Evaluation of Marker Tracking Using Mono and Stereo Vision in Microsoft HoloLens for Surgical Navigation

Thabit, Abdullah; Niessen, Wiro J.; Wolvius, Eppo B.; Van Walsum, Theo

DOI
10.1117/12.2607262

Publication date
2022

Document Version
Final published version

Published in
Medical Imaging 2022

Evaluation of marker tracking using mono and stereo vision in Microsoft HoloLens for surgical navigation

Abdullah Thabit, Wiro Niessen, Eppo Wolvius, Theo van Walsum

Abdullah Thabit, Wiro J. Niessen, Eppo B. Wolvius, Theo van Walsum, "Evaluation of marker tracking using mono and stereo vision in Microsoft HoloLens for surgical navigation," Proc. SPIE 12034, Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, 1203415 (4 April 2022); doi: 10.1117/12.2607262

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Evaluation of Marker Tracking Using Mono and Stereo Vision in Microsoft HoloLens for Surgical Navigation

Abdullah Thabit[1 2], Wiro J. Niessen[1 3], Eppo B. Wolvius[2], and Theo van Walsum[1]

[1]Biomedical Imaging Group Rotterdam, Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands
[2]Department of Oral & Maxillofacial Surgery, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands
[3]Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

ABSTRACT

Optical-based navigation systems are widely used in surgical interventions. However, despite their great utility and accuracy, they are expensive and require time and effort to setup for surgeries. Moreover, traditional navigation systems use 2D screens to display instrument positions causing the surgeons to look away from the operative field. Head mounted displays such as the Microsoft HoloLens may provide an attractive alternative for surgical navigation that also permits augmented reality visualization. The HoloLens is equipped with multiple sensors for tracking and scene understanding. Mono and stereo-vision in the HoloLens have been both reported to be used for marker tracking, but no extensive evaluation on accuracy has been done to compare the two approaches. The objective of our work is to investigate the tracking performance of various camera setups in the HoloLens, and to study the effect of the marker size, marker distance from camera, and camera resolution on marker locating accuracy. We also investigate the speed and stability of marker pose for each camera setup. The tracking approaches are evaluated using ArUco markers. Our results show that mono-vision is more accurate in marker locating than stereo-vision when high resolution is used. However, this comes at the expense of higher frame processing time. Alternatively, we propose a combined low-resolution mono-stereo tracking setup that outperforms each tracking approach individually and is comparable to high resolution mono tracking, with a mean translational error of $1.8 \pm 0.6mm$ for 10cm marker size at 50cm distance. We further discuss our findings and their implications for navigation in surgical interventions.

Keywords: surgical navigation, marker tracking, Microsoft HoloLens, ArUco

1. INTRODUCTION

Over the past years, traditional navigation systems have been widely adopted in the operating room for many types of surgeries.[1] These navigation systems are used to align the patient's preoperative model to the patient's target anatomy, allowing the surgeons to navigate and track their instruments intraopertively with respect to the preoperative model. Many of these navigation systems are based on optical tracking, which uses an optical camera to track rigid-bodies with reflective spheres that are attached to the patient and surgical instruments during the surgery. These systems display the preoperative model and instruments' tracking data on a 2D display beside the surgery table. This type of visualization suffers from poor hand-eye coordination as it forces the surgeon to look away from the operative field.[2] Augmented reality (AR) with a Head Mounted Display (HMD) permits in-situ visualization of the preoperative model as a 3D overlay aligned with the target anatomy. Therefore alleviating the issues of 2D screens and allowing the surgeon to focus at the task at hand.[3]

HMD headsets, such as the Microsoft HoloLens can be combined with optical tracking systems for surgical navigation,[4] where the optical tracking system provides reliable tracking and the HoloLens provides 3D overlay of the preoperative model. However, connecting the HoloLens to an external tracking system requires calibration and can increase the setup time and system complexity. Moreover, these optical systems require a direct line

Send correspondence to: a.thabit@erasmusmc.nl

Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by Cristian A. Linte, Jeffrey H. Siewerdsen, Proc. of SPIE Vol. 12034, 1203415
 · 1605-7422 · doi: 10.1117/12.2607262

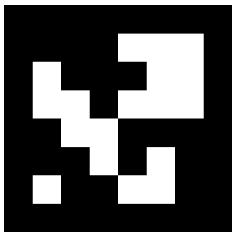

Figure 1. ArUco marker

of sight of the rigid-bodies with the tracking camera, which can limit the surgeon movements during surgery. Recently, much research has been done to study the possibility of using HMD headsets as a stand-alone navigation system,[5–7] where the headset is used for both tracking and visualization without relying on any external tracking device. This approach takes leverage of the built-in cameras in the headset to track 2D markers and therefore reduces the system costs and complexity. Several libraries in the past few years have been developed and proposed for 2D marker tracking, most notably ArUco,[8] AprilTag,[9] and Vuforia.[10] ArUco and AprilTag are open-source marker tracking libraries and have been reported to have good tracking accuracies with webcam cameras.[11] Vuforia, on the other hand, is an augmented reality multi-platform SDK that can detect salient features in a marker-image and estimate its pose, but it is not open-source and offers limited flexibility.[12] For surgical navigation, inside-out tracking of 2D markers using AR headsets is still being validated to ensure its clinical applicability.

One of the most popular AR headsets is The Microsoft HoloLens, where the interest in using it for surgical navigation has been increasing over the past few years. Several studies have leveraged the optical cameras in the HoloLens for marker tracking in surgical applications.[13,14] The HoloLens has two sets of camera sensors: 1) The RGB Photo-Video (PV) camera at the front, which is of high resolution and can be used for taking photos and recording videos, and 2) The gray-scale environmental cameras, two on the side and two at the front, which are of low resolution and can be accessed in research mode.[15] In the literature, some articles suggested using the PV mono camera for tracking,[16] while others used the Front-facing Environment (FE) stereo cameras.[7] Different accuracy values have been reported for each camera setup, but no study has been made to compare which is more accurate and more suitable for surgical navigation, and how these values depend on various parameters, such as marker size and tracking distance.

The aim of this work is to evaluate the Microsoft HoloLens in marker tracking for surgical applications. To that end, we compare the performance of two tracking approaches, one using the PV mono-vision camera, and another using the FE stereo-vision cameras. We evaluate the two tracking setups in terms of accuracy, speed, and stability. We further investigate how the pose estimation accuracy for each camera setup is affected by the size of the marker, distance from camera and camera resolution. We then propose a combined mono-stereo tracking approach for better marker pose estimation, and assess its performance. Finally, we conclude with recommendations for using the HoloLens for marker tracking in surgical settings.

2. METHODS

In this work, We choose ArUco for marker detection since it is an open source library and allows for both mono and stereo tracking to be implemented. It is also widely used and relatively easy to integrate and build for the HoloLens. ArUco detects square binary markers using basic image processing techniques. It first detects the square shaped marker candidates in the image using adaptive thresholding. Then, it identifies the binary code of the marker by classifying its inner-bits to get the marker ID and reject false candidates. Figure 1 shows a sample ArUco marker.

2.1 Marker pose estimation

In PV mono-vision camera tracking, the 3D transformation from the marker coordinate system to the camera coordinate system $T^C_{M_{PV}}$ can be estimated by solving a perspective-n-point (PnP) problem, given the camera

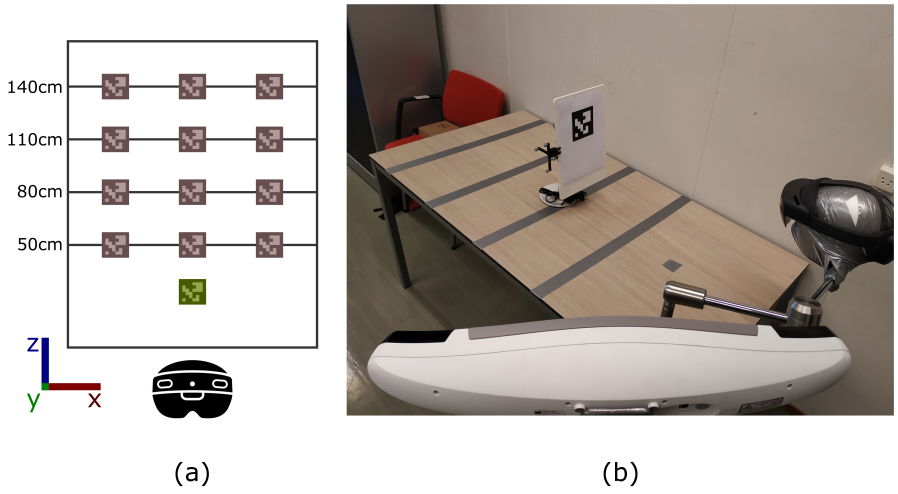

Figure 2. (a) The marker predefined positions, with the green marker position being the reference marker pose and red markers positions being the target marker poses to be evaluated. Coordinate system (CS) of the measurements is shown in the left-bottom corner, which is the CS of the reference marker. (b) The experimental setup, showing the optical tracker, the HoloLens and a tracked ArUco marker

calibration parameters and the marker corner points in the image plane.[17] The marker pose in the HoloLens-space $T^{HL}_{M_{PV}}$ can then be determined by multiplying the marker pose in camera coordinates $T^{C}_{M_{PV}}$ by the camera pose in HoloLens-space T^{HL}_{C}:

$$T^{HL}_{M_{PV}} = T^{HL}_{C} * T^{C}_{M_{PV}} \tag{1}$$

In Microsoft HoloLens (version 2) the camera matrix and distortion coefficients can be obtained directly from the HoloLens, and therefore no further camera calibration is required.

In FE stereo-vision camera tracking, the pose of the marker can be determined by triangulation as proposed by Liebmann et al.[7] To that end, the marker corners are detected in the image plane of both left and right cameras. Then, giving the pose of each camera in the HoloLens-space T^{HL}_{Cr} and T^{HL}_{Cl}, a ray from the camera centers to the marker corner points in the image plane is determined, and the 3D position of each corner point is found as the intersection of the two rays (i.e. by determining the point where the root mean square of the distances to both rays is smallest). The marker pose $T^{HL}_{M_{FE}}$ is then obtained by finding the rigid transformation that aligns the 3D positions of the marker corners to their respective positions in the marker model:[18]

$$T^{HL}_{M_{FE}} = [R|T] \quad . \tag{2}$$

R and T are the marker orientation and translation determined by finding the least-square solution for:

$$P^{R} = P^{M} * R + T \quad , \tag{3}$$

where P^R is the point-set of marker corners' 3D positions (i.e. intersections of camera rays), and P^M is the point-set of corner positions in marker Model.

To increase the robustness and stabilize marker pose estimation, linear Kalman filters are applied for both camera setups. Therefore, when a marker is detected in a new frame, the Kalman filter predicts its pose based on a state-space model and then update its prediction with each new estimated pose. For PV mono tracking, the Kalman filtering was applied to the final pose of the marker (position and orientation), similar to Kam et al.[19] For FE stereo tracking, the Kalman filters were applied to the 3D positions of the four marker corners instead.

2.2 Evaluation metrics

The tracking performance of the two tracking approaches were evaluated in terms of: locating accuracy, speed and pose stability.

Accuracy: For a tracking system to be used in surgical applications, it is of great importance for it to be able to accurately locate and track markers in the operative field. There are mainly two approaches for evaluating marker tracking accuracy. Either by directly assessing the marker pose, which requires the knowledge of the ground truth marker pose, or by measuring a relative change in the marker pose and assess it using an external ground truth tracking system. In this work, we opted for the later, where an NDI optical tracker was used to track and record the marker pose at the same time as the HoloLens, serving as the ground truth tracking system.

For relative pose assessment to be feasible, a calibration process was first performed to get the marker pose directly in the optical tracker's coordinate system. This was done by rigidly attaching a tracked reference-star to the AruCo marker-board and then locating the marker corners using a tracked pointer. These 3D corner positions were then registered to the marker model to get the marker calibration-pose in the optical tracker's coordinate system $T^{OT}_{M_{calib}}$. The rigid transformation of ArUco marker coordinate system to the reference-star coordinate system can then be determined as:

$$T^{RS}_{M} = (T^{OT}_{RS})^{-1} * T^{OT}_{M_{calib}} \quad , \tag{4}$$

which in turn can be used to find the marker pose $T^{OT}_{M_{NDI}}$ in the optical tracker's coordinate system:

$$T^{OT}_{M_{NDI}} = T^{OT}_{RS} * T^{RS}_{M} \quad . \tag{5}$$

The accuracy of marker pose estimation for HoloLens's mono and stereo tracking setups was determined by measuring the relative change in position and orientation of the ArUco marker with respect to a reference position and comparing it to that of the NDI optical tracker. Hence, given two marker poses, reference pose $T^{S}_{M_R}$ and target pose $T^{S}_{M_T}$, the relative change in pose $T^{M_R}_{M_T}$ can be found as follow:

$$T^{M_R}_{M_T} = (T^{S}_{M_R})^{-1} * T^{S}_{M_T}, \tag{6}$$

where S represents any tracking system (can be the NDI optical tracker OT or HoloLens HL). Then, from $T^{M_R}_{M_T}$, the translational and rotational errors can be calculated as:

$$E_T = d(T^{OT}, T^{HL}) \tag{7}$$

$$E_R = m(|R^{OT} - R^{HL}|) \tag{8}$$

where T and R are the translation vector (x, y, z) and rotational angles around the three axes $(\theta_x, \theta_y, \theta_z)$ respectively. d is the euclidean distance, and m is the mean.

There are many detection parameters that may contribute to the marker tracking performance; in this work, we focus on evaluating the two tracking approaches with the following detection parameters: marker size, distance from camera, and camera resolution. Other detection parameters, such as the size of marker margin, were chosen before the evaluation experiments for optimal detection rate, and therefore their affect to the pose estimation accuracy is assumed to be negligible.

Speed: It is important for marker tracking to be applied in real time, as lagging can hinder the applicability of utilizing the tracking system in surgical application. We assess the speed of the two camera tracking setups in terms of the time required to process one frame to get the marker pose.

Stability: for AR applications, markers are usually used as anchors for positioning virtual objects, and therefore stability of the marker pose is paramount for stable visualization. We assess the marker pose stability in terms of jittering, that is the standard deviation in marker position over a number of frames.

Table 1. Pose estimation translational error for marker size of 6cm in [mm]

Distance	FE640	PV640	PV760	PV960	PV1280	PV1920
50cm	4.0±1.2	11.5±2.5	7.5±2.4	7.6±1.7	7.6±1.2	2.9±2.0
80cm	14.0±1.7	24.0±2.2	22.3±1.3	18.7±1.6	17.0±1.5	7.9±3.3
110cm	25.3±5.7	32.3±1.6	29.1±3.9	30.3±4.5	24.3±2.4	8.7±2.5
140cm	52.7±2.2	53.2±12.9	42.7±4.0	39.3±3.7	38.1±2.6	15.2±2.2

Table 2. Pose estimation translational error for marker size of 8cm in [mm]

Distance	FE640	PV640	PV760	PV960	PV1280	PV1920
50cm	2.7±0.4	8.4±1.4	5.8±1.8	6.7±2.8	5.3±3.5	3.7±1.7
80cm	13.0±1.2	20.5±3.0	14.6±1.6	15.3±0.9	11.5±1.1	4.8±0.6
110cm	24.2±3.9	33.0±3.2	21.9±4.7	22.3±0.5	16.4±2.6	3.9±2.1
140cm	43.6±6.9	32.4±2.7	30.5±5.5	25.3±2.0	22.3±3.4	7.9±0.9

Table 3. Pose estimation translational error for marker size of 10cm in [mm]

Distance	FE640	PV640	PV760	PV960	PV1280	PV1920
50cm	3.3±0.2	5.6±1.8	6.3±1.6	5.4±2.1	4.6±1.3	2.8±1.7
80cm	7.4±1.5	15.6±0.3	15.8±1.4	14.9±1.2	13.4±0.3	5.4±1.1
110cm	30.0±5.0	20.9±1.5	22.1±3.8	18.6±1.7	19.7±1.9	8.3±0.7
140cm	56.9±5.0	35.7±3.9	27.1±0.9	27.4±3.9	27.9±2.2	15.8±2.4

3. EXPERIMENTS AND RESULTS

In this study, the HoloLens 2 (Microsoft corporation) headset was used. Marker tracking was implemented using ArUco and OpenCV (version 4.5.1). For both mono and stereo tracking, StereoKit (version 0.2.1) was used for creating the Holographic app, and all poses were reported in the StereoKit app world-coordinate system.

3.1 Marker pose accuracy

We study the effect of the marker size and its distance from the HoloLens in estimating the marker pose for both mono and stereo tracking setups. The following PV camera resolutions were considered: 640x360 (PV640), 760x428 (PV760), 960x540 (PV960), 1280x760 (PV1280) and 1920x1080 (PV1920), while the FE stereo cameras have only one resolution of 640x480 (FE640). For this experiment, three marker sizes were evaluated (6cm, 8cm and 10cm) at four distances from the HoloLens (50cm, 80cm, 110cm, 140cm) and at three different positions for each distance. A reference position was chosen at 30cm for measuring the relative pose error as discussed in 2.2. Figure 2a shows the marker positions considered for the experiment.

The HoloLens was mounted on a table and fixed in a static position and the ArUco marker was mounted on a stand and allowed to move across the table. Marker poses were acquired statically at the predefined positions for each PV camera resolution as well as for the FE stereo cameras and the optical tracker. For all positions, marker detection rate was ensured to be at 100% for all recorded frames. Figure 2b shows the experimental setup. Marker poses were recorded at the predefined positions for 100 consecutive frames. The mean pose of the marker at each position was then used for evaluation.

Tables 1-3 show the mean translational error at each distance for the different camera resolutions of the PV mono tracking and for the FE stereo tracking, for marker sizes of 6cm, 8cm and 10cm respectively. The results show that as the marker size increases and distance from camera decreases, the translational error decreases for all camera setups, this is more obvious in Figure 3, where we can see that PV1920 has the lowest translational error at all distances. FE stereo tracking also shows relatively low translational error at close distances (40cm - 80cm) but it increases exponentially as the distance increases.

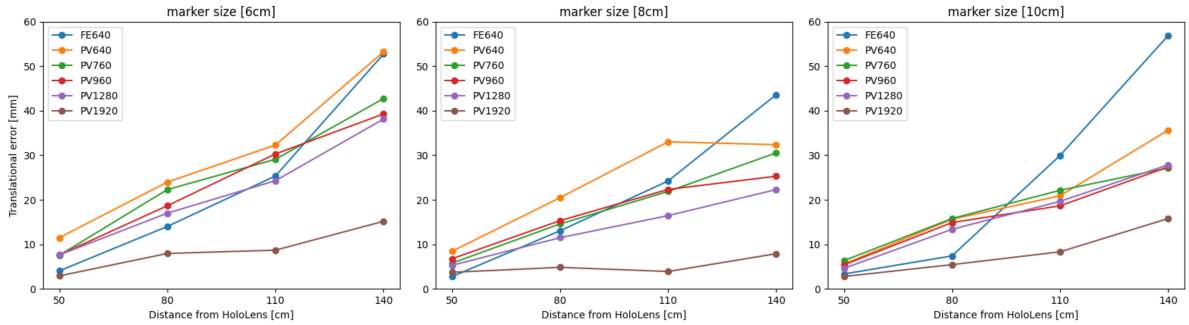

Figure 3. Transnational error in [mm] vs marker distance from the HoloLens, for 6cm, 8cm and 10cm marker sizes.

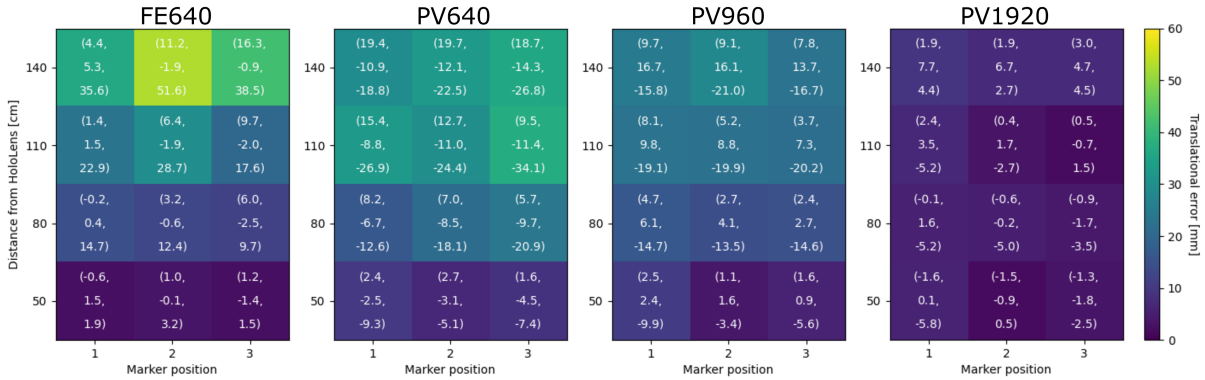

Figure 4. Heatmaps of the translational error for marker pose estimation at each marker position, for marker size 8cm. The values in each cell represent the (x, y, z) components of the translational error in [mm].

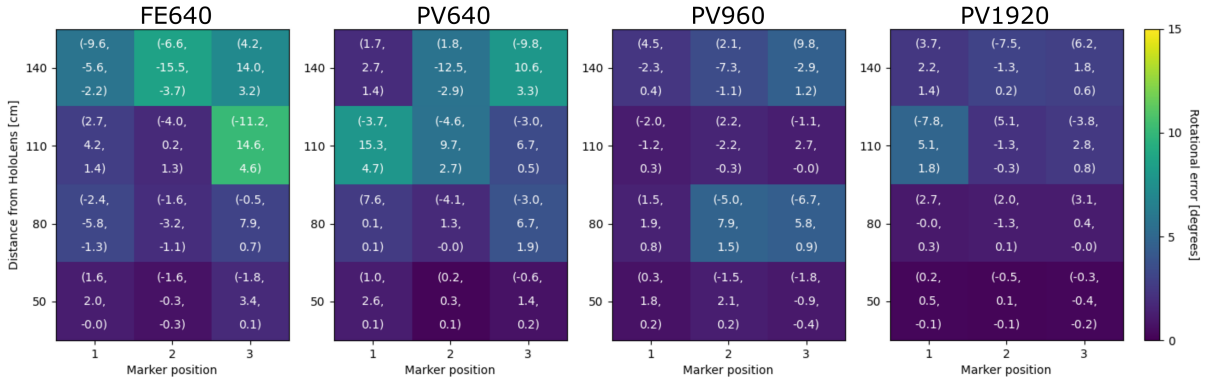

Figure 5. Heatmaps of the rotational error for marker pose estimation at each marker position, for marker size 8cm. The values in each cell represent the (x, y, z) components of the rotational error in [degrees].

To have a closer look at the result at each marker position, figure 4 shows heatmaps of the translational error for FE stereo tracking, and PV mono tracking for camera resolutions (PV640, PV960 and PV1920), for marker size 8cm. At each position the relative error with respect to the optical tracker measurement is represented as (x, y, z), while the color bar shows the overall translational error. Similarly, figure 5 shows the rotational errors for the same marker size and camera resolutions.

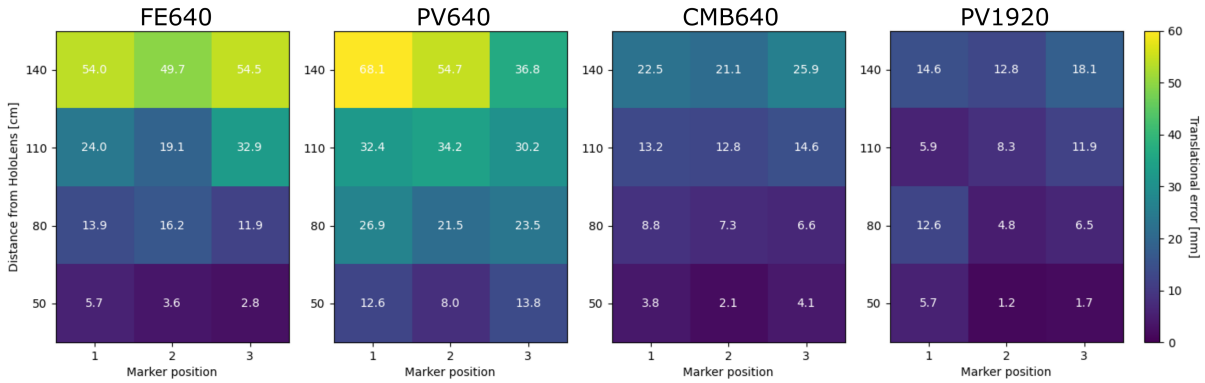

Figure 6. Heatmaps of the translational error for marker size 6cm comparing the locating error of the proposed combined mono-stereo tracking to each tracking approach individually as well as to the highest PV camera resolution. Error is in [mm].

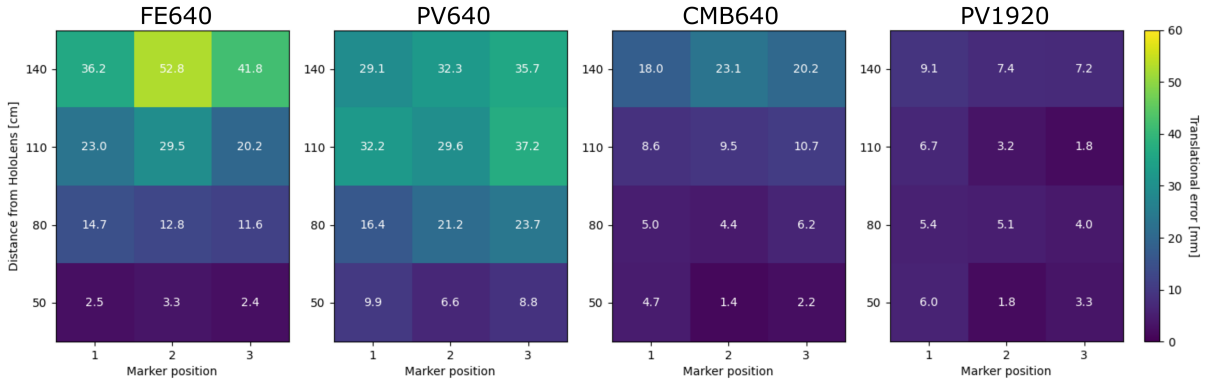

Figure 7. Heatmaps of the translational error for marker size 8cm comparing the locating error of the proposed combined mono-stereo tracking to each tracking approach individually as well as to the highest PV camera resolution. Error is in [mm].

3.2 Combined mono and stereo tracking

The results from evaluating each of the mono PV and FE stereo tracking suggested that a combination of the two tracking approaches could improve the pose locating accuracy when low PV camera resolution is used (PV640). In this experiment we evaluate the possibility of combining the two tracking approaches by taking the average of the z-axis transnational component of the marker pose for mono and stereo tracking approaches, while keeping the orientation and x-y translational components of the FE stereo tracking.

Figures 6-8 show the heatmaps of the translational error for PV640, FE640, combined mono-stereo tracking (CMB640), as well as for the highest PV resolution (PV1920), for marker sizes of 6cm, 8cm and 10cm respectively. The result show that the combined mono-stereo tracking approach outperforms each tracking approach individually, with a locating accuracy comparable to that of PV1920.

3.3 Processing time

For evaluating the speed of each tracking approach, the elapsed time for processing one frame to get the marker pose was recorded for 200 consecutive frames for both camera setups. The experiment was conducted at normal lighting conditions. The average processing time of each camera resolution of the PV camera and that of the FE stereo cameras were then compared.

Table 4 shows the processing time for each camera setup. We see that the processing time increases as the camera resolution increases.

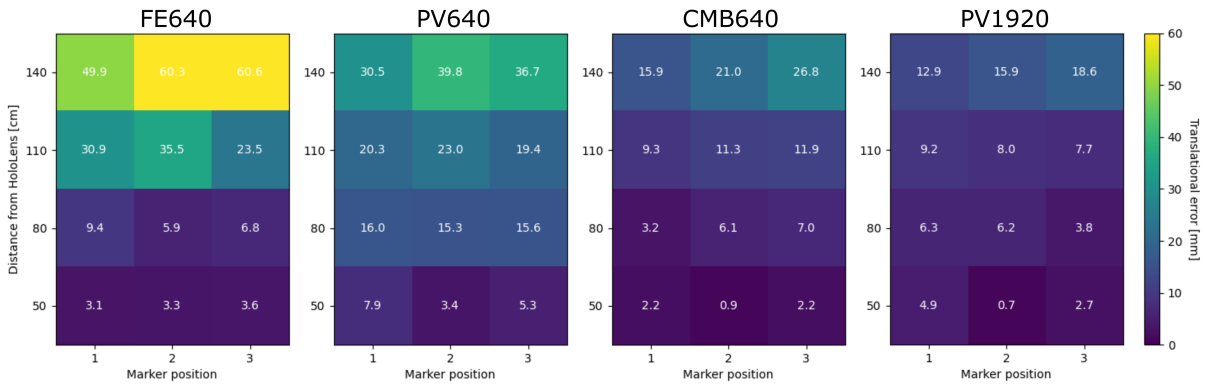

Figure 8. Heatmaps of the translational error for marker size 10cm comparing the locating error of the proposed combined mono-stereo tracking to each tracking approach individually as well as to the highest PV camera resolution. Error is in [mm].

Table 4. Processing time of one frame for each camera setup in [ms].

	EF stereo	PV640	PV760	PV960	PV1280	PV1920
processing time (ms)	46.7	43.7	52.4	69.5	106.9	228.1

3.4 Jittering in pose estimation

To evaluate the stability of the reported poses for both tracking setups, the marker poses of 200 consecutive frames were recorded and the standard deviation from the mean pose was measured. For this experiment, we compared the marker pose stability with and without applying the Kalman filter. A marker of size 6cm was positioned at a distance of 100cm.

Table 5 shows the standard deviation from mean marker pose for translation, with and without applying kalman filtering for the different camera setups.

4. DISCUSSION

From the tables 1-3 and figure 3, we see that PV640 mono tracking has the highest locating error. This can be attributed to its low resolution and poor locating of the marker corners. FE640 stereo tracking while also has low resolution, it showed lower locating error at close distances (40cm - 80cm) due to its stereo vision, however as the distance increases, the locating error increases exponentially. We can see from figure 4 that as the marker distance from camera increases, FE stereo tracking tends to underestimate the depth component of the marker pose (i.e. z-axis along the viewing angle of the HoloLens), while PV640 tends to overestimate the depth component. This difference in estimating the marker depth between mono and stereo tracking could increase up to 8cm for 6cm marker size at 140cm distance. This difference could be leveraged to even out the marker depth estimation of the two approaches and resulting in a more accurate locating of the marker as proposed in the combined mono-stereo tracking (see figures 6-8), which has a locating accuracy that is comparable to that of the highest PV resolution (PV1920).

Table 5. translational standard deviation from mean marker pose in [mm].

Jittering [mm]	EF stereo	PV640	PV760	PV960	PV1280	PV1920
without Kalman filter	1.10	0.08	0.10	0.65	0.87	0.90
With Kalman filter	0.27	0.13	0.10	0.42	0.29	0.52

For most surgical applications that use head mounted devices for tracking, such as the HoloLens, the tracking area of interest is mostly between 30cm-80cm, where the surgeons can track the patient, surgical tools or other instruments in the operative field. For this tracking area, our results show that the HoloLens can have a tracking error in the range of 0.5mm - 6mm for PV1920 or combined mono-stereo tracking when a large marker is used (8cm-10cm). This tracking accuracy is still relatively high for surgical navigation, but it has the potential to allow the utilization of HoloLens for application where perpendicular view can be maintained and therefore neglect the translational error along the viewing angle.

It is worth mentioning that our assessment approach is based on relative pose evaluation, which only takes into account the pose estimation error of the target marker with respect to the reference marker, assuming the locating of the reference marker does not introduce additional error. Furthermore, only twelve marker positions were considered for this evaluation, therefore, due diligence is needed when interpreting the results as absolute values.

5. CONCLUSION

In this work we evaluated the mono-vision and stereo-vision tracking approaches in the HoloLens, and we compared their performance in terms of locating accuracy, processing time and pose stability. We evaluated the locating accuracy of each tracking setup in relation to the marker size, its distance from the camera, and the camera resolution used. We showed that mono-vision tracking have higher locating accuracy than stereo-vision tracking when high resolution of the PV camera is used. However, due to the longer processing time needed, we proposed a combined low-resolution mono-stereo tracking approach that corrects for the high error in depth estimation when each tracking approach is used individually. Our results suggest that the proposed combined mono-stereo tracking have the potential to be used for marker tracking at close distances (30-80cm) with good accuracy that is comparable to marker tracking using the high resolution setup of the PV camera. Further investigation is required to study the robustness of this approach and its applicability in surgical settings.

REFERENCES

[1] Cleary, K. and Peters, T. M., "Image-guided interventions: technology review and clinical applications," *Annual review of biomedical engineering* **12**, 119–142 (2010).

[2] Herrlich, M., Tavakol, P., Black, D., Wenig, D., Rieder, C., Malaka, R., and Kikinis, R., "Instrument-mounted displays for reducing cognitive load during surgical navigation," *International journal of computer assisted radiology and surgery* **12**(9), 1599–1605 (2017).

[3] Yoon, J. W., Chen, R. E., Kim, E. J., Akinduro, O. O., Kerezoudis, P., Han, P. K., Si, P., Freeman, W. D., Diaz, R. J., Komotar, R. J., et al., "Augmented reality for the surgeon: systematic review," *The International Journal of Medical Robotics and Computer Assisted Surgery* **14**(4), e1914 (2018).

[4] Meulstee, J. W., Nijsink, J., Schreurs, R., Verhamme, L. M., Xi, T., Delye, H. H., Borstlap, W. A., and Maal, T. J., "Toward holographic-guided surgery," *Surgical innovation* **26**(1), 86–94 (2019).

[5] van Doormaal, T. P., van Doormaal, J. A., and Mensink, T., "Clinical accuracy of holographic navigation using point-based registration on augmented-reality glasses," *Operative Neurosurgery* **17**(6), 588–593 (2019).

[6] Frantz, T., Jansen, B., Duerinck, J., and Vandemeulebroucke, J., "Augmenting microsoft's hololens with vuforia tracking for neuronavigation," *Healthcare technology letters* **5**(5), 221–225 (2018).

[7] Liebmann, F., Roner, S., von Atzigen, M., Scaramuzza, D., Sutter, R., Snedeker, J., Farshad, M., and Fürnstahl, P., "Pedicle screw navigation using surface digitization on the microsoft hololens," *International journal of computer assisted radiology and surgery* **14**(7), 1157–1165 (2019).

[8] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J., "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition* **47**(6), 2280–2292 (2014).

[9] Olson, E., "Apriltag: A robust and flexible visual fiducial system," in [*2011 IEEE International Conference on Robotics and Automation*], 3400–3407, IEEE (2011).

[10] Incorporated, Q., "Qualcomm vuforia."

[11] Kunz, C., Genten, V., Meißner, P., and Hein, B., "Metric-based evaluation of fiducial markers for medical procedures," in [*Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*], **10951**, 109512O, International Society for Optics and Photonics (2019).

[12] La Delfa, G. C., Catania, V., Monteleone, S., De Paz, J. F., and Bajo, J., "Computer vision based indoor navigation: a visual markers evaluation," in [*Ambient Intelligence-Software and Applications*], 165–173, Springer (2015).

[13] Perkins, S. L., Lin, M. A., Srinivasan, S., Wheeler, A. J., Hargreaves, B. A., and Daniel, B. L., "A mixed-reality system for breast surgical planning," in [*2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*], 269–274, IEEE (2017).

[14] Luzon, J. A., Stimec, B. V., Bakka, A. O., Edwin, B., and Ignjatovic, D., "Value of the surgeon's sightline on hologram registration and targeting in mixed reality," *International Journal of Computer Assisted Radiology and Surgery* **15**(12), 2027–2039 (2020).

[15] Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Meekhof, C., Stühmer, J., Cashman, T. J., Tekin, B., Schönberger, J. L., Olszta, P., et al., "Hololens 2 research mode as a tool for computer vision research," *arXiv preprint arXiv:2008.11239* (2020).

[16] Hu, X., y Baena, F. R., and Cutolo, F., "Head-mounted augmented reality platform for markerless orthopaedic navigation," *IEEE Journal of Biomedical and Health Informatics* (2021).

[17] Collins, T. and Bartoli, A., "Infinitesimal plane-based pose estimation," *International journal of computer vision* **109**(3), 252–286 (2014).

[18] Besl, P. J. and McKay, N. D., "Method for registration of 3-d shapes," in [*Sensor fusion IV: control paradigms and data structures*], **1611**, 586–606, International Society for Optics and Photonics (1992).

[19] Kam, H. C., Yu, Y. K., and Wong, K. H., "An improvement on aruco marker for pose tracking using kalman filter," in [*2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*], 65–69, IEEE (2018).