



UltrARsound: in situ visualization of live ultrasound images using HoloLens 2

Felix von Haxthausen^{1,2} · Rafael Moreta-Martinez¹ · Alicia Pose Díez de la Lastra¹ · Javier Pascau¹ · Floris Ernst²

Received: 13 January 2022 / Accepted: 31 May 2022
© The Author(s) 2022

Abstract

Purpose Augmented Reality (AR) has the potential to simplify ultrasound (US) examinations which usually require a skilled and experienced sonographer to mentally align narrow 2D cross-sectional US images in the 3D anatomy of the patient. This work describes and evaluates a novel approach to track retroreflective spheres attached to the US probe using an inside-out technique with the AR glasses HoloLens 2. Finally, live US images are displayed in situ on the imaged anatomy.

Methods The Unity application UltrARsound performs spatial tracking of the US probe and attached retroreflective markers using the depth camera integrated into the AR glasses—thus eliminating the need for an external tracking system. Additionally, a Kalman filter is implemented to improve the noisy measurements of the camera. US images are streamed wirelessly via the PLUS toolkit to HoloLens 2. The technical evaluation comprises static and dynamic tracking accuracy, frequency and latency of displayed images.

Results Tracking is performed with a median accuracy of 1.98 mm/1.81° for the static setting when using the Kalman filter. In a dynamic scenario, the median error was 2.81 mm/1.70°. The tracking frequency is currently limited to 20 Hz. 83% of the displayed US images had a latency lower than 16 ms.

Conclusions In this work, we showed that spatial tracking of retroreflective spheres with the depth camera of HoloLens 2 is feasible, achieving a promising accuracy for in situ visualization of live US images. For tracking, no additional hardware nor modifications to HoloLens 2 are required making it a cheap and easy-to-use approach. Moreover, a minimal latency of displayed images enables a real-time perception for the sonographer.

Keywords Augmented reality · Tracking · Ultrasound guidance · Retroreflective spheres

Introduction

Ultrasound (US) as a radiation-free, portable, widely available, and real-time capable imaging modality has become one of the core diagnostic tools for a variety of diseases. It has also proved application possibilities in therapeutic areas ranging from bone fracture healing to cancer treatment [1]. In addition, general practitioners, usually less experienced with this imaging modality, are increasingly using US in their daily practice [2].

However, US requires an experienced and skilled sonographer mentally aligning the narrow cross-sectional 2D US images in the 3D anatomy of the patient to perform the right diagnosis and treatment. The conventional image display on a screen is neither intuitive nor ergonomically friendly in this context. This inconvenient examination may also be associated with work-related musculoskeletal disorders [3]. Hand-eye coordination is another critical skill for sonogra-

✉ Felix von Haxthausen
vonhaxthausen@rob.uni-luebeck.de

Rafael Moreta-Martinez
rmoreta@pa.uc3m.es

Alicia Pose Díez de la Lastra
apose@ing.uc3m.es

Javier Pascau
jpascau@ing.uc3m.es

Floris Ernst
ernst@rob.uni-luebeck.de

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain

² Institute for Robotics and Cognitive Systems, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Schleswig-Holstein, Germany

phers as the gaze is directed to the screen, while the hand needs to adjust the position and orientation of the US probe. An intuitive and straightforward way of visualization could be beneficial, especially for novices, for whom a higher visuo-spatial skill is related to improved US examination outcomes [4].

The augmented reality (AR) glasses HoloLens 2 (Microsoft, Redmond, WA, USA), a head mounted computer with an optical see-through display, may overcome the aforementioned disadvantages by displaying the US images in situ below the US probe—superimposed on the imaged anatomy. This could improve understanding of anatomy and spatial orientation of structures. In addition to imaging, it is possible to overlay important information such as vital signs or tracked surgical instruments during procedures.

Therefore, in this work, we describe a method for tracking a rigid body based on retroreflective spheres by using the integrated depth camera of the AR glasses. Herewith, we demonstrate the first inside-out six degrees of freedom (DoF) tracking approach with HoloLens 2 for passive marker spheres that neither requires additional expensive tracking hardware nor illumination of the spheres. Additionally, the application is able to stream live images from an US station and display them in situ—resulting in an intuitive and realistic visualization. To provide other researchers in this field with an easy-to-set-up tracking approach, we implemented the algorithm with the Unity game engine (Unity Technologies, San Francisco, CA, USA). The code will be available at <https://github.com/BIIG-UC3M/IGT-UltrARsound>. A further aim of this study is to provide thorough technical details of this approach such as tracking accuracy in a static and dynamic setting, frequency and latency of displayed US images.

Related work

The idea of using AR to support US examinations was already described in 1996 by Fuchs et al. [5] demonstrating its research interest. Due to its early stage development, the AR glasses weighed nearly 3 kg, and the probe was spatially tracked using a mechanical arm. Thus, the whole setup was still far from being usable in daily practice. More than two decades later, the tetherless standalone AR device HoloLens 2, with a weight of 566 g, improved image resolution and self-localization, hand gestures and voice commands for interactions could bring the initial idea closer to the actual clinical application. In situ visualization of US with AR is certainly possible with other state-of-the-art visors [6]; however, in the following section, we want to concentrate on HoloLens as it provides the previously mentioned key features.

Tracking with HoloLens

Two approaches for instrument tracking with AR glasses exist. On the one hand, outside-in techniques utilize external sensors (e.g., optical or electromagnetic tracking) placed statically in the room to track the instrument as well as the HoloLens. On the other hand, inside-out techniques perform tracking via cameras and sensors integrated into the AR glasses—thus no additional hardware nor a cumbersome calibration would be necessary. Furthermore, incorporating an external tracking system leads to a more laborious system setup and workflow.

Outside-in tracking

A clear advantage of outside-in techniques is the achievable accuracy and precision that certain medical interventions require. However, the use of an external tracking system in combination with HoloLens requires an often cumbersome registration to have a common coordinate system. In many cases, optical tracking systems (OTS) facilitate tracking of medical instruments [7–10]. In this scenario, line of sight between tracked tools (e.g., AR glasses and US probe) and OTS needs to be ensured at all times which limits the possibilities of movement. Electromagnetic tracking, on the other hand, does not have this disadvantage but suffers from interference with metallic parts. Nevertheless, in scenarios such as AR-guided endovascular interventions [11,12], accurate tracking without line of sight is only provided by this modality.

Inside-out tracking

The most common way of tracking surgical tools is to utilize the front-facing central camera for detecting AR markers [13–17]. A general drawback, nonetheless, is that the markers constantly need to point towards the front camera, limiting the movement and rotation of the tracked tool. Additionally, the tracking process depends heavily on light conditions. As an alternative, retroreflective spheres can be used for rigid body tracking with HoloLens. Kunz et al. [18] showed the feasibility of this approach with HoloLens 1 by either using the integrated depth camera or two of the four front facing cameras via triangulation. However, tracking with the front-facing cameras requires placing an IR light emitter on the HoloLens. Additionally, this work only evaluated relative translations (three DoF) of the spheres, thus no conclusions can be made regarding the six DoF rigid body tracking. A recent proposal by Gsaxner et al. [19] showed promising results of rigid body tracking with retroreflective markers via triangulation, obtaining an accuracy of 1.70 mm/1.11° with an extended Kalman filter. However, this approach again involves a supplementary light source on HoloLens to illuminate the spheres which might impair the mobility of the AR glasses. Considering the previous approaches, in this work, we want to exploit the advantage of tracking retroreflective spheres with an inside-out approach, namely less light depen-

gency and no need for external expensive tracking systems, without the necessity of additional hardware to illuminate the spheres.

Ultrasound visualization with HoloLens

Only a few studies have investigated the combination of US imaging and HoloLens. Some of them rely on outside-in approaches to track the US probe and display the image [8,9,20]. Existing works utilizing inside-out approaches are on the other hand performed using an AR marker [11,21,22]. To acquire and stream images from the US station to HoloLens, one can either use a frame grabber at the display output of the US station or, if available, a software development kit (SDK) providing access to the image data. Nguyen et al. [22] followed the latter approach and then downsampled the images to 100×100 pixels to send them as a single UDP package. However, this resizing comes with an image quality loss. On the other hand, R uger et al. [9] and Kuzhagaliyev et al. [8] grabbed the screen output of the US stations and sent the trimmed images to HoloLens. Here again, depending on the image output, an image quality loss is possible. This loss of quality may become relevant when the US image in HoloLens is enlarged *ex situ* to see finer structures. Our approach overcomes these disadvantages by streaming US images in original image quality with minimal latency and a commonly used open-source network communication interface (OpenIGTLink).

Materials and methods

The developed application runs directly on HoloLens 2 and was built with Unity 2019.4.22. The research mode was incorporated within Unity [23] allowing access to the following sensors: four grayscale cameras, the inertial measuring unit and the depth camera which operates in two modes for different distances—each one providing a depth image and an active brightness (AB) image [24]. On top, OpenCV was integrated within Unity for image processing tasks. The depth camera offers two different modes, whereas in this work, only the AHAT (Articulated HAnd Tracking) mode was used because it provides more frames per second and the maximum distance of 1 m is sufficient for tracking a handheld US probe. The depth and AB image of the AHAT mode have a resolution of 512×512 pixels with 16 bits per pixel acquired at 45 frames per second. To track the pose of our rigid body composed of retroreflective spheres, the workflow described in detail in Sects. 2.1 and 2.2 consists of the following steps: (1) Measure the position of each sphere with respect to our world coordinate system using the depth camera. (2) Apply some filtering to reduce the noisy measurements of the depth

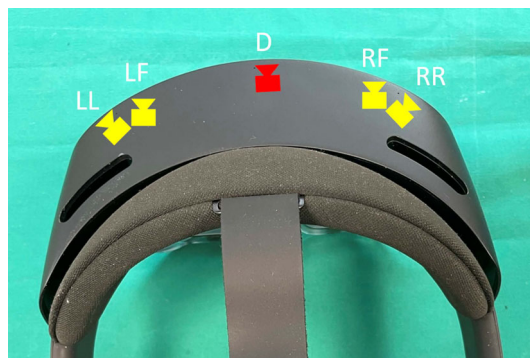


Fig. 1 Top view of HoloLens 2, the different visible light cameras (yellow), and the depth camera (red). The research mode provides for each acquired frame the pose of the left-front (LF) visible light camera. The extrinsic transform from the left-front camera to the depth camera can also be accessed via the research mode

camera. (3) Calculate the pose based on the resulting positions.

For US acquisitions, we used the portable system MicrUs from TELEMED (Vilnius, Lithuania) in combination with the linear US probe L12-5L40S-3. This system offers access to the US images without requiring a frame grabber. In Sect. 2.3, we describe the network communication interface and the registration approach required to find the spatial transformation from the tracked marker to the origin of the image.

Retroreflective sphere tracking

In order to compute ${}^W\mathbf{s} = [x, y, z, 1]^T$, which is the surface position of a retroreflective sphere with respect to a world coordinate system W , the following equation is applied:

$${}^W\mathbf{s} = {}^W\mathbf{T}_L {}^L\mathbf{T}_D {}^D\mathbf{s} \tag{1}$$

where ${}^W\mathbf{T}_L \in \mathbb{R}^{4 \times 4}$ is the pose of the left-front camera of HoloLens 2 (Fig. 1) with respect to our world coordinate system, ${}^L\mathbf{T}_D$ is the extrinsic camera matrix of the depth camera which can be accessed via the research mode and ${}^D\mathbf{s}$ is the surface position of a sphere with respect to the depth camera coordinate system. ${}^W\mathbf{T}_L$ is provided by the research mode for each frame, while the origin of the world coordinate system is placed at the device’s pose at application start. To track the retroreflective spheres and hereby the marker in the camera space, one first needs to detect the spheres in the AB image of the AHAT mode, which then allows finding the corresponding points in the depth image of the AHAT mode.

A blob detection algorithm provided by OpenCV is utilized to detect the spheres in the AB image (Fig. 2a). For each of the $i = 1 \dots 4$ spheres in the image, the method returns the pixel coordinates \mathbf{p}_i corresponding to the center of the circular image of the sphere. Then, the corresponding pixel value

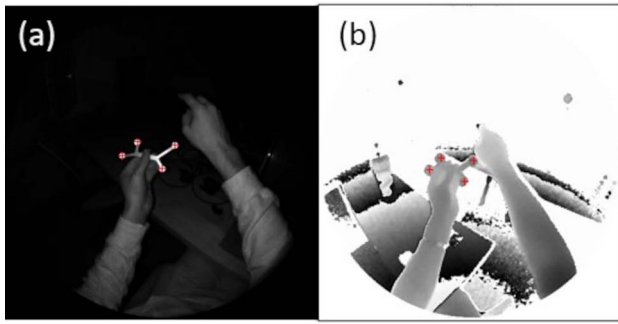


Fig. 2 **a** AB image with the detected centers of the spheres (red). **b** The corresponding depth image. The corresponding pixel values are used to calculate the position of the spheres using a lookup table

d_p in the depth image (Fig. 2b) is multiplied by the appropriate vector from a static lookup table $\mathbf{L} \in \mathbb{R}^{512 \times 512 \times 3}$ (also accessible via the research mode) to compute the initial guess of the surface sphere position ${}^D\mathbf{s} = [x, y, z]^T$ in the depth camera coordinate system:

$${}^D\mathbf{s} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = d_p \mathbf{L}(\mathbf{p}) = d_p \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad (2)$$

where $[u, v, w]^T$ is the entry of the lookup table corresponding to the image position \mathbf{p} .

As our approach does not directly allow tracking the center position of the spheres, but just a point on the surface, a correction vector is added to the surface position ${}^W\mathbf{s}$. The direction and length of this vector can be calculated by subtracting the position of the measured surface point ${}^W\mathbf{s}$ and the depth camera ${}^W\mathbf{c}_d$ with a length equal to the sphere radius r_s . Thus, the final center sphere position \mathbf{s}_c is calculated as:

$${}^W\mathbf{s}_c = {}^W\mathbf{s} + \frac{{}^W\mathbf{s} - {}^W\mathbf{c}_d}{\|{}^W\mathbf{s} - {}^W\mathbf{c}_d\|} r_s \quad (3)$$

Position estimation using a Kalman filter

A Kalman filter [25] is a mathematical method for the iterative estimation of parameters describing system states on the basis of potentially erroneous observations. As the measurements of the depth camera of HoloLens 2, which is hardware-wise the same as the Azure Kinect camera, show a noisy behavior [26], we decided to implement a Kalman filter to reduce the influence of erroneous measurements. To this end, a 3D constant acceleration model served as underlying motion model for each measured sphere position. We determine the optimal values of the process noise covariance σ_Q and measurement noise covariance σ_R through testing on a prerecorded sequence obtaining $\sigma_Q = 0.0001$ and $\sigma_R = 0.001$ for the static setting. In a dynamic environment, the optimal values were $\sigma_Q = 0.0001$ and $\sigma_R = 0.0001$.

Marker pose calculation

The four sphere center positions ${}^W\mathbf{s}_{c,i}$ in the world coordinate system can be used to calculate the pose ${}^W\mathbf{T}_M$. Therefore, to avoid an ambiguous identification of the order, the relative distances between the spheres need to be unique. The 3D coordinates of the sphere centers are expressed in a common reference frame whose origin can be placed at a desired position and orientation. In this work, we used an OTS, more precisely Polaris Spectra from NDI, to measure the distances and create the origin of our reference frame including the positions ${}^{RF}\mathbf{s}_{c,i}$ of the sphere centers. The tool definition is performed beforehand and provided to the tracking algorithm. Within the tracking algorithm and for each cycle, we first measure the distances of the spheres to each other to identify corresponding points and then perform a landmark-based registration [27] using ${}^{RF}\mathbf{s}_{c,i}$ and ${}^W\mathbf{s}_{c,i}$ to compute ${}^W\mathbf{T}_M$.

Ultrasound streaming and calibration

The US images were sent via the open-source-toolkit PLUS [28] from the computer to the Unity application running on HoloLens 2. As no library exists in Unity supporting the OpenIGTLink protocol [29] used by PLUS, we developed a custom-made client side. In order to display the US images in situ on the AR glasses, it is necessary to find the spatial transformation from the tracked marker to the origin of the image ${}^M\mathbf{T}_{US}$. To this end, we used a phantom-free approach [30] that requires an external OTS. We opted for this approach as it shows better results than the commonly used N-wire phantom [31]. For the calibration, we recorded the static pose of the optically tracked US probe marker ${}^{OT}\mathbf{T}_M$ immersed in saltwater. Using a tracked and pivot-calibrated stylus, we successively measured eight stylus tip positions in the US beam (Fig. 3). Identifying the positions of the stylus tip in the eight US images and simultaneously in the OTS space allows for the calculation of the transformation of the US image in the OTS coordinate space ${}^{OT}\mathbf{T}_{US}$ by means of a landmark-based registration [27]. Finally, we can compute the desired transformation by closing the transformation loop with:

$${}^M\mathbf{T}_{US} = {}^M\mathbf{T}_{OT} {}^{OT}\mathbf{T}_{US} \quad (4)$$

The 3D models for the attachment of the markers used in this work were retrieved from PLUS [32].

Evaluation

Apart from testing the feasibility, the aim of this work was to perform a technical evaluation of the proposed approach. To this end, we measured the tracking accuracy of the sphere's positions and the pose estimation in a static scenario. To

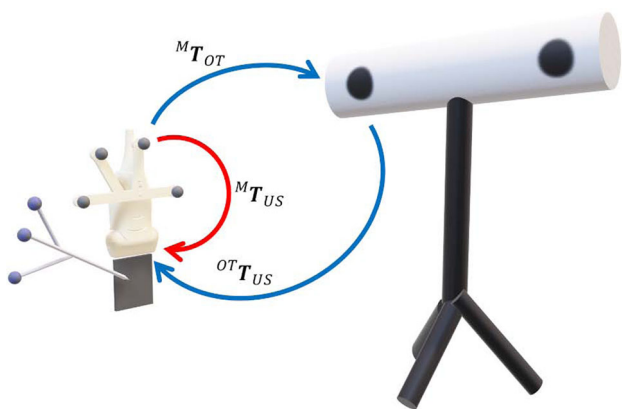


Fig. 3 Transformations (arrows) available and calculated for the spatial calibration between the US marker and the US image. The measured positions of the stylus tip in the US images and tracking space allow the calculation of ${}^{OT}\mathbf{T}_{US}$ and thereafter close the loop to compute ${}^M\mathbf{T}_{US}$ (red arrow)

also evaluate the system in a realistic environment, the pose tracking accuracy was measured in a dynamic scenario. Additionally, we measured the time duration of one tracking cycle and the latency of the displayed US images on HoloLens 2.

Retroreflective sphere tracking

To quantify the accuracy of our tracking method in a static scenario, we compared relative movements of the marker measured with the AR glasses and, as a ground truth, the highly accurate OTS (Fig. 4). Twenty different poses of the marker in a range from approximately 0.3 to 1 m were recorded acquiring 30 samples for each pose. Thereafter, for each sample and each pose, the relative distance of each sphere to the corresponding sphere in the first pose was computed and compared to the value obtained with OTS. For example, to calculate the distance error d_e between the second and first pose for one sphere, we compute:

$$d_e = \| {}^W\mathbf{s}_{c,p_2} - {}^W\mathbf{s}_{c,p_1} \| - \| {}^{OT}\mathbf{s}_{c,p_2} - {}^{OT}\mathbf{s}_{c,p_1} \| \quad (5)$$

where ${}^{OT}\mathbf{s}_{c,p_1}$ is a sphere center position for the first pose in the OTS coordinate system. To additionally evaluate the actual marker pose tracking accuracy, the pose change of each marker, which is based on the four sphere positions, was also analyzed. Here, we compared the relative translational and rotational movements between poses measured with HoloLens 2 and the OTS. Again, for example, to measure the translation error \mathbf{t}_e and rotational error \mathbf{r}_e between the second and first pose, we first measure the pose error \mathbf{T}_e :

$$\mathbf{T}_e = (({}^W\mathbf{T}_{M,p_1})^{-1} {}^W\mathbf{T}_{M,p_2})^{-1} (({}^{OT}\mathbf{T}_{M,p_1})^{-1} {}^{OT}\mathbf{T}_{M,p_2}) \quad (6)$$

which we use to extract \mathbf{t}_e and \mathbf{r}_e in Euler angles.



Fig. 4 Setup for the static tracking accuracy evaluation. HoloLens 2 and the OTS are fixed, while the US probe with its marker is moved to 20 different poses for measurement acquisition

For the dynamic scenario on the other hand, we acquired a trajectory of our marker with the statically placed AR glasses and the OTS (trajectory illustrated in Fig. 5). As line of sight of the spheres is necessary, the possible rotation around the x- and y-axis was limited. For the temporal calibration between the measured trajectories of HoloLens 2 and the tracking system, we acquired a sinusoidal trajectory before the actual acquisition. Thereafter, the sinusoidal signals were down-sampled to have matching frequencies. Finally, the maximum position of the cross-correlation of both signals allows to calculate the time shift between both measurements. As in the static scenario, we compute \mathbf{t}_e and \mathbf{r}_e with the initial pose as our reference. To measure the tracking frequency, the time interval of 500 tracking cycles was obtained.

Ultrasound streaming

Apart from a correct in situ placement of the US image, it is essential that the latency of displayed US images on HoloLens 2 is low enough for a real-time perception. To measure the latency, a slow motion video with 120Hz is recorded which results in a temporal resolution of 8 ms. The video is recorded through the lenses of the AR glasses such that the virtual canvas including the US image and the US image on the real screen are visible. To measure the delay, the depth of the US image is changed and the frames are counted until the change becomes visible on the virtual canvas. For the evaluation, the latency was measured 30 times. Images with a matrix size of 512×512 pixels and 8 bits per pixel were sent. Note that the image matrix size does not change even if the depth is modified.

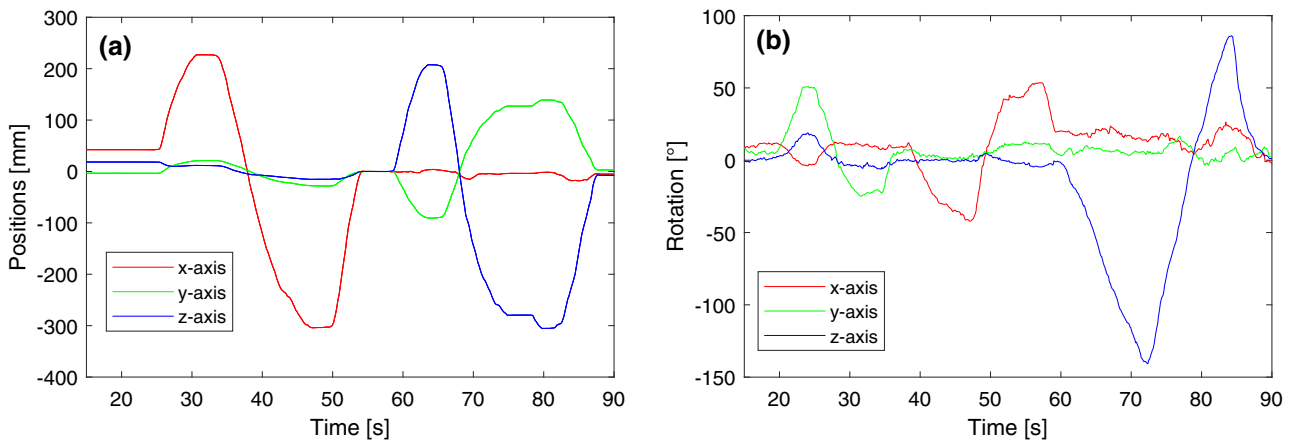


Fig. 5 Trajectories obtained with the OTS for the dynamic analysis. **a** Relative positions and **b** relative rotations for each axis with the respect to the initial pose

Results

Following, the evaluation results of the tracking accuracy, tracking frequency and the US streaming latency are described. Within the tracking accuracy evaluation, we first present the results of the static and dynamic scenario separately and then compare them .

Retroreflective sphere tracking

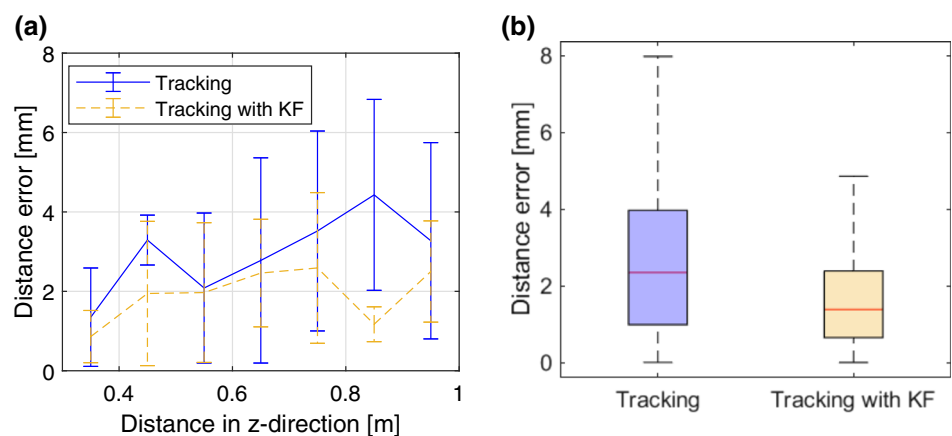
We compared the accuracy of the spheres position tracking with and without the implemented Kalman filter. The mean distance error and standard deviation without filter was $d_{e,RMS} = 3.27 \pm 1.90$ mm. On the other hand, the Kalman filter lowered the distance error to $d_{e,RMS} = 2.23 \pm 1.38$ mm (Fig. 6b). As shown in Fig. 6a, the error depends on the distance to the depth camera and increases roughly by 1.5 mm between 0.3 and 1 m. For all distances except one, the standard deviation decreases once the filter is applied.

In a second step, we compared the accuracy of the pose tracking with and without Kalman filter for a static scenario.

Again, only the relative movements to the first obtained pose are evaluated. In Fig. 7, $\mathbf{t}_{e,RMS}$ and $\mathbf{r}_{e,RMS}$ are visualized for each axis. When looking at $\mathbf{t}_{e,RMS}$, we see that applying a Kalman filter lowers the median error for each axis. Except for the x -axis, the 75th percentile and the maximum error are lower when using a Kalman filter. For the rotational error $\mathbf{r}_{e,RMS}$, only a slight improvement can be achieved with the filter. In all cases, the median is close to or lower than 1.

In a third step, we compared the accuracy of the pose tracking in a dynamic scenario. Figure 8 shows the distribution of $\mathbf{t}_{e,RMS}$ and $\mathbf{r}_{e,RMS}$ for each axis. We can clearly see that in a dynamic scenario, we obtain errors higher than roughly 3 mm for one fourth of the measurements with maximum errors of around 15 mm. On the other side, the median values are below 2 mm . The Kalman filter is able to lower the higher errors to a maximum value of 10 mm, while the median values are similar in both cases. Overall, we see a similar behavior for the rotational error $\mathbf{r}_{e,RMS}$, namely partially higher errors with low median errors. In this case, however, the Kalman filter brings only a slight improvement regarding the accuracy.

Fig. 6 Results of the accuracy evaluation for position tracking with HoloLens 2. **a** Mean and standard deviation of $d_{e,RMS}$ versus the distance to the depth camera. **b** Distribution (median, 25th and 75th percentile, minimal and maximal errors) of $d_{e,RMS}$ for tracking with and without Kalman filter



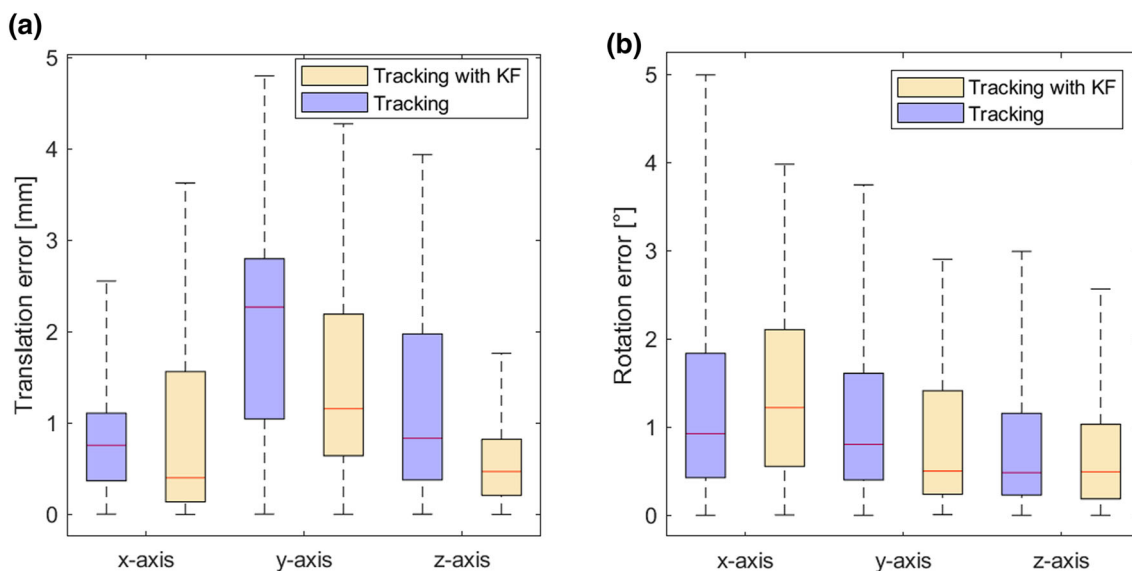


Fig. 7 Results of the accuracy evaluation for pose tracking with HoloLens 2 in a static scenario. **a** The distribution of the translational error and **b** the distribution of the rotational error with and without Kalman filter

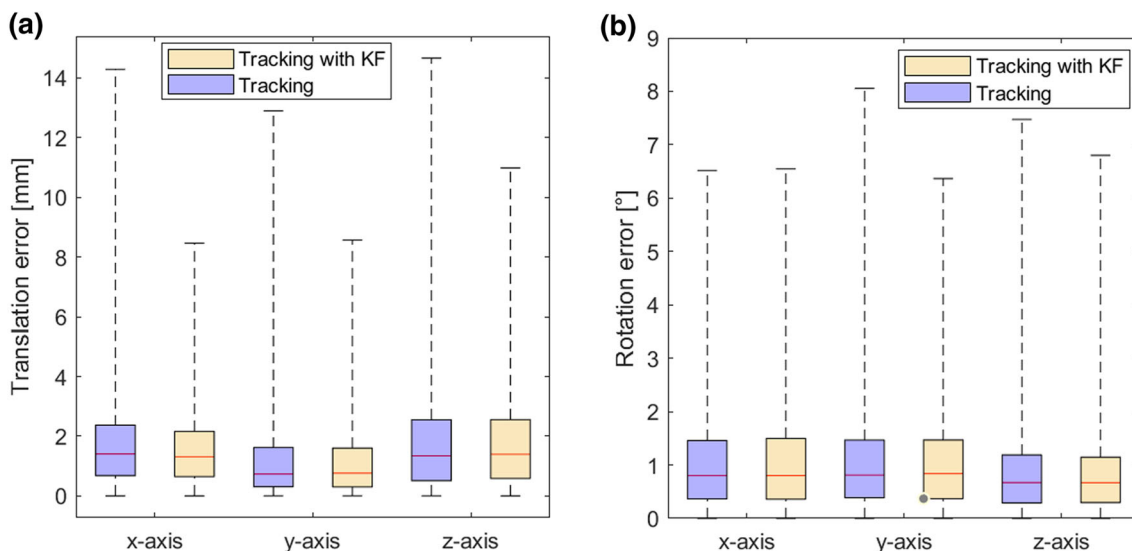


Fig. 8 Results of the accuracy evaluation for pose tracking with HoloLens 2 in a dynamic scenario. **a** The distribution of the translational error with and without Kalman filter. **b** The distribution of the rotational error for each axis with and without Kalman filter

Table 1 shows the statistics of the errors obtained in both scenarios for a comparative analysis. Note that, to this end, we computed the magnitude of \mathbf{t}_e and \mathbf{r}_e which is a valid option in our case [33]. When looking at $\|\mathbf{t}_e\|$, we see that the median only slightly increases without any filter. The RMS value decreases by roughly 1 mm when applying the filter for both scenarios. For $\|\mathbf{r}_e\|$, the median as well as the RMS values remain similar for all cases. To summarize, we see that the Kalman filter improves the accuracy regarding the translation measurements, while it does not have a noticeable effect on the rotational values.

Finally, the average time and standard deviation for one tracking cycle was 50.64 ± 7.11 ms resulting in a frequency of around 20Hz.

Ultrasound streaming

83% of the frames had a delay of 2 or less frames and thus a maximum delay of 16 ms (Fig. 9). Figure 10 shows the application running on HoloLens 2. The US image of a breast phantom is visible on the computer screen, while the user can see the image in situ through the AR glasses.

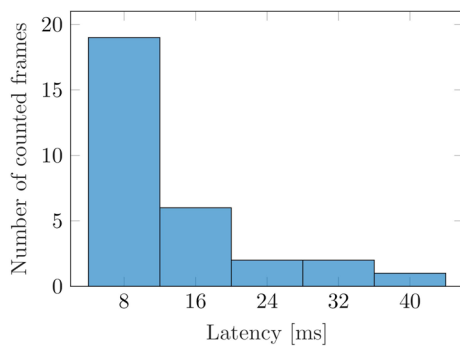


Fig. 9 Results of the latency measurement of displayed US images on HoloLens. A careful examination of the slow-motion video with 120Hz allows to count the frames until a change in the US image is visible on HoloLens 2

Discussion

US requires skilled and experienced sonographers for a correct and accurate diagnosis. This is partially explained by the mental workload of aligning small cross-sectional 2D images in the 3D anatomy of the patient. Our application UltrAR-sound overcomes this disadvantage by displaying live US images in situ using the AR glasses HoloLens 2. In this work, we describe an approach to track the US probe with retroreflective spheres with an inside-out technique, utilizing solely the depth camera of HoloLens 2. Accordingly, no expensive

additional hardware is necessary. Further, we show how to stream US images from an US station with minimal latency. For a final evaluation, we perform a technical characterization of the application by measuring spatial tracking accuracy and latency of displayed US images.

We measured the accuracy by comparing relative movements obtained with HoloLens 2 and a high accuracy OTS. The results show that the position tracking accuracy improves when applying a Kalman filter. Not only does the median of $d_{e,RMS}$ decrease but so do the 75th percentile and maximum values. Consequently, we also see a decrease in $t_{e,RMS}$ for each axis. The rotation error, otherwise, remains similar. A possible explanation is that if we measure the same incorrect position shift for each sphere, the translation error increases, while the rotation error stays the same. Indeed, our measured errors align with previously published works on the accuracy of the depth camera [26] in which errors between 1.5 to 2.5 mm are stated for distances of 0.5 to 1 m in the binned wide mode. A study published by Microsoft [34] states slightly lower values (1–2 mm). However, they do not differentiate between different modes which makes a comparison difficult. To conclude, we assume that a majority of the measured error in a static setting is due to hardware limitations. Nevertheless, the following software-related steps could improve the results. It is assumed that the static lookup table L used to calculate the positions is perfect. A re-calibration might lower $d_{e,RMS}$ and therefore T_e . Addi-

Fig. 10 HoloLens 2 user holding the US probe with the attached retroreflective spheres during a breast phantom acquisition, showing the US image on the computer screen (left). The view through the AR glasses enables to see the image in situ (right)

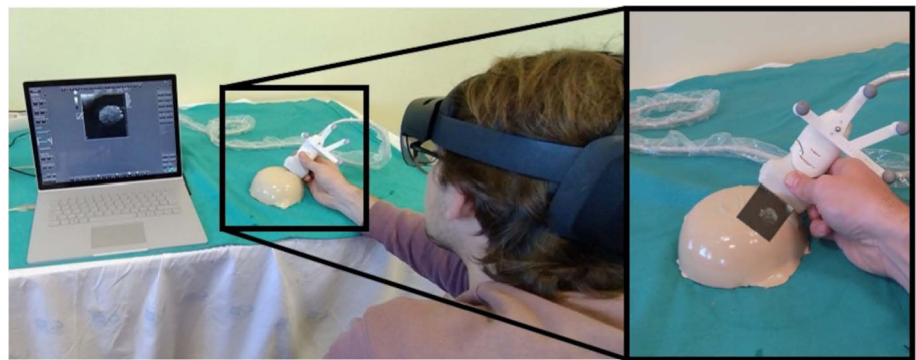


Table 1 Error statistics of $\|t_e\|$ and $\|r_e\|$ for dynamic and static scenarios with and without Kalman filter

		Min	25th p	Median	75th p	Max	RMS
		Translation error [mm]					
Static	Tracking only	0.03	1.73	2.71	3.50	5.93	3.07
	With Kalman filter	0.02	1.24	1.98	2.58	4.70	2.14
Dynamic	Tracking only	0.15	2.02	3.01	4.42	17.58	4.79
	With Kalman filter	0.08	1.96	2.81	3.79	16.15	3.76
		Rotation error [°]					
Static	Tracking only	0.10	1.21	2.01	2.74	4.32	2.24
	With Kalman filter	0.01	1.11	1.81	2.67	4.11	2.16
Dynamic	Tracking only	0.19	1.17	1.75	2.49	8.51	2.25
	With Kalman filter	0.23	1.18	1.70	2.46	10.07	2.19

tionally, our approach assumes that measurement vector of the Kalman filter consists of the calculated positions in all three dimensions. Another option that could potentially further improve the results is to use the measured blob location \mathbf{p} in the AB image as measurement vector instead. However, to calculate the predicted measurement from the state prediction, one would need to know the inverse version \mathbf{L} which is not available to our knowledge. Therefore, we relied on a simplified model for our Kalman filter.

Regarding the dynamic measurements, it is to be expected that the errors increase. However, it is spiking that a quarter of the measurements (Fig. 8) range from 4 to 15 mm. In this case, as no literature exists that investigates the accuracy of the depth camera in a dynamic scenario, it is difficult to determine if the error is hardware- or software-related. Applying a Kalman filter improves the accuracy regarding the translation measurements in both settings. To obtain the best results, it was necessary to apply different values of the noise covariance and measurement noise covariance for the static and the dynamic scenario, respectively.

Gsaxner et al. [19] showed slightly lower errors for translation and rotation with 1.70 mm and 1.11° in a static setup. In a dynamic setting, their approach outperforms ours with 1.90 mm and 1.18° . However, we would like to note that our approach does not need an additional light source to illuminate the retroreflective spheres. Additionally, Gsaxner et al. [19] compared absolute positions and rotations of HoloLens 2 and an OTS in the same coordinate system which was achieved by means of a hand-eye-calibration. Thus, a direct comparison between the two approaches is challenging. We decided to compare only relative movements of the marker as it is not clear how much the hand-eye-calibration influences the measured error.

With the current arrangement of the spheres, it is inevitable that there are some blind spots such that the rigid body cannot be tracked. For the proposed application, we assume that the possible rotations are sufficient as the sonographer will not look at the US image turned by 90° so that only the edge of the image is visible. It is possible to add more spheres to the rigid body which would allow to track the body with less or even no movement constraints. However, this would impair the handiness and maneuverability of the US probe with the attached rigid body. Therefore, we decided to use the rigid body provided by the PLUS library and accept to lose some movement freedom.

Overall, the tracking accuracy reported here is promising for training cases in ultrasound diagnostics in which a sub-millimeter accurate overlay of real and imaged anatomy is not required. It is worth mentioning that the accuracy might be lower in a real scenario in which the HoloLens will also move. In this work, the AR glasses are placed statically throughout the measurements. Thus, the error of the self-localization of HoloLens is not taken into account. A previous study showed

that the error of the self-localization may be up to 3 cm in a room-scale environment [35]. Consequently, the accuracy of our approach may be greatly influenced by a movement of the HoloLens.

Using the proposed approach, a tracking frequency of 20 Hz was achieved. We believe that tracking frequency should ideally be ≥ 60 Hz to match the frame rate of HoloLens 2 and provide a smooth visualization. Therefore, future work will consist of an analysis of the algorithm to identify the most time-consuming parts. An initial improvement might be achieved by performing the blob detection only in a region of interest defined by the previous found locations of the blobs.

Low latencies are crucial for in situ visualization of US images especially when guiding interventions such as biopsies. In this work, the latency of displayed US images was sufficiently low for a real-time perception with a majority of frames being < 16 ms delayed, proving that our approach is faster than the 80 ms reported by Ngyuen et al. [22]. García-Vázquez et al. [11] stated a latency of 259 ms for US volumes, correspondingly a significantly higher amount of data was transmitted which makes a comparison difficult. Other studies mentioned in Sect. 1.1.2 did not measure any latencies. The development of the client-side based on the OpenIGTLink protocol additionally allows connecting any other US station that supports streaming via the PLUS toolkit. Nevertheless, the US calibration needs to be performed for each new US probe.

Summary

This work presents an approach to spatially track retroreflective spheres with HoloLens 2 which in turn allows to track the position and orientation of an US probe with attached spheres. For in situ visualization, live US images are streamed from the US station to the AR glasses wirelessly. The inside-out tracking is performed through the depth camera of HoloLens 2. Accordingly, neither an external tracking system nor illumination of the spheres with additional hardware is required and, thus, it provides a cost and space efficient alternative to other approaches. The achieved accuracy is promising for diagnostic US examinations and partially for image-guided therapies in which no sub-millimeter accuracy is necessary. We believe that sub-millimeter accuracy may be achievable if both the depth camera and the stereo setup of the visible light cameras are taken into account, and both measurements are inputs to the Kalman filter.

Overall, our open-source application UltrARsound may allow sonographers to learn and perform US examinations faster and more effectively by providing an intuitive and

ergonomically friendly in situ visualization of the 2D US image.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by a fellowship within the IFI programme of the German Academic Exchange Service (DAAD).

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical approval No ethical approval was needed for this phantom study.

Informed consent For this type of study, formal consent is not required.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Escoffre J-M, Bouakaz A (2015) Therapeutic ultrasound. Springer, Berlin
- Andersen CA, Holden S, Vela J, Rathleff MS, Jensen MB (2019) Point-of-care ultrasound in general practice: a systematic review. *Ann Fam Med* 17(1):61–69. <https://doi.org/10.1370/afm.2330>
- Evans K, Roll S, Baker J (2009) Work-related musculoskeletal disorders (WRMSD) among registered diagnostic medical sonographers and vascular technologists: A representative sample. *J Diagn Med Sonogr* 25(6):287–299. <https://doi.org/10.1177/8756479309351748>
- Shafiqat A, Ferguson E, Thanawala V, Bedfordth NM, Hardman JG, McCahon RA (2015) Visuospatial ability as a predictor of novice performance in ultrasound-guided regional anesthesia. *Anesthesiology* 123(5):1188–1197. <https://doi.org/10.1097/ALN.0000000000000870>
- Fuchs H, State A, Pisano ED, Garrett WF, Hirota G, Livingston M, Whitton MC, Pizer SM (1996) Towards performing ultrasound-guided needle biopsies from within a head-mounted display. In: Höhne KH, Kikinis R (eds) *Visualization in biomedical computing*. Springer, Berlin, Heidelberg, pp 591–600
- Cattari N, Condino S, Cutolo F, Ferrari M, Ferrari V (2021) In situ visualization for 3d ultrasound-guided interventions with augmented reality headset. *Bioengineering*. <https://doi.org/10.3390/bioengineering8100131>
- El-Hariri H (2018) Augmented reality visualisation for orthopaedic surgical guidance with pre- and intra-operative multimodal image data fusion. *Healthcare Technol Lett* 5:189–1934
- Kuzhagaliyev T, Clancy NT, Janatka M, Tchaka K, Vasconcelos F, Clarkson MJ, Gurusamy K, Hawkes DJ, Davidson B, Stoyanov D (2018) Augmented reality needle ablation guidance tool for irreversible electroporation in the pancreas. In: Fei B, III, RJW (eds.) *Medical imaging 2018: Image-guided procedures, robotic interventions, and modeling*, vol. 10576, pp. 260–265. SPIE, ??? (2018). <https://doi.org/10.1117/12.2293671>. International Society for Optics and Photonics
- Rüger C, Feufel MA, Moosburner S, Özbek C, Pratschke J, Sauer IM (2020) Ultrasound in augmented reality: a mixed-methods evaluation of head-mounted displays in image-guided interventions. *Int J Comput Assist Radiol Surg* 15(11):1895–1905. <https://doi.org/10.1007/s11548-020-02236-6>
- Meulstee JW, Nijsink J, Schreurs R, Verhamme LM, Xi T, Delye HHK, Borstlap WA, Maal TJJ (2019) Toward holographic-guided surgery. *Surgical Innov* 26(1):86–94. <https://doi.org/10.1177/1553350618799552> (PMID: 30261829)
- García-Vázquez V, von Haxthausen F, Jäckle S, Schumann C, Kuhlemann I, Bouchagiar J, Höfer A-C, Matysiak F, Hüttmann G, Goltz JP, Kleemann M, Ernst F, Horn M (2018) Navigation and visualisation with hololens in endovascular aortic repair. *Innov Surg Sci* 3(3):167–177. <https://doi.org/10.1515/iss-2018-2001>
- Kuhlemann I, Kleemann M, Jauer P, Schweikard A, Ernst F (2017) Towards x-ray free endovascular interventions - using hololens for on-line holographic visualisation. *Healthcare Technol Lett* 4(5):184–187. <https://doi.org/10.1049/htl.2017.0061>
- Liebmann F, Roner S, von Atzigen M, Scaramuzza D, Sutter R, Snedeker J, Farshad M, Fünstahl P (2019) Pedicle screw navigation using surface digitization on the microsoft hololens. *Int J Comput Assist Radiol Surg* 14(7):1157–1165. <https://doi.org/10.1007/s11548-019-01973-7>
- Moreta-Martinez R, García-Mato D, García-Sevilla M, Pérez-Mañanes R, Calvo-Haro J, Pascau J (2018) Augmented reality in computer-assisted interventions based on patient-specific 3D printed reference. *Healthc Technol Lett* 5(5):162–166. <https://doi.org/10.1049/htl.2018.5072>
- Kriechling P, Roner S, Liebmann F, Casari F, Fünstahl P, Wieser K (2021) Augmented reality for base plate component placement in reverse total shoulder arthroplasty: a feasibility study. *Arch Orthop Trauma Surg* 141(9):1447–1453. <https://doi.org/10.1007/s00402-020-03542-z>
- Qian L, Zhang X, Deguet A, Kazanzides P (2019) Aramis: augmented reality assistance for minimally invasive surgery using a head-mounted display. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A (eds) *Medical image computing and computer assisted intervention - MICCAI 2019*. Springer, Cham, pp 74–82
- Carbone M, Condino S, Cutolo F, Vigliani RM, Kaschke O, Thomale UW, Ferrari V (2018) Proof of concept: wearable augmented reality video see-through display for neuro-endoscopy. In: De Paolis LT, Bourdot P (eds.) *Augmented reality, virtual reality, and computer graphics*, pp. 95–104. Springer, Cham. https://doi.org/10.1007/978-3-319-95282-6_7
- Kunz C, Maurer P, Kees F, Henrich P, Marzi C, Hlaváč M, Schneider M, Mathis-Ullrich F (2020) Infrared marker tracking with the hololens for neurosurgical interventions. *Curr Dir Biomed Eng* 6(1):20200027. <https://doi.org/10.1515/cdbme-2020-0027>
- Gsaxner C, Li J, Pepe A, Schmalstieg D, Egger J (2021) Inside-out instrument tracking for surgical navigation in augmented reality. In: *Proceedings of the 27th ACM symposium on virtual reality software and technology. VRST '21*. Association for computing machinery, New York, NY, USA. <https://doi.org/10.1145/3489849.3489863>

20. Mahmood F, Mahmood E, Dorfman RG, Mitchell J, Mahmood FU, Jones SB, Matyal R (2018) Augmented reality and ultrasound education: initial experience. *J Cardiothorac Vasc Anesth* 32(3):1363–1367. <https://doi.org/10.1053/j.jvca.2017.12.006>
21. Farshad-Amacker NA, Bay T, Rosskopf AB, Spirig JM, Wanivenhaus F, Pfirrmann CWA, Farshad M (2020) Ultrasound-guided interventions with augmented reality in situ visualisation: a proof-of-mechanism phantom study. *Eur Radiol Exp*. <https://doi.org/10.1186/s41747-019-0129-y>
22. Nguyen T, Plishker W, Matisoff A, Sharma K, Shekhar R (2021) HoloLens: augmented reality visualization of live ultrasound images using holoLens for ultrasound-guided procedures. *Int J Comput Assist Radiol Surg*. <https://doi.org/10.1007/s11548-021-02526-7>
23. Gu W (2021) HoloLens2-ResearchMode-Unity: Unity Plugin for using research mode functionality in HoloLens 2. Modified based on HoloLens2ForCV. <https://github.com/petergu684/HoloLens2-ResearchMode-Unity>. Accessed: 25 May 2021
24. Ungureanu D, Bogo F, Galliani S, Sama P, Duan X, Meekhof C, Stühmer J, Cashman TJ, Tekin B, Schönberger JL, Olszta P, Pollefeys M (2020) HoloLens 2 research mode as a tool for computer vision research. arXiv
25. Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82(1):35–45. <https://doi.org/10.1115/1.3662552>
26. Tölgyessy M, Dekan M, Chovanec L, Hubinský P (2021) Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors* 21(2):1–25. <https://doi.org/10.3390/s21020413>
27. Arun KS, Huang TS, Blostein SD (1987) Least-squares fitting of two 3-d point sets. *IEEE Trans Pattern Anal Mach Intell* 9(5):698–700. <https://doi.org/10.1109/TPAMI.1987.4767965>
28. Lasso A, Heffter T, Rankin A, Pinter C, Ungi T, Fichtinger G (2014) Plus: open-source toolkit for ultrasound-guided intervention systems. *IEEE Trans Biomed Eng* 61(10):2527–2537. <https://doi.org/10.1109/TBME.2014.2322864>
29. Tokuda J, Fischer GS, Papademetris X, Yaniv Z, Ibanez L, Cheng P, Liu H, Blevins J, Arata J, Golby AJ, Kapur T, Pieper S, Burdette EC, Fichtinger G, Tempny CM, Hata N (2009) Openiglink: an open network protocol for image-guided therapy environment. *Int J Med Robot* 5(4):423–434. <https://doi.org/10.1002/rcs.274>
30. SlicerIGT-community: User tutorial U31 - Tracked ultrasound calibration. <http://www.slicerigt.org/wp/user-tutorial/>. Accessed: 25 May 2021 (2016)
31. Welch M, Andrea J, Ungi T, Fichtinger G (2013) Freehand ultrasound calibration: phantom versus tracked pointer. In: III DRH, Yaniv ZR (eds) *Medical imaging 2013: image-guided procedures, robotic interventions, and modeling*, vol. 8671, pp. 348–353. SPIE. <https://doi.org/10.1117/12.2007922>. International Society for Optics and Photonics
32. Plus: Plus toolkit - printable 3D models catalog. <http://perk-software.cs.queensu.ca/plus/doc/nightly/modelcatalog/>. Accessed: 25 May 2021 (2020)
33. Huynh DQ (2009) Metrics for 3D rotations: comparison and analysis. *J Math Imag Vis* 35(2):155–164. <https://doi.org/10.1007/s10851-009-0161-2>
34. Bamji CS, Mehta S, Thompson B, Elkhatib T, Wurster S, Akkaya O, Payne A, Godbaz J, Fenton M, Rajasekaran V, Prather L, Nagaraja S, Mogallapu V, Snow D, McCauley R, Mukadam M, Agi I, McCarthy S, Xu Z, Perry T, Qian W, Chan V-H, Adepu P, Ali G, Ahmed M, Mukherjee A, Nayak S, Gampell D, Acharya S, Kordus L, O'Connor P (2018) Impixel 65nm bsi 320mhz demodulated tof image sensor with 3 μ m global shutter pixels and analog binning. In: 2018 IEEE international solid - state circuits conference - (ISSCC), pp. 94–96. <https://doi.org/10.1109/ISSCC.2018.8310200>
35. Hübner P, Clintworth K, Liu Q, Weinmann M, Wursthorn S (2020) Evaluation of holoLens tracking and depth sensing for indoor mapping applications. *Sensors* 20(4):1021. <https://doi.org/10.3390/s20041021>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.