ORIGINAL ARTICLE

Journal Section

A Hand-Eye Calibration Method for Augmented Reality Applied to Computer-Assisted Orthopedic Surgery

M. E. de Oliveira PhD[1] | H. G. Debarba PhD[1] |
A. Lädermann MD[2,3] | S. Chagué MSc[1] |
C. Charbonnier PhD[1,4]

[1] Medical Research Department, Artanim Foundation, Geneva, 1217, Switzerland

[2] Division of Orthopedics and Trauma Surgery, La Tour Hospital, Geneva, 1217, Switzerland

[3] Department of Orthopedic Surgery and Traumatology, Geneva University Hospital, Geneva, 1205, Switzerland

[4] Faculty of Medicine, University of Geneva, Geneva, Switzerland

Correspondence
M. E. de Oliveira PhD, Medical Research Department, Artanim Foundation, Geneva, 1217, Switzerland
Email: marcelo.eliasoliveira@artanim.ch

Funding information

Background: Augmented reality (AR) allows the surgeon to represent holographic patient-specific anatomical information and surgical instruments in the physical world. To correctly superimpose virtual and physical objects, a hand-eye calibration method for mapping the virtual and physical spaces was proposed.
Methods: Mathematical relationships between the virtual camera and the physical space were derived. Finally, the accuracy and robustness of the proposed HE calibration method were qualitatively and quantitatively evaluated.
Results: The proposed calibration method allows us to determine an optimal invariant spatiotemporal mapping between the virtual camera and the physical space.
Conclusion: Qualitatively and quantitatively reliable and accurate estimates for the physical-virtual mapping transformation were verified. Consequently, imaging data and surgical instruments holograms can be precisely represented in the physical space.

KEYWORDS
Computer Assisted Surgery, Augmented Reality, Calibration

1 | INTRODUCTION

Since its introduction about two decades ago, computer-assisted surgery (CAS) has become available for a wide range of surgical interventions due to increased availability of different imaging modalities, and advances in surgical instruments and tracking hardware. Single or multimodal imaging datasets, such as computed tomography (CT), magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), and ultrasound (US) can, therefore, be used to generate patient-specific anatomical and functional models. Once the correspondences between anatomical landmarks or implanted fiducial are established in the physical and imaging reference frames, preoperative planning tools, and computer-aided surgery systems can be used to represent positions and orientations of anatomical structures and of surgical instruments into a common reference frame, with a high degree of accuracy and consistency [1].

Current commercial surgical navigation platforms are non-immersive systems (e.g., BrainLAB, Germany; Medtronic, Switzerland; and Medacta International, Switzerland), where the surgical planning, patient information, and tracking data are displayed on nearby large-screen surgical displays located at or near the surgeon's working space. Therefore, attention disruptions are more likely to occur in terms of the surgeon's gaze, and lack of intuitiveness may possibly exist, especially among inexperienced surgeons [2, 3].

Due to the remarkable advances in both surgical instruments and techniques, as well as in functional and structural multimodality imaging, it becomes highly important to provide the clinical staff with all the relevant patient-specific information and preoperative planning in real-time, aiming at improving surgical outcomes, while decreasing postoperative surgical complications and patient care costs. However, in order to take full advantage of all these recent advances, new visualization and display technologies as well as human-machine interaction interfaces were needed, which led to the introduction of new AR paradigms in the field of Computer Assisted Surgery (CAS). Some of these advantages have been stated in recent studies in the fields of laparoscopic surgery [4, 5], orthopedics [6], neurosurgery [7], and in oral and craniomaxillofacial surgery [8, 9, 10, 11].

Over the past few years, the computer science community has made considerable efforts towards the development of calibration techniques for wearable optical see-through head-mounted displays (OST-HMDs), which still remains a challenging and tailored problem due to the complexities associated with specific image projection systems, optics, as well as to neurophysiological features of the human visual system. Currently, available solutions to determine computational relationships between the OST-HMDs imaging system and the user's eyes are established based on image-vision methods and external tracking systems, which are cumbersome and error-prone, largely relying on subjective assessments provided by the feedback from users [12, 13, 14, 15, 16].

Recent technology advances in AR is creating an opportunity for a paradigm shift in the field of Computer Assisted Surgery (CAS). In March 2007, Microsoft released the HoloLens Development Edition. This platform is a wearable high-definition stereoscopic 3D optical head-mounted system equipped with near-eye multifocus dichromated gelatin holographic lenses, two pairs of stereoscopic gray-scale cameras, one depth camera, an RGB camera, and an inertial measurement unit (IMU). Multisensor data fusion techniques combined with sophisticated hardware- and software-assisted methods were developed by the Microsoft research group to solve important subjectivity-related issues involved in the calibration process of OST-HMDs [17, 18, 19]. These methods are accomplished by propagating features existing in the physical space into the virtual world, and by mimicking biological principles of stereo vision in the human visual system in the absence of an eye-tracking system, where the user-specific inter-pupillary distance (IPD) is the only needed information. Moreover, a sophisticated hardware-assisted holographic computation technique called *stabilization plane* is employed based on the user's gaze[1] [20]. These combined techniques provide precise holographic

---

[1] In the current Microsoft HoloLens development version, the user's gaze is determined by computing the intersection between a projected ray from the middle of the camera viewport with the surface mesh model representing the physical space.

rendering capabilities of the virtual content, where the latter is represented with respect to its poses in the physical space and not in relation to the position of the user's eyes in relation to the headset reference frame [21, 22].

The pose of the HoloLens is determined by a proprietary algorithm for Simultaneous Localization and Mapping (SLAM), which estimates the estates (i.e., position, orientation, linear accelerations, and angular velocities) of the headset with respect to the physical space by combining multisensor data fusion techniques, e.g., inertial, depth, and visual cues). Recently, several studies have explored the possibility of using the Microsoft HoloLens in neurosurgery [10, 23, 24], reconstructive and plastic surgery [25], liver cancer surgery [26], endovascular surgery [27], and in telemedicine [28]. In all these previously mentioned studies, the authors have not estimated the transformation between the physical and virtual reference frames. The virtual content representing a physical object and/or a surgical instrument was manually placed and represented with respect to the virtual reference frame. These studies were mainly focused on ergonomics, usability, and computer-machine interfaces rather than quantitative and methodological aspects.

Although indoor/outdoor SLAM-based methods have reached a considerable state of maturity, and that the implemented HoloLens SLAM-based method being able to accurately and consistently estimate the states of the virtual camera in the context of recreational applications, additional efforts are needed in order to establish a highly precise and stable representation of holograms with respect to their respective counterparts in the physical space, which is a highly desirable characteristic, especially in the field of CAS.

Recently, Long et al. [29] have proposed two different hybrid-based methods to address the HoloLens calibration problem. The first method is based on a head-anchored technique, while the second one is a world-anchored method relying on an external optical tracking system to determine the pose of a reference object with respect to the physical space, which has its counterpart representation in the virtual space. In both evaluated methods, the pose of the HoloLens's virtual camera was obtained by using its own SLAM-based method and, it has also been assumed that the pose of the HoloLens's virtual camera could be determined with a higher degree of reliability and accuracy.

Despite the Microsoft HoloLens being a promising technology in the field of CAS, the accuracy, robustness, and performances of its SLAM-based method under specific environmental conditions have not been fully investigated and evaluated. In our work, we have quantitatively investigated the accuracy of the HoloLens Microsoft SLAM-based algorithm, supporting the need for the development of more suitable and accurate calibration methods. Finally, we proposed a Hand-Eye (HE) based approach combined with a commercial optical tracking system to determine a transformation mapping between the physical world and the virtual camera model. Although we considered an external optical tracking system combined with vision-based methods, our work fundamentally differs from Long et al. [29] in its formulation and implementation, since only the holographic near-eye display of the Microsoft HoloLens is used and, therefore, the inherent uncertainties arising from SLAM-based methods do not play any role in our transformation matrices estimates.

2 | METHODS

2.1 | Assessing the accuracy and reliability of the HoloLens SLAM-based method

In this work, we initially created an objective benchmark for evaluating the precision of the SLAM method employed by the Microsoft HoloLens. Initially, we investigated the convergence of the reconstructed surface meshes starting from a naive physical model $\mathcal{M}_0$. A subject wearing the HoloLens headset was asked to walk inside a rectangular region of 10 m long and 5 m wide at a self-selected walking speed describing an elliptical trajectory. When completing each lap n, the surface mesh $\mathcal{M}_n$ representing the n-th mesh model of the physical space was stored for offline post-processing. Finally, the model convergence, accuracy, and reliability were assessed by evaluating the similarities between consecutive mappings of the same physical space by using nearest neighbor analysis. Subject-motion tracking data was simultaneously acquired by the HoloLens tracking middleware, as well as by an external optical motion capture system (VICON, Oxford Metrics, UK) by using a set of non-collinear retro-reflective markers rigidly attached to it. Consequently, the trajectories datasets represented in these two distinct reference frames (i.e., VICON and HoloLens) were transformed to the principal component space by applying the transformations T_w^λ and T_s^λ, as shown in Figure 1, where the subscripts s and w denote the SLAM and the VICON reference frames, respectively, and the subscript λ denotes the principal component space. Finally, a transformation matrix T_c^h relating the headset reference frame and the OST-HMD's virtual camera was derived by solving the following matrix equation: $T_c^h = T_w^h T_\lambda^w T_s^\lambda T_c^s$, as graphically represented in Figure 1, where the transformation T_c^s represent the pose of the virtual camera represented in the virtual space and the T_h^w the pose of the orthonormal basis h defined by the set of non-collinear retro-reflective markers attached to the HoloLens headset with respect to the physical space. It is important to note that the HoloLens headset has its representation in the physical and in the virtual reference frames, and that each point in the elliptical motion trajectory has an orthonormal transformation mapping T_c^h associated to it, where the latter would obviously be identical in an idealized error-free tracking process.

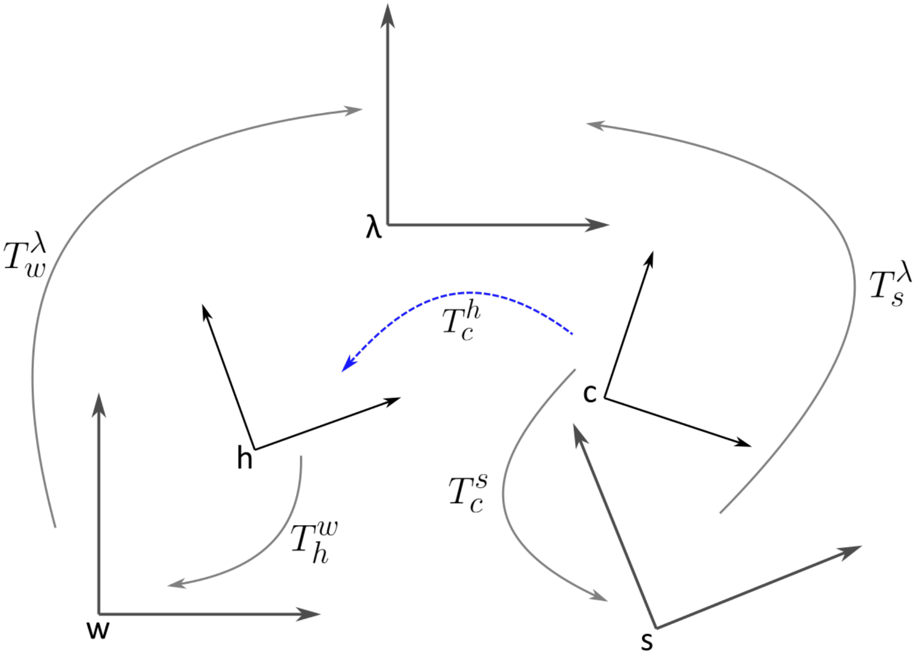

FIGURE 1 SLAM-PCA. ***v***: VICON reference frame (physical space). ***s***: SLAM reference frame. ***h***: HoloLens headset physically represented into the VICON reference frame. ***c***: HoloLens virtual camera represented into the SLAM reference frame (virtual space). ***λ***: Principal component space. T_c^h: Computed calibration mapping represented by a dashed blue line, where the known transformations are represented by continuous gray lines.

2.2 | Latency estimation

The latency of the developed application is attributed to different factors, such as network streaming, hardware computing capabilities, physical material characteristics of the near-eye displays, as well as different challenges in multitarget tracking. The system's latency was computed based on the mid-stance walking pattern, which was accomplished by asking a subject wearing the HoloLens headset to walk at a self-selected walking speed in a predefined space. Time-series representing the HoloLens headset were simultaneously acquired in the physical and virtual spaces by the optical motion capture and by the HoloLens SLAM-based localization systems, respectively. These two time-series with homologous spatiotemporal characteristics were then mapped into the principal component space. The mid-stance walking pattern is characterized by a tibiofemoral extension combined with a ankle dorsiflexion, moving the body over the stationary foot, which can then be easily detected by finding local maxima and minima of the gait time-series in the third component, where the third component is mainly associated with the perpendicular motion of the headset with respect to the walking surface [30, 31].

2.3 | Pointer calibration and checkerboard-pattern digitization

A set of non-collinear retro-reflective markers were rigidly attached to the Microsoft HoloLens headset and to the surgical instruments, as shown in Figure 2. The surgical pointer was calibrated by requesting a subject to place the tip of the pointer on a rigid surface and to perform randomly selected pure 3D rotations, ensuring that any point represented with respect to this rigid body is characterized by an invariant locus. During this calibration step, the pointer reference frame was tracked by the optical motion capture system, and the tip of the surgical pointer was determined by representing locally the center of the parameterized sphere with respect to the surgical pointer reference frame by using a parametric least-square approach.

Checkerboard patterns were digitized by using the calibrated optically tracked pointer. Before acquiring the positions of each individual feature on the checkerboard calibration border, the tip of the pointer was carefully positioned at each individual corner and remained stationary. Acquired features were locally represented with respect to the calibration board coordinate frame T_p^b, and the random sample consensus algorithm (RANSAC) was applied to accurately detect the estimation of an ideal checkerboard-pattern, ensuring the detection of parallel and perpendicular lines [32]. Finally, the calibration pattern with known geometrical characteristics was used to determine the HoloLens's relative pose with respect to the calibration object T_c^p, which was then used to determine its absolute pose with respect to the physical space [32], as shown in Figure 3.

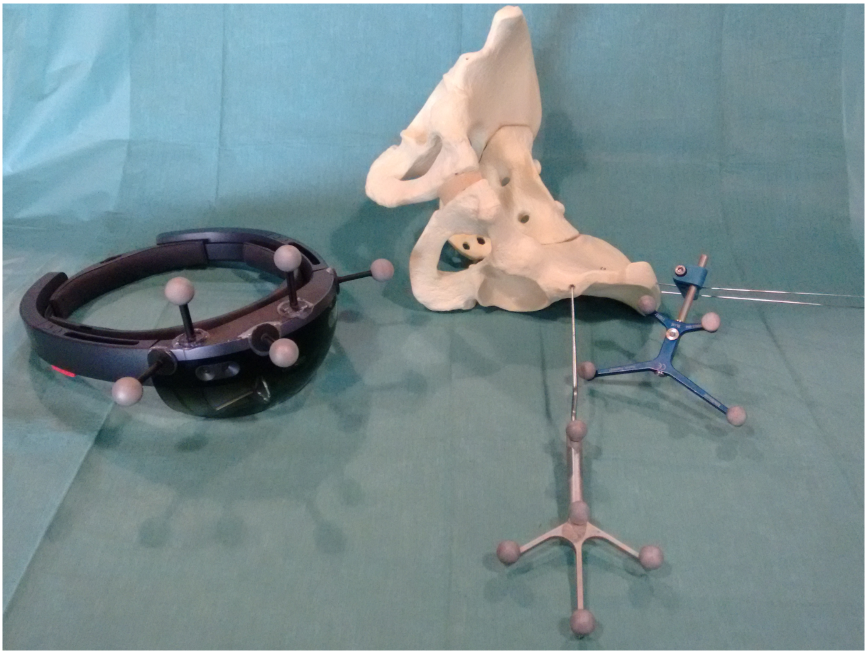

FIGURE 2 Experimental setup consisting of a Microsoft HoloLens with a set of non-collinear retro-reflective markers (⌀ 19 mm) rigidly attached to it, organ of interest, surgical pointer, surgical reference frame anchored at the organ of interest by using two Kirschner wires, and a VICON optical motion capture system (VICON, Oxford Metrics, UK) that is not depicted in this image.

2.4 | HE-based calibration method applied to AR in CAS

HE calibration method has been originally developed for serial manipulators by the robotics community, where a camera was mounted on an end effector (e.g., gripper) and were abstractly related to eye and hand, respectively. In the robotics context, the goal of this method is to determine the unknown transformation from the camera (eye) reference frame to the end effector (hand) reference frame [33, 34]. On other hand, in our proposed formulation, the digitized checkerboard patterns were used to create a reference frame $\boldsymbol{p}$, which has then been locally represented with respect to the reference frame of a rigid and flat surface $\boldsymbol{b}$. Therefore, the checkerboard pattern reference frame can be represented with respect to the absolute reference frame defined by the optical tracking motion camera system $\boldsymbol{w}$, as shown in Figure 3.

HoloLens RGB intrinsic camera parameters, as well as their radial and tangential lenses distortion coefficients were determined by simultaneously solving a set of homogeneous system of equations by using nonlinear optimization techniques [34]. This information were then posteriorly used in our application to remove image distortions, as well as to redefine the frustum of the HoloLens virtual camera. For more detailed information, we refer the reader to the Chapters 6 and 9 of [35].

By using the computed HoloLens RGB camera extrinsic camera parameters, the relationships between the physical and virtual reference frames T_c^h can be derived by the following matrix equation: $T_c^h = T_w^h T_b^w T_p^b T_c^p$, as graphically represented in Figure 3. Note that the calibration matrix T_c^h can be determined from a single image and, therefore, the checkerboard-pattern is obviously not required to be in the operating room, since this procedure can be conducted prior to the clinical application of the AR system. During the validation process, multiple images of the calibration board were acquired to verify the variances of the computed transformations.

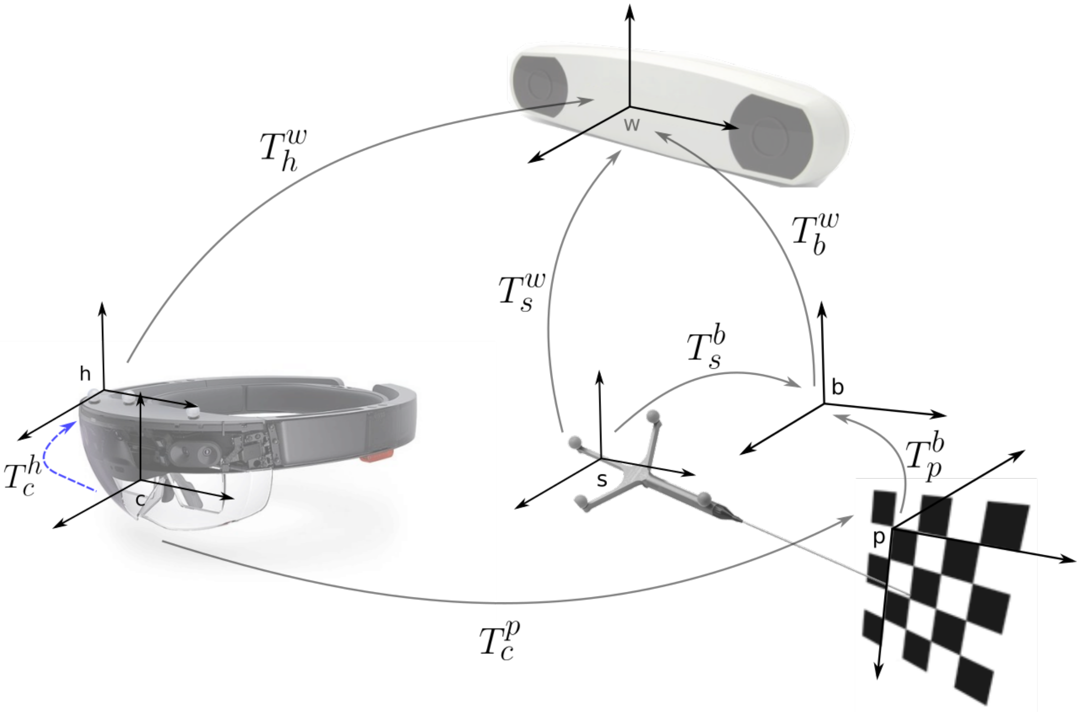

FIGURE 3 Orthonormal Basis (OB) ***h***: HoloLens headset with passive markers attached to it in order to define its position with respect to the physical world. OB ***s***: surgical pointers reference frame used to digitize the checkerboard-pattern and to define fiducial and/or anatomical landmarks, which are then used to represent the organ of interest (e.g., femur and pelvis) with respect to the surgical frame. The transformation mapping T_c^h needed to be determined is highlighted by a blue dashed line. Note that the HoloLens virtual camera is intrinsically related to the user's visual system (i.e., OB ***c***). T_c^p: Transformation from the checkerboard-pattern to the HoloLens RGB camera. T_b^w: Transformation mapping the calibration board with respect to the physical world. T_p^b: Transformation mapping between the digitized checkerboard-pattern with respect to the calibration board represented into the physical space. T_h^w: Transformation mapping the HoloLens headset with respect to the physical world. T_s^w: Mapping between the surgical pointer and world reference frames. T_s^b: Mapping between the surgical pointer and the calibration board reference frames.

2.5 | Application of the HE-based method to computer-assisted orthopaedic surgery

In order to validate the proposed HE calibration method for AR in computer-assisted orthopaedic surgery (CAOS), we considered a full male pelvis model with acetabulum and cancellous inner structures (SKU 1301-1, Sawbones, Pacific Research Laboratories, Vashon, Washington, USA) and a femur-tibia articulated with ligaments and cancellous inner structures (SKU 1145-1, Sawbones, Pacific Research Laboratories, Vashon, Washington, USA). Titanium fiducials screws were implanted in these models as shown in Figures 4(a) and 4(b). These two models underwent computed tomography using a VCT scanner (General Electric Lightspeed VCT 64 rows system, Milwaukee, WI) assuming an interplanar spacing resolution of 0.63 mm. Subsequently, these two datasets were segmented by using the commercial Mimics software (Materialise NV, Leuven, Belgium). Isosurface meshes representing the cortical structures and their respective fiducial markers were generated and were then exported in stereolithography format (STL, 3D Systems Inc, Valencia, Calif). Finally, these isosurface meshes were imported and processed using the open source visualization toolkit library (VTK 8.1.0, Kitware, `https://www.vtk.org/`). In order to establish a correspondence between the physical and imaging spaces, the following methodology was applied: first, the nodes of the fiducial markers isosurface meshes were extracted and clustered by using the K-mean clustering method [36], where the total number of clusters to be detected was equal to the number of fiducial markers. Second, the detected clusters were mapped into the principal component space by solving an eigenvalue decomposition problem based on the nodes' spatial positions [36], where the directions of the fiducial markers were associated with the first eigenvector. Last, the main axes directions were parameterized and their respective intersections with the surface meshes representing the cortical structures were determined, as shown in Figures 4(a)-4(c).

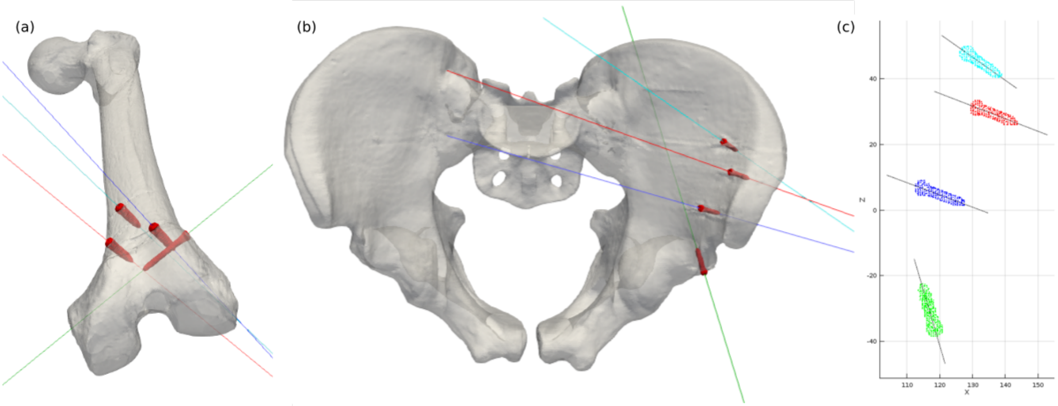

FIGURE 4 (a) and (b): Extracted cortical structures and fiducial markers meshes isosurfaces. The computed fiducial markers main directions are represented by the four parametrized lines. (c) A X-Z projection of the clustered nodes with their respective main axes are represented for the pelvis model. It is important to note that the color identification for each detected cluster has its parametrized line in the same color. The intersections between these main directions and their respective surface meshes were then used to determine an orthonormal basis in the image space, which has its equivalent in the physical world represented into the surgical frame coordinate system.

2.6 | Establishment of correspondences between the physical and imaging spaces

Initially, a surgical frame was anchored to the organ of interest by using K-wires as shown in Figure 10. The developed AR CAOS application for Microsoft HoloLens was implemented using the Mixed Reality Toolkit Library - Unity (`https://github.com/Microsoft/MixedRealityToolkit-Unity`). This application allows the surgeon to navigate among different options, such as loading patient data, conducting calibration, establishing correspondences between the physical and the virtual organs of interest based on fiducial or anatomical landmarks, and tracking rigid bodies by using the extended Kalman filter. After loading the desired virtual content, a hologram of the desired model emerges at one meter from the middle of the camera viewport (user's visual system). The user can then move the object in the physical space by using simple hand gestures (i.e., by pinching the index finger and thumb together) and move its hand towards the desired position. Therefore, the model can be placed into a convenient position for the surgeon. When selecting the option *Establishment of Correspondences* on the main menu, consecutive landmarks representing fiducial or anatomical markers will start to blink in the loaded model as previously defined in the preoperative phase and, by using the tracked surgical pointer and a double-click gesture, each individual correspondence is established. The surgeon is also provided with an audio feedback, ensuring that he/she would move the surgical pointer to the next highlighted fiducial marker. These fiducial markers are then mapped with respect to the surgical frame anchored into the organ of interest. After finishing this procedure, an orthonormal basis is created in the physical space, which has its counterpart in the image space and, therefore, a transformation can be determined to map the hologram with respect to the surgical frame.

2.7 | Accuracy assessment

The accuracy of the proposed HE-based method was evaluated by using a retro-projection error estimation (RPEE) approach. As previously described, the checkerboard pattern was digitized by using the surgical navigation pointer (Figure 2) and therefore, the position of each individual pattern (i.e., internal corners of the checkerboard pattern) was determined with respect to the world reference frame by considering the estimated transformation matrix T_c^h defined in Section 2.4. Holograms representing the internal edges were created and mapped into the physical space. Since these holograms representing the internal edges of the assumed checkerboard calibration object have their counterparts in the real world, the root mean square error (RMSE) of the 2D representations of these holograms were compared with the acquired checkerboard images by the HoloLens RGB camera, where the latter is in agreement with the HoloLens's virtual camera. The inner checkerboard corners and the centers of the holograms representing the projected patterns in the physical space were detected by using the opensource computer vision library (OpenCV, http://opencv.org). For more details, we refer the reader to [37].

Finally, the variabilities of the estimated transformation mappings were assessed by using directional descriptive statistics to quantitatively evaluate the accuracy and robustness of the HoloLens-SLAM algorithm and of the proposed HE-based method. The latter was considered since commonly used descriptive statistics on Euclidean space are not appropriate for describing orthonormal rotations matrices. In this context, the orthonormal basis were assumed to be lying on a compact Riemannian manifold, which is the natural generalization of Euclidean spaces to locally Euclidean spaces [38, 39, 40]. In this context, the uncertainties resulting from the computed transformation matrices T_c^h for both methods can be indirectly assessed by the circular and offset variances.

3 | RESULTS

3.1 | Quantitative assessment of the HoloLens SLAM-based method

Preliminary investigations on mesh stability and reliability for the HoloLens SLAM-based method were qualitatively and quantitatively evaluated by conducting Nearest Neighbour Distances (NNDs) analysis and the Hausdorff distances (H_d). The experiments were conducted on a rectangular room 16.9 m long, 8.2 m wide, and 3.7 m height, where the mapping was conducted along a rectangular region of 10 m long and 5 m wide under quasi-static dynamic consideration, as described in Section 2.1. Similarities between two consecutive surface meshes with the same inter-lap interval are illustrated in Figure 5. Hausdorff distances as well as the minimum Euclidean distances and their statistical properties were computed and, it was verified that no significant differences were observed after eight completed laps ($\mathcal{M}_{10}$, $\mathcal{M}_7$), as shown in Figure 6.

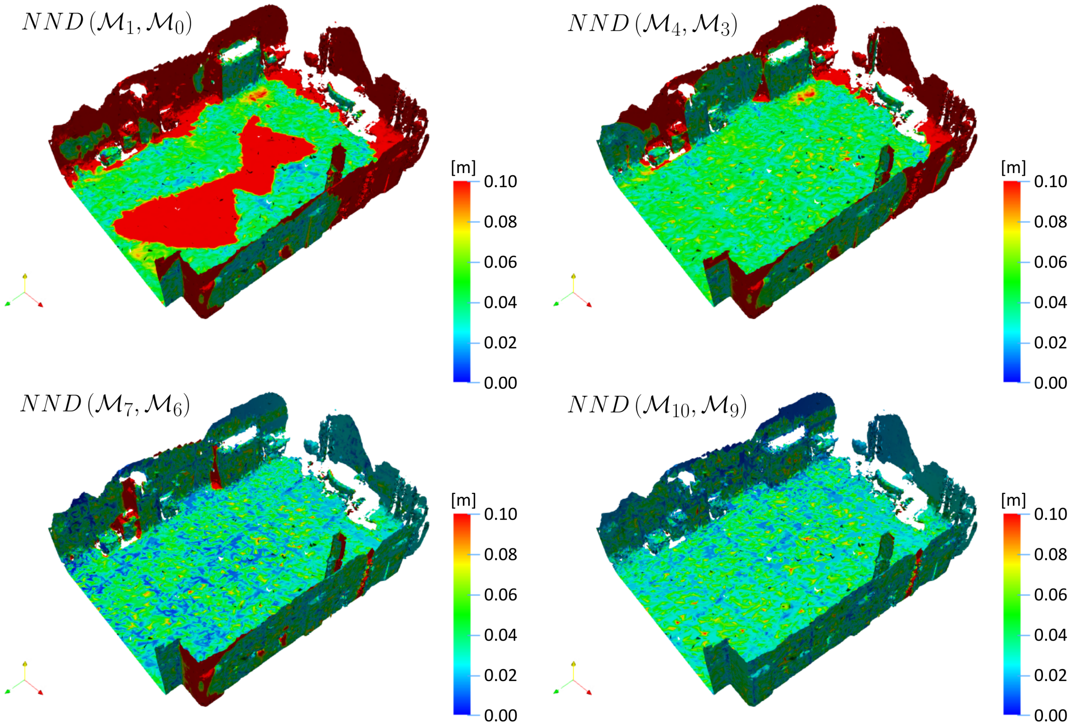

FIGURE 5 Qualitative assessment of convergences for mesh surfaces. It is important to note that only a specific region of interest has been considered in this analysis. NNDs between two consecutive surface meshes are represented by NND($\mathcal{M}_n - \mathcal{M}_{n-1}$). No significant differences among consecutive models were found after eight completed laps. The minimum Euclidean distances are represented by scalars ranging from 0 to 10 centimeters for a better representation of their variabilities.

All estimates for both evaluated methods were computed after ensuring that the HoloLens SLAM-based spatial mapping converged to a high-quality triangulation surface model. The convergence criterion was defined as the minimum numbers of complete laps capable of preserving mesh model geometric characteristics which are similar to the characteristics observed in consecutively generated surface models. In order to avoid the effects of new learned features outside the region of interest (i.e, rectangular trajectory performed by the user wearing the HoloLens), all mesh elements outside this region were not considered in the analysis, as shown in Figure 5. No significant differences were observed after the 8th completed lap ($\mathcal{M}_{10}$, $\mathcal{M}_7$). For illustrative purposes, the H_d and NNDs similarities measures are illustrated for the first 10 completed laps, and all similarities were computed based on the 11th mesh model $\mathcal{M}_{11}$, where the NNDs are represented as box-plot distributions, while the H_d is represented by a cyan dashed-line, as shown in Figure 6.

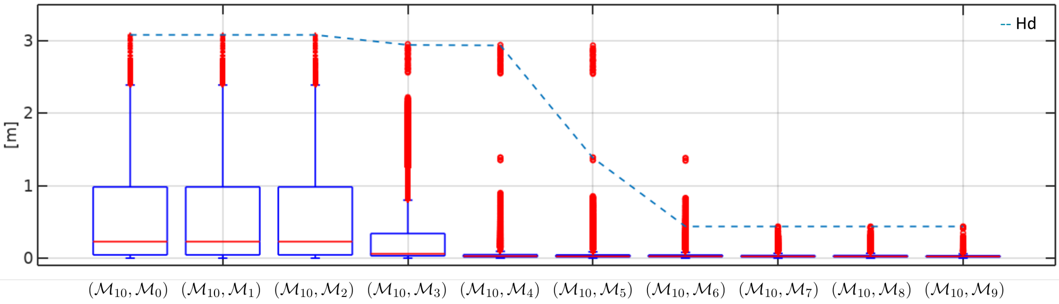

FIGURE 6 Convergence analysis of the HoloLens SLAM-based spatial-mapping. The box-plots represent the NNDs computed between the mesh models f and i, where $(\mathcal{M}_f, \mathcal{M}_i)$ represents the similarities computed between the final state f and an intermediate state i. Each box-plot contains $N = 53616$ observations, which corresponds to the number of nodes in the mesh model f representing the converged region of interest surface model. The dashed cyan line represents the Hd distances. After convergence a mean and standard deviation of 0.032 ± 0.03 m for the NNDs and a Hd of 0.44 m were verified.

3.2 | Latency estimation

The procedure described in Section 2.2 for estimating the latency of the developed AR application is shown in Figures 7(a)-7(c). As it can seen, these results demonstrate that the proposed method of feature dimensionality reduction is able to capture the mid-stance walking pattern spatiotemporal features in the third principal component, when a self-selected comfortable walking speed is assumed. The peaks with greater amplitudes represent randomly performed squats by the subject, which were aimed at improving the representation of the data into the principal component space. The latency of the system was estimated by computing the differences between corresponding spatiotemporal events detected in the optical-tracking and SLAM-based time-series, W and S, respectively. In our experiments, a latency of 49.84±0.98 ms was measured.

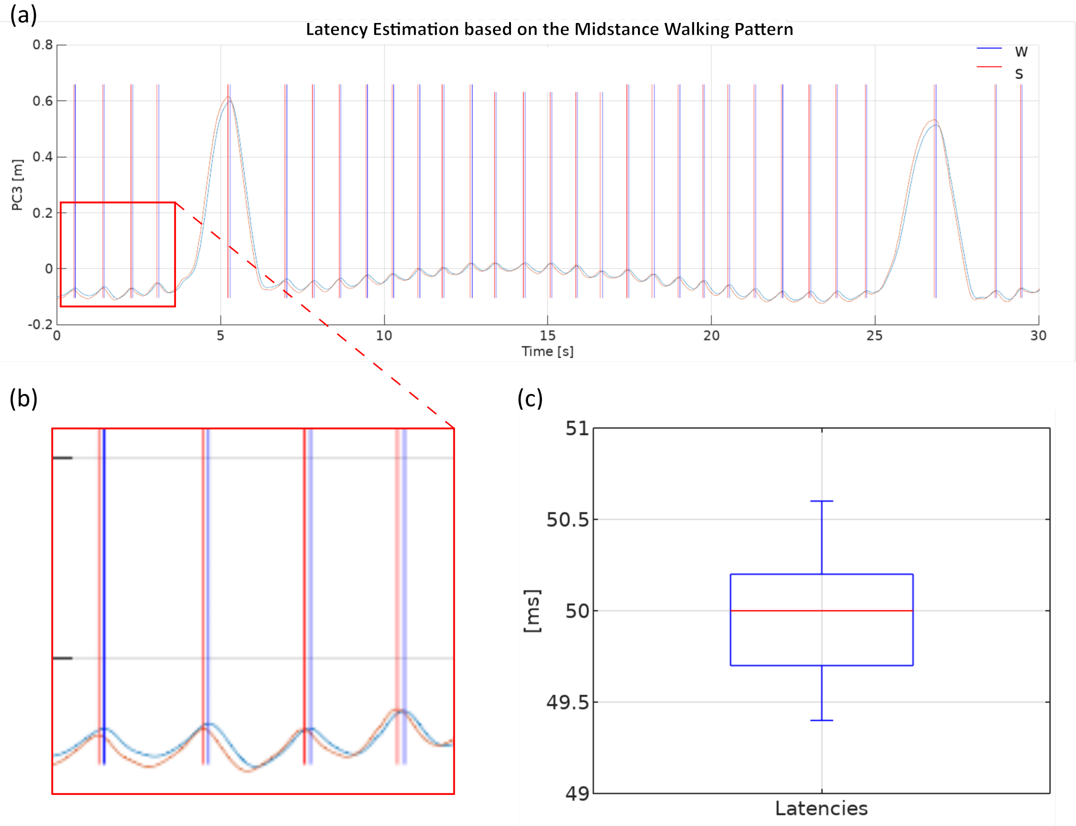

FIGURE 7 (a): Representation of the first 30 seconds of a randomly selected mid-stance walking pattern in the third principal component (PC3). The blue and red time-series represent the third component (PC3) of the optical-tracking W and HoloLens SLAM-based S position time-series and, the time-series were acquired at a mean frequency of 58.76 Hz. The vertical blue and red lines represent the automatically detected spatiotemporal features for the optical-tracking and SLAM-based times-series, respectively. (b) Inset highlighting the detected spatiotemporal W and S events. (c) Box-plot representation of the estimated latencies $N = 121$.

3.3 | Quantitative reliability assessment of the virtual-physical mappings T_c^h

The robustness and accuracy of the proposed HE-based and of the HoloLens SLAM–based calibration methods can be observed qualitatively in the two-unit spheres shown in Figures 8(a) and 8(b), respectively. Mean axes deviations distributions in degrees along the X-, Y,and Z-axes coordinates are represented in box-plot graphs in Figures 9(b) and 9(b). Considerably higher standard deviations for the offsets along the X-, Y-, and Z-axes were observed in the HoloLens SLAM method when compared with the proposed HE-based calibration method. These observations are quantitatively summarized in Tables 1 and 2.

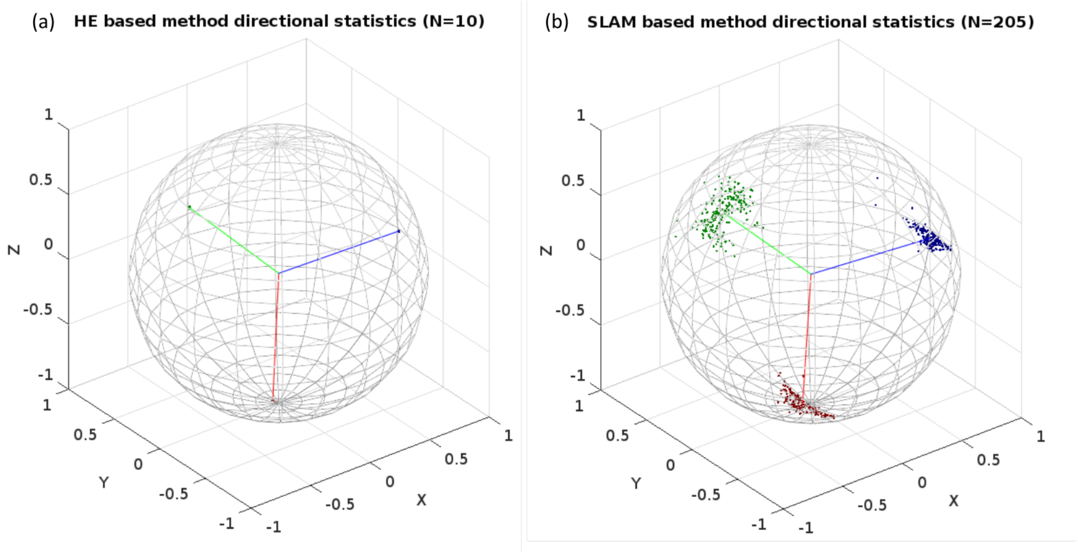

FIGURE 8 (a): Directional statistics for the proposed HE-based method. Randomly selected poses at different radial distances from the calibration board, ranging from 0.5 to 1.5 meters (N=10). (b): Directional statistics for the HoloLens SLAM-based calibration method assuming equally spaced time observations (N=205). A more detailed description can be found in Tables 1 and 2.

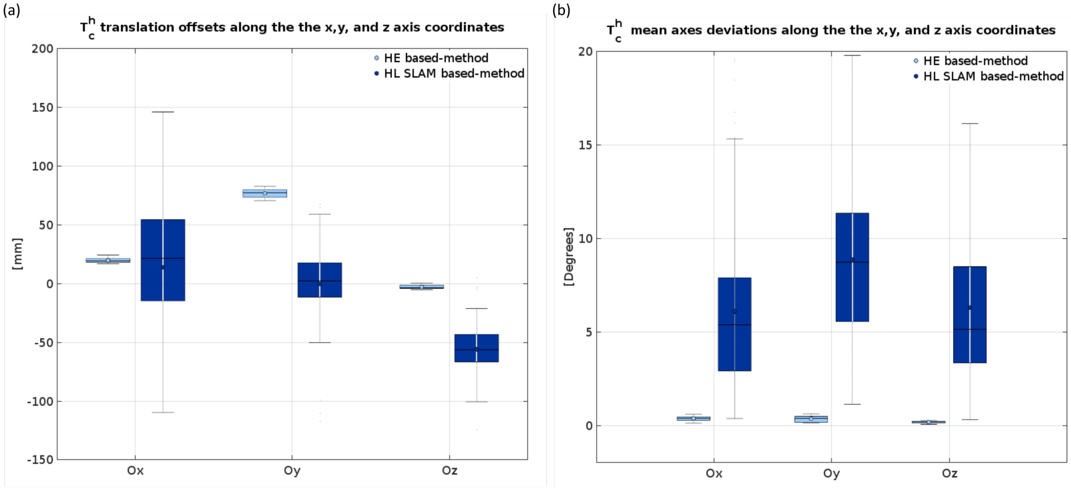

FIGURE 9 The light blue box-plots represent the variability observed in the proposed HE-based method, while the dark blue box-plots represent the variability of the HoloLens SLAM-based method. (a): T_c^h translations Ox, Oy, and Oz offsets in millimeters along the x,y, and z axis coordinates, respectively. (b) T_c^h: Mean axes deviations along the X-, Y, and Z-axes coordinates represented in degrees. The light blue box-plots are composed by ten random selected poses at different radial distances from the calibration board. The dark blue box-plots are constituted by 205 equally spaced time observations.

	HE-based calibration method		
	$X - axis$	$Y - axis$	$Z - axis$
Axis offsets [mm]	19.74 ± 2.38	76.82 ± 3.83	-2.74 ± 1.96
Axis deviations [degrees]	0.37 ± 0.15	0.35 ± 0.17	0.17 ± 0.07
Circular $\boldsymbol{var}$ [rad]	3.18×10^{-6}	4.18×10^{-6}	6.25×10^{-7}
Circular $\boldsymbol{std}$ [rad]	2.52×10^{-3}	2.89×10^{-3}	1.11×10^{-3}

TABLE 1 Descriptive statistics for the proposed HE-based calibration method (N=10), where $\boldsymbol{var}$ and $\boldsymbol{std}$ stand for variance and standard deviation, respectively.

	HoloLens SLAM method		
	$X - axis$	$Y - axis$	$Z - axis$
Axis offsets [mm]	13.80 ± 56.37	-0.22 ± 40.77	-55.86 ± 20.75
Axis deviations [degrees]	6.10 ± 4.15	8.87 ± 4.68	6.30 ± 4.22
Circular $\boldsymbol{var}$ [rad]	2.60×10^{-3}	3.31×10^{-3}	2.69×10^{-3}
Circular $\boldsymbol{std}$ [rad]	7.22×10^{-2}	8.14×10^{-2}	7.34×10^{-2}

TABLE 2 HoloLens SLAM-based calibration method descriptive statistics (N=205), where $\boldsymbol{var}$ and $\boldsymbol{std}$ stand for variance and standard deviation, respectively.

Directional descriptive statistics are given for the proposed HE-based calibration and HoloLens SLAM methods in Tables 1 and 2, respectively. The HoloLens SLAM-based presented with a circular variance within an order of magnitude of −3 radians along the X-, Y-, and Z-axes, while the proposed HE-based method presented with a significantly lower circular variance with magnitude order less or equal ≤ -6 radians along the X-, Y-, and Z-axes. These measures describe the spread of all possible estimated set of orthonormal basis representing a specific calibration matrix T_c^h. In this sense, the robustness and accuracy of these methods can be interpreted as a function of its computed circular variances and standard deviations. Therefore, lower circular variances indicate that the obtained basis vectors are concentrated around their respective mean directions, which is a highly desirable characteristic when determining an orthonormal basis, whereas values close to 1 indicate orthonormal basis uniformly distributed in the unit sphere, which inevitably results in less reliable estimates.

3.3.1 | Qualitative and quantitative assessment of the superimposed holograms in the physical space

The surgical reference frame was rigidly anchored to the organ of interest by using Kirschner wires (K-wires). Consequently, the implanted fiducial markers were mapped with respect to the surgical reference frame, allowing us to determine a transformation mapping between the imaging (virtual) and physical spaces. Therefore, holograms representing patient-specific anatomical models and surgical instruments can be represented in the physical space, independently of patient's position and orientation and in real time, as shown in Figure 10.

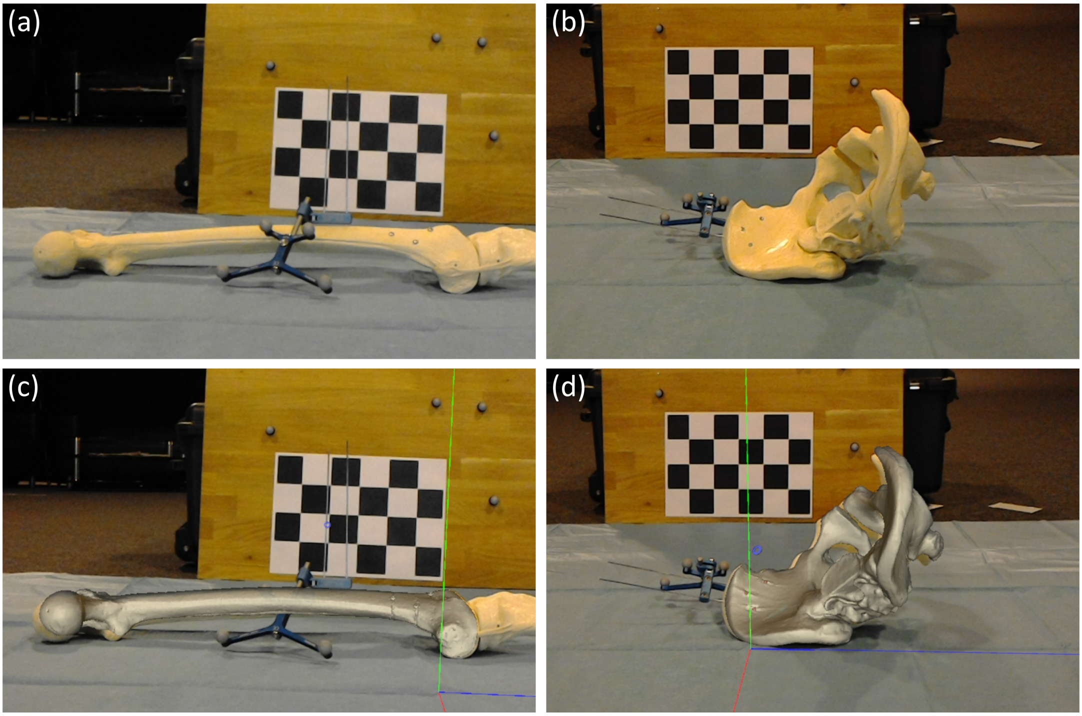

FIGURE 10 (a): Full male femur-tibia with acetabulum and cancellous inner structures (Sawbones, Pacific Research Laboratories, Vashon, Washington, USA). (b) Randomly selected pose from a full male pelvis model with acetabulum and cancellous inner structures (Sawbones, Pacific Research Laboratories, Vashon, Washington, USA). (c) and (d): Superimposed holograms with their corresponding physical organs. The holograms were generated from the extracted isosurface meshes, which were then subsequently mapped with respect to their respective surgical frames.

The accuracy of the proposed HE-based calibration method was assessed based on the RPEE approach described in Section 2.7 by representing the detected inner edges of the checkerboard pattern with respect to the physical space. Ten randomly selected poses at different radial distances from the calibration board, ranging from 0.5 to 1.5 meters were considered. This analysis showed a RMSE of 3.2 ± 1.6 mm for a total of 150 different patterns (3x5 grid size and 10 poses). For illustrative purposes, a randomly selected calibration frame is shown Figures 11(a) and 11(b).

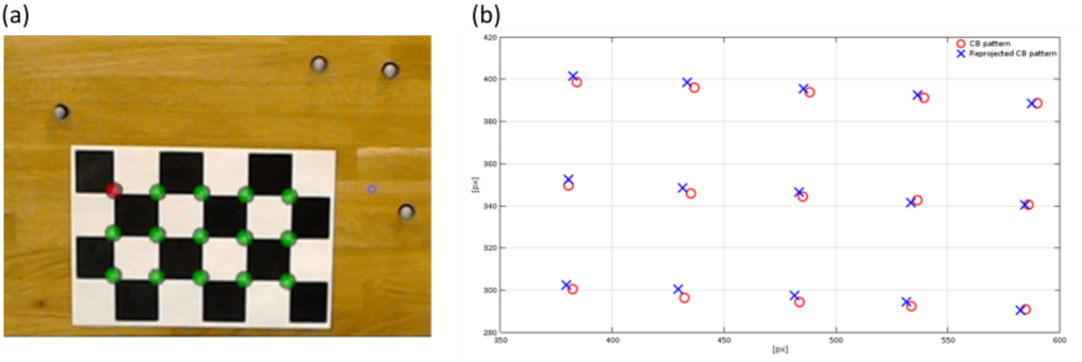

FIGURE 11 Retro-projection error estimation. (a): The green spheres are holograms mapped into the physical space by using the estimated mapping T_h^c, where the red sphere is representing the origin of the checkerboard pattern coordinate system. (b): Red circles represent the inner edges of the physical checkerboard calibration board. The blue crosses represent the centers of the holographic spheres represented in the physical world. 2D holograms' positions were determined by extracting the green color spectrum based on RGB band-pass filter, consequently we have converted the initial mask into a gray image, followed by an edge detection approach and finally the centers were estimated by using the Hough transform.

4 | DISCUSSION

Despite not being the primary objective of this work, to the authors' best knowledge, no other studies have systematically investigated neither the accuracy of the existing HoloLens-SLAM proprietary algorithm nor the system's latency, which we believe to be of special interest, since the Microsoft HoloLens is currently considered to be the most suitable Holographic headset for mixed reality applied to CAS [29]. It is important to note that the estimated estates of the HoloLens virtual camera, as well as the environment modeling are based on stochastic multisensory data fusion: inertial data (i.e., accelerometer, gyroscope, and magnetometer) combined with the vision sensors (i.e., depth and environmental understanding cameras). Consequently, its performance depends on the quality and amount of detected visual features and assumed feature selection criteria, as well as the illumination conditions, geometry of the physical space, and dynamic characteristics of the user wearing the device. For all these reasons, the accuracy of a specific hardware can be determined only for a very specific set of variable conditions, and therefore, it cannot be easily generalized and compared across different studies. A more accurate and precise assessment requires further studies involving real and synthetic data, as well as a more detailed theoretical and practical analysis, which would be prohibitive since the HoloLens-SLAM method developed by the Microsoft is not freely and publicly available.

The reliability and accuracy of the estimated transformation matrices were quantified by linear and circular variances, for the offsets and orthonormal basis, respectively. However, it is important to note that the two evaluated calibration methods fundamentally differ from each other. The SLAM-based method relies upon the pose information of the virtual camera provided by the HoloLens, where on the proposed HE-based method, the pose of the virtual camera in the physical space is determined by using a vision-based approach. Subsequently, all physical objects are represented with respect to the absolute reference frame h defined by the retro-reflective markers placed on the HoloLens headset (Figure 2) and are then mapped to the virtual camera by applying the estimated calibration matrix T_c^h. The evaluated two methods are not homologous and, therefore, the estimated rotations and offsets cannot be compared against each other.

Based on our results obtained for the SLAM-based method, the most significant variabilities for the offsets were found in the X-, Y-, and Z-axes, subsequently ranked in descending order, which refer to the first, second, and third principal components of the virtual trajectories mapped into the physical space by performing the principal component analysis. The HoloLens virtual camera is a left-handed coordinate system, however, these results should not be interpreted with respect to the virtual camera left-handed coordinate system, but rather with respect to the physical space right-handed coordinate system, where X-, Y-, and Z-axes are represented by green, red, and yellow arrows, respectively, as shown in Figure 5. Our experiments were conducted in the environment shown in Figure 5, which is a rectangular carpeted room with considerably larger dimensions when compared with common residential buildings and without visual features on the floor, resulting in a sparse representation of near visual features. Therefore, the localization of the virtual camera and the space mapping were mainly determined by the depth sensor and, therefore, the greater variabilities for X-, Y-, and Z-axes ranked in descending order are likely to be explained by the practical range limitations of the depth sensing camera, which has an optimal range of 0.85-3 m. These variabilities can be qualitatively observed in Figure 8(a) and are quantitatively summarized in Table 2. It is important to observe that these results are associated to the reliability and accuracy of the HoloLens SLAM-based method under very specific conditions and, therefore, any calibration method relying on the SLAM-based trajectories would lead to lower accuracy of the transformation matrices estimates, which would then be revealed in the descriptive statistics analysis, as verified in our results.

Geometrical distortions of the physical trajectories could possibly be induced indirectly by the system's latency and, consequently, resulting in non-optimal calibration matrices. For this reason, a quasi-static calibration approach was considered, i.e., the user was instructed to walk at a self-selected comfortable walking speed. From the latency quantification approach, we can infer that the evaluated physical and virtual trajectories are not distorted by the latency and, therefore, ensuring spatiotemporal correspondences between the trajectories in the physical and virtual spaces. For all these previously mentioned reasons, nonlinear inherited properties due to system's latency can be neglected for the estimates.

Optical multi-object tracking is an interactive process involving a number of different imaging and optimization steps, such as background subtraction, image filtering, blob detection, 3D reconstruction from multiple images based on epipolar geometry, establishment of correspondences between different images and among consecutive frames, occlusion handling, and data network streaming. For these reasons, the use of an external optical tracking combined with OST-HMDs is inherently an iterative process, which may significantly affect the user's experience based on very specific conditions and applications and, consequently, its use in some clinical applications may possibly be diminished, especially in rapid surgical maneuvers, where the actions of the user may not be time-locked with the renderization of the near-eye multifocus dichromated-gelatin holographic lenses. The observed overall system's latency was qualitatively satisfactory and promising based on the evaluated theoretical premise. In order to reduce the system's latency and to improve tracking robustness, an extended Kalman filter (EKF) [41] was implemented, however, it has not been fully validated yet, since the predictions of estates are sensitive to model parameters, e.g., state-transition and observation models and their respective covariance matrices, which should be further investigated.

The current version of the HoloLens is not equipped with eye-tracking technology and, therefore, after the user-specific calibration process has been performed, the HoloLens headset uses the spatial mapping and localization information to correctly display the digital content on its near-eye multifocus dichromated gelatin holographic lenses and, therefore, the immersive experience of the user is not drastically affected as in previously developed systems, since both the depth camera and the IMU can accurately estimate the virtual camera's pose with respect to near detected large planes (e.g., floor, ceiling, and walls), which are accomplished by the spatial understanding module API. Moreover, a sophisticated hardware-assisted holographic technique called stabilization plane is performed, which is associated with motion and change of the point of view when the user moves its head or when the headset moves with respect to the user's head. This stabilization plane is automatically chosen by using the depth camera information to localize the gaze intersection point with a physical or user-defined virtual object and, therefore, enhancing the immersive experience. It is important to note that the relationships between the near-eye multifocus dichromated gelatin holographic lenses and the user's eyes are not addressed in this work, assuming that this step has already been accomplished by the proprietary calibration routine developed by Microsoft Research, which is user-specific, and is tailored to a specific hardware by the manufacturer [17, 18, 19, 20, 21, 22]. Recently, small gaze tracking cameras became available, which are expected to be incorporated into AR OST-HMDs in the near future, thus allowing a precise and robust estimation of the eyes' localization and gaze directions. This additional information can then be used to improve the calibration process, since no additional information or assumptions are needed to correctly display a hologram into the physical space with respect to the user's view.

5 | CONCLUSION

An important aspect to be emphasized in the use of AR in CAS is that such technology provides an instantaneous qualitative assessment of the accuracy of the registration process between the surgical and imaging reference frames, which is a highly desirable and needed feature that cannot be achieved by using the current available non-immersive surgical navigation platforms. Anatomical landmarks or implanted fiducial markers are manually digitized by a surgical pointer during the surgical procedure, while in the imaging space are determined preoperatively by using different semi-automated image-processing techniques. Therefore, the establishment of correspondences between the physical and imaging spaces is cumbersome and prone to different sources of errors. These sources of errors cannot be decoupled in the registration process and are subject to inter- and intra-observer variability [42]. Consequently, the qualitative mismatch observed between the physical anatomies of the considered phantom models and their respective holograms cannot be attributed solely to the proposed HE-based calibration method (Figures 10(a)-10(d)).

Our results shown that the developed platform is very intuitive, especially among inexperienced surgeons, since surgical instruments and maneuvers are represented with respect to the surgeon's visual system and, therefore, attention disruptions are more likely to be reduced, since patient-specific imaging and preoperative planning data can be displayed on virtual monitors that can be placed and dynamically allocated according to the surgeon's needs, allowing the surgeon to be focused on the surgical field. In addition, this type of platform allows the surgeon to interact with patient's data and preoperative planning in a more natural fashion by using speech, hand gestures, and eye gazing.

The proposed HE-based method provides reliable and accurate estimates for the transformation matrix T_c^h, which were qualitatively and quantitatively evaluated. Although the results are promising, the proposed method needs to be further improved and clinically validated and, for these reasons, the authors believe that the use of the Microsoft HoloLens development edition should be carefully considered before it can be translated to a specific clinical setting. However, the achieved precision by the proposed HE-based calibration method makes it suitable for a wider range of clinical, educational, and industrial applications.

ACKNOWLEDGEMENTS

The authors would like to thank Ms. Sandra Carraud from the Department of Orthopaedic Surgery, Hôpital de La Tour, Geneve, Switzerland for many helpful suggestions regarding the placement and use of surgical instruments and, Dr. med. Frank C. Kolo from the Rive Droite Radiology Center, Geneva, Switzerland for the CT data acquisition. We would like also to thank the anonymous reviewers for their helpful comments and suggestions, which have considerably improved the quality of our manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Simon DA, Hebert M, Kanade T. Techniques for fast and accurate intrasurgical registration. Journal of Image Guided Surgery 1995;1(1):17–29.

[2] Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Computer Graphics and Applications 2001;21(6):34–47.

[3] Vávra P, Roman J, Zonča P, Ihnát P, Němec M, Kumar J, et al. Recent Development of Augmented Reality in Surgery: A Review. Journal of Healthcare Engineering 2017 aug;2017:1–9.

[4] Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. Surgical Oncology 2011 sep;20(3):189–201.

[5] Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. Medical Image Analysis 2017 apr;37:66–90.

[6] Andress S, Johnson A, Unberath M, Winkler AF, Yu K, Fotouhi J, et al. On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. Journal of Medical Imaging 2018 jan;5(02):1.

[7] Besharati Tabrizi L, Mahvash M. Augmented reality–guided neurosurgery: accuracy and intraoperative application of an image projection technique. Journal of Neurosurgery 2015 jul;123(1):206–211.

[8] Badiali G, Ferrari V, Cutolo F, Freschi C, Caramella D, Bianchi A, et al. Augmented reality as an aid in maxillofacial surgery: Validation of a wearable system allowing maxillary repositioning. Journal of Cranio-Maxillofacial Surgery 2014 dec;42(8):1970–1976.

[9] Lin L, Shi Y, Tan A, Bogari M, Zhu M, Xin Y, et al. Mandibular angle split osteotomy based on a novel augmented reality navigation using specialized robot-assisted arms—A feasibility study. Journal of Cranio-Maxillofacial Surgery 2016 feb;44(2):215–223.

[10] Wang J, Suenaga H, Yang L, Kobayashi E, Sakuma I. Video see-through augmented reality for oral and maxillofacial surgery. The International Journal of Medical Robotics and Computer Assisted Surgery 2017 jun;13(2):e1754.

[11] Landaeta-Quinones CG, Hernandez N, Zarroug NK. Computer-Assisted Surgery: Applications in Dentistry and Oral and Maxillofacial Surgery. Dental Clinics of North America 2018 jul;62(3):403–420.

[12] Janin AL, Mizell DW, Caudell TP. Calibration of head-mounted displays for augmented reality applications. In: Proceedings of IEEE Virtual Reality Annual International Symposium IEEE; 1993. p. 246–255.

[13] Itoh Y, Klinker G. Interaction-free calibration for optical see-through head-mounted displays based on 3D Eye localization. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI) IEEE; 2014. p. 75–82.

[14] Zhang Z, Weng D, Liu Y, Xiang L. 3D optical see-through head-mounted display based augmented reality system and its application. vol. 9524 International Society for Optics and Photonics; 2015. p. 952428.

[15] Plopski A, Itoh Y, Nitschke C, Kiyokawa K, Klinker G, Takemura H. Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays. IEEE Transactions on Visualization and Computer Graphics 2015 apr;21(4):481–490.

[16] Grubert J, Itoh Y, Moser KR, Swan II JE. A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. IEEE Transactions on Visualization and Computer Graphics 2017;.

[17] John R Lewis,Yichen Wei,Robert L Crocco,Benjamin I Vaught,Alex Aben-Athar Kipman KSP, Gaze detection in a see-through, near-eye, mixed reality display, US 8487838B2; 2011-08-29.

[18] John R Lewis,Yichen Wei,Robert L Crocco,Benjamin I Vaught,Kathryn Stone Perez,Alex Aben-Athar Kipman , Aligning inter-pupillary distance in a near-eye display system, US 20130050642A1; 2011-08-30.

[19] John R LewisKathryn Stone PerezRobert L CroccoAlex Aben-Athar Kipman, Adjustment of a mixed reality display for inter-pupillary distance alignment, US 20130050833A1; 2011-08-30.

[20] Ashraf Ayman Michail,Roger Sebastian Kevin Sylvan,Quentin Simon Charles Miller,Alex Aben-Athar Kipman , Stabilization plane determination based on gaze location, US 20150310665A1; 2014-04-29.

[21] Nathan Ackerman,Drew Steedly, Andy Hodge,Alex Aben-Athar Kipman, Tracking head movement when wearing mobile device, US 20140375680A1; 2013-06-24.

[22] Mathew J Lamb,Ben J Sugden,Robert L Crocco, JR Brian E Keane,Christopher E Miles,Kathryn Stone Perez,Laura K Massey,Alex Aben-Athar Kipman,Tom G Salter , Propagation of real world properties into augmented reality images, US 20140002492A1; 2014-04-29.

[23] Tepper OM, Rudy HL, Lefkowitz A, Weimer KA, Marks SM, Stern CS, et al. Mixed Reality with HoloLens: Where Virtual Reality Meets Augmented Reality in the Operating Room. Plastic and reconstructive surgery 2017 nov;140(5):1066–1070.

[24] Incekara F, Smits M, Dirven C, Vincent A. Clinical Feasibility of a Wearable Mixed-Reality Device in Neurosurgery. World Neurosurgery 2018 jul;.

[25] Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, et al. Through the HoloLens^{TM} looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. European radiology experimental 2018 dec;2(1):2.

[26] Shi L, Luo T, Zhang L, Kang Z, Chen J, Wu F, et al. [Preliminary use of HoloLens glasses in surgery of liver cancer]. Zhong nan da xue xue bao Yi xue ban = Journal of Central South University Medical sciences 2018 may;43(5):500–504.

[27] Kuhlemann I, Kleemann M, Jauer P, Schweikard A, Ernst F. Towards X-ray free endovascular interventions - using HoloLens for on-line holographic visualisation. Healthcare technology letters 2017 oct;4(5):184–187.

[28] Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, et al. Augmented Reality as a Telemedicine Platform for Remote Procedural Training. Sensors (Basel, Switzerland) 2017 oct;17(10).

[29] Qian L, Azimi E, Kazanzides P, Navab N. Comprehensive Tracker Based Display Calibration for Holographic Optical See-Through Head-Mounted Display 2017 mar;.

[30] Daunoravičienė K, Apanskienė V, Žižienė J, Ovčinikova A, Kizlaitienė R, Sereikė I, et al. Differences in temporal gait parameters between multiple sclerosis and healthy people. Springer, Singapore; 2017.p. 31–35.

[31] Speciali DS, de Oliveira EM, dos Santos NM, Pereira FV, Fracini AC, Fukuda TY, et al. Use of the Gait Deviation Index and spatiotemporal variables for the assessment of dual task interference paradigm. Journal of bodywork and movement therapies 2013 jan;17(1):19–27.

[32] Hanqi Zhuang, Roth ZS, Sudhakar R. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form AX=YB. IEEE Transactions on Robotics and Automation 1994;10(4):549–554.

[33] Loughlin C, Hudson E. Eye·in·hand robot vision scores over fixed camera. Sensor Review 1983 jan;3(1):23–26.

[34] Shiu YC, Ahmad S. Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form AX = XB. IEEE Transactions on Robotics and Automation 1989;.

[35] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision, vol. 2. Cambridge University Press; 2004.

[36] Macqueen JB. Some Methods for classification and analysis of multivariate observations. In: Procedings of the Fifth Berkeley Symposium on Math, Statistics, and Probability; 1967. .

[37] Community O. The OpenCV Reference Manual. October 2010;.

[38] Downs TD. Orientation statistics. Biometrika 1972;.

[39] Bingham C. An Antipodally Symmetric Distribution on the Sphere. The Annals of Statistics 1974;.

[40] Fisher NI. Statistical analysis of circular data. Cambridge University Press; 1995.

[41] Wan EA, Van Der Merwe R. The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373) IEEE;. p. 153–158.

[42] de Oliveira ME, Netto LMG, Kistler M, Brandenberger D, Büchler P, Hasler CC. An image-based method to automatically propagate bony landmarks: application to computational spine biomechanics. Computer Methods in Biomechanics and Biomedical Engineering 2015 oct;18(14):1535–1542.