens has been used to track fiducial image markers affixed to surgical tools [11, 20, 35, 40, 50, 55, 56]. However, as already mentioned, the error of commonly used fiducial tracking approaches is too high for most medical applications. Using the two front-facing grey-scale tracking cameras of the HoloLens to triangulate the marker pose might improve tracking accuracy [43]. Independent of the camera hardware, a common drawback of using flat image markers is that they need to constantly face the camera, which restricts movement of the instrument, especially in terms of rotation. Further, image fiducials are sensitive to occlusions and varying lighting conditions. The HoloLens also features a time-of-flight (ToF) camera operating in the IR range. Kunz et al. [38] showed how this camera can be used for tracking spherical markers. Unfortunately, they only evaluated the relative translation error, and it remains unclear if their approach actually supports six degrees of freedom. Moreover, the HoloLens 1 depth sensor is known to suffer from noise and distortions [23, 30].

3 METHOD

Our novel marker tracking runs directly on the HoloLens 2. It computes the 6DOF pose of a marked medical instrument M with respect to a stationary world frame W. The HoloLens 2 is equipped with four built-in tracking cameras, which deliver grey-scale frames of 480×640 resolution at approximately 20-30 Hz and can be accessed through the devices *research mode* [65]. These cameras feed into the built-in SLAM algorithm of the HoloLens. We re-purpose the left and right front-facing cameras, which we will denote L and R, respectively, as a stereo setup for the tracking of instruments.

The computation of 3D positions from stereo image pairs assumes that measurements in the two cameras are collected at the same time. Unfortunately, L and R are not synchronized. Hence, our tracking algorithm consists of two paths: (1) In case two sufficiently synchronized frames are available, a stereo vision algorithm based on 3D triangulation can be executed. (2) Otherwise, a recursive single-constraint-at-a-time Extended Kalman Filter (SCAAT-EKF) [68] is used to obtain an estimate of the instrument's pose. Since we found that SCAAT-EKF is also more accurate and computationally efficient, we use it as our main tracking method. We use the stereo vision path only for initialization and re-initialization when tracking through SCAAT-EKF is lost.

We implemented the proposed tracking system in C++/WinRT using Direct3D 11, HoloLens 2 Research Mode API [65] and OpenCV library [7]. Our implementation will be available at https://github.com/cgsaxner/HL2ToolTracking.

3.1 Stereo calibration

For both tracking paths, accurate calibration of the stereo setup within the HoloLens 2 is a prerequisite. While the HoloLens 2 does expose sensor coordinates and mapping functions [65], from which intrinsic and extrinsic camera matrices could be obtained, we found these parameters to be too inaccurate. Therefore, we use the Camera Calibration Toolbox for MATLAB [6] together with a calibration target designed for cameras with low resolution [18] to compute the intrinsic parameters $\mathbf{K}_L$ and $\mathbf{K}_R$, the distortion coefficients k_L and k_R, and the transformation of the right camera with respect to the left, ${}^{L}\mathbf{T}_R$. To fully characterize our stereo setup, we further compute the fundamental matrix $\mathbf{F} = \mathbf{K}_R^{-T}\mathbf{E}\mathbf{K}_L^{-1}$, with $\mathbf{E} = {}^{L}\mathbf{R}_R[{}^{L}\mathbf{t}_R]_x$ being the essential matrix, ${}^{L}\mathbf{R}_R$, the rotation matrix, and $[{}^{L}\mathbf{t}_R]_x$, the matrix representation of the cross product with the translation vector ${}^{L}\mathbf{t}_R$ of ${}^{L}\mathbf{T}_R$. The calibration and computation of camera parameters is performed once for each device in an offline step.

3.2 Trackable instrument definition

We define a trackable medical instrument by a set of marker spheres $\mathbf{M} = \{\mathbf{m}_1, ..., \mathbf{m}_n\}$ attached to it, constituting a rigid body. While, theoretically, three markers are sufficient to estimate the pose of a rigid body, we use four or five marker spheres, so that occlusions or overlaps of individual markers can be handled. To allow an unambiguous identification of marker spheres, we require the relative, pairwise distances between markers $d(\mathbf{m}_i, \mathbf{m}_j)$, for $1 \leq i < j \leq n$, to be unique. The 3D marker coordinates $\mathbf{m}_1, ..., \mathbf{m}_n$ are expressed in a common reference frame, which can be chosen to have its origin at a point of interest on the instrument, such as the tip of a tracked needle. The configuration of $\mathbf{M}$ can be measured manually, with an optical measuring system, or, for commercial tools, may be provided by the manufacturer. Instrument definition is performed in an offline step and provided to the tracking algorithm.

3.3 Self-localization

We seek to compute ${}^{W}\mathbf{T}_M$, the transformation of the marked instrument M with respect to a stationary world coordinate system W. The HoloLens 2 runs an algorithm for SLAM [51], to build a map of the environment and simultaneously locate the device within this environment. The HoloLens SLAM uses L as a reference frame [65], therefore, we obtain the pose ${}^{W}\mathbf{T}_L$, which rigidly transforms from the left environmental tracking camera to the stationary world coordinate system W for every frame. We map the environment with the HoloLens SLAM before starting tracking, and set W at the devices' position at the application start. Using ${}^{L}\mathbf{T}_R$ obtained through stereo calibration in step 3.1, ${}^{W}\mathbf{T}_R = {}^{W}\mathbf{T}_L\ {}^{L}\mathbf{T}_R$ is also known. Thus, we require ${}^{L/R}\mathbf{T}_M$ to calculate the instrument's world pose by

$$ {}^{W}\mathbf{T}_M = {}^{W}\mathbf{T}_{L/R}\ {}^{L/R}\mathbf{T}_M. \tag{1} $$

3.4 Marker detection

While the reflective spheres commonly used for tracked surgical instruments are not as distinguishable in the gray-scale images from L and R as in IR, a similar effect can be achieved by illuminating

them with an additional light source, such as a surgical headlight mounted on the HMD. Then, we segment images into background and marker regions by applying a median blur filter and an adaptive thresholding algorithm, followed by a connected component search to extract potential marker contours. These contours are filtered based on their included area and their elliptical shape to eliminate wrong detections, before the weighted center of mass of each potential marker is computed. This results in a set of image-space marker candidate coordinates, $\mathbf{x}'_L$ and $\mathbf{x}'_R$.

Operations such as adaptive thresholding and connected component search are computationally expensive when applied to full resolution images, which is problematic especially on mobile devices with limited computing capabilities. Therefore, we implemented several strategies to improve the runtime of this step. We compute the median filtering and adaptive thresholding algorithm on the HoloLens' GPU. The latter is done by constructing a Gaussian image pyramid from input images, and then scanning the lowest level of the pyramid for potential regions of interest (ROI) around markers using a global threshold. If such a ROI is found, an adaptive threshold is computed from its neighboring pixels. The region is then propagated to the full resolution image, where binarization using the new threshold and connected component search is carried out only in this ROI. After initialization, we use a linear Kalman filter [31] to track a ROI around known image positions per marker, and we apply the marker detection algorithm only to these regions.

3.5 Pose estimation through stereo vision

If sufficient potential markers have been identified in L and R frames, and if the time difference between the two frames is below one millisecond, the frames are candidates for instrument pose estimation through stereo vision. As mentioned, this step is only run in the beginning of tracking, when SCAAT-EKF is not yet initialized, or if tracking by SCAAT-EKF has been lost. An overview of the process is shown in Figure 2.

Marker matching. As a first step, we use the epipolar constraint $\mathbf{x}_R'^T \mathbf{F} \mathbf{x}'_L = 0$ to find inter-view correspondences between left and right camera images. Since measurement errors in the marker detection are unavoidable, the epipolar constraint is not fulfilled precisely. Therefore, all markers lying within a tolerance around the epipolar lines $\mathbf{l}_R = \mathbf{F}\mathbf{x}'_L$ and $\mathbf{l}_L = \mathbf{F}^T \mathbf{x}'_R$ are assigned as possible matches. This can lead to several matching options, which we all add to sets of possibly matched marker points, $\mathbf{x}_L$ and $\mathbf{x}_R$. We rely on subsequent steps to filter out erroneous correspondences.

3D triangulation. If sufficient potentially matching markers have been identified in L and R, their 3D positions are recovered using triangulation. The triangulation step reconstructs the set of 3D points $\mathbf{Y}$ that project on both $\mathbf{x}_L$ and $\mathbf{x}_R$. We compute the projection matrices $\mathbf{P}_L = \mathbf{K}_L\mathbf{I}$ and $\mathbf{P}_R = \mathbf{K}_R[{}^L\mathbf{R}_R|{}^L\mathbf{t}_R]$, with $\mathbf{I}$ being the identity matrix, which effectively sets the coordinate system of L as the reference coordinate system of the stereo setup. By the optimal triangulation algorithm of Hartley and Sturm [28], the 3D points $\mathbf{Y}$ can be estimated. Due to our definition of the projection matrices, $\mathbf{Y}$ is expressed with respect to L.

Rigid body fitting and pose estimation. After having recovered potential 3D marker points $\mathbf{Y}$, the last step is to estimate the 6DOF pose of the instrument with respect to the world coordinate frame. This can be accomplished by rigidly fitting the known 3D marker points on the medical instrument $\mathbf{M} = \{\mathbf{m}_1, ..., \mathbf{m}_n\}$ to the point cloud from the previous step, $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_m\}$. The main challenge in this step is to assign to each marker on the instrument a corresponding point in $\mathbf{Y}$. We perform this assignment based on the relative pairwise distances d between markers [60, 61]. Therefore, instrument markers must be arranged in a way that $d(\mathbf{m}_i, \mathbf{m}_j)$ are pairwise different for $1 \leq i < j \leq n$. For each new set of world points, we compute $d(\mathbf{y}_i, \mathbf{y}_j)$ for $1 \leq i < j \leq m$ and compare the distances with those of the instrument. The geometric difference of each pair is added into a correspondence matrix $\mathbf{C}^{m \times n}$, contributing to the score of each involved marker. From $\mathbf{C}$, the optimal assignment is computed by minimum-weight matching [37]. Finally, all assignments for which $|d_Y - d_M| > \delta_d$, where δ_d is a distance tolerance lower than the smallest difference between any two d_M, are rejected. If $N \geq 3$ matching point pairs between $\mathbf{Y}$ and $\mathbf{M}$ can be found, the instrument has successfully been detected, and its pose ${}^L\mathbf{T}_M$ can be computed by solving the absolute orientation problem using Umeyama's method [64].

Finally, the instrument's pose is transformed to world space using ${}^W\mathbf{T}_L$ from the HoloLens SLAM.

3.6 Pose estimation through SCAAT-EKF

An EKF represents a system state $\mathbf{z}$ based on measurements $\mathbf{x}$ observed over time in a predictor-corrector fashion, where the state at a given time is first estimated using an underlying nonlinear dynamic model and then corrected with a measurement function $\mathbf{h}$ and its Jacobian $\mathbf{H}$. While EKF relies on a full state description at every correction step, SCAAT-EKF [68] updates the state with partial or incomplete measurements as soon as they are available.

Since we track instruments with respect to a stationary world frame and, thus, head motion needs not be accounted for, we assume that the dynamics of the movement can be described by a Gaussian position-velocity model. Therefore, the filter's state $\mathbf{z}$ contains the instruments' current 6DOF pose ${}^W\mathbf{T}_M$ as position and orientation components, as well as the linear and angular velocity of the instrument. The state vector contains $n = 12$ elements $\mathbf{z} = [x, y, z, \dot{x}, \dot{y}, \dot{z}, \varphi, \theta, \psi, \dot{\varphi}, \dot{\theta}, \dot{\psi}]$, where (x, y, z) describe the position in Cartesian coordinates, and (φ, θ, ψ) describe incremental rotations in axis-angle representation. Similar to Welch [68], to avoid axis-angle singularities, we maintain the overall orientation of the instrument using an external quaternion representation $\mathbf{q} = (q_w, q_x, q_y, q_z)$, in which the incremental rotation is factored in after every update step and then reset. As a measurement, we use individual detected marker points in either one of the camera frames $\mathbf{x}'$ from step 3.4. Therefore, our measurement vector contains $m = 2$ elements, for the 2D image coordinates of identified markers. The implementation of a SCAAT-EKF does not only allow us to overcome a lack of synchronization between L and R, but also enables the estimation and update of ${}^W\mathbf{T}_M$ for every frame, as long as at least one measurement is available. The algorithm operates in four steps, as visualized in Figure 3.

State prediction. SCAAT-EKF is applied to each new image from L or R, for which at least one measurement $\mathbf{x}'$ is available from the marker detection (step 3.4). The current filter state $\mathbf{z}$,

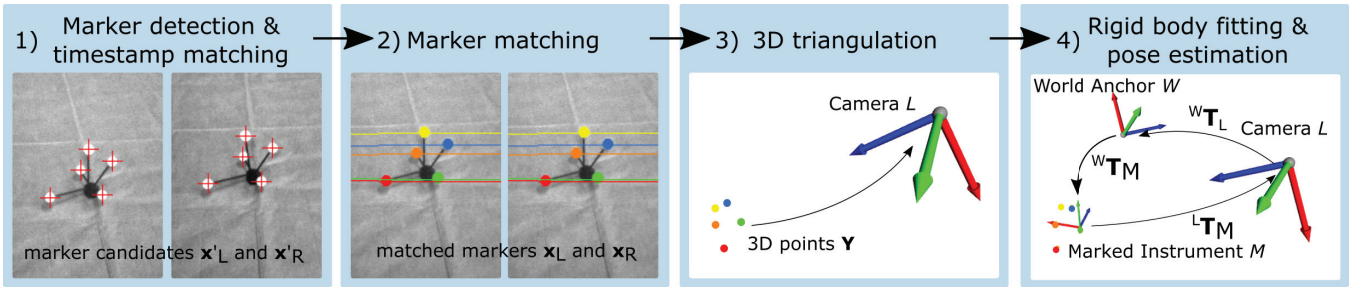

Figure 2: Workflow for pose estimation through stereo vision. (1) Marker candidates x'_L and x'_R are identified in closely acquired left and right frames and (2) matched via the epipolar constraint, yielding sets of matching marker candidates x_L and x_R. (3) From x_L and x_R, a set of 3D points Y is reconstructed via 3D triangulation. Since we choose L as reference coordinate frame, Y is expressed with respect to L. (4) Finally, known instrument markers are fitted to Y, yielding ${}^{L}\mathrm{T}_M$, and the transformation ${}^{W}\mathrm{T}_M$ from the marked instrument to the world frame is calculated using ${}^{W}\mathrm{T}_L$ from the HoloLens SLAM.

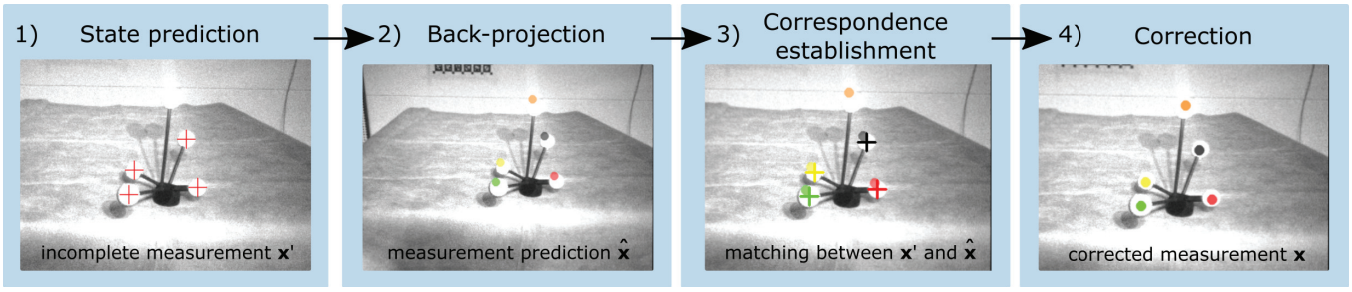

Figure 3: Pose estimation workflow through SCAAT-EKF. (1) For an incomplete measurement $\mathbf{x}'$ at a new time, the predicted state of the instrument is estimated. (2) This state is projected onto the camera frame, yielding a measurement prediction $\hat{\mathbf{x}}$. (3) Correspondence between candidate measurements $\mathbf{x}'$ and estimation $\hat{\mathbf{x}}$ are established. (4) The state is corrected using the residual $\Delta\mathbf{x} = \mathbf{x}' - \hat{\mathbf{x}}$.

describing the instrument's pose, is extrapolated to the new images' acquisition time, yielding the predicted state $\hat{\mathbf{z}} = \mathbf{A}(\delta t) \cdot \mathbf{z}(t - \delta t)$, where δt is the time since the last sample, and $\mathbf{A}(\delta t)$ is the $n \times n$ state transition matrix.

Back-projection. From the predicted filter state, an estimated measurement $\hat{\mathbf{x}}$ is computed using the measurement function $\mathbf{h}$ as $\hat{\mathbf{x}} = \mathbf{h}(\hat{\mathbf{z}}, \alpha)$, where α are additional system parameters. In our case, the measurement function is defined by the back-projection of the predicted instrument pose to the current frame, therefore $\hat{\mathbf{x}} = \mathbf{K}_{L/R}[{}^{L/R}\mathbf{R}_W|{}^{L/R}\mathbf{t}_W][{}^{W}\hat{\mathbf{R}}_M|{}^{W}\hat{\mathbf{t}}_M]\mathbf{M}$, using the intrinsics matrix of the current camera $\mathbf{K}_{L/R}$ as estimated in step 3.1, and the rotation matrix and translation vector describing the inverse camera pose ${}^{L/R}\mathbf{R}_W$ and ${}^{L/R}\mathbf{t}_W$ obtained from the HoloLens self-localization and stereo calibration, and the marker configuration $\mathbf{M}$ from step 3.2. ${}^{W}\hat{\mathbf{R}}_M$ and ${}^{W}\hat{\mathbf{t}}_M$ describe the current predicted pose of the instrument M, as described by the predicted filter state $\hat{\mathbf{z}}$.

Correspondence establishment. We now have a set of predicted, back-projected measurements $\hat{\mathbf{x}}$ and actual candidate measurements $\mathbf{x}'$. Correspondences between these two sets are established based on their distance in image space, by building a cost matrix of the Euclidean distances between any two possible matches and solving the assignment problem [37].

Correction. In the final correction step, we obtain an updated filter state $\mathbf{z}$ based on the measurement residual between correspondences $\Delta\mathbf{x} = \mathbf{x}' - \hat{\mathbf{x}}$ and the Kalman gain $\mathbf{K}$. The Kalman gain is computed from the $n \times n$ state error covariance matrix $\mathbf{P}(t)$, modelling the uncertainty in the current estimated state, the $m \times m$ measurement error covariance matrix $\mathbf{R}$, modelling the uncertainty in the obtained measurements and assumed constant in our case, and the measurement Jacobian $\mathbf{H}$ as $\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}$. The measurement Jacobian is the $m \times n$ matrix containing the first-order partial derivatives of the measurement function $\mathbf{h}$. Finally, the corrected filter state $\mathbf{z} = \hat{\mathbf{z}} + \mathbf{K} \cdot \Delta\mathbf{x}$ fully describes the desired world-space pose of the instrument ${}^{W}\mathbf{T}_M$ at the current time.

4 EXAMPLE: SURGICAL NAVIGATION

Tracking as described in this paper enables the localization of tools with respect to an arbitrary world frame, such as a patient's anatomy in case of a SNS. Consequently, for surgical navigation, we must also track the patient, who cannot be assumed to be stationary at a known location, even if sedated. Localization of the patient further permits "X-ray vision" of the patient anatomy in an AR display. Existing SNS usually rely on auxiliary methods, such as screwing a rigid marker to the bone of the patient followed by another medical scan [47] or digitizing landmarks using tracked tools [49]. Not only

are such method invasive or cumbersome, they may also incur additional radiation exposure or significant manual overhead.

In this example for end-to-end tracking of instrument and patient, we avoid such auxiliary methods entirely and combine the instrument tracking with the automatic, markerless patient localization proposed by Gsaxner et al. [23], which also runs on the HoloLens 2. The patient registration system acquires a point cloud representing the patient's skin surface P from the HoloLens' built-in depth sensor D and determines ${}^{D}\mathbf{T}_{P}$. We calibrate the transformation of D to the reference camera L using the same principles as described in Section 3.1 and subsequently use SLAM to localize D in the world space W. An automatic point cloud registration algorithm registers P with a model of the outer anatomy of the patient, denoted as I, resulting in the rigid transformation from imaging to patient ${}^{P}\mathbf{T}_{I}$. I is routinely obtained in pre-operative medical imaging, such as CT or magnetic resonance imaging. With these steps, the position of I with respect to the world coordinate system can be calculated as ${}^{W}\mathbf{T}_{I} = {}^{W}\mathbf{T}_{D}\ {}^{D}\mathbf{T}_{P}\ {}^{P}\mathbf{T}_{I}$, and the position of the instrument with respect to the patient imaging can be computed as ${}^{I}\mathbf{T}_{M} = ({}^{W}\mathbf{T}_{I})^{-1}\ {}^{W}\mathbf{T}_{M}$. Note that all three tracking systems (SLAM, instrument tracking, patient localization) run solely on the HoloLens and, thereby, deliver untethered end-to-end surgical navigation (Figure 4).

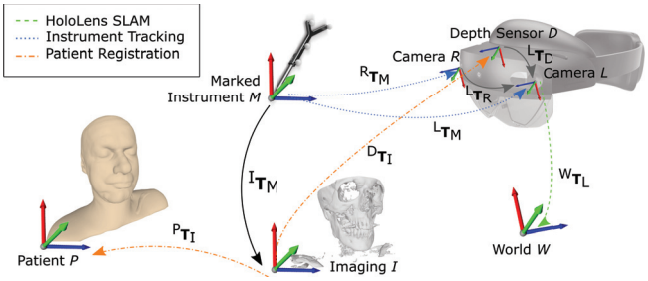

Figure 4: To compute the pose of a medical instrument in relation to the patient's head ${}^{I}\mathrm{T}_{M}$, we combine three localization systems on the HoloLens 2: HoloLens SLAM for self-localization, instrument tracking as described in this paper, and patient registration proposed by Gsaxner et al. [23]. ${}^{L}\mathrm{T}_{R}$ and ${}^{L}\mathrm{T}_{D}$ are obtained offline during calibration.

5 EXPERIMENTS AND RESULTS

5.1 Tracking Accuracy

To assess the accuracy of the proposed tracking algorithm, we compare the final transformation ${}^{W}\mathbf{T}_{M}$ obtained through our application with the equivalent ground truth transformation ${}^{W}\mathbf{S}_{M}$ computed by a high-precision, outside-in optical infrared motion capturing system. The tracking system consists of 15 OptiTrack Flex 13 cameras (NaturalPoint, Inc., Corvallis, OR, USA) and operates at sub-millimeter precision comparable to commercial SNS, thus, serving as our baseline. We mark an instrument with five 7.9 mm reflective marker spheres and rigidly attach a set of spheres H to the HoloLens. With these preparations, we measure ${}^{H}T_{M}$ using the OptiTrack. H needs to be calibrated with the HoloLens' world coordinate system to obtain the corresponding transformations. This can be formulated as a hand-eye calibration problem between the outside-in and inside-out tracking systems. We capture the same calibration target as employed in stereo calibration (see Section 3.1) with the left stereo camera L while moving the tracked HoloLens around the calibration target. Then, we use Daniilidis' algorithm [14] to solve for ${}^{L}\mathbf{T}_{H}$ and compute the equivalent transformation using ${}^{W}\mathbf{S}_{M} = {}^{W}\mathbf{T}_{L}\ {}^{L}\mathbf{T}_{H}\ {}^{H}\mathbf{T}_{M}$. The experimental setup is shown in Figure 5 (a).

Static and dynamic tracking. We perform static and dynamic experiments with this setup. Since medical instruments can be assumed to be operated at arm's length, we limit the tracking distance to 20-80 cm along the viewing direction of the headset. For the static experiment, the instrument and HoloLens were placed at 30 distinct poses. For the dynamic test, we evaluated five sequences while wearing the HoloLens and moving the instrument in front of the camera. For both scenarios, we compute the average root mean square (RMS) translation and rotation differences, T_{RMS} and R_{RMS}, between 50 (static) and 500 (dynamic) corresponding measurements of ${}^{W}\mathbf{T}_{M}$ from our tracking system and ${}^{W}\mathbf{S}_{M}$ from the OptiTrack.

For static experiments, the mean tracking error and standard deviation was $\mathbf{T_{RMS}}$**=1.70±0.81 mm** and $\mathbf{R_{RMS}}$**=1.11±0.39**°, while, for the dynamic experiments, the error was $\mathbf{T_{RMS}}$**=1.90±0.88 mm** and $\mathbf{R_{RMS}}$**=1.18±0.59** °. Error distributions are provided in Figure 6 (a, b). Since tracking accuracy is often highly dependent on the distance between target and camera [1, 8], we further analyze the tracking error in relation to the distance between instrument and HoloLens, as shown in Figure 6 (c, d).

Comparison to other tracking approaches. Tracking of fiducial image markers using the front-facing video camera of the HoloLens is a commonly employed technique in medical AR systems. Therefore, we compare our tracking system with monoscopic ArUco marker [22] tracking in a static scenario. We detect ArUco markers and recover their poses using the ArUco module in OpenCV [7] to obtain ${}^{V}\mathbf{T}_{A}$, the transformation of the ArUco marker A to the video camera V, which is transformed to world space using the pose information from the HoloLens SLAM.

The visual light video camera of the HoloLens 2 delivers frames at a resolution of 760×428 px at 30 fps. First, we follow the same calibration procedure as described in Section 3.1 to obtain the intrinsic matrix $\mathbf{K}_{V}$ and distortion coefficients k_{V} of this camera. Again, we use the OptiTrack for acquiring ground truth transformations and consequently need to relate the transformations acquired via inside-out tracking with this outside-in baseline. To this end, we rigidly attach a set of five retro-reflective spheres, trackable with the OptiTrack, to a planar ArUco marker of known size. To compare the transformation ${}^{W}\mathbf{T}_{A}$ with the equivalent transformation ${}^{W}\mathbf{S}_{A}$ from the OptiTrack, we require a further hand-eye type calibration between the coordinate frame of the retro-reflective spheres T and the coordinate frame of the marker. This time, we use a calibrated, high-resolution stationary camera and move the marker in front of it, capturing ArUco and OptiTrack poses at the same time. Finally, we again use Daniilidis' algorithm [14] to solve for ${}^{T}\mathbf{T}_{A}$. Furthermore, the hand-eye calibration of the HoloLens markers H must be repeated for the video camera V, yielding ${}^{V}\mathbf{T}_{H}$. Now, we can compute ${}^{W}\mathbf{S}_{A} = {}^{W}\mathbf{T}_{V}\ {}^{V}\mathbf{T}_{H}\ {}^{H}\mathbf{T}_{T}\ {}^{T}\mathbf{T}_{A}$, see Figure 5 (b).

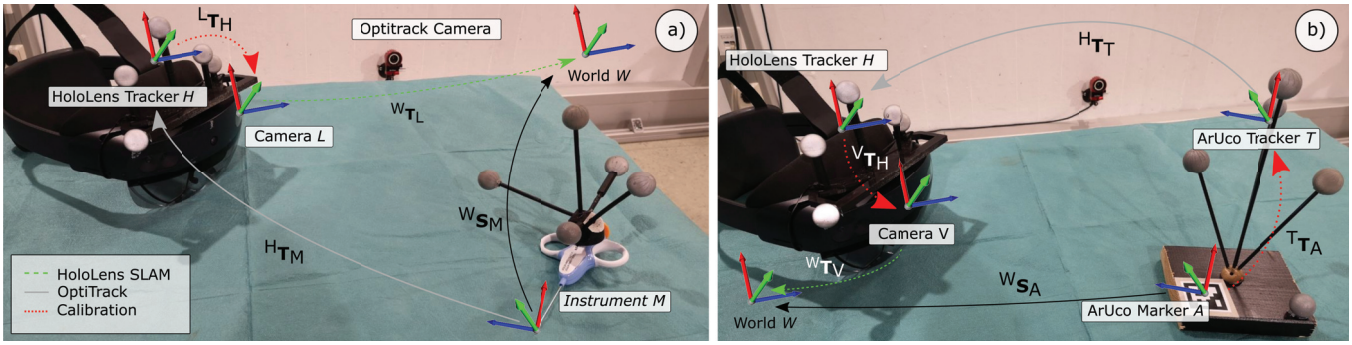

Figure 5: Experimental setup for accuracy evaluation. (a) Both HoloLens H and instrument M are tracked using the OptiTrack system, yielding ${}^{H}\mathrm{T}_{M}$. After calibrating for ${}^{L}\mathrm{T}_{H}$, the ground truth transformation ${}^{W}\mathrm{S}_{M}$ can be computed. (b) For the ArUco experiment, an additional calibration from fiducial to tracker, ${}^{T}\mathrm{T}_{A}$, is required to compute the ground truth ${}^{W}\mathrm{S}_{A}$.

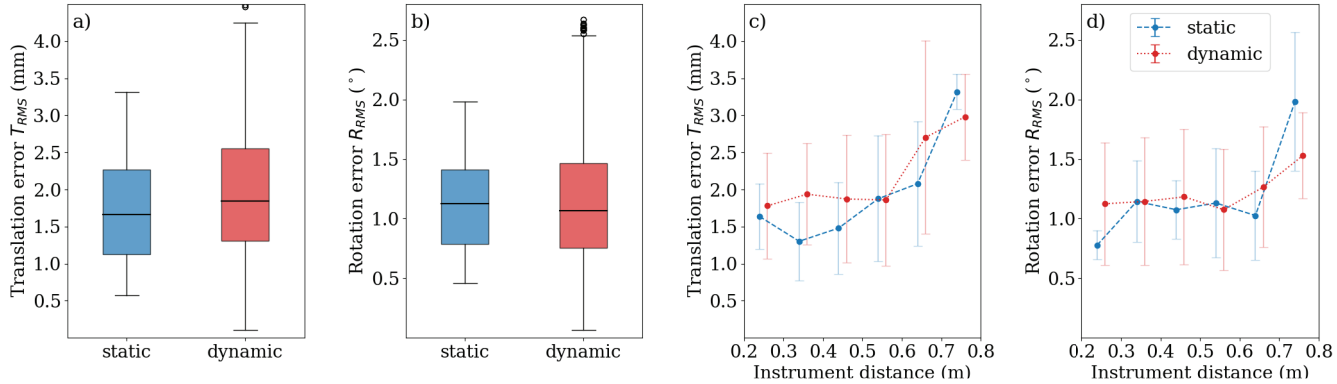

Figure 6: Experimental results of the accuracy evaluation of our SCAAT-EKF-based instrument tracker. (a, b) Distribution (median, 25th and 75th percentile, minimal and maximal errors) of translation error T_{RMS} and rotation error R_{RMS} compared to high-precision, outside-in tracking. (c, d) Mean and standard deviation of T_{RMS} and R_{RMS} over discrete distance intervals between instrument and camera. The error increases with distance.

For completeness, we also report the tracking error using stereo vision alone (no SCAAT), as described in Section 3.5 and used in our system for initialization. For both the ArUco and stereo tracking, we again place the marker in 30 distinct poses with respect to the HoloLens, and we average T_{RMS} and R_{RMS} over 50 frames for each pose. The results of this comparative study are provided in Table 1 and visualized in Figure 7.

Table 1: Root mean square (RMS) translation and rotation errors, T_{RMS} and R_{RMS}, for static tracking using our SCAAT-EKF based tracker, our stereo-vision only tracker, and monoscopic ArUco tracking, all in comparison to a high-precision, outside-in tracking system.

Method	$\mathrm{T}_{\mathrm{RMS}}$ (mm)	$\mathrm{R}_{\mathrm{RMS}}$ (°)
SCAAT	1.70 ± 0.81	1.11 ± 0.39
Stereo vision only	8.07 ± 0.91	0.90 ± 0.18
ArUco	6.09 ± 1.15	6.73 ± 3.47

5.2 Runtime

The runtimes of the system components, as well as the runtime of ArUco tracking, averaged over 500 frames, are reported in Table 2. All components were implemented and measured directly on a HoloLens 2. Overall, one pass through our tracking system takes approximately **27.1 ms** (~37 Hz) for initialization using full resolution (480×640) image pairs from L and R and stereo pose estimation, and **1.48 ms** (> 600 Hz) for incremental tracking using a region of interest (50×50) around known markers and SCAAT-EKF. For ArUco tracking using the 760×428 video camera stream of the HoloLens, one pass takes approximately 38.7 ms, or around 26 Hz.

5.3 Surgical Navigation

As a proof-of-concept for end-to-end surgical navigation (Section 4), we simulated a core needle biopsy of the skull base. Diagnosis of pathologies via biopsy is gold standard in the treatment of lesions in the head and neck area [71]. Usually, ultrasound-guidance is recommended for core needle biopsies in this region [52]. However, for deep lesions, such as at the skull base, ultrasound cannot penetrate to the required depth. CT guidance is a viable alternative [25], but

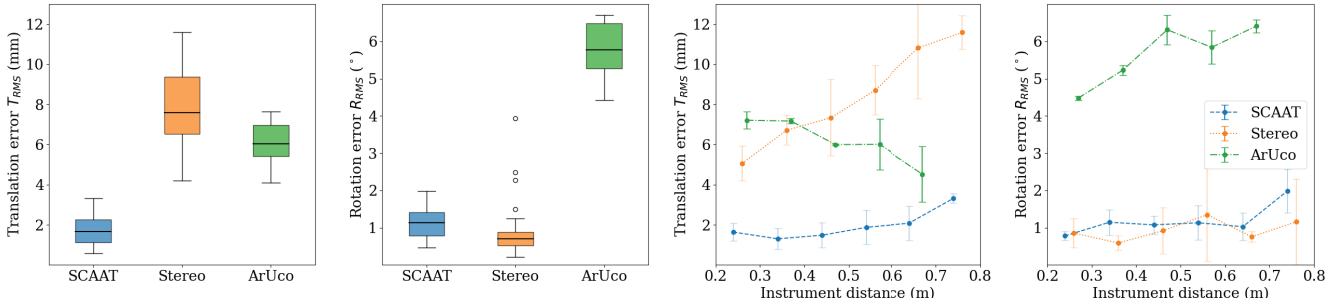

Figure 7: Visualization of results of the comparative study between our SCAAT-EKF based tracker (blue), our stereo-vision only tracker (orange), and monoscopic ArUco tracking (green). (a, b) Distribution (median, 25th and 75th percentile, minimal and maximal errors) of translation error T_{RMS} and rotation error R_{RMS}. (c, d) Mean and standard deviation of T_{RMS} and R_{RMS} over discrete distance intervals between instrument and camera.

Table 2: Average runtimes of our algorithm components on the HoloLens 2. Times are averaged over 500 frames and given in milliseconds.

Task	Explanation	Detect	Pose	Total
Initialization	Full images, stereo	25.58	1.50	27.09
Incremental track	ROI, SCAAT-EKF	0.97	0.51	1.48
ArUco tracking	RGB images, OpenCV	37.15	1.58	38.73

implies additional radiation exposure for both patient and operator. If an open biopsy should be avoided, image guidance using an SNS offers an alternative [71].

To simulate such an intervention, we use a 3D printed patient phantom and corresponding medical imaging [24]. The phantom contains an occluded, spherical target with a diameter of 7.9 mm, accessible through a hole in the phantom's cheek, replicating a tumor in the skull base. The user first calibrates the display of the HoloLens 2 with the on-board calibration application. Then, the phantom is registered using the automatic image-to-patient registration, and the location of the tumor is visualized in the HoloLens 2. The user is tasked to align the tip of the biopsy needle, tracked and visualized through the HoloLens, with the tumor. This experimental setup is shown in Figure 1. Once the user is satisfied with the alignment, we again use the OptiTrack to measure the distance between the center of the spherical target and the needle tip, adjusted by 3.95 mm along the direction of the needle to account for the spherical target's diameter.

Table 3: Root mean square (RMS) distances between needle tip and target during navigation experiment. We compare a blind condition, in which the target is occluded, with a direct condition, in which the user is allowed to see the target.

Condition	E_{RMS} (mm)			
	Mean	SD	Min	Max
Blind	4.93	2.24	1.78	10.71
Direct	2.57	1.25	0.96	5.71

We performed this experiment in two conditions, a *blind* condition where the user could not see the tumor target inside the skull, and a *direct* condition, where the user could see the target, thereby largely eliminating human performance as a source of tracking error. We repeated the task 10 times per condition with different phantom and target positions, and we measured the average RMS difference between needle tip and target. The results of these experiments are reported in Table 3. The mean translation difference of the direct condition is 2.57 mm, compared to 1.70 mm for the instrument tracking alone, which suggests that the hybrid tracking of patient and instrument increases the error by only 0.87 mm.

6 DISCUSSION

We track instruments with a precision of 1.70 mm/1.11° in static poses, outperforming monoscopic image fiducial tracking by a large margin (see Table 1). We would like to point out that tracking of 3D sphere targets has several benefits over flat markers, aside from accuracy. First, a 3D target better leverages our stereo camera configuration. Second, it is more robust to partial occlusions and cannot suffer from poor viewing angles. Third, retro-reflective marker detection is less affected by challenging lighting conditions. Table 2 also shows that our tracking implementations are faster than naive ArUco tracking using OpenCV library, which, at ~26 Hz, does not reach the camera refresh rate of 30 Hz. We note that the HoloLens 2 offers access to several video profiles of its RGB camera in terms of resolution and frame rate – we choose a profile for our comparison which offers a fair trade-off between accuracy and run time. Higher resolution might allow a more accurate tracking of markers, but processing times increase dramatically, while a lower resolution improves runtime, at the expense of less reliable marker detection and pose estimation.

The tracking accuracy achievable with our method is partly restricted by HoloLens hardware. The geometric depth resolution of stereo systems is proportional to the squared distance d of the measured object and inversely proportional to stereo baseline and focal length [19]. As the HoloLens cameras were not designed for short-distance stereo tracking, the rather short baseline between L and R of ~97 mm and short focal length (~366 mm) limit the accuracy achievable by triangulation. We calculated a geometric resolution of 1.2 mm for near (d=20 cm) and 18 mm for far range (d=80 cm)

with our setup, which agrees with the average error we obtain with stereo triangulation only, as shown in Figure 7. SCAAT-EKF largely overcomes this limitation, since it minimizes the re-projection error of every single measurement, while implicitly filtering noise in both measurement and process (i.e., instrument movement). Still, as seen in Figures 6 and 7, the tracking error with SCAAT-EKF is proportional to the distance of the instrument, showing that the limited geometric depth resolution is still a considerable factor. As our application domain typically involves settings were instruments need to be tracked in the immediate vicinity of the headset, we consider the most accurate tracking range up to 70 cm as feasible. Since ArUco does not rely on stereo, it is largely unaffected by variations in instrument distance. However, it should be noted that with ArUco, we were only able to reliably track the chosen marker at distances of up to 70 cm. At larger distances, marker detection becomes too brittle with the available camera resolution.

For the dynamic experiments, the error was slightly higher at T_{RMS}=1.90 mm and R_{RMS}=1.18°. This is partly caused by self-occlusions of markers, which lead to wrong predictions by SCAAT-EKF based on erroneous correspondences, in particular, if they persist over several frames. Based on informal experiments, we estimate that, by improving the correspondence matching within SCAAT-EKF, e.g., by incorporating 3D information and information from previous frames, the dynamic error could be improved by about 0.1 mm/0.1°. Another problem is an under-sampling of motion, which can occur in cases of abrupt movement changes, when only few measurements are available. We plan on improving robustness in these cases by using a higher order motion model in SCAAT-EKF. It should be pointed out, however, that we found the frame rates of L and R delivered by the HoloLens 2 to be inconsistent - while they were on average between 20-30 Hz, they occasionally drop to as low as 1 Hz. In such cases, fast motion cannot be captured anymore, regardless of the motion model.

In our surgical navigation scenario, an RMS error of 2.6 mm between the tip of a tracked needle and the target was achieved in the direct condition. In the blind condition, which corresponds to the intended usage of the end-to-end system, the error was 4.9 mm. A large contributing factor to this error is likely the patient registration pipeline, which was reported to have an error of 3.9 mm/5.0° on a HoloLens 1, mainly caused by depth data quality and instabilities in the SLAM system [23]. While it can be expected that the improved hardware of the HoloLens 2 mitigates this error sources to a certain extent, registration accuracy remains an important factor. The larger error in the blind condition is also caused by human factors. First, although the HoloLens 2 performs automatic display calibration through eye tracking, the display was not designed for close interactions, and a slight perceived offset between real and virtual content remains, even in the case of perfect registration. Second, it has been shown that the success of targeting tasks in AR is heavily influenced by aspects such as the strategy of displaying navigation information [29], perception of virtual content [4], or haptics [12]. We consider our experiment a proof of concept, meant to demonstrate the basic technical capabilities of our system. Investigating ways of effectively dealing with human factors in medical applications of our system remains an important area of future work.

While inside-out *self-tracking* of AR devices, e.g., using SLAM, has been studied extensively, and strong benchmarks exist for its evaluation [9, 33, 62], there is no gold standard for evaluating *object tracking* in AR, which makes a direct comparison to related work challenging. The experimental setup and acquired metrics are usually highly dependent on the available hardware and application domain. We chose an outside-in optical tracking system as the baseline for our accuracy evaluation, which is comparable in accuracy to commercial SNS (<1 mm, 1°) and an approach also found in current literature [5]. Although the accuracy of OptiTrack is in the sub-millimeter range, for our AR application, the outside-in system and HoloLens need to be brought into the same coordinate frame. This calibration procedure introduces some uncertainty to our baseline. To evaluate its influence, another method for acquiring ground truth would be required, e.g., a high precision measuring board for static experiments and robotic arm for dynamic experiments [38, 48, 57], which are both currently not available to us.

7 CONCLUSION AND OUTLOOK

We present an inside-out 6DOF tracking method using a commercial, consumer-grade OST-HMD, the Microsoft HoloLens 2. Our tracking algorithm is based on the detection of reflective spheres, which are commonly used in optical SNS. The system requires no bulky hardware, such as external monitors or tracking cameras. Thus, it is efficient in cost and space consumption. After an initial calibration of the device is performed once, the setup requires no technical knowledge or lengthy preparation. We further describe an example application of our tracking method for surgical navigation during core needle biopsies of the skull base.

Our work shows that inside-out tracking can approach the performance of commercial SNS despite the limitations of mobile hardware. While it does not quite reach the reference precision of SNS (<1 mm, 1°), it comes close even when running on less-than-optimal hardware and without low level system access (i.e., with unsynchronized cameras). Lately, there have been efforts towards OST-HMD designs specifically for medical applications. By integrating a capable stereo setup into such devices, inside-out instrument tracking with sub-millimeter accuracy could be achievable with the presented approach. Hence, we plan to make our implementation publicly available. We think that the reported accuracy, together with our high flexibility compared to an SNS, already allows an application of our system for image-guided interventions were sub-millimeter preciseness is not required, such as core needle biopsy, needle ablation [39] or ventriculostomy [2]. In this context, we plan to further develop our biopsy application and conduct a user study with medical experts, investigating factors such as suitable visualization and usability in a clinical environment. Outside of medical settings, our tracking system can be seen as a drop-in replacement for applications where optical outside-in tracking, e.g., with OptiTrack, is used.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (KLI 678-B31 "enFaced") and the Austrian Research Promotion Agency (COMET K-Project 871132 "CAMed").

REFERENCES

[1] Daniel F Abawi, Joachim Bienwald, and Ralf Dorner. 2004. Accuracy in optical tracking with fiducial markers: an accuracy function for ARToolKit. In *International Symposium on Mixed and Augmented Reality*. IEEE, 260–261.

[2] Ehsan Azimi, Zhiyuan Niu, Maia Stiber, Nicholas Greene, Ruby Liu, Camilo Molina, Judy Huang, Chien-Ming Huang, and Peter Kazanzides. 2020. An Interactive Mixed Reality Platform for Bedside Surgical Procedures. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 65–75.

[3] Ronald Azuma and Gary Bishop. 1994. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*. 197–204.

[4] Christoph Bichlmeier, Felix Wimmer, Sandro Michael Heining, and Nassir Navab. 2007. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *International Symposium on Mixed and Augmented Reality*. IEEE, 129–138.

[5] Alireza Bilesan, Mohammadhasan Owlia, Saeed Behzadipour, Shuhei Ogawa, Teppei Tsujita, Shunsuke Komizunai, and Atsushi Konno. 2018. Marker-based motion tracking using Microsoft Kinect. *IFAC-PapersOnLine* 51, 22 (2018), 399–404.

[6] Jean-Yves Bouguet. 2004. Camera calibration toolbox for matlab. *http://www.vision.caltech.edu/bouguetj/calib_doc/index.html* (2004). accessed May 2021.

[7] Gary Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[8] Michael Brand, Lukas Antonio Wulff, Yogi Hamdani, and Thorsten Schüppstuhl. 2020. Accuracy of Marker Tracking on an Optical See-Through Head Mounted Display. In *Annals of Scientific Society for Assembly, Handling and Industrial Robotics*. Springer, 21–31.

[9] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics* 32, 6 (2016), 1309–1332.

[10] Andong Cao, Ali Dhanaliwala, Jianbo Shi, Terence P Gade, and Brian J Park. 2020. Image-based marker tracking and registration for intraoperative 3D image-guided interventions using augmented reality. In *SPIE Medical Imaging*, Vol. 11318.

[11] Marina Carbone, Sara Condino, Fabrizio Cutolo, Rosanna Maria Viglialoro, Oliver Kaschke, Ulrich W Thomale, and Vincenzo Ferrari. 2018. Proof of concept: wearable augmented reality video see-through display for neuro-endoscopy. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 95–104.

[12] Long Chen, Thomas W Day, Wen Tang, and Nigel W John. 2017. Recent developments and future challenges in medical mixed reality. In *International Symposium on Mixed and Augmented Reality*. IEEE, 123–135.

[13] Jaeyong Chung, Namgyu Kim, Jounghyun Kim, and Chan-Mo Park. 2001. Postrack: A low cost real-time motion tracking system for vr application. In *International Conference on Virtual Systems and Multimedia*. IEEE, 383–392.

[14] Konstantinos Daniilidis. 1999. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research* 18, 3 (1999), 286–298.

[15] Marcelo E de Oliveira, Henrique G Debarba, Alexandre Lädermann, Sylvain Chagué, and Caecilia Charbonnier. 2019. A hand-eye calibration method for augmented reality applied to computer-assisted orthopedic surgery. *International Journal of Medical Robotics and Computer Assisted Surgery* 15, 2 (2019), e1969.

[16] Klaus Dorfmüller. 1999. Robust tracking for augmented reality using retroreflective markers. *Computers & Graphics* 23, 6 (1999), 795–800.

[17] Houssam El-Hariri, Prashant Pandey, Antony J Hodgson, and Rafeef Garbi. 2018. Augmented reality visualisation for orthopaedic surgical guidance with pre-and intra-operative multimodal image data fusion. *Healthcare Technology Letters* 5, 5 (2018), 189–193.

[18] David Ferstl, Christian Reinbacher, Gernot Riegler, Matthias Rüther, and Horst Bischof. 2015. Learning Depth Calibration of Time-of-Flight Cameras. In *British Machine Vision Conference*. 102–1.

[19] David Gallup, Jan-Michael Frahm, Philippos Mordohai, and Marc Pollefeys. 2008. Variable baseline/resolution stereo. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[20] Yuan Gao, Li Lin, Gang Chai, and Le Xie. 2019. A feasibility study of a new method to enhance the augmented reality navigation effect in mandibular angle split osteotomy. *Journal of Cranio-Maxillofacial Surgery* 47, 8 (2019), 1242–1248.

[21] Verónica García-Vázquez, Felix Von Haxthausen, Sonja Jäckle, Christian Schumann, Ivo Kuhlemann, Juljan Bouchagiar, Anna-Catharina Höfer, Florian Matysiak, Gereon Hüttmann, Jan Peter Goltz, et al. 2018. Navigation and visualisation with HoloLens in endovascular aortic repair. *Innovative Surgical Sciences* 3, 3 (2018), 167–177.

[22] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.

[23] Christina Gsaxner, Antonio Pepe, Jürgen Wallner, Dieter Schmalstieg, and Jan Egger. 2019. Markerless image-to-face registration for untethered augmented reality in head and neck surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 236–244.

[24] Christina Gsaxner, Jürgen Wallner, Xiaojun Chen, Wolfgang Zemann, and Jan Egger. 2019. Facial model collection for medical augmented reality in oncologic cranio-maxillofacial surgery. *Scientific Data* 6, 1 (2019), 1–7.

[25] Sanjay Gupta, Joy A Henningsen, Michael J Wallace, David C Madoff, Frank A Morello Jr, Kamran Ahrar, Ravi Murthy, and Marshall E Hicks. 2007. Percutaneous biopsy of head and neck lesions with CT guidance: various approaches and relevant anatomic and technical considerations. *Radiographics* 27, 2 (2007), 371–390.

[26] Jonas Hajek, Mathias Unberath, Javad Fotouhi, Bastian Bier, Sing Chun Lee, Greg Osgood, Andreas Maier, Mehran Armand, and Nassir Navab. 2018. Closing the calibration loop: an inside-out-tracking paradigm for augmented reality in orthopedic surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 299–306.

[27] Christian Hansen, David Black, Christoph Lange, Fabian Rieber, Wolfram Lamadé, Marcello Donati, Karl J Oldhafer, and Horst K Hahn. 2013. Auditory support for resection guidance in navigated liver surgery. *International Journal of Medical Robotics and Computer Assisted Surgery* 9, 1 (2013), 36–43.

[28] Richard I Hartley and Peter Sturm. 1997. Triangulation. *Computer Vision and Image Understanding* 68, 2 (1997), 146–157.

[29] Florian Heinrich, Fabian Joeres, Kai Lawonn, and Christian Hansen. 2019. Comparison of projective augmented reality concepts to support medical needle insertion. *Transactions on Visualization and Computer Graphics* 25, 6 (2019), 2157–2167.

[30] Patrick Hübner, Kate Clintworth, Qingyi Liu, Martin Weinmann, and Sven Wursthorn. 2020. Evaluation of HoloLens tracking and depth sensing for indoor mapping applications. *Sensors* 20, 4 (2020), 1021.

[31] Rudolph Emil Kalman. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82, 1 (1960), 35–45.

[32] Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *International Workshop on Augmented Reality*. IEEE, 85–94.

[33] Kangsoo Kim, Mark Billinghurst, Gerd Bruder, Henry Been-Lirn Duh, and Gregory F Welch. 2018. Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017). *Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2947–2962.

[34] Florian Kral, Elisabeth J Puschban, Herbert Riechelmann, and Wolfgang Freysinger. 2013. Comparison of optical and electromagnetic tracking for navigated lateral skull base surgery. *International Journal of Medical Robotics and Computer Assisted Surgery* 9, 2 (2013), 247–252.

[35] Philipp Kriechling, Simon Roner, Florentin Liebmann, Fabio Casari, Philipp Fürnstahl, and Karl Wieser. 2020. Augmented reality for base plate component placement in reverse total shoulder arthroplasty: a feasibility study. *Archives of Orthopaedic and Trauma Surgery* (2020), 1–7.

[36] Ivo Kuhlemann, Markus Kleemann, Philipp Jauer, Achim Schweikard, and Floris Ernst. 2017. Towards X-ray free endovascular interventions–using HoloLens for on-line holographic visualisation. *Healthcare Technology Letters* 4, 5 (2017), 184–187.

[37] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2, 1-2 (1955), 83–97.

[38] Christian Kunz, Paulina Maurer, Fabian Kees, Pit Henrich, Christian Marzi, Michal Hlaváč, Max Schneider, and Franziska Mathis-Ullrich. 2020. Infrared marker tracking with the HoloLens for neurosurgical interventions. *Current Directions in Biomedical Engineering* 6, 1 (2020).

[39] Timur Kuzhagaliyev, Neil T Clancy, Mirek Janatka, Kevin Tchaka, Francisco Vasconcelos, Matthew J Clarkson, Kurinchi Gurusamy, David J Hawkes, Brian Davidson, and Danail Stoyanov. 2018. Augmented reality needle ablation guidance tool for irreversible electroporation in the pancreas. In *SPIE Medical Imaging*, Vol. 10576. International Society for Optics and Photonics, 1057613.

[40] Christoph Leuze, Grant Yang, Brian Hargreaves, Bruce Daniel, and Jennifer A McNab. 2018. Mixed-reality guidance for brain stimulation treatment of depression. In *International Symposium on Mixed and Augmented Reality Adjunct*. 377–380.

[41] Bo Li, Lionel Heng, Kevin Koser, and Marc Pollefeys. 2013. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In *International Conference on Intelligent Robots and Systems*. IEEE, 1301–1307.

[42] Ruotong Li, Tianpei Yang, Weixin Si, Xiangyun Liao, Qiong Wang, Reinhard Klein, and Pheng-Ann Heng. 2019. Augmented reality guided respiratory liver tumors punctures: A preliminary feasibility study. In *SIGGRAPH Asia Technical Briefs*. 114–117.

[43] Florentin Liebmann, Simon Roner, Marco von Atzigen, Davide Scaramuzza, Reto Sutter, Jess Snedeker, Mazda Farshad, and Philipp Fürnstahl. 2019. Pedicle screw navigation using surface digitization on the Microsoft HoloLens. *International journal of Computer Assisted Radiology and Surgery* 14, 7 (2019), 1157–1165.

[44] Jun Liu, Subhi J Al'Aref, Gurpreet Singh, Alexandre Caprio, Amir Ali Amiri Moghadam, Sun-Joo Jang, S Chiu Wong, James K Min, Simon Dunham, and

Bobak Mosadegh. 2019. An augmented reality system for image guidance of transcatheter procedures for structural heart disease. *PloS one* 14, 7 (2019), e0219174.

[45] Ran Long, Christian Rauch, Tianwei Zhang, Vladimir Ivan, and Sethu Vijayakumar. 2021. RigidFusion: Robot Localisation and Mapping in Environments With Large Dynamic Rigid Objects. *IEEE Robotics and Automation Letters* 6, 2 (2021), 3703–3710.

[46] Dietrich Manzey, Stefan Röttger, J Elin Bahner-Heyne, Dirk Schulze-Kissing, Andreas Dietz, Jürgen Meixensberger, and Gero Strauss. 2009. Image-guided navigation: the surgeon's perspective on performance consequences and human factors issues. *International Journal of Medical Robotics and Computer Assisted Surgery* 5, 3 (2009), 297–308.

[47] Primoz Markelj, Dejan Tomaževič, Bostjan Likar, and Franjo Pernuš. 2012. A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis* 16, 3 (2012), 642–661.

[48] Jene W Meulstee, Johan Nijsink, Ruud Schreurs, Luc M Verhamme, Tong Xi, Hans HK Delye, Wilfred A Borstlap, and Thomas JJ Maal. 2019. Toward holographic-guided surgery. *Surgical Innovation* 26, 1 (2019), 86–94.

[49] Uli Mezger, Claudia Jendrewski, and Michael Bartels. 2013. Navigation in surgery. *Langenbeck's Archives of Surgery* 398, 4 (2013), 501–514.

[50] Fabio Müller, Simon Roner, Florentin Liebmann, José M Spirig, Philipp Fürnstahl, and Mazda Farshad. 2020. Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging. *The Spine Journal* 20, 4 (2020), 621–628.

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

[52] Eva Novoa, Nicolas Gürtler, André Arnoux, and Marcel Kraft. 2012. Role of ultrasound-guided core-needle biopsy in the assessment of head and neck lesions: a meta-analysis and systematic review of the literature. *Head & Neck* 34, 10 (2012), 1497–1503.

[53] Axel Pinz, Markus Brandner, Harald Ganster, Albert Kusej, Peter Lang, and Miguel Ribo. 2002. Hybrid tracking for augmented reality. *Ögai Journal* 21, 1 (2002), 17–24.

[54] PTC. 2021. Vuforia. *https://library.vuforia.com/* (2021). accessed September 2021.

[55] Long Qian, Anton Deguet, and Peter Kazanzides. 2018. ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery. *Healthcare Technology Letters* 5, 5 (2018), 194–200.

[56] Long Qian, Xiran Zhang, Anton Deguet, and Peter Kazanzides. 2019. Aramis: Augmented reality assistance for minimally invasive surgery using a head-mounted display. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 74–82.

[57] Niels Tjørnly Rasmussen, Moritz Störring, Thomas B Moeslund, and Erik Granum. 2006. Real-time tracking for virtual environments using SCAAT Kalman filtering and unsynchronized cameras. In *International Conference on Computer Vision Theory and Applications*. 25–28.

[58] HS Rayt, AJ Sutton, NJM London, RD Sayers, and MJ Bown. 2008. A systematic review and meta-analysis of endovascular repair (EVAR) for ruptured abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery* 36, 5 (2008), 536–544.

[59] Miguel Ribo, Axel Pinz, and Anton L Fuhrmann. 2001. A new optical tracking system for virtual and augmented reality applications. In *Instrumentation and Measurement Technology Conference*, Vol. 3. IEEE, 1932–1936.

[60] Bernd Schwald. 2005. A tracking algorithm for rigid point-based marker models. In *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Václav Skala-UNION Agency.

[61] Frank Steinicke, Christian P Jansen, Klaus H Hinrichs, Jan Vahrenhold, and Bernd Schwald. 2007. Generating optimized marker-based rigid bodies for optical tracking systems.. In *International Conference on Computer Vision Theory and Applications*. 387–395.

[62] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. 2012. A benchmark for the evaluation of RGB-D SLAM systems. In *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 573–580.

[63] Qichang Sun, Yongfeng Mai, Rong Yang, Tong Ji, Xiaoyi Jiang, and Xiaojun Chen. 2020. Fast and accurate online calibration of optical see-through head-mounted display for AR-based surgical navigation using Microsoft HoloLens. *International Journal of Computer Assisted Radiology and Surgery* 15, 11 (2020), 1907–1919.

[64] Shinji Umeyama. 1991. Least-squares estimation of transformation parameters between two point patterns. *Computer Architecture Letters* 13, 04 (1991), 376–380.

[65] Dorin Ungureanu, Federica Bogo, Silvano Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Bugra Tekin, Pawel Olszta, and Marc Pollefeys. 2020. HoloLens 2 Research Mode as a Tool for Computer Vision Research. *arXiv:2008.11239* (2020).

[66] Petr Vávra, Jan Roman, Pavel Zonča, Peter Ihnát, Martin Němec, Jayant Kumar, Nagy Habib, and Ahmed El-Gendi. 2017. Recent development of augmented reality in surgery: a review. *Journal of Healthcare Engineering* 2017 (2017).

[67] Jianren Wang, Long Qian, Ehsan Azimi, and Peter Kazanzides. 2017. Prioritization and static error compensation for multi-camera collaborative tracking in augmented reality. In *Virtual Reality*. IEEE, 335–336.

[68] Greg Welch and Gary Bishop. 1997. SCAAT: Incremental tracking with incomplete information. In *Conference on Computer Graphics and Interactive Techniques*. ACM, 333–344.

[69] Greg Welch, Gary Bishop, Leandra Vicci, Stephen Brumback, Kurtis Keller, and D'nardo Colucci. 1999. The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments. In *Symposium on Virtual Reality Software and Technology*. ACM, 1–ff.

[70] Suya You, Ulrich Neumann, and Ronald Azuma. 1999. Hybrid inertial and vision tracking for augmented reality registration. In *Virtual Reality*. IEEE, 260–267.

[71] J-H Zhu, R Yang, Y-X Guo, J Wang, X-J Liu, and C-B Guo. 2021. Navigation-guided core needle biopsy for skull base and parapharyngeal lesions: a five-year experience. *International Journal of Oral and Maxillofacial Surgery* 50, 1 (2021), 7–13.

[72] Yan Zuo, Taoran Jiang, Jiansheng Dou, Dewang Yu, Zaphlene Nyakuru Ndaro, Yunxiao Du, Qingfeng Li, Shuyi Wang, and Gang Huang. 2020. A novel evaluation model for a mixed-reality surgical navigation system: where microsoft hololens meets the operating room. *Surgical Innovation* 27, 2 (2020), 193–202.