

ZHEJIANG
UNIVERSITY PRESSThe Institution of
Engineering and Technology

WILEY

ORIGINAL RESEARCH

Automatic feature-based markerless calibration and navigation method for augmented reality assisted dental treatment

Faizan Ahmad^{1,2} | Jing Xiong¹ | Zeyang Xia³ ¹Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China²University of Chinese Academy of Sciences, Beijing,
China³School of Mechanical Engineering, Shanghai Jiao
Tong University, Shanghai, China**Correspondence**Zeyang Xia and Jing Xiong,
Email: zxia@sjtu.edu.cn and
jing.xiong@siat.ac.cn**Funding information**National Natural Science Foundation of China,
Grant/Award Numbers: U2013205, 62073309;
Guangdong Basic and Applied Basic Research
Foundation, Grant/Award Number:
2022B1515020042; Shenzhen Science and
Technology Program, Grant/Award Number:
JCYJ20220818101603008**Abstract**

Augmented reality (AR) is gaining traction in the field of computer-assisted treatment (CAT). Head-mounted display (HMD)-based AR in CAT provides dentists with enhanced visualisation by directly overlaying a three-dimensional (3D) model on a real patient during dental treatment. However, conventional AR-based treatments rely on optical markers and trackers, which makes them tedious, expensive, and uncomfortable for dentists. Therefore, a markerless image-to-patient tracking system is necessary to overcome these challenges and enhance system efficiency. This paper proposes a novel feature-based markerless calibration and navigation method for an HMD-based AR visualisation system. The authors address three sub-challenges: firstly, synthetic RGB-D data for anatomical landmark detection is generated to train a deep convolutional neural network (DCNN); secondly, the HMD is automatically calibrated using detected anatomical landmarks, eliminating the need for user input or optical trackers; and thirdly, a multi-iterative closest point (ICP) algorithm is developed for effective 3D-3D real-time navigation. The authors conduct several experiments on a commercially available HMD (HoloLens 2). Finally, the authors compare and evaluate the approach against state-of-the-art methods that employ HoloLens. The proposed method achieves a calibration virtual-to-real re-projection distance of (1.09 ± 0.23) mm and navigation projection errors and accuracies of approximately (0.53 ± 0.19) mm and 93.87%, respectively.

KEYWORDS

motion estimation, navigation, sensor calibration

1 | INTRODUCTION

Computer-assisted treatment (CAT) is commonly used, which transforms most aspects of different fields of medicine, including digital dentistry, for better efficiency, accuracy, minimal impact on the patient, and pre-operative surgical planning and assistance [1–3]. However, traditional CAT relies on two-dimensional (2D) imaging instead of perceptive three-dimensional (3D) images for guidance and navigation, which causes surgeons to have a constrained field-of-view (FOV), a lack of depth perception, and hand-eye co-ordination problems [4, 5]. In dentistry, many complex surgical procedures, such as tooth extraction, dental implantation, gum grafting, and jaw surgery, can be challenging for clinicians because of the

smaller viewing area. Therefore, a 3D-based immersive surgical navigation system is needed to improve the effectiveness of CAT.

Virtual reality (VR) and augmented reality (AR) were developed to address these challenges by superimposing a virtual environment with naked-eye anatomical structures onto real patients for surgical navigation [6, 7]. Integrating AR into CAT reduces the cognitive load and enhances surgical navigation [8, 9]. Head-mounted displays (HMDs) offer a more egocentric view than standard monitor-based 2D visualisations do, reducing the cognitive effort needed to correct perspective differences. Furthermore, optical-see through (OST) displays eliminate the need for a camera to capture the real environment; instead, they use an optical combiner to overlay virtual

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *IET Cyber-Systems and Robotics* published by John Wiley & Sons Ltd on behalf of Zhejiang University Press.

data directly onto the real patient [10]. Consequently, the real target (anatomical structure) can be precisely tracked and aligned with virtual coordinates, enabling smooth and efficient treatment. Therefore, OST-HMD display methods in AR surgical guidance and navigation systems require calibration to ensure proper alignment between virtual and real coordinates and to retain no-latency real-world perception [11]. Furthermore, an OST-HMD-based AR system enhances the overlay of preoperative plans onto patient anatomy, providing real-time guidance during surgery [12].

Several methods have been developed to increase the precision and reliability of the display calibration procedure [13–17]. Surgical tool tracking is essential in clinical scenarios where the main aim is to reduce the tracking error in image-guided surgery [18]. Most of the work is based on an optical tracking system using multiple markers for a dynamic reference frame. These markers are firmly fixed on the target anatomy to track the targeted orientation and pose during dynamic reference frame movement [19]. It is often necessary to calibrate the relative pose between the dynamic reference frame and the anatomy offline, utilising error-prone point-based or contour-based registration because errors in marker-to-target registration can be distributed throughout the process [20], and unnecessary injury can be caused during marker insertion. Therefore, it is challenging to achieve maximum safety and minimal invasiveness when anatomical markers are used in CAT.

In the field of digital dentistry, particularly in oral and maxillofacial surgery [21, 22], the current research highlights the use of fiducial markers for cone beam computed tomography (CBCT) and patient alignment, large devices for motion tracking, and dental casts. However, these casts are prone to errors in retaining reference markers for patient tracking. Using these fiducial markers and bulky tracking devices presents significant challenges, such as creating line-of-sight limitations and generating inaccuracies due to marker displacement, which increase the degree of technical intricacy and necessitate significant modifications to the surgical arrangement [23]. Specifically, the confined operative field and the mobile nature of temporomandibular joints render the stable attachment of both fiducial and markers to the patient's mandible and maxilla regions challenging without engendering invasiveness. Therefore, surgical navigation is currently challenging in dentistry. To address the challenges inherent in dental AR applications, the markerless registration and tracking techniques have emerged as indispensable tools [24]. The point set registration, a fundamental methodology in computer graphics, facilitates the precise alignment of 3D shapes or objects, thereby serving as a cornerstone for accurate spatial representation [25, 26]. This process is particularly important in medical AR and 3D reconstruction, where the precise registration of 3D points is critical for successful implementation. Among the extant methodologies, the iterative closest point (ICP) algorithm has garnered widespread recognition for its efficacy in managing rigid transformations, wherein the geometric structure remains invariant. The applicability of ICP to AR-based dental model registration is particularly noteworthy, as it efficiently aligns 3D

dental scans with virtual models, enabling high-precision overlays in real-time scenarios. The simplicity, fast convergence, and robustness to minimum initial misalignments render it ideal for accurately aligning rigid dental structures.

In contrast, the coherent point drift (CPD) algorithm [27] adopts a probabilistic approach that accommodates non-rigid transformations by modelling one point set as moving coherently towards another. The CPD allows for smooth and flexible deformations, which may be advantageous in soft tissue modelling. It is computationally more intensive than ICP and may be superfluous for rigid structures such as dental models. As a result, the efficiency and accuracy of ICP make it the most suitable for AR-based dental applications.

The ultimate objective in OST-HMD-based dental treatment assistance (DTA) systems is to achieve precise calibration and effective surgical navigation. This goal underscores the critical importance of robust registration techniques in ensuring the accuracy and reliability of these advanced dental treatment modalities.

The major contributions of our work are given as follows:

1. We propose a novel feature-based reconfigurable method for HMD calibration and navigation. The main innovation is that our method uses a depth-RGB sensing stereo camera to automatically detect the feature points of teeth as regions of interest (RoIs), thereby avoiding the use of optical markers and sensors to track the 3D overlay between the virtual and real counterparts.
2. To overcome the data issue, we developed a large-scale synthetic RGB-D dataset with multiple planned occlusions and random regions with different poses that conform to real scenarios.
3. We propose a method that learns features through tooth exposure and enables automatic alignment with their real counterparts.
4. We design and build a highly reconfigurable OST-HMD-based application for AR-assisted DTA in real time via HoloLens 2.

The rest of the paper is organised as follows: Section 2 describes the system design for the OST-HMD. Section 3 explains the proposed markerless calibration and navigation method. Finally, Section 4 discusses the experimental setup and presents the results.

2 | SYSTEM DESIGN

Our novel feature-based approach is based on an AR-assisted markerless calibration and navigation system for real-time DTA. This work aims to calibrate a virtual 3D mandible model automatically for OST-HMD using an attached depth-RGB stereo camera as shown in Figure 1.

There are three main tasks involved in our workflow. First, we utilise a pre-trained deep convolutional neural network (DCNN) model to detect anatomical landmarks (teeth) in RGB-D images. Second, the OST-HMD is calibrated to

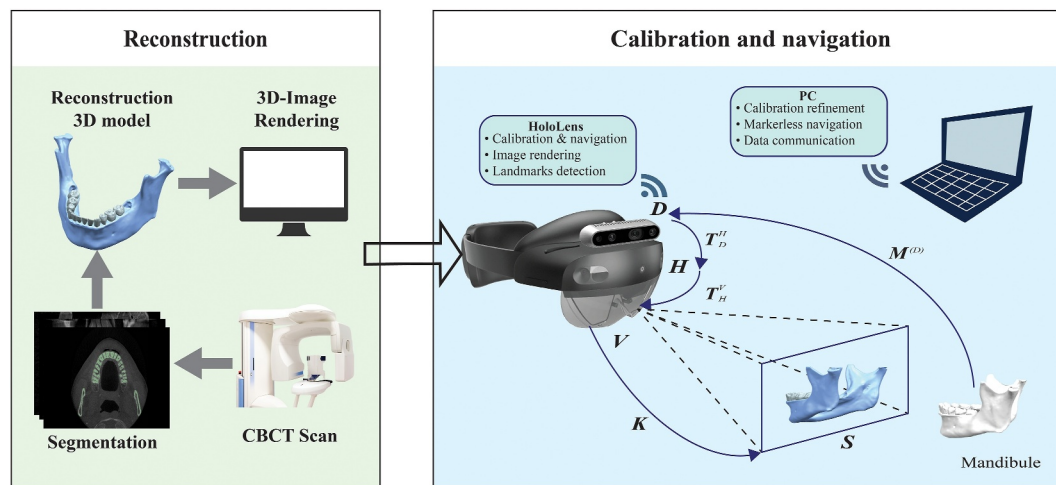


FIGURE 1 Proposed learning-based markerless calibration and navigation system for dental treatment assistants. Preoperatively acquiring and pre-processing patient data, initialising the system prior to surgery, estimating anatomical landmarks, and overlaying 3D images during surgery.

automatically align virtual coordinates with real-world coordinate frames, using these landmarks as reference points. Third, we employ markerless navigation to superimpose the virtual model onto the real model (phantom), enabling precise surgical navigation for dental treatments. Our system comprises two primary components: (1) a Microsoft HoloLens (second generation) and (2) an Intel RealSense depth stereo camera (D435i). This research aims to eliminate the use of various sensors that can introduce noise during treatment, thereby increasing the accuracy and robustness of the system. To the best of the author's knowledge, this approach represents a novel development in digital dentistry.

2.1 | Display

Most commercial AR-assisted headsets share several common components. For instance, the rendering engine generates and projects virtual images onto the real image plane. Also, the onboard sensors are used for scene capture, detection, tracking, and user interaction [28]. However, the details of hardware characteristics (quality of sensors) may differ from device to device. In some AR headsets, there are no or low-quality depth sensors, which makes the device inadequate for markerless anatomical target tracking, in which high-resolution and depth information are essential.

We employed HoloLens 2, which has an average angular resolution (θ) of approximately 1.4 arcmin per pixel [29]. An RGB camera (H) is integrated within the headset, providing stabilised video streaming at a resolution of 2048×1080 at an average angular resolution of approximately 2.1 arcmin per pixel. Although HoloLens 2 has a time-of-flight (ToF) depth camera sensor that enables 3D world anchoring and gesture recognition, it is not suitable for tracking small-scale objects due to inadequate resolution [1]. Additionally, its accuracy decreases over time with temperature drift, approaching 8–10 mm after just an hour of operation. Therefore, the

HoloLens built-in depth sensor is inadequate for tracking the complex dental anatomical structure for markerless navigation.

Furthermore, the HoloLens has a limited FOV. By using an external depth camera, users can easily capture depth data from a wider or different angle than can built-in sensors. Therefore, our study processed real-time depth data acquired from the RealSense (D435i) camera to provide high-resolution 3D data to the HoloLens in real time. These processed data were subsequently transmitted to the HoloLens via WiFi. We sent video frames of 640×480 pixels from the HoloLens 2 RGB camera at a rate of 30 frames per second (FPS) to the application, which was developed via C# in the Unity platform [30]. For our model, we used dimensions of 640×480 for all the variables. Our video and data communication protocol, which uses the WebRTC library [31], enables low-latency communication between the HoloLens 2 and the remote desktop. The WebRTC library uses the user datagram protocol (UDP) and H.264 codec to compress and stream the video frames in YUV420 format from the HoloLens 2 to the workstation. It is generally recommended that the application be implemented for specific tasks and that the results be sent back to HoloLens rather than transmitting the raw depth data. A depth-RGB camera is used to enhance depth navigation and tracking of the 3D model. We integrated an Intel RealSense D435i depth sensor on top of HoloLens 2. There are several methods for tracking 3D objects. However, several limitations need to be addressed: the RGB-D camera must be tightly integrated with the HoloLens and pre-calibrated, the camera must be able to detect reflections, and both RGB and RGB-D cameras satisfy these requirements [32]. Since the HMD can access only the RGB camera source, a 3D model navigation approach relies on monocular RGB input images [33, 34]. However, implementing such techniques on the HoloLens 2 is not currently viable because of the variability and robustness of these novel approaches and hardware limitations that limit their real-time execution. Our prototype uses an external depth-RGB sensor permanently mounted to HoloLens 2.

The D435i camera comprises RGB and depth sensors (D) that leverage unstructured light-based active stereoscopy technology to enable precise reconstructions within a confined FOV. Notably, the stereo-depth camera achieves an average $\theta_D \approx 3$ arcmin per pixel in high-definition conditions and exhibits superior depth quality owing to considerable improvements in stereo alignment, reflected in the minimal disparity error $\varepsilon_d \leq 0.1$ of the RGB-D camera in pixels. To estimate the hypothetical depth tracking errors ε_z and $\varepsilon_{(x,y)}$, appropriate parameters need to be calculated as follows [35]:

$$\varepsilon_z = \frac{z^2}{f \cdot B} \varepsilon_d, \varepsilon_{(x,y)} = \frac{z \cdot \varepsilon_d}{f} \quad (1)$$

where f and B are the camera focal length in pixels and the camera baseline, respectively. Based on manufacturer specifications, for a target-to-camera depth of $z = 500$ mm, we have $\varepsilon_z \leq 0.5$ mm and $\varepsilon_{(x,y)} \leq 0.05$ mm. As a result, we can estimate that any error caused by the quantisation of disparity will be in the sub-millimetre range, which is suitable for manual processing that requires accurate tracking within the depth range. Furthermore, unlike other cameras, the D435i camera is not susceptible to temperature drifts during its typical ToF [36], ensuring that the tracking accuracy remains high over an extended period.

2.2 | Synthetic data generation

We obtain patient volumetric data via CBCT during the pre-operative surgical planning stage. The data are segmented and reconstructed on the basis of the RoI, depending on the diagnosis and treatment plan. After that, a depth-RGB stereo camera is used to record and generate the synthetic data to train a DCNN for anatomical landmark detection. To capture the diverse dataset, we generated 2D images of 640×480 pixels and depth maps at a rate of 30 FPS. The Unity [30] game engine was utilised to render the synthetic environment on the basis of a CAD model and provided texture and material details. The Intel RealSense D435i camera [35] was utilised, and its intrinsic parameters were incorporated into the Unity camera component to ensure that aspects such as the FOV were consistent between real and synthetic data. Multiple texture and material variations in the scene geometry enhanced the dataset's variety. Additionally, various lighting configurations comprising different intensities, colour temperatures, and positions were implemented to maximise dataset diversity.

In order to enable occlusion interactions among multiple postures, it is crucial to use the depth buffer from a rendered scene, which provides depth information for each pixel. However, the depth buffers in digital imaging technologies like Unity are non-linear and enable greater depth resolution over shorter distances. Thus, it is necessary to linearise and convert the depth buffer into metres before recording it as a synthetic depth map. To acquire data in Unity, an animation script was used to play the animation for each environment setup and collect data at regular intervals. After the animation for a

specific configuration was completed, the environment changed to a new one, and data gathering started after a short delay to ensure consistent lighting conditions. A modest random position and rotation offset were incorporated into the camera before each image was captured to ensure that the training dataset encompassed a diverse range of camera postures. This strategy enhances the variety in camera posture and further enriches the training dataset.

Two files in EXR [37] format were saved, one containing the 8-bit per channel colour buffer and the other containing linearised 16-bit depth buffer contents. The resolution of both files was 640×480 pixels. To ensure consistency between the colour and depth data for each simulation, the Unity timescale was set at 0 during the recording process to prevent object motion from causing differences. The synthetic data generation process resulted in a total of 5000 samples, with 4000 allocated for training and 1000 allocated for validation. RGB-D samples were extracted from the synthetic training dataset as shown in Figure 2.

2.3 | Landmark detection network

The anatomical landmark (tooth) detection network consists of a fully convolutional network (FCN) [38] as shown in Figure 3. The purpose of landmark detection is to identify essential features of dental teeth. The recognition of different landmarks is required for different fields of computer vision applications. Thus, we utilise anatomical landmarks to develop a DTA-based AR application to calibrate the OST-HMD device automatically.

We implement an FCN that has been optimised for anatomical landmark detection and used to predict various aspects of a 3D dental model by generating dense probability maps after separating the foreground from the background. In accordance with Ref. [39], the outputs of the FCN are then used to identify the landmarks on teeth and construct pixel-wise labels for real-time tracking. The FCN architecture generally comprises an encoder that computes feature maps with diminishing spatial resolution and increasing depth information via convolution and pooling operations. The decoder uses transposed convolutions and element-wise fusion to generate class score maps corresponding to the spatial dimensions of the input image.

To perform the initial calibration of the HoloLens, the depth-RGB camera mounted on the HMD captures and recognises the features of the mandible (phantom). The captured frames are processed to conform to the input modality required by the FCN, which produces a 3-channel probability map indicating the probability values for each class (background = 0, mandible = 1, and landmarks = 2) at each pixel location. The teeth feature points can be densely labelled by combining the probabilities with depth information from the input frame. These labelled pixels are then used to create data correspondences for model-based tracking. This tracking method fits an object model to the labelled data, enabling the 6D pose registration of the object. In the

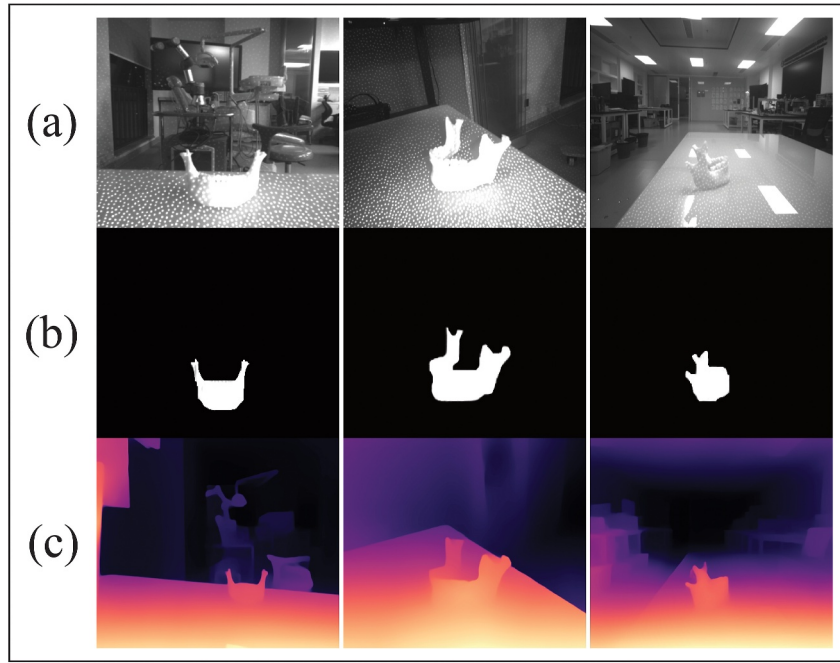


FIGURE 2 Samples of synthetic data. (a) RGB images, (b) binary mask, and (c) depth map.

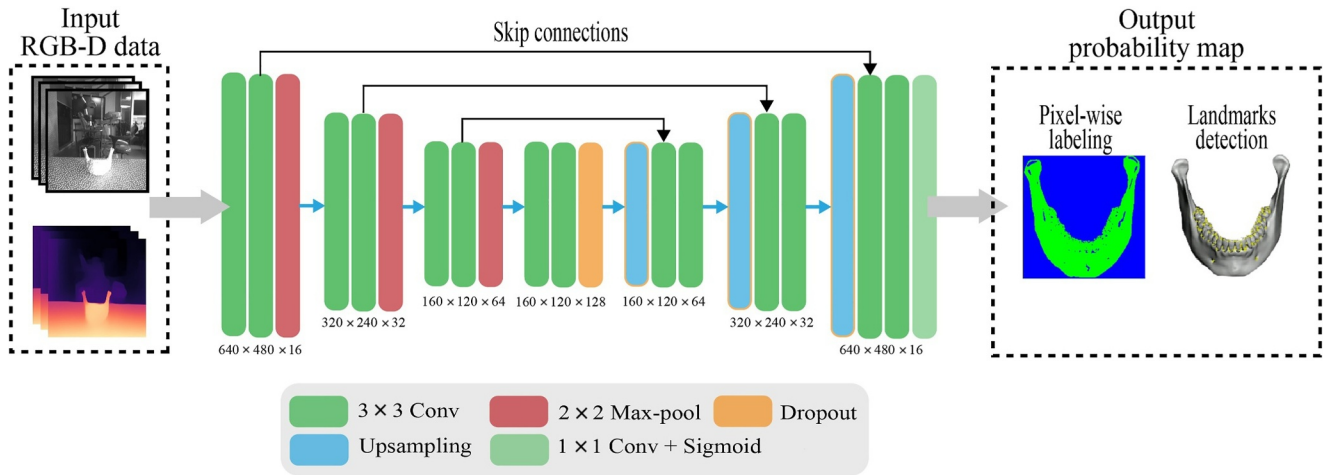


FIGURE 3 Pixel-wise landmark detection network using RGB-D data.

dataset, the pixel locations corresponding to landmarks on teeth are designated with a ground truth (GT) label classified as class = 2.

2.3.1 | Dense pixel-wise classification and detection

The output from the FCN is utilised to generate pixel-wise labels for the mandible, enabling us to detect and track the landmarks of a 3D model for 6D pose estimation. Generally, pixel-wise classification involves selecting the class with the highest probability. However, owing to the relatively coarse and

uncertain nature of FCN predictions, this approach's accuracy of region outlines is limited. To overcome this limitation, we label only those pixels whose depth values fall within a specific interaction volume. This approach results in more accurate outlines, although ambiguity may persist in areas where landmarks overlap. At pixel i , we estimate the class label f_i for a given number N of classes as probabilities $\{p_{c,i}\}$, where $c \in \{1, \dots, N\}$ classes at the current pixel. After that, the maximum labelling probability is calculated for each sample via the following equation:

$$\hat{f}_i = \underset{c \in \{1, \dots, N\}}{\text{sample}} \{p_{c,i}\} \quad (2)$$

where the sample returns the class index $c \in \{1, \dots, N\}$ at probability $\{p_{c,i}\}$. Finally, we use these detected landmarks to track the 3D mandible model. Given the class probability to estimate the object 6D pose, the depth-RGB camera is used to calibrate the HoloLens and navigate the virtual object with respect to the real object (Phantom).

2.3.2 | Network training

The network was trained on synthetic RGB-D images generated from a 3D model. A total of 5000 images were used for training and validation, whereas a real 3D model was used for testing. In addition, we employed data augmentation techniques to enhance the network's performance. We increased the dataset by flipping the images horizontally and vertically, intensifying the light, and resizing each image. Moreover, we carried out random image perturbations with a 0.3 probability, which involved (1) constraining colour values between arbitrary minimum and maximum limits, (2) converting the image to grayscale, (3) introducing uniform noise, and (4) adjusting hue, saturation, and value with random multipliers. After the perturbation and with a 0.5 probability, we incorporated RGB noise.

Moreover, after data augmentation, the training dataset was increased to 20,000 RGB-D images, which were trained for 100 epochs with a batch size of 8. We resized the RGB image and depth map during training to 640×480 . We used Reverse Huber as the training loss [40] and Adam as the optimiser [41] with $\alpha = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. PyTorch was used to implement the network. The performance parameters of the landmark detection network achieve comparable segmentation, detection, and inference time results of 93%, 88%, and 0.03 s, respectively, making it fast and robust online tracking for HMD.

3 | PROPOSED MARKERLESS CALIBRATION AND NAVIGATION METHOD

The detected landmarks expose the features of the mandible and have spatial resolution over the mandibular area. This helps to detect the main features of the mandible, such as the head, neck, ramus, and base, before the transformation matrices are calculated for calibration. We calibrate the OST-HMD for the AR-based DTA system using these detected anatomical landmarks. Jun et al. [14] proposed a calibration approach in which the user is asked to point to virtual circles using fingertips. The method is tedious, and the chance of drawing the wrong virtual circles is high, which reduces the accuracy of the calibration. Therefore, we develop an intuitive approach to calibrate the OST-HMD automatically without user input. Our method first detects the landmarks on real 3D objects using a depth stereo camera. Afterwards, it aligns multiple feature points of a real 3D object with its virtual model, which facilitates and speeds up the calibration

procedure. With this approach, the distance between virtual and real feature points is preserved, and an accurate calibration is achieved. As shown in Figure 4, the detected landmarks on the teeth automatically align with the phantom in a virtual environment.

To determine the pose of a 3D object, 3 non-collinear points in space are sufficient, but we choose a single feature point on each tooth to make alignment more accessible, provide a better depth cue, and collect additional points. Finally, we compute the corresponding 3D-3D transformations and apply the random sample consensus (RANSAC) [42] algorithm to reject the outliers with re-projection errors and refine the most accurate transformation.

3.1 | Calibration

The HoloLens transformation H acts as an interface between depth camera D and the rendering camera frustum V . To ensure precise image rendering, it is essential to calibrate the projection matrix K of the rendering camera on the basis of the intrinsic parameters of the eye-display pinhole camera system [28], as the user's perception directly reflects reality.

Failure to account for parallax may result in misalignment, owing to the difference in depth between the target and the virtual display [43]. Notably, the render focal distance for HoloLens 2 is approximately 2.0 mm ($\pm 5\%$), rendering the OST display susceptible to parallax-induced misalignment, particularly in near-field manual tasks, such as surgical assistance [44]. To achieve spatial consistency between the display pixels $p^{(S)}$ and an observed 3D real point $v^{(M)}$, we use the following equation:

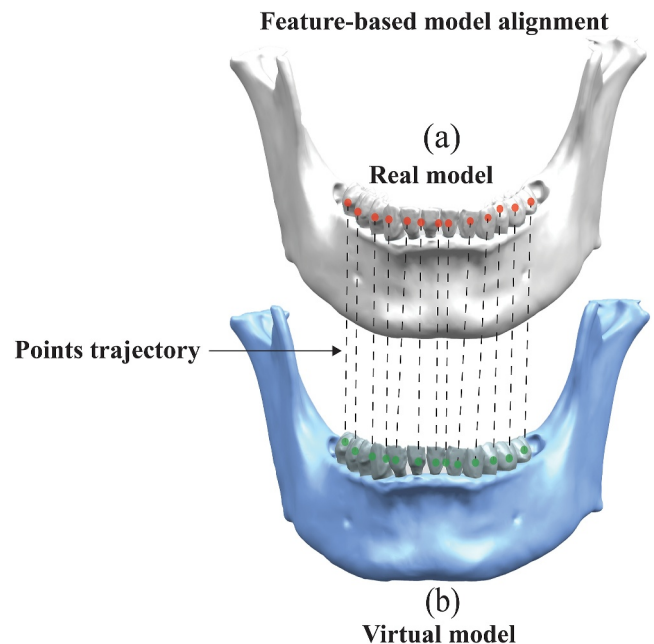


FIGURE 4 3D model alignment. (a) Phantom and (b) virtual model.

$$\mathbf{p}^{(S)} = \mathbf{K} \times \mathbf{T}_D^H \times \mathbf{T}_H^V \times \mathbf{M}^{(D)} \times \mathbf{v}^{(M)} \quad (3)$$

The intrinsic camera parameter $\mathbf{K}_H$ can easily be calibrated via a typical plane-based approach [43]. As illustrated in Figure 1, the stereo pose $\mathbf{T}_D^H$ and the viewpoint-dependent projection matrix $\mathbf{K}$ are associated with the transformations. The transformation matrices should always be carefully calibrated for a precise OST overlay. Our method calibrates HoloLens 2 automatically via an external depth-RGB camera, which is integrated onto a headset for the immersive AR platform. Similarly, in Ref. [45], we used the depth-RGB camera to transfer the data through the wireless network to HoloLens and calculated the transformation matrix. The $\mathbf{T}_D^H$ is calculated in two steps via $\mathbf{T}_D^H = \mathbf{T}_{\text{RGB}}^H \times \mathbf{T}_D^{\text{RGB}}$; initially, the pose transformation between the cameras $\mathbf{T}_{\text{RGB}}^H$ is calibrated via the standard procedure [46]. After that, the intrinsic transformation between $\mathbf{T}_D^{\text{RGB}}$ was determined from the depth-RGB camera's application programming interface (API). On the other hand, our calibration work does not require any internal transformation $\mathbf{T}_D^{\text{RGB}}$. Furthermore, when the transformation $\mathbf{T}_D^H$ is needed, it provides a fast check-and-update method for refining the alignment. This is due to the relative movement of the depth-RGB camera caused by the system "add-on" environment. The calibration model is tracked as follows: initially, the pose of the mandible is estimated in RGB coordinates, and the positions of the detected landmarks are calculated. After that, the pixels associated with the detected landmarks of $\mathbf{M}^{(\text{RGB})}$ from the depth frame are converted into 3D points. The generated 3D point cloud is processed via a depth threshold to eliminate noise and unwanted background. Moreover, the model is registered with $\mathbf{M}^{(D)}$ as an initial global alignment. In order to remove the occlusion noise from the 3D landmarks, we compare $\mathbf{M}^{(D)}$ with $\mathbf{M}^{(\text{RGB})}$ and then remove the pose difference higher than the selected threshold (p value ≤ 0.7). After that, the accepted $\mathbf{M}^{(D)}$ landmarks are computed and forwarded to the Kalman filter [47] to remove the processing and measurement noise with covariances of 10^{-4} and 10^{-5} , respectively. We assumed that the movement of the calibration model was constant under 30 FPS for HoloLens 2 [29]. We calculate the feature points as $\mathbf{f}_i^{(D)}$.

Moreover, these feature points are transferred to HoloLens H , and the landmarks are extracted $\mathbf{I}_i^{(H)}$ for virtual-real alignment. After that, pairwise 3D-3D corresponding points for the HoloLens $\mathbf{I}_i^{(H)}$ and depth-RGB camera $\mathbf{f}_i^{(D)}$ are collected. Finally, we optimise all the corresponding points via perspective-n-point (PnP) optimisation [48] via the following equation:

$$\mathbf{T}_D^H = \text{argmin} \left\| \mathbf{I}_i^{(H)} - \hat{\mathbf{I}}_i^{(H)} \left(\mathbf{f}_i^{(D)}, \mathbf{K}_H, \hat{\mathbf{T}}_D^H \right) \right\| \quad (4)$$

where $\hat{\mathbf{I}}_i^{(H)}$ represents the 3D estimated re-projected landmarks from the collected 3D $\mathbf{f}_i^{(D)}$, $\mathbf{K}_H$ represents the camera intrinsic parameters, and $\hat{\mathbf{T}}_D^H$ represents the current estimated pose.

3.2 | Markerless navigation

Markerless navigation is an important step in developing an OST-HMD-based application to acquire a transformation matrix between the stereo camera and the 3D overlay in the real world. It is challenging to obtain the re-projection matrix for HMD. Thus, we can obtain the overall transformation and re-projection matrix via the feature point detection technique from the depth-RGB camera. We then identify the location and posture of a 3D model and reflect it directly onto the phantom to obtain the overall transformation and re-projection matrix.

We have already calculated the transformation from depth-RGB to HoloLens and the respective landmarks to automate the calibration process without user interaction. Initially, an ICP method can be used to calculate the rotational misalignment [49]. This paper proposes a multi-feature ICP method based on detected landmarks for point-to-point navigation. The method uses the corresponding landmarks of the 3D mandible model $\mathbf{I}_i^{(H)}$ to automatically estimate the navigation error between the exposed phantom and detected 3D landmarks from the depth-RGB camera $\mathbf{f}_i^{(D)}$. The feature point coordinates of the 3D model are derived from the depth camera coordinate frame, facilitating precise positioning and alignment of the 3D model over the phantom to enable markerless navigation. We use the transformation matrix [50] as follows:

$$\mathbf{I}_i^{(H)} = \mathbf{T}_H^V \left(\mathbf{f}_i^{(D)} \right) \quad i = 1, 2, \dots, n \quad (5)$$

where $\mathbf{T}_H^V$ is the transformation from the HoloLens to the virtual display (S). For the detected landmarks $\mathbf{f}_i^{(D)} = \{ \mathbf{f}_1^{(D)}, \mathbf{f}_2^{(D)}, \dots, \mathbf{f}_n^{(D)} \}$, we calculate the transformation $\mathbf{I}_i^{(H)} = \{ \mathbf{I}_1^{(H)}, \mathbf{I}_2^{(H)}, \dots, \mathbf{I}_n^{(H)} \}$. Both $\mathbf{f}_i^{(D)}$ and $\mathbf{I}_i^{(H)} \in \mathbb{R}^3$, and the objective is to map the efficient 3D-3D transformation between the set of points.

In addition, we assume that the transformation $\mathbf{T}_H^V$ is linear, and our goal is to estimate the transformation between the HMD 3D tracking sensor coordinates and 3D scene camera (visualisation) coordinates. Additionally, owing to the automatic detection of landmarks, the unknown parameters have less re-projection error. Therefore, we employed direct linear transformation (DLT) [51] to minimise the re-projection error. Furthermore, we use the least-square error to correctly register orientations on 3D sets of points to decrease the overlay error.

The error associated with each of these transformations is analysed to determine the best representation of the model transformation. However, the alignment task is automatic, but the chance of an error always exists; therefore, we use the RANSAC [42] algorithm to reject outlier samples based on re-projection error for more accurate 3D-3D transformation. We calculate the re-projection error as follows:

$$E_{\text{re-proj}} = \frac{\sum_{i=1}^n \sqrt{I_i^{(H)} - T_H^V(f_i^{(D)})^2}}{n} \quad (6)$$

Finally, the landmarks on the detected phantom model are aligned with the user perspective, and the HoloLens is calibrated with the real model.

4 | EXPERIMENT AND RESULTS

To assess the effectiveness of our proposed markerless calibration and navigation method, we conducted experiments using head-mounted setups as illustrated in Figure 5. We attached a depth-RGB camera (D) to detect the landmarks on the phantom since the alignment is performed relative to the phantom. It can provide high-resolution depth maps of 120×72 pixels at 30 FPS.

Evaluating the calibration of an OST-HMD is always challenging since only the user wearing the device can observe the superimposed objects resulting from the HMD. To make this process more objective, some studies have used a camera instead of the user's eye [52]. These measurements were conducted on images captured via an eye camera. As a result, using a camera will inevitably impact the depth information. Therefore, in our study, we used HoloLens 2 to create an immersive environment for surgical navigation.

Algorithm OST-HMD navigation.

Input: Feature points (landmarks) $f_i^{(D)}, i = 1, 2, \dots, n$

Output: Calculate the re-projection error based on detected landmarks $f_i^{(D)}$

1. **while** $i \leq n$
2. **if** $I_i^{(H)} = f_i^{(D)} \rightarrow$ corresponding landmark is matched with depth-RGB camera
3. Initialisation: $\{HMD\}_i$
4. **for** $j \in L_i, \rightarrow L = \text{landmarks } \{L_1, L_2, \dots, L_n\}$
5. $I_i^{(H)} = T_H^V(f_i^{(D)})$
6. **if** HMD move (translation or rotation)
7. calculate the cumulative re-projection error $+= E_{\text{re-proj}}$
8. **end if**
9. **end for**
10. **end if**
11. **end while**

4.1 | System configuration

We implemented the OST-HMD-based DTA application on the game engine Unity3D (v2019.3.15f1) and deployed it on

HoloLens 2. An application was developed via Python 3.7 to track depth and was executed on a PC with an Intel® Core™ i9-12900K processor and 32 GB of memory. The OpenCV library was used to build both applications. To enable real-time wireless communication between the PC and HMD, the UDP was utilised. The CAD model was displayed to the combined camera system for stereo-pose calibration. The system simultaneously tracks the phantom and sends the detected 3D landmarks $I_i^{(H)}$ to the PC. $f_i^{(D)}$ and $I_i^{(H)}$ were synchronised and used for stereo-pose optimisation. After that, the OST-HMD projection was refined with a phantom and displayed 0.5 m away from the observer. The multi-point landmark detection using a single 3D model for calibration is adopted [1]. The displayed virtual object is aligned with the phantom based on detected feature points of the teeth. Therefore, these detected landmarks are essential in validating the 3D virtual-to-real alignment. Afterwards, these landmarks are passed through the RANSAC and DLT point-matching processes, which refine the necessary transformations and reject the outliers, respectively [50].

4.2 | Validation

First, we perform calibration experiments from HoloLens to phantom misalignment. The overall hardware display (HD) error e_{HD} and pixel-wise error e_p for OST-HMD are calculated. Since the pixel error depends on the device resolution and the target distance, the overlay error is device dependent but depth independent and measured in mm, and the visual angle error is device independent and depth independent and measured in arcmin. We estimated the e_p for the OST-HMD display when it was re-projected onto the phantom between 0.2 and 0.5 m. Additionally, the average calibration time that the users observed was 26.3 s for stereo poses and 56.7 s for OST projection refinement. The overlay errors e_{HD} and e_p are calculated as (6.47 pixels, 1.21 mm, 8.31 arcmin) and (4.37 pixels, 1.03 mm, 5.39 arcmin), respectively.

Second, we validated OST-HMD-based navigation, and a quantitative analysis of motion parameters, that is, translation and rotation, was performed as shown in Figure 6. The motion parameters include x, y, z linear translations and angular rotations θ . The phantom was placed on the table in a clinical environment without any markers. Initially, we evaluated the translation of the HoloLens to estimate the pose of the stereo camera H . After that, the user slightly moved the head between 1 and 10 mm in x, y, z , and rotated the head between -10° and $+10^\circ$. The motion estimation error can be defined as the difference between the initial reference frame and the transformed frame of environmental change. Within the measurement range, the motion platform performs multiple round trips as well as fixed-point measurements. To evaluate the overall moving parameters for the OST-HMD quantitatively, we used the 3D landmarks of mandibular features as the display features because these points can easily extract the main

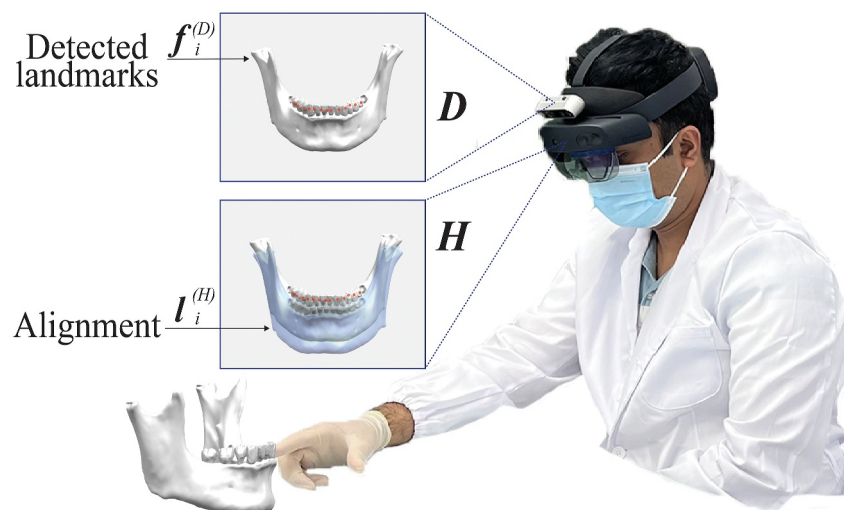


FIGURE 5 3D-3D calibration and navigation of the exposed 3D printed mandibular phantom for both D and H .

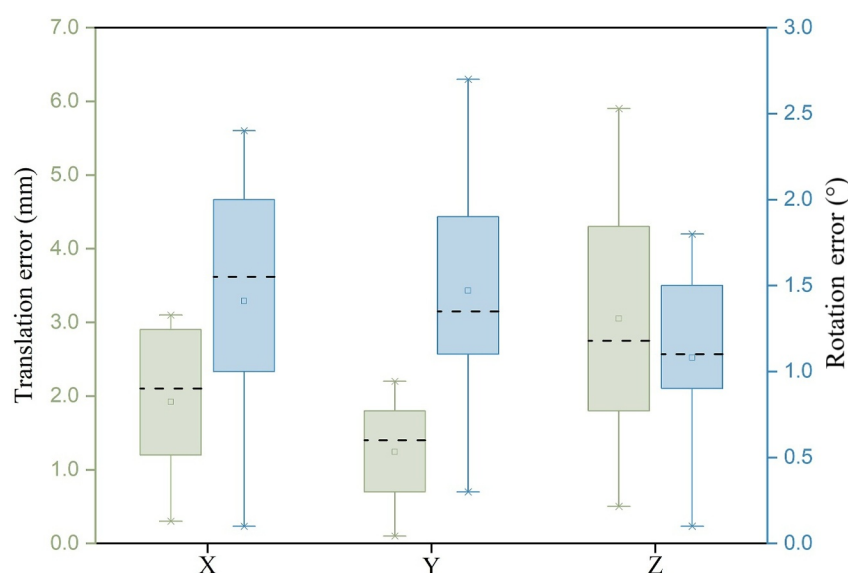


FIGURE 6 Box and whisker plot of 3D navigational errors in a dental AR system, showing translation (mm) and rotation ($^{\circ}$) errors across the X , Y , and Z axes. This illustrates the error distribution and median values, which are used to assess the navigational accuracy of the system and its acceptability for clinical use.

features of the desired dental mandible. Therefore, we can accurately obtain the desired position of the 3D image in the world coordinate system.

The markerless navigation performance was evaluated quantitatively and achieved average translation and rotation errors of (3.85 ± 0.62) mm and $2.65^{\circ} \pm 1.47^{\circ}$, respectively. Additionally, we calculated an alignment time of (17.43 ± 5.68) s. The user's experience wearing HoloLens 2 does not affect the navigational accuracy ($p \geq 0.5$). The best value is achieved for OST-HMD ($p = 0.87$), showing that 3D OST-HMD navigation is intuitive for novice users. In addition, the time delay of our system is associated with FPS, as our system achieves 30 FPS, which delivers a seamless and immersive experience while

minimising the delay between user movement and the corresponding feedback in the virtual environment.

As anticipated, using OST-HMD effectively reduces attentional distractions by eliminating the need for eye movements. Moreover, users have attested to the reliability of AR visualisation for dental treatment in clinical settings for mandible-based studies. Table 1 presents a comparative analysis of our approach with other state-of-the-art methods that employ the HoloLens in their research.

The navigation projection error was examined via the phantom, and the results were compared with those obtained via traditional methods. Similarly, we calculated and evaluated the precision of our model in comparison to the AR-based

method as demonstrated in Table 2. After several trials, the average projection error and accuracy are (0.53 ± 0.19) mm and 93.87%, respectively. The integrated image marker and target model were obtained through the corresponding point registration via the point registration method. However, searching for the corresponding points depends on the user's experience. The alignment of the virtual and real markers in the 3D scanner method was manually adjusted, which also relies on the user's input. The OST-based method introduces tracking error into the optical tracking device. The markerless shape-tracking method has high requirements for the contour characteristics of the target. The FMR and PointDSC algorithms have significant registration errors. However, our proposed method uses a landmark-based feature detection approach to automatically align and accurately calibrate real-virtual environments and does not rely on the user's experience or input. This method can easily track images in complex clinical settings.

Finally, we tested our proposed method by performing phantom-based experiments for guided 3D image navigation. Several trials were conducted to verify the feasibility of the DTA system for digital dentistry as shown in Figure 7. The

TABLE 1 Quantitative analysis of display calibration with state-of-the-art methods using HoloLens.

Method	Virtual–real distance (mm)	Anatomy	Study
Marker-based RGB-D	1.27 ± 0.69	Knee	Xue et al. [1]
Optical marker	1.30 ± 0.39	Skull	Sun et al. [53]
Markerless RGB-D	1.09 ± 0.23	Mandible	Ours proposed

TABLE 2 Evaluation parameters of calibration and navigation.

Calibration		Navigation	
Display (arcmin)	Time (s)	Translation (mm)	Rotation (°)
5.39 ± 1.89	26.3	3.85 ± 0.62	2.65 ± 1.47

motion parameters (i.e. translation and rotation) of HoloLens 2 were verified for calibration performance. The guided image was superimposed onto the phantom in situ without any optical tracker. We identified 6 mandible landmarks and tracked their translation and rotation motion between (1–4) mm and $(\pm 10)^\circ$, respectively. Table 3 shows the calibration and navigation results. Importantly, the calibration display and time are greatly improved for a single 3D object. According to the demonstrated phantom-based experimental results, the initial calibration and markerless navigation are valuable to clinicians in digital dentistry. Specifically, automatic calibration provides an enhanced and time-efficient AR-assisted experience during dental treatment.

4.3 | Ablation study

We conducted an ablation study by expanding the experiments and performing a quantitative analysis to evaluate the strengths and limitations of the key components in our system. An in-house dataset was used to assess the calibration and navigation framework. During the landmark detection process, the background was masked to eliminate inconsistencies caused by head movements (translational and rotational).

We compared our method with ICP [54], which estimates the camera pose via global image retrieval and local feature matching. Our state-of-the-art multi-ICP approach integrates a learning-based landmark descriptor with depth-based pixel matching, utilising depth data from the RealSense camera to automatically calibrate and align the virtual model in the HoloLens for pose estimation. To validate the effectiveness of our optimisation approach, we evaluated the performance of using multi-ICP alone versus using both multi-ICP and depth information. We followed the evaluation metrics outlined in Ref. [55], which estimated the position and orientation errors between the predicted and ground truth data.

We compared and evaluated the FPS versus orientation (translation and rotation) error estimation for global virtual and real-world navigation via a 3D model captured using

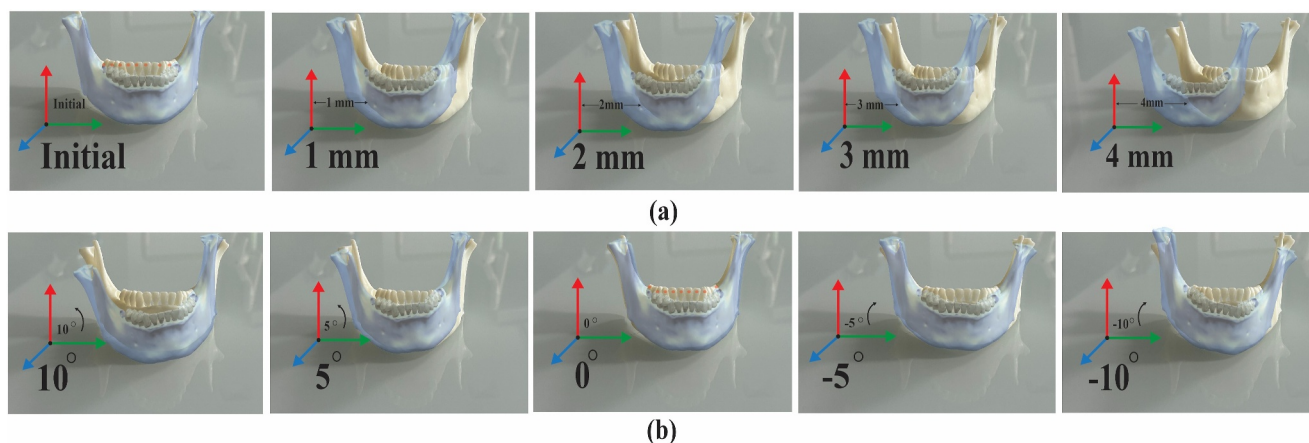
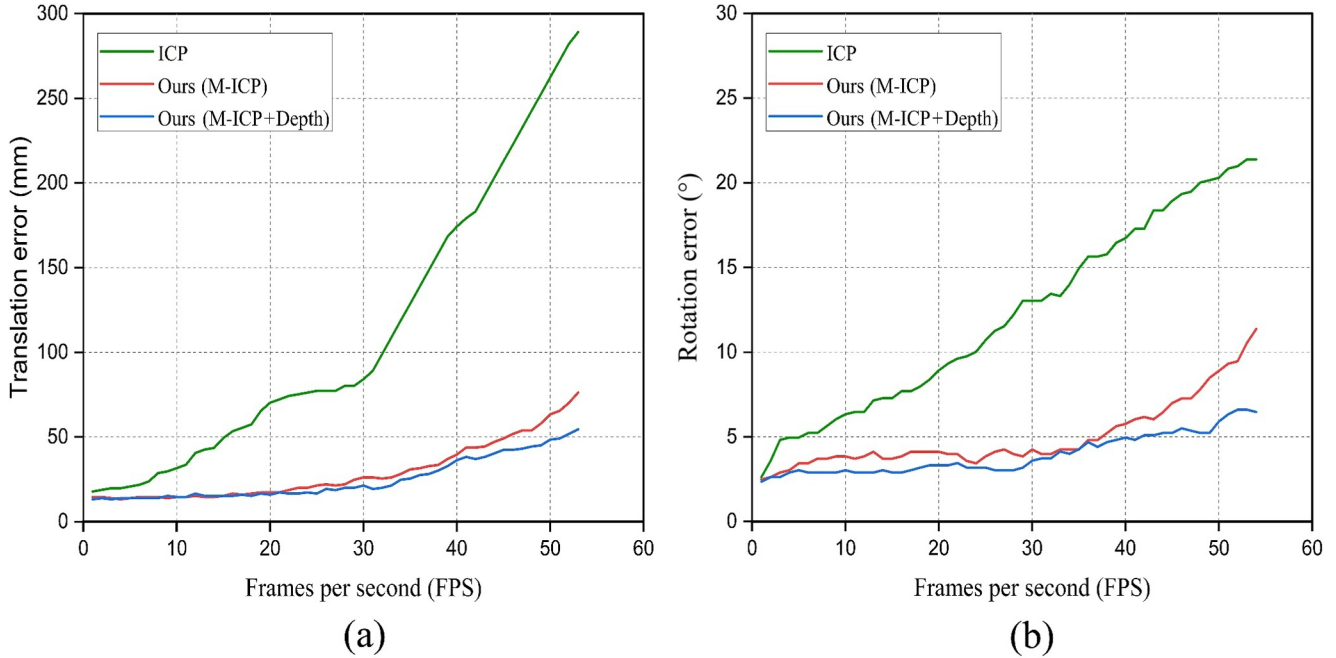


FIGURE 7 3D virtual model alignment with the phantom using HoloLens 2. (a) Translational movement and (b) Rotational movement. Initially, the world coordinates were aligned with the camera coordinates, and slight translation movement from 1 to 5 mm and rotational movement $\pm 10^\circ$ were performed.

TABLE 3 Quantitative analysis of the navigation projection error and accuracy of AR-based methods.

Parameters	Ours proposed	Iteration based [56]	Point based [57]	Shape based [58]	PointDSC [59]
Projection error (mm)	0.53 ± 0.19	0.62 ± 0.14	1.45 ± 0.13	1.65 ± 0.20	4.94 ± 1.67
Accuracy (%)	93.87%	54.25%	54.25%	55.70%	84.85%

**FIGURE 8** OST-HMD movement error at different FPSs. (a) Translation error and (b) rotation error.

HoloLens 2. The 3D model was captured at different locations and orientations. We initialised the pose by determining an approximate transformation and introduced random noise within the range of $[-5, 5]$ mm for translation and $[-10, 10]$ degrees for rotation in three directions. To assess the localisation accuracy of our model, we measured the relative translation and rotation errors as follows:

$$R_t = |t - \hat{t}| \quad (7)$$

$$R_R = \cos^{-1} \left(\frac{R\hat{R} - 1}{2} \right) \quad (8)$$

where $\hat{t}$ and $\hat{R}$ are the ground truth of translation and rotation, respectively, and t and R are the estimated values. For the ground truth, we first estimate the static pose through world coordinates and then scale the pose according to the real scanning of the 3D model. We demonstrated the translation and rotation errors at different FPSs taken from different viewpoints as illustrated in Figure 8. Our method demonstrated real-time navigation capabilities and outperformed ICP. The translation and rotation errors of the multi-ICP approach, which included depth information at 30 FPS, were 2 mm and 3°, respectively. These results indicated that the method could be feasibly implemented in a dental setting. However, the fixed-focus camera in the HMD

glasses introduced limitations; blurring occurred at short distances, and the depth measurements at longer distances were unreliable, leading to decreased navigation accuracy from various orientations.

5 | DISCUSSION

We propose a highly reconfigurable markerless calibration and navigation system for OST-HMD, which can automatically use anatomical landmarks of teeth as a feature of the mandible to align the HoloLens for the DTA system. First, we generated a synthetic dataset from a reconstructed 3D model to train the DCNN algorithm and extract the desired features before performing a calibration. Our main contribution to this work is to address the OST-HMD calibration problem, in which external sensors and user input are tedious and difficult. We jointly perform calibration and navigation to make our approach efficient during surgical navigation, resulting in an automatic system that can output highly reliable, stable, and efficient results for CAT. We have performed a phantom-based prospective study to evaluate the clinical potential of our framework thoroughly. The ultimate aim is to develop a robust real-time system for clinical dental applications. This approach bridges the gap between controlled laboratory studies and practical implementation in clinical dentistry. By refining our

techniques on phantom models, we establish a foundation for accurate landmark detection that can be translated into real-time patient scenarios. In addition, the system is designed to provide real-time assistance during dental procedures, offering immediate and precise identification of critical dental anatomical structures. This capability is particularly valuable in time-sensitive situations such as implant placement or orthodontic assessments. The transition from phantom-based testing to real-time clinical use underscores our commitment to creating a system that enhances dental practice efficiency and patient outcomes.

The 3D alignment between the target calibration and perceived virtual model is unconstrained in general MPAAM solutions [50]; for example, users need to align the corresponding 6 DOFs (degree of freedom), and the accuracy of the overlay is entirely dependent upon user perception. Notably, rotational alignment is prone to errors because of the reduced depth perception of OST-based displays. Since the OST-HMD update problem has 3 DOFs (i.e. viewpoint shift) by nature, the additional DOFs in rotation degrade the calibration accuracy and are thus undesirable. In our setup, users must maintain the landmark's orientation while translating it into 3 DOFs. Furthermore, RANSAC optimisation can tolerate up to 10° rotational misalignment, making the overall calibration more robust and effective for surgical navigation. The initial landmarks allow the OST-HMD to align the specified features automatically with real-world coordinates. According to our test, the landmarks can be easily retained during the translation and rotation of the user's head. To the best of our knowledge, we are the first to use the raw RGB-D camera output (landmark detection) for OST-HMD calibration.

6 | CONCLUSIONS

In this study, we propose a novel markerless OST-HMD AR system for CAT in digital dentistry. The system aims to utilise AR-based application in the DTA system to achieve real-time and robust feature-based markerless navigation. We aim to fully automate the OST-HMD calibration process by increasing the depth map quality between the virtual and real distances. The depth-RGB camera is attached to the HoloLens to detect the anatomical landmarks and identify the mandible in real-time. This information is then sent to the HoloLens for calibration. This novel approach eliminates the user's input and avoids an external sensor. Our system achieved a virtual-to-real distance re-projection error in pixels of (1.09 ± 0.23) mm for calibration, and the average calibration time obtained by the user for HoloLens 2 was 26.3 s. Moreover, our calibration achieved a display error of (5.39 ± 1.89) arcmin.

Furthermore, we perform markerless navigation for a phantom-based experiment for dental treatment (mandible) to verify the use of our system in digital dentistry. According to a user study, HoloLens-based surgical navigation is reliable, has stable tracking, displays latency, and is fast to align with real anatomical structures (phantom). We achieved overall translational and rotational surgical navigation errors of

(3.85 ± 0.62) mm and $(2.65^\circ \pm 1.47^\circ)$, respectively. In addition, the overall navigation projection error and accuracy are (0.53 ± 0.19) mm and 93.87%, respectively.

Finally, we can conclude the study by mentioning the potential and future of AR in digital dentistry. With the advancement of technology and increasing accessibility, we can expect more dentists to utilise AR tools and techniques to improve patient outcomes, enhance the overall dental experience, and transform traditional dental treatment into a fully immersive experience.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Nos. U2013205 and 62073309), the Guangdong Basic and Applied Basic Research Foundation (No. 2022B1515020042), and the Shenzhen Science and Technology Program (No. JCYJ20220818101603008).

CONFLICT OF INTEREST STATEMENT

All the authors disclosed no relevant relationships. No potential conflicts of interest were reported by the authors.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

CONSENT FOR PUBLICATION

All authors of this manuscript approved the publication.

ORCID

Zeyang Xia  <https://orcid.org/0000-0002-0075-7949>

REFERENCES

- Hu, X., y Baena, F.R., Cutolo, F.: Head-mounted augmented reality platform for markerless orthopaedic navigation. *IEEE J. Biomed. Heal. Inf.* 26(2), 910–921 (2022). <https://doi.org/10.1109/JBHI.2021.3088442>
- Chen, Y.-W., et al.: Computer-assisted surgery in medical and dental applications. *Exp. Rev. Med. Dev.* 18(7), 669–696 (2021). <https://doi.org/10.1080/17434440.2021.1886075>
- Collins, T., et al.: Augmented reality guided laparoscopic surgery of the uterus. *IEEE Trans. Med. Imag.* 40(1), 371–380 (2021). <https://doi.org/10.1109/TMI.2020.3027442>
- Nicolau, S., et al.: Augmented reality in laparoscopic surgical oncology. *Surg. Oncol.* 20(3), 189–201 (2011). <https://doi.org/10.1016/j.suronc.2011.07.002>
- Taylor, R.H., et al.: Medical robotics and computer-integrated surgery. *Springer Handbooks*, 1657–1684 (2016). https://doi.org/10.1007/978-3-319-32552-1_63
- Fida, B., et al.: Augmented reality in open surgery. *Updates Surg.* 70(3), 389–400 (2018). <https://doi.org/10.1007/s13304-018-0567-8>
- Zhang, N., et al.: A VR environment for human anatomical variation education: modeling, visualization and interaction. *IEEE Trans. Learn. Technol.* 17, 391–403 (2024). <https://doi.org/10.1109/TLT.2022.3227100>
- Joda, T., et al.: Augmented and virtual reality in dental medicine: a systematic review. *Comput. Biol. Med.* 108, 93–100 (2019). <https://doi.org/10.1016/j.combiomed.2019.03.012>
- Yang, L., Etsuko, K.: Review on vision-based tracking in surgical navigation. *IET Cyber-Syst. Robot.* 2(3), 107–121 (2020). <https://doi.org/10.1049/iet-csr.2020.0013>
- Sielhorst, T., Feuerstein, M., Navab, N.: Advanced medical displays: a literature review of augmented reality. *J. Disp. Technol.* 4(4), 451–467 (2008). <https://doi.org/10.1109/JDT.2008.2001575>

11. Li, R., et al.: A comparative evaluation of optical see-through augmented reality in surgical guidance. *IEEE Trans. Vis. Comput. Graph.* 30(7), 1–13 (2024). <https://doi.org/10.1109/TVCG.2023.3260001>
12. Zhang, A., et al.: Towards an accurate augmented-reality-assisted orthopedic surgical robotic system using bidirectional generalized point set registration. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4600–4607 (2023). <https://doi.org/10.1109/IROS55552.2023.10341401>
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), pp. 75–82 (2014). <https://doi.org/10.1109/3DUI.2014.6798846>
14. Jun, H., Kim, G.: A calibration method for optical see-through head-mounted displays with a depth camera. In: 2016 IEEE Virtual Reality (VR), pp. 103–111 (2016). <https://doi.org/10.1109/VR.2016.7504693>
15. Plopski, A., et al.: Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE Trans. Vis. Comput. Graph.* 21(4), 481–490 (2015). <https://doi.org/10.1109/TVCG.2015.2391857>
16. Min, Z., Zhu, D., Meng, M.Q.-H.: Accuracy assessment of an N-ocular motion capture system for surgical tool tip tracking using pivot calibration. In: 2016 IEEE International Conference on Information and Automation (ICIA), pp. 1630–1634 (2016). <https://doi.org/10.1109/ICInfA.2016.7832079>
17. Fan, G., et al.: Sampling visual SLAM with a wide-angle camera for legged mobile robots. *IET Cyber-Syst. Robot.* 4(4), 356–375 (2022). <https://doi.org/10.1049/csy2.12074>
18. Min, Z., Ren, H., Meng, M.Q.-H.: Estimation of surgical tool-tip tracking error distribution in coordinate reference frame involving pivot calibration uncertainty. *Healthcare Technol. Lett.* 4(5), 193–198 (2017). <https://doi.org/10.1049/hlt.2017.0065>
19. Sugano, N.: Computer-assisted orthopedic surgery. *J. Orthop. Sci.* 8(3), 442–448 (2003). <https://doi.org/10.1007/s10776-002-0623-6>
20. Hussain, R., et al.: Contribution of augmented reality to minimally invasive computer-assisted cranial base surgery. *IEEE J. Biomed. Heal. Inf.* 1 (2019). <https://doi.org/10.1109/JBHI.2019.2954003>
21. Wang, J., et al.: Video see-through augmented reality for oral and maxillofacial surgery. *Int. J. Med. Robot. Comput. Assist. Surg.* 13(2), e1754 (2017). <https://doi.org/10.1002/rcs.1754>
22. Badiali, G., et al.: Augmented reality as an aid in maxillofacial surgery: validation of a wearable system allowing maxillary repositioning. *J. Cranio-Maxillofacial Surg.* 42(8), 1970–1976 (2014). <https://doi.org/10.1016/j.jcms.2014.09.001>
23. Huang, T., et al.: Augmented reality (AR) and virtual reality (VR) applied in dentistry. *Kaohsiung J. Med. Sci.* 34(4), 243–248 (2018). <https://doi.org/10.1016/j.kjms.2018.01.009>
24. Wang, J., et al.: Augmented reality navigation with automatic marker-free image registration using 3-D image overlay for dental surgery. *IEEE Trans. Biomed. Eng.* 61(4), 1295–1304 (2014). <https://doi.org/10.1109/TBME.2014.2301191>
25. Maiseli, B., Gu, Y., Gao, H.: Recent developments and trends in point set registration methods. *J. Vis. Commun. Image Represent.* 46, 95–106 (2017). <https://doi.org/10.1016/j.jvcir.2017.03.012>
26. Zhang, Z., et al.: GHMM: learning generative hybrid mixture models for generalized point set registration in computer-assisted orthopedic surgery. *IEEE Trans. Med. Robot. Bionics* 6(3), 1285–1295 (2024). <https://doi.org/10.1109/TMRB.2024.3407362>
27. Myronenko, A., Song, X.: Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(12), 2262–2275 (2010). <https://doi.org/10.1109/TPAMI.2010.46>
28. Grubert, J., et al.: A survey of calibration methods for optical see-through head-mounted displays. *IEEE Trans. Vis. Comput. Graph.* 24(9), 2649–2662 (2018). <https://doi.org/10.1109/TVCG.2017.2754257>
29. Ungureanu, D., et al.: HoloLens 2 research mode as a tool for computer vision research. *arXiv* (2020). <https://doi.org/10.48550/arXiv.2008.11239>
30. Unity: Game engine. Available at: <https://unity.com/>. Accessed 4 March 2023
31. Microsoft: MixedReality-WebRTC (2023). Available at: <https://github.com/microsoft/MixedReality-WebRTC>. Accessed 28 July 2024
32. Habert, S., et al.: RGBDX: first design and experimental validation of a mirror-based RGBD X-ray imaging system. In: 2015 IEEE International Symposium on Mixed and Augmented Reality, pp. 13–18 (2015). <https://doi.org/10.1109/ISMAR.2015.17>
33. Zhou, X., et al.: Sparseness meets deepness: 3D human pose estimation from monocular video. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4966–4975 (2016). <https://doi.org/10.1109/CVPR.2016.537>
34. Zhang, X., Yu, H., Zhuang, Y.: A robust RGB-D visual odometry with moving object detection in dynamic indoor scenes. *IET Cyber-Syst. Robot.* 5(1), e12079 (2023). <https://doi.org/10.1049/csy2.12079>
35. Keselman, L., et al.: Intel(R) RealSense(TM) stereoscopic depth cameras. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1267–1276 (2017). <https://doi.org/10.1109/CVPRW.2017.167>
36. Tsykunov, E., et al.: Coupling of localization and depth data for mapping using Intel RealSense T265 and D435i cameras. *arXiv* (2020). <https://doi.org/10.48550/arXiv.2004.00269>
37. Openexr: OpenEXR. Available at: <https://openexr.com/en/latest/>. Accessed 28 July 2024
38. Schroder, M., Ritter, H.: Hand-object interaction detection with fully convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1236–1243 (2017). <https://doi.org/10.1109/CVPRW.2017.163>
39. Zeng, A., et al.: Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 1383–1386 (2017). <https://doi.org/10.1109/ICRA.2017.7989165>
40. Laina, I., et al.: Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 239–248 (2016). <https://doi.org/10.1109/3DV.2016.32>
41. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv* (2014). <https://doi.org/10.48550/arXiv.1412.6980>
42. Fischler, M.A., Bolles, R.C.: Random sample consensus. *Commun. ACM* 24(6), 381–395 (1981). <https://doi.org/10.1145/358669.358692>
43. Owen, C.B., et al.: Display-relative calibration for optical see-through head-mounted displays. In: Third IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 70–78 (2004). <https://doi.org/10.1109/ISMAR.2004.28>
44. Cutolo, F., et al.: Optical see-through head-mounted displays with short focal distance: conditions for mitigating parallax-related registration error. *Front. Robot. AI* 7 (2020). <https://doi.org/10.3389/frobt.2020.572001>
45. Garon, M., et al.: Real-time high resolution 3D data on the HoloLens. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 189–191 (2016). <https://doi.org/10.1109/ISMAR-Adjunct.2016.0073>
46. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(11), 1330–1334 (2000). <https://doi.org/10.1109/34.888718>
47. Kalman, R.E.: A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82(1), 35–45 (1960). <https://doi.org/10.1115/1.3662552>
48. Genc, Y., Tuceryan, M., Navab, N.: Practical solutions for calibration of optical see-through devices. In: Proceedings. International Symposium on Mixed and Augmented Reality, pp. 169–175 (2002). <https://doi.org/10.1109/ISMAR.2002.1115086>
49. Marlière, D.A.A., et al.: Feasibility of iterative closest point algorithm for accuracy between virtual surgical planning and orthognathic surgery outcomes. *J. Cranio-Maxillofacial Surg.* 47(7), 1031–1040 (2019). <https://doi.org/10.1016/j.jcms.2019.03.025>
50. Azimi, E., et al.: Alignment of the virtual scene to the tracking space of a mixed reality head-mounted display. *arXiv* (2017). <https://doi.org/10.48550/arXiv.1703.05834>

51. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2004). <https://doi.org/10.1017/CBO9780511811685>
52. Makibuchi, N., Kato, H., Yoneyama, A.: Vision-based robust calibration for optical see-through head-mounted displays. In: 2013 IEEE International Conference on Image Processing, pp. 2177–2181 (2013). <https://doi.org/10.1109/ICIP.2013.6738449>
53. Sun, Q., et al.: Fast and accurate online calibration of optical see-through head-mounted display for AR-based surgical navigation using Microsoft HoloLens. *Int. J. Comput. Assist. Radiol. Surg.* 15(11), 1907–1919 (2020). <https://doi.org/10.1007/s11548-020-02246-4>
54. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-9(5), 698–700 (1987). <https://doi.org/10.1109/TPAMI.1987.4767965>
55. Sarlin, P.-E., et al.: From coarse to fine: robust hierarchical localization at large scale. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12708–12717 (2019). <https://doi.org/10.1109/CVPR.2019.01300>
56. Shao, L., et al.: Augmented reality calibration using feature triangulation iteration-based registration for surgical navigation. *Comput. Biol. Med.* 148, 105826 (2022). <https://doi.org/10.1016/j.compbimed.2022.105826>
57. Zhu, M., et al.: A novel augmented reality system for displaying inferior alveolar nerve bundles in maxillofacial surgery. *Sci. Rep.* 7(1), 42365 (2017). <https://doi.org/10.1038/srep42365>
58. Vuforia: Shape based tracking. Available at: <https://library.vuforia.com/objects/model-target-generator-user-guide>. Accessed 28 July 2024
59. Bai, X., et al.: PointDSC: robust point cloud registration using deep spatial consistency. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, 15854–15864 (2021). <https://doi.org/10.1109/CVPR46437.2021.01560>

How to cite this article: Ahmad, F., Xiong, J., Xia, Z.: Automatic feature-based markerless calibration and navigation method for augmented reality assisted dental treatment. *IET Cyber-Syst. Robot.* e70003 (2024). <https://doi.org/10.1049/csy2.70003>