



# A practical marker-less image registration method for augmented reality oral and maxillofacial surgery

Junchen Wang<sup>1,2</sup> · Yu Shen<sup>1,2</sup> · Shuo Yang<sup>3</sup>

Received: 15 May 2018 / Accepted: 29 January 2019 / Published online: 1 March 2019  
© CARS 2019

## Abstract

**Background** Image registration lies in the core of augmented reality (AR), which aligns the virtual scene with the reality. In AR surgical navigation, the performance of image registration is vital to the surgical outcome.

**Methods** This paper presents a practical marker-less image registration method for AR-guided oral and maxillofacial surgery where a virtual scene is generated and mixed with reality to guide surgical operation or provide surgical outcome visualization in the manner of video see-through overlay. An intraoral 3D scanner is employed to acquire the patient's teeth shape model intraoperatively. The shape model is then registered with a custom-made stereo camera system using a novel 3D stereo matching algorithm and with the patient's CT-derived 3D model using an iterative closest point scheme, respectively. By leveraging the intraoral 3D scanner, the CT space and the stereo camera space are associated so that surrounding anatomical models and virtual implants could be overlaid on the camera's view to achieve AR surgical navigation.

**Results** Jaw phantom experiments were performed to evaluate the target registration error of the overlay, which yielded an average error of less than 0.50 mm with the time cost less than 0.5 s. Volunteer trial was also conducted to show the clinical feasibility.

**Conclusions** The proposed registration method does not rely on any external fiducial markers attached to the patient. It performs automatically so as to maintain a correct AR scene, overcoming the misalignment difficulty caused by patient's movement. Therefore, it is noninvasive and practical in oral and maxillofacial surgery.

**Keywords** Augmented reality · Oral and maxillofacial surgery · Video see-through · Image registration

## Introduction

Augmented reality (AR) provides in situ augmented visualization which can enhance the understanding toward a scene. This feature is especially useful for minimally invasive surgery that has inherent challenges in surgical visualization due to the limited surgical access and indirect vision to surgical field. Over the past decade, AR techniques have

emerged in medical areas such as for medical education [1], surgical training [2, 3] and surgical interventions [4–9] and are attracting more and more attentions from both industrials and academia. In AR surgical navigation, anatomical models containing rich and detailed information are generated from the patient's CT or MR data preoperatively. During the surgery, the models are registered to the surgical site and mixed with the reality by one of the means of 3D image overlay [6, 10], optic see-through display [11–14], video see-through display [15–17] and projector-based mapping [18, 19], to provide surgeons with augmented vision on surrounding anatomy and critical structures, which could enhance the safety of the operation.

Image registration [20] lies in the core of AR, which aligns the virtual scene with the reality. Especially in a nonstructured natural environment, the image registration has become the bottleneck limiting the practical application of AR. In the context of AR surgical navigation, the performance of image registration is vital to the surgical

✉ Shuo Yang  
alex2005191007@163.com

Junchen Wang  
wangjunchen@buaa.edu.cn

<sup>1</sup> School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

<sup>2</sup> Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100086, China

<sup>3</sup> Stomatological Hospital, Southern Medical University, Guangzhou, China

outcome. Many image registration methods for AR surgery have been proposed in the literature, most of which focus on laparoscopic surgery. Souzaki et al. [21] presented an AR navigation system for pediatric oncologic surgery using a fiducial marker-based registration algorithm with an external tracking device. Su et al. [17] proposed an AR surgical navigation system for partial nephrectomy using a 3D point cloud registration method with manual initialization. Puerto-Souza et al. [22] presented an AR display system for minimally invasive surgery using image-based tracking. Initial manual adjustment is needed to align the 3D anatomical model with the video stream, and feature tracking is applied in successive video frames to maintain the correct overlay. One means of achieving marker-less image registration for abdominal surgery is using a stereoscopic laparoscope to reconstruct 3D surface of the organ which is further registered to its preoperative CT model using a rigid/nonrigid shape matching scheme. Nevertheless, feature-based sparse reconstruction usually has insufficient number of 3D points for shape matching, while texture-poor appearance and specular reflection bring great challenges to pixelwise dense reconstruction. Stoyanov et al. [23–25] have proposed some dense reconstruction algorithms using a stereoscopic laparoscope. These algorithms require the organ surface of interest have rich textures, and only a small piece of the organ surface can be reconstructed due to the limited endoscopic view. The automatic segmentation of the reconstructed organ surface from the background is still unsolved, and manual initial alignment is often required when matching the partially reconstructed surface with its global preoperative counterpart.

To transfer the technology to clinical use, the image registration is expected to have the following characteristics in addition to accuracy requirement: (1) It should not introduce too much extra work into the current surgical flow. (2) It should not introduce too much extra invasiveness (including X-ray radiation) to the patient. (3) It should be performed quickly and updated in real time. Although the image registration in surgical navigation has been being a hot topic in the area and has drawn intensive attention in the academic community, existing solutions still fail to fulfill all the mentioned requirements, which hampers its clinical application.

In the context of oral and maxillofacial surgery, in the current practice a gap indeed exists between surgical planning stages and interventional procedures. Mature medical CAD/CAM technologies ensure a precise surgical plan based on the patient's CT data, whereas the surgical outcome usually cannot be guaranteed to be the same as what the preoperative plan indicates. The reason is the lack of accurate surgical guidance and anatomical visualization during the procedure. Surgical navigation for oral and maxillofacial surgery has the potential to fill the gap [26–29]. However, many challenges arise in the image registration step. Most

of the existing solutions rely on a 'wearing marker' that is a 3D printing dental cast for holding reference markers. Visual markers are fixed in the specified positions on the dental cast so that the pose of the dental cast can be tracked by an external optical tracker. This kind of image registration has several disadvantages: (1) It takes much time and labor to prepare such a dental cast. (2) The mismatch between the actual positions and the designed positions of the markers contributes to the tracking error. (3) For the visibility of the markers, the dental cast has an extending part which is out of the mouth to hold the markers. The tracking error will be significantly amplified due to the long error propagation path and the low stiffness of the extending part.

In our previous work [16, 30], we proposed an AR-based surgical navigation solution for oral and maxillofacial surgery. An automatic real-time marker-less image registration method was presented to match the patient's teeth model derived from the CT data with its 2D images captured by a single video camera. However, the real teeth are partially covered by the soft tissue gingiva which cannot be accurately identified and reconstructed from the CT data. In contrast, the gingiva will be present clearly in the 2D image, which means that the boundary shape between the gingiva and the teeth in the image is missing from the CT-derived model. This kind of shape inconsistency adversely affects the registration accuracy.

In this study, to overcome the shape inconsistency between the real teeth and the CT-derived teeth model, we propose to use an intraoral 3D scanner to acquire the 3D optical reconstruction of the teeth with the gingiva. A semi-automatic segmentation tool based on curvature analysis is used to quickly extract the exposed teeth part without the gingiva. The exposed teeth model from the intraoral scanner is then matched with the preoperative CT data using an iterative closest point (ICP) scheme and with an intraoperative stereo video camera system automatically using a novel 3D stereo shape matching algorithm, respectively. Therefore, surrounding anatomical structures, virtual implants and informative models from the CT data can be overlaid on the stereo video stream for surgical visualization and surgical outcome confirmation. The novelty and improvement of this study compared with existing and our previous methods are clarified as follows: (1) The registration is directly performed by matching the 3D model with the stereo images. Therefore, it does not introduce any invasiveness or discomfort to the patient. (2) The registration does not need tedious preparation and can be performed in nearly real time by just letting the camera see the patient's teeth, minimizing the interference to the current surgical workflow. (3) A novel 3D stereo matching algorithm is proposed to improve the shape matching accuracy by taking stereo images into account.

## Materials and methods

### Method overview

Figure 1 illustrates the proposed image registration procedure. A customized stereo camera is built to track the patient's teeth (shape tracking).

Before surgery, the patient's CT data are acquired and processed to extract his/her 3D models (e.g., jaw and teeth) with surrounding critical anatomy (e.g., nerve channels and tooth roots). Virtual implants (e.g., dental implant and drilling/cutting positions) are designed and placed appropriately on the patient's model.

Intraoperatively, a dental clamp is used to expand the patient's mouth so that the teeth area is exposed to the camera as largely as possible. A commercial intraoral 3D scanner is employed to scan the teeth and produce a 3D model (M\_B). The model M\_B is matched with the corresponding CT-derived model M\_A. This matching is only done once at the beginning of the surgery. Afterward, the pose of the patient's teeth with respect to the stereo camera is obtained by matching the model M\_B with the stereo images of the camera using a 3D stereo shape matching algorithm. According to the transformation chain, the spatial relationship between the CT space and the stereo camera is resolved dynamically. The anatomical models and virtual implants from the CT space can be overlaid on the stereo stream of the camera, to guide the surgical operation.

### Data acquisition and model generation

#### Preoperative CT data

The patient undergoes CT scan for diagnosis and surgical planning purpose. According to the position of the surgical site, the maxilla or mandible with teeth (e.g., see M\_A in Fig. 1) is segmented out from the CT data using a thresholding-based segmentation method. The teeth model is reconstructed for image registration purpose. It will be matched with its counterpart obtained by an intraoral 3D scanner. In addition, virtual informative information such as implants' appearance, positions and orientations can also be modeled in the same CT space. Because they are from the same CT space, these informative models will be automatically aligned with the stereo stream after the image registration to provide an AR scene. Take the mandible for example; Fig. 2 shows the CT model generation process.

#### Intraoperative intraoral data

Note that the teeth model generated from CT data loses the soft gingiva structure. Therefore, the teeth part covered by the gingiva cannot be identified and removed accordingly from the CT-derived model. In contrast, the teeth part covered by the gingiva is hidden from the stereo camera's view. This will cause shape inconsistency when matching the CT-derived teeth model with the stereo images, making the matching process unstable and inaccurate. To overcome this difficulty, an intraoral 3D scanner is employed to acquire

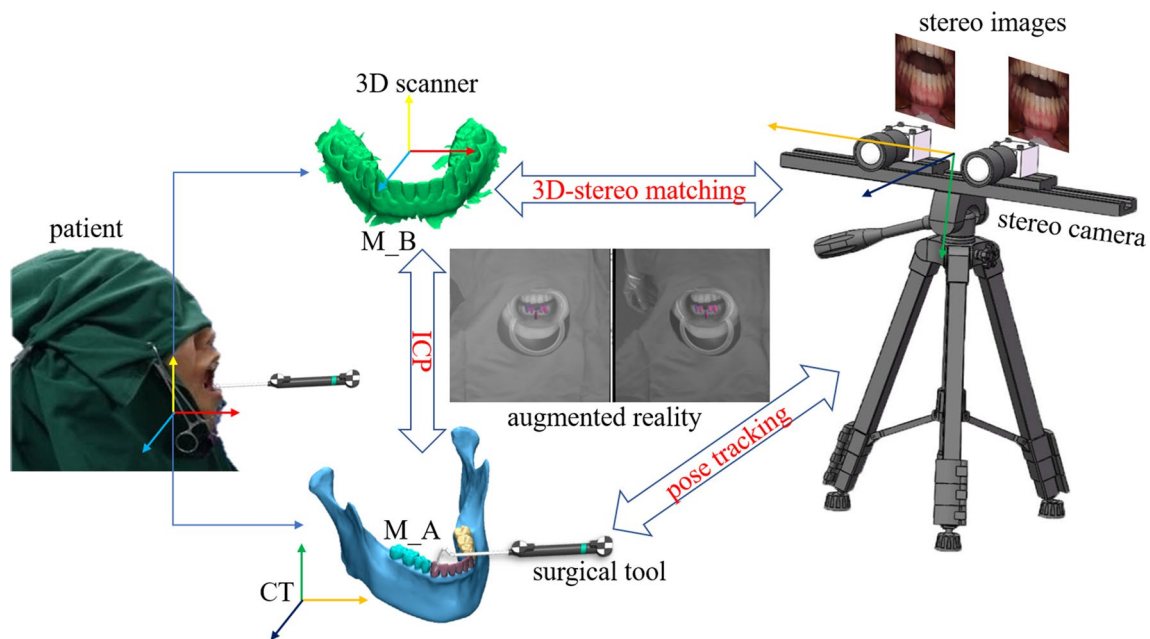
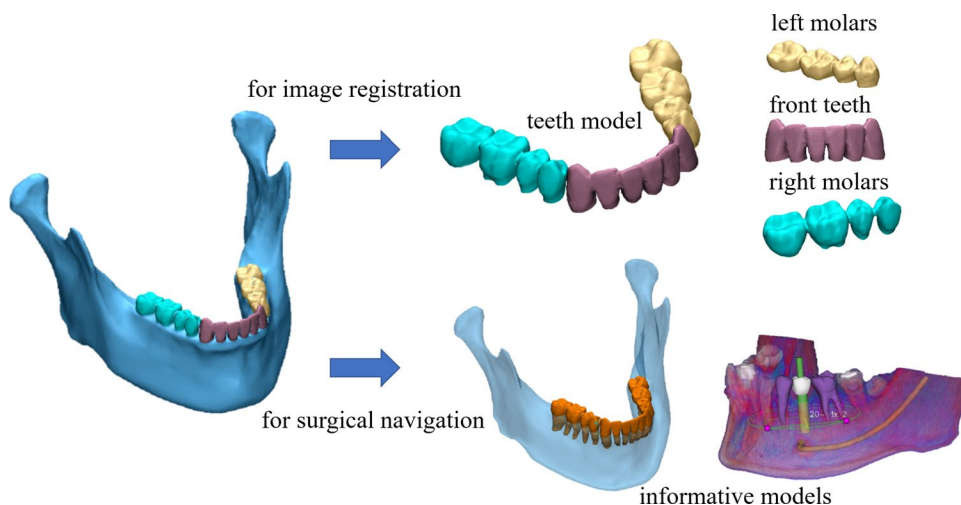


Fig. 1 Registration method overview

**Fig. 2** Model generation from CT data. Teeth models (left/right molars and front teeth) are generated from the CT data for image registration purpose. Which teeth model is used for the registration depends on the location of the surgical site. The teeth model will be matched with its counterpart from the 3D scanner so that the exposed teeth model (see Fig. 3c) can be transformed into the CT space. Informative models are also generated from the CT data for image overlay

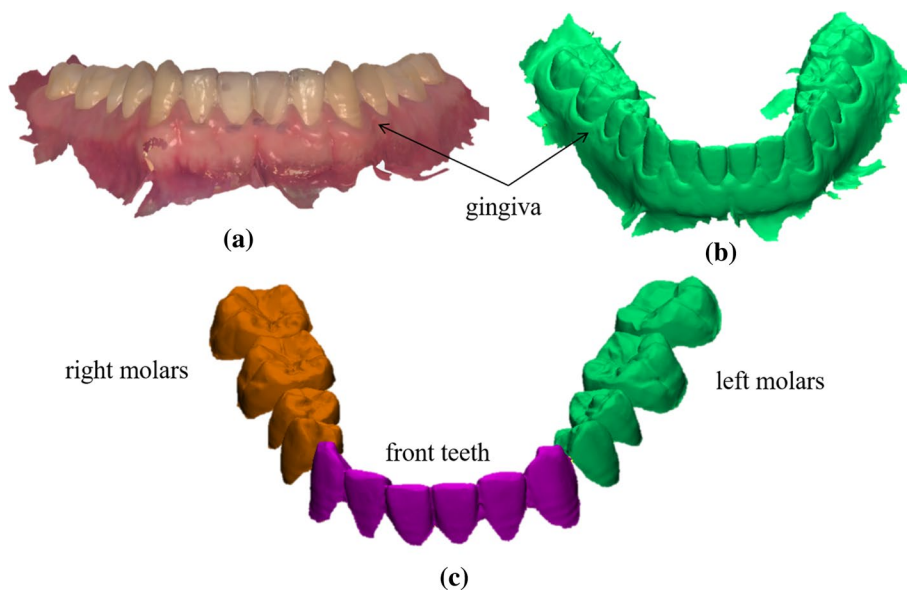


the 3D model of the teeth, as shown in Fig. 3a. Figure 3b shows its underlying 3D mesh model. Because the 3D scanner works in an optical 3D reconstruction manner, the gingiva structure is clearly present. The boundary of the teeth and the gingiva exhibits high curvature characteristic. We can use a semiautomatic tool (e.g., 3-matic or Geomagic) to quickly depict the boundary lines and extract the exposed teeth part (i.e., removing the part covered by gingiva), as shown in Fig. 3c. The 2D projection of the exposed teeth model should coincide well with the camera's view if its pose with respect to the camera is known. By leveraging a 3D scanner, we have addressed the shape inconsistency issue when directly registering the CT-derived model with the stereo camera.

### Stereo camera system

A stereo camera system is developed to track the patient's teeth. Virtual scenes can be rendered and mixed with the camera's stereo views using OpenGL APIs for AR surgical navigation. The spatial relationship between the camera system and the CT space is determined by the image registration procedure so that the CT-derived models can be overlaid on the stereo views correctly to provide a virtual reality mixed scene for surgical guidance. The stereo camera consists of two optical cameras. Stereo camera calibration and stereo rectification are performed to obtain the camera parameters and eliminate the vertical disparity of the stereo images.

**Fig. 3** Model generation from intraoral 3D scanner. **a** 3D model with texture. **b** 3D mesh model. **c** Exposed teeth model



### 3D scan to CT registration

3D scan to CT registration is performed by matching the CT-derived teeth model with the exposed teeth model from the 3D scanner. An ICP algorithm [31] is used to perform the alignment. Before the ICP algorithm is applied, the two models should be coarsely aligned. Observing that the 3D bounding box of the teeth has distinct magnitudes in width, length and height, principle component analysis (PCA) [32] is performed to obtain the three main directions of the teeth model, which is given by the singular value decomposition (SVD) of the zero-mean model data matrix. Together with the gravity center, four-point correspondences can be established between the two models. There are existing algorithms to solve the point correspondence registration problem. After the initial alignment, the ICP algorithm is performed to refine the alignment. The whole registration procedure is performed automatically and only is done once. Once the 3D scan to CT registration is done, the geometry of the exposed teeth model is transformed to the CT space so that its base coordinate system is the same as the CT's. Next, we will match the exposed teeth model (after transformation) with the stereo images of the camera to associate the CT space with the stereo camera's space.

### 3D scan to camera registration

3D scan to camera registration is the process of matching the transformed exposed teeth model (based on the CT space) with the camera's stereo images. Because the camera is looking at the patient's teeth, this process is also the tracking of the patient's teeth so that the virtual models from the CT space can be aligned with the images correctly. The underlying idea is 3D–2D shape matching [33]. Assume there is a virtual stereo camera whose camera parameters are set to be the same as the real ones, a pair of virtual stereo images can be rendered using the graphics APIs (e.g., OpenGL). Given different poses of the model with respect to the virtual camera, the rendered virtual stereo images will change, just like you are looking at the same object from different viewpoints. If we could find a viewpoint from which the rendered virtual stereo images are the most consistent with the real stereo images, the model's pose can be determined from that viewpoint.

#### Problem formulation

Assume  $P_l = K(I, \mathbf{0})$  and  $P_r = K(I, \mathbf{b})$  denote the projection matrix of the left and right cameras, respectively, where  $K$  is the camera's intrinsic matrix,  $I$  is the identity matrix,  $\mathbf{b}$  has a form of  $(-b, 0, 0)^T$  with  $b$  the length of the stereo camera's

baseline. Above parameters are obtained by stereo camera calibration. The task is to solve the following maximization problem:

$$\max_T \frac{1}{2} (s(\text{Proj}(\text{obj}, P_l T), I_l) + s(\text{Proj}(\text{obj}, P_r T), I_r)) \quad (1)$$

where  $\text{Proj}(\text{obj}, P_i T)$ ,  $i = l, r$  is the 2D projected shape of the 3D model  $\text{obj}$  using the projection matrix  $P_i T$ ,  $i = l, r$ ;  $T = (R, t; \mathbf{0}, 1)$  is the pose of the  $\text{obj}$  with respect to the left camera;  $s(\cdot, \cdot)$  is the metric measuring the similarity between the projected 2D shape and the image;  $I_l$  and  $I_r$  represent the left and right image of the camera.

The 2D projected shape of a 3D model consists of a set of edge feature points  $(x_i, y_i)$  with the associated direction vectors  $\mathbf{d}_i$  representing the normal of the 2D shape at  $(x_i, y_i)$ . The edge feature points  $(x_i, y_i)$  are the projection points of the 3D model's edges whose face angles are beyond a certain threshold. The similarity metric  $s$  between a 2D projected shape with  $N$  points and an image  $I$  is defined as follows:

$$s = \frac{1}{N} \sum_{i=1}^N \frac{|\mathbf{d}_i^T \nabla I(x_i, y_i)|}{\|\nabla I(x_i, y_i)\| \|\mathbf{d}_i\|} \quad (2)$$

where  $\nabla I(x_i, y_i)$  represents the image gradient at  $(x_i, y_i)$ . Please refer to our previous work [16] for the implementation details of how to extract the 2D projected shape given a rendered view of a 3D model using OpenGL.

#### Search algorithm

Direct optimization of Eq. (1) is iterative in nature and takes significant time to converge with the state-of-the-art computing hardware because it does not have an analytic form. Instead, a coarse-to-fine scheme is proposed to solve the problem. We first match the model with the left and right images separately to find a good initial pose. The first step casts the problem into the same problem as the one in our previous work [16]. The search strategy is briefly described as follows: Thousands of views of the 3D model are rendered regularly in a spherical coordinate system whose origin is located at the centroid of the 3D model, by setting viewpoints of a virtual camera which has the same intrinsic parameters as the real cameras. The generated views are clustered into aspects according to their mutual similarities. The similarity of two rendered views is calculated by first extracting the 2D projected shapes and then calculating the average absolute value of the dot product of the direction vectors on the overlapped pixels. The aspect here is a cluster of views whose mutual similarities are high. After clustering is finished, the aspect is downsampled to the next higher image pyramid level and the clustering process is repeated. In the online search phase, the aspects on the top

level are searched for in the top level of the image, and all the aspects exceeding the similarity threshold will be added into a candidate list. All candidates are tracked down along the hierarchical image level until reaching the bottom. The aspect with the highest similarity score is the viewpoint we are looking for.

### Refinement algorithm

After the initial matching, the 3D model is coarsely aligned with the stereo images. Next, we will perform pose refinement. Assume the current pose of the model with respect to the left camera is  $T_l = (\mathbf{R}, \mathbf{t}; \mathbf{0}, 1)$ , its pose with respect to the right camera is calculated by  $T_r = (\mathbf{R}, \mathbf{t} + \mathbf{b}; \mathbf{0}, 1)$ . With  $T_l$  and  $T_r$ , we can project the 3D model onto the left and right images as 2D contour shapes, denoted by  $\Gamma^l$  and  $\Gamma^r$ .  $\Gamma^l$  and  $\Gamma^r$  consist of 2D points which are the projections of salient 3D edge points on the model. For each point  $\mathbf{x}_i = (x_i, y_i)^T$  in  $\Gamma$ , the nearest edge point  $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i)^T$  in the image is searched for along the direction of  $\nabla \mathbf{x}_i$ . This can be achieved by first convolving the image with a Laplacian of Gaussian filter and then finding the zero crossing of the filtered image along  $\nabla \mathbf{x}_i$  [34]. Denote by  $\mathbf{X}_i = (X_i, Y_i, Z_i)^T$  the corresponding 3D point of  $\mathbf{x}_i$  on the model; we optimize the following function to update the current pose:

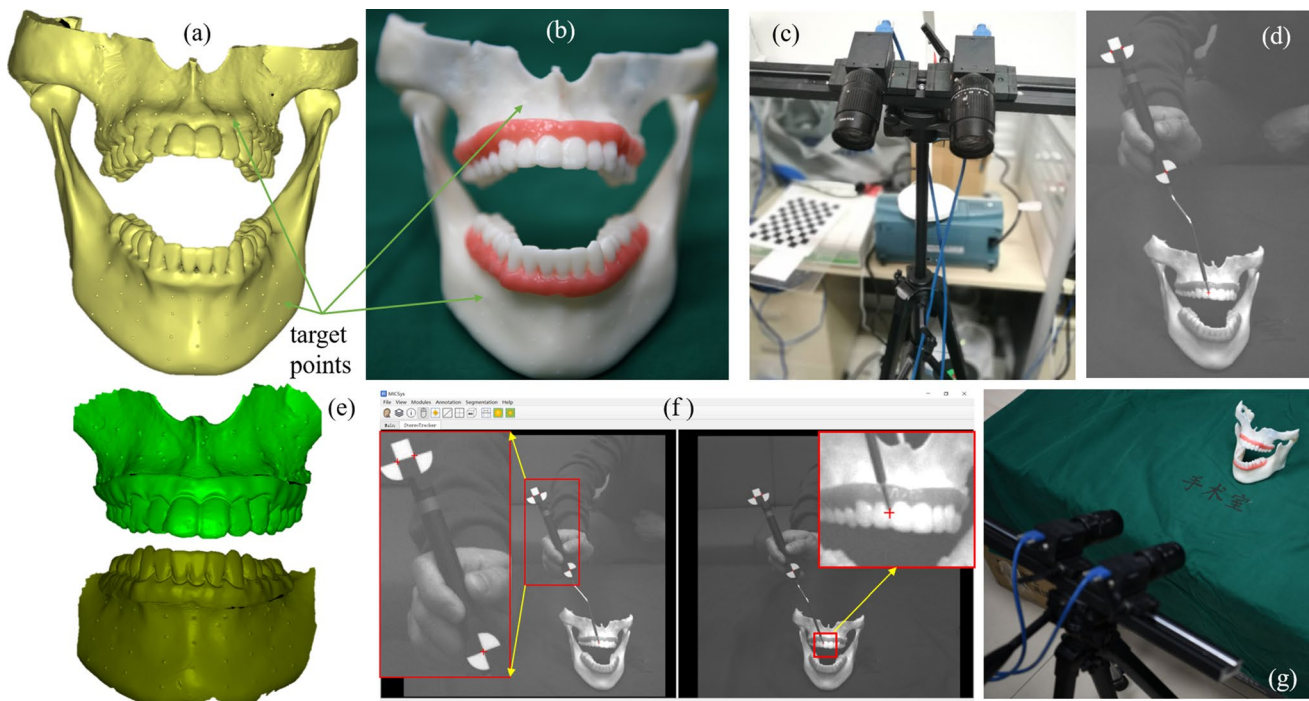
$$\min_{\mathbf{R}, \mathbf{t}} \frac{1}{N_l + N_r} \left( \sum_{\mathbf{x}_i \in \Gamma^l} \text{dist}(\mathbf{K}(\mathbf{R}\mathbf{X}_i^l + \mathbf{t}), \hat{\mathbf{x}}_i^l)^2 + \sum_{\mathbf{x}_i \in \Gamma^r} \text{dist}(\mathbf{K}(\mathbf{R}\mathbf{X}_i^r + \mathbf{t} + \mathbf{b}), \hat{\mathbf{x}}_i^r)^2 \right) \quad (3)$$

where  $\text{dist}(\mathbf{x}, \mathbf{y})$  represents the Euclidean distance between the inhomogeneous coordinates of  $\mathbf{x}$  and  $\mathbf{y}$ ; and  $N_l$  and  $N_r$  are the point number of  $\Gamma^l$  and  $\Gamma^r$ , respectively. Equation (3) can be efficiently solved using the BFGS algorithm [35]. Once the pose is updated, the next iteration is performed until the change of the pose is less than a threshold.

## Experiments and results

### Experimental setup

A jaw resin phantom (including the maxilla and mandible) was fabricated using a 3D printer from a volunteer's CT dataset. Red wax was attached to the phantom to mimic the gingiva. For accuracy evaluation, small solid balls with a radius of 0.5 mm were made on the phantom's



**Fig. 4** Experiments. **a** Jaw CT model (including the maxilla and mandible) with target points. **b** 3D printed phantom. **c** Stereo camera system. **d** Tracked stylus mimicking a surgical tool. **e** Optical 3D

reconstruction of the upper and lower teeth using an intraoral scanner. **f** Software interface. **g** Experimental scene

surface as target points whose 3D coordinates in the CT space are known. Figure 4a shows the CT model with target points, and Fig. 4b shows the picture of the corresponding phantom. Figure 4c shows the stereo camera system which consists of two USB3.0 industrial cameras with the image resolution of  $2048 \times 2048$  (GS3-U3-41C6M-C, FLIR Systems). The length of the baseline was approximately 110 mm. The stereo camera was calibrated and rectified using the OpenCV library. We have developed a GPU-based robust and fast  $x$  point feature detection and localization method based on which the 6DoF pose of a surgical tool can be tracked with the tracking error of less than 0.25 mm at a frame rate of 30 Hz [36]. Figure 4d shows a stylus mimicking the surgical tool being tracked. A commercial intraoral 3D scanner (TRIOS, 3Shape, Denmark) was employed to acquire the optical 3D reconstruction of the upper and lower teeth as shown in Fig. 4e. It took about 5 min to scan the phantom and acquire the 3D model of the exposed teeth. Figure 4f shows our developed software for 3D stereo image registration and AR visualization. Figure 4g shows the experimental scene.

## Experimental protocol

The purpose of the experiment was to evaluate the accuracy of the proposed image registration. To evaluate the accuracy, we projected the target points (small balls) from the CT space to the stereo video stream using the registration matrix. It was expected that the overlaid target points coincided with their counterparts on the phantom. We measured the distance between the real target points on the phantom's surface and the indicated target points by the virtual balls as the target registration error (TRE) using the tracked stylus. The stylus was used to mimic the surgical tool. The final error of AR guidance in real surgery comes from two sources. One is the AR image registration error, and the other is the surgical tool tracking error. Using a stylus to evaluate the TRE can show the most realistic error taking both error sources into account. Choosing either the upper or lower teeth as the registration target, the details of the accuracy evaluation procedure are described as follows.

- The exposed teeth model of the 3D scanner reconstruction was extracted by delineating the boundary curve using the 3-matic software. The extraction was finished within 2 min.
- The exposed teeth model was matched with the corresponding CT model and then transformed to the CT space.
- Since the TRE increases along with the distance to the registration feature, we divided the exposed teeth model

into three parts: front teeth, left molars and right molars. For evaluating the TRE in different areas (i.e., front teeth area, left molars area and right molars area), the nearest corresponding model would be chosen. In our experiment, we chose the front teeth area as the evaluation area.

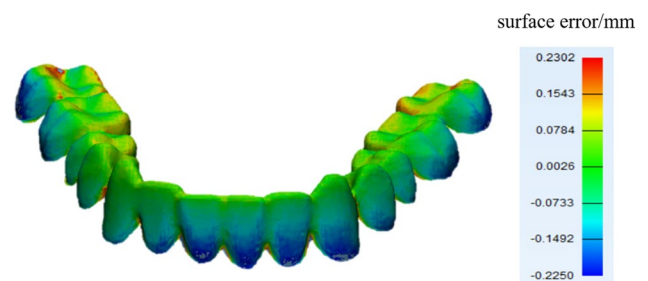
- The stereo camera was used to capture the stereo video of the phantom as shown in Fig. 4g. For each stereo image pair, teeth tracking was performed to outline the bounding box of the teeth part. Within the restricted area, 3D–2D matching was performed to obtain the initial pose of the exposed teeth model. Next, the 3D stereo matching was performed to obtain the refined pose with respect to the stereo camera.
- The target balls were overlaid on the stereo views of the camera using the registration matrix. The TRE was measured on each target position using the tracked stylus shown in Fig. 4d.

## Experimental results

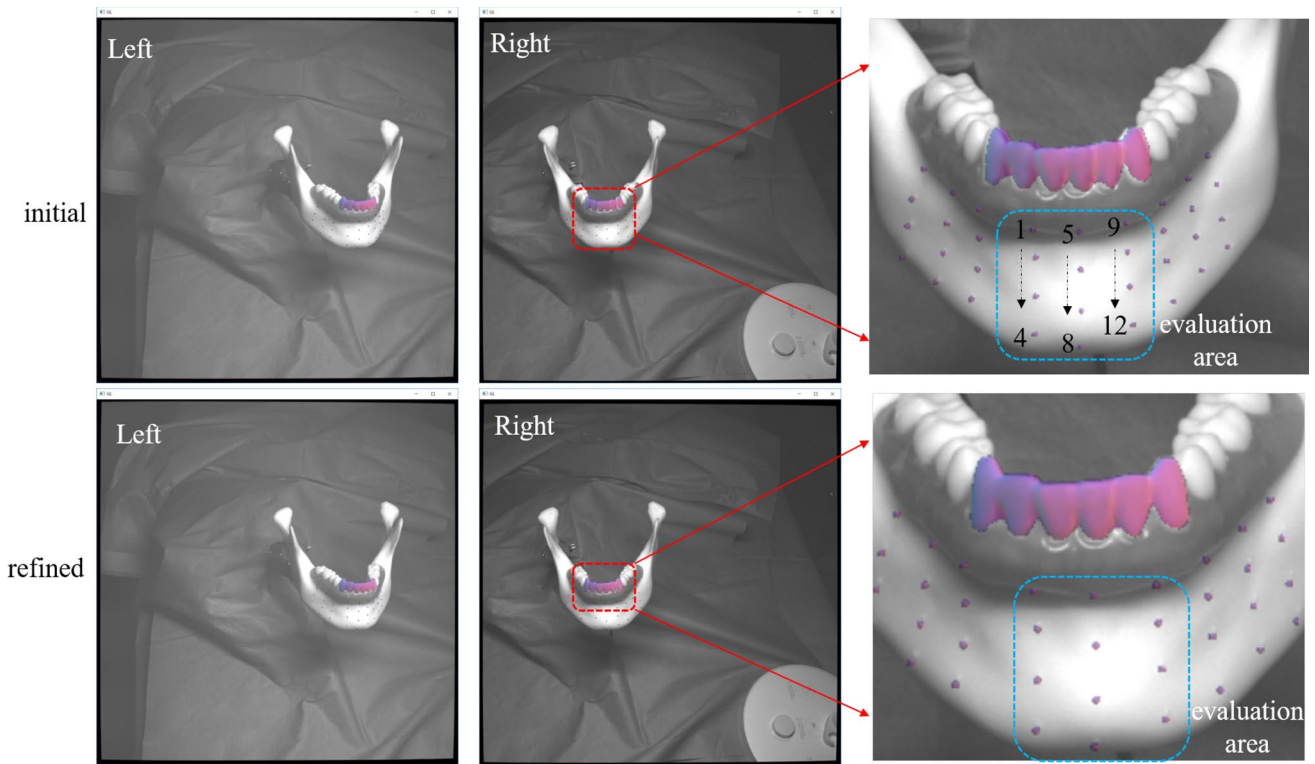
### Mandible

We first chose the lower teeth as the registration target and evaluated the TRE on the mandible. The 3D scan to CT registration yielded a maximum registration error of 0.23 mm. Figure 5 shows the surface error distribution of the exposed teeth model after the registration. The most part of the surface had very small alignment error.

The results of 3D scan to camera registration using the front teeth model are shown in Fig. 6. It took about 0.1 s to finish the initial matching and another 0.3 s to finish the 3D stereo matching. The first row of Fig. 6 shows the initial matching result. After the pose refinement, the matching accuracy was obviously improved which is indicated by the rightmost picture of the second row in the figure. The registration was repeated by ten times, and the TRE was measured in the evaluation area. The results are shown in Table 1. An average error of 0.42 mm was obtained.



**Fig. 5** 3D scan to CT registration error of the lower exposed teeth model

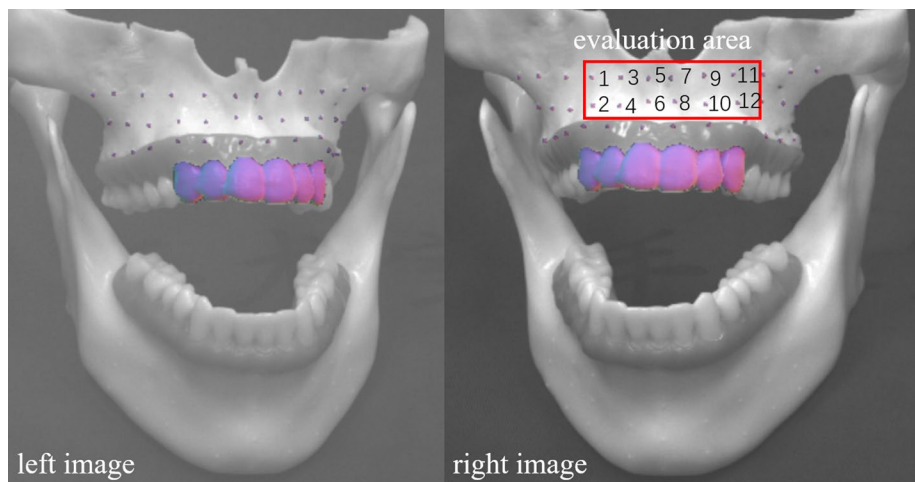


**Fig. 6** Evaluation of 3D scan to camera registration using lower teeth model

**Table 1** Target registration error on mandible

Index	1	2	3	4	5	6	7	8	9	10	11	12
Error/mm	0.32	0.51	0.43	0.66	0.25	0.31	0.39	0.71	0.24	0.26	0.31	0.63

**Fig. 7** Evaluation of 3D scan to camera registration using upper teeth model



**Table 2** Target registration error on maxilla

Index	1	2	3	4	5	6	7	8	9	10	11	12
Error/mm	0.41	0.35	0.33	0.30	0.38	0.37	0.39	0.33	0.34	0.32	0.43	0.34



## Maxilla

We next chose the upper teeth as the registration target and evaluated the TRE on the maxilla. The 3D scan to CT registration yielded a similar result with that of the mandible. The results of 3D scan to camera registration using the front teeth model are shown in Fig. 7, and the TREs are given in Table 2. An average error of 0.36 mm was obtained.

## Volunteer trial

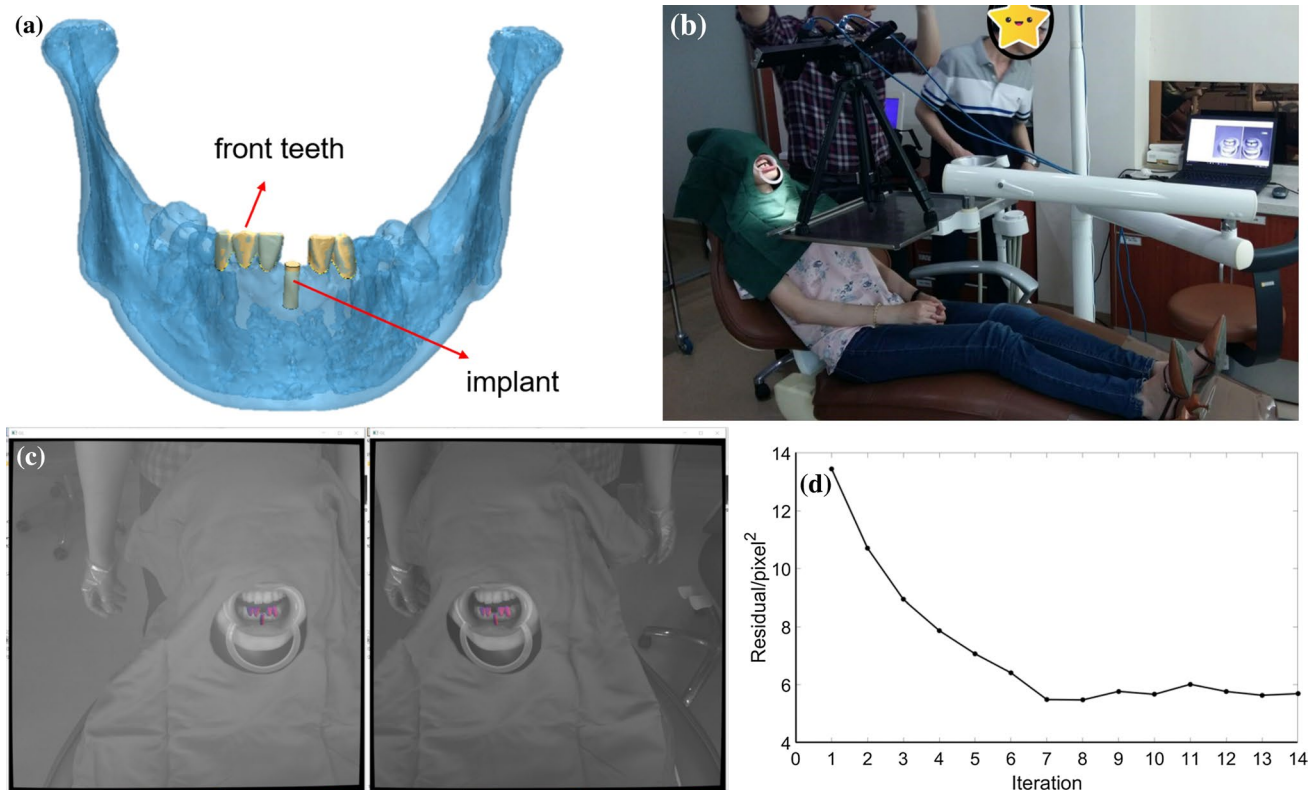
We also performed experiments on a volunteer to confirm the practical feasibility of the method. A woman who suffered from front tooth loss agreed to be our volunteer. The aim of the surgery was to place a dental implant in her lower jaw. With the proposed image registration method, the procedure could be guided using AR. The surgical planning results including the implant's position and orientation were overlaid on the stereo views of the camera. The surgeon was able to operate, avoid critical structures and confirm surgical outcomes with the help of the AR scene. In addition, the stereo AR views provided stereo parallax which gave the surgeon more accurate spatial perception compared with a single view. Figure 8 shows the experimental scene and AR scene of the volunteer

trial. Because the proposed image registration method does not introduce invasiveness to the patient, we tested the AR guidance before the clinical procedure on the volunteer. In the following real surgery, the patient was fully disinfected and covered with sterile drapes. Sterile operation was strictly applied.

## Conclusion and discussion

The proposed registration method has adequate accuracy and does not rely on any external fiducial markers attached to the patient. It performs automatically so as to maintain a correct AR scene, overcoming the misalignment difficulty caused by patient's movement. Therefore, it is noninvasive and practical in oral and maxillofacial surgery and may achieve a seamless integration between computer-aided surgical simulation and the interventional procedure. The stereo camera can also serve as a tracking device for measuring the pose of the surgical instrument, leading to a virtual reality surgical navigation paradigm where the instrument is visualized with respect to the patient's anatomy.

Because the teeth have less texture and only occupy a small portion of the camera view, it is very challenging to reconstruct only the exposed 3D surface of the teeth in real



**Fig. 8** Volunteer trial. **a** Surgical planning. **b** Experimental scene. **c** Stereo AR views. **d** Loss function [value of Eq. (3)] improving along with the iteration number

time without background noise using either structured light or stereo vision. Therefore, we propose to match the 3D teeth model with the stereo video stream (left and right images) directly. Experiments were performed to evaluate the registration accuracy which was less than 0.5 mm. Note that the TRE was measured on the phantom surface rather than in the free 3D space. It is meaningful because in the real surgery, the surgeon will find the entry points on the patient's jaw according to the AR guidance. The entry points are indeed planned on the surface of the jaw, rather than in a free 3D space. On the real-time performance of the registration, at present it takes approximately 0.5 s to finish the registration. The time cost could be further reduced by using a GPU.

One concern may be about the artifacts in CT images induced by metal implants for some patients. The artifacts bring challenges to image segmentation. For modern advanced CT machines, effective artifacts reduction algorithms (such as single-energy metal artifact reduction) have already been integrated; therefore, we can take the scanning protocol that can suppress metal artifacts when taking the CT images. For residual light artifacts, in the stage of pre-operative model generation it is possible to refine the segmentation by manual delineation. For very severe artifacts that make a large portion of teeth of interest unrecognizable, this will lead to an inaccurate teeth mesh model which may adversely affect the image registration. The proposed method in this paper is supposed to be used in AR-guided oral and maxillofacial surgery for those patients who have small number of missing teeth with moderate metal artifacts.

In the future work, we will develop a navigation system for oral and maxillofacial surgery by integrating the proposed registration method. The system will fill the gap between the CAD/CAM-based surgical planning and the intraoperative guidance. Just like the car navigation where the GPS satellites should see the car, in our navigation system the only requirement to generate correct AR scenes is to expose the patient's teeth to the camera.

**Funding** This work was partially supported by National Natural Science Foundation of China (Grant No. 61701014).

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of our institutional review board and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

### References

- Kamphuis C, Barsom E, Schijven M, Christoph N (2014) Augmented reality in medical education? *Perspect Med Educ* 3(4):300–311
- Loukas C, Lahanas V, Georgiou E (2013) An integrated approach to endoscopic instrument tracking for augmented reality applications in surgical simulation training. *Int J Med Robot Comput Assist Surg* 9(4):e34–e51
- Rhiemora P, Gajananan K, Haddawy P, Dailey MN, Suebnukarn S (2010) Augmented reality haptics system for dental surgical skills training. In: *ACM symposium on virtual reality software and technology*, 2010, pp 97–98
- Edwards PJ, Johnson LG, Hawkes DJ, Fenlon MR, Strong AJ, Gleeson MJ (2004) Clinical experience and perception in stereo augmented reality surgical navigation. In: Yang G-Z, Jiang T-Z (eds) *Medical imaging and augmented reality*, Berlin, Heidelberg, Springer, Berlin, pp 369–376
- Fallavollita P, Wang L, Weidert S, Navab N (2016) Augmented reality in orthopaedic interventions and education. In: Zheng G, Li S (eds) *Computational radiology for orthopaedic interventions*. Springer, Cham, pp 251–269. [https://doi.org/10.1007/978-3-319-23482-3\\_13](https://doi.org/10.1007/978-3-319-23482-3_13)
- Liao H, Inomata T, Sakuma I, Dohi T (2010) 3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay. *IEEE Trans Biomed Eng* 57(6):1476–1486
- Tardif JP, Roy S, Meunier J (2003) Projector-based augmented reality in surgery without calibration. In: *Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439)*, 17–21 Sept. 2003, vol 541, pp 548–551. <https://doi.org/10.1109/iembs.2003.1279797>
- Wang J, Suenaga H, Hoshi K, Yang L, Kobayashi E, Sakuma I, Liao H (2014) 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. <https://doi.org/10.1109/TBME.2014.2301191>
- Wang J, Suenaga H, Liao H, Hoshi K, Yang L, Kobayashi E, Sakuma I (2015) Real-time computer-generated integral imaging and 3D image calibration for augmented reality surgical navigation. *Comput Med Imaging Graph* 40:147–159. <https://doi.org/10.1016/j.compmedimag.2014.11.003>
- Liao H, Ishihara H, Tran HH, Masamune K, Sakuma I, Dohi T (2010) Precision-guided surgical navigation system using laser guidance and 3D autostereoscopic image overlay. *Comput Med Imaging Graph* 34(1):46–54. <https://doi.org/10.1016/j.compmedimag.2009.07.003>
- Kiyokawa K, Kurata Y, Ohno H (2001) An optical see-through display for mutual occlusion with a real-time stereovision system. *Comput Graph* 25(5):765–779. [https://doi.org/10.1016/S0097-8493\(01\)00119-4](https://doi.org/10.1016/S0097-8493(01)00119-4)
- Qian L, Barthel A, Johnson A, Osgood G, Kazanzides P, Navab N, Fuerst B (2017) Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display. *Int J Comput Assist Radiol Surg* 12(6):901–910. <https://doi.org/10.1007/s11548-017-1564-y>
- Rodrigues DG, Jain A, Rick SR, Shangley L, Suresh P, Weibel N (2017) Exploring mixed reality in specialized surgical environments. Paper presented at the proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems, Denver, Colorado, USA
- Chen X, Xu L, Wang Y, Wang H, Wang F, Zeng X, Wang Q, Egger J (2015) Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted

- display. *J Biomed Inform* 55:124–131. <https://doi.org/10.1016/j.jbi.2015.04.003>
15. Teber D, Guven S, Simpfendorfer T, Baumhauer M, Güven EO, Yencilek F, Gözen AS, Rassweiler J (2009) Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results. *Eur Urol* 56(2):332–338. <https://doi.org/10.1016/j.eururo.2009.05.017>
  16. Wang J, Suenaga H, Yang L, Kobayashi E, Sakuma I, Wang J, Suenaga H, Yang L, Kobayashi E, Sakuma I (2017) Video see-through augmented reality for oral and maxillofacial surgery. *Int J Med Robot Comput Assist Surg* 13(2):e1754–e1767. <https://doi.org/10.1002/rcs.1754>
  17. Su L-M, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor RH, Hager GD (2009) Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration. *Urology* 73(4):896–900. <https://doi.org/10.1016/j.urology.2008.11.040>
  18. Volonte F, Fo Pugin, Bucher P, Sugimoto M, Ratib O, Morel P (2011) Augmented reality and image overlay navigation with OsiriX in laparoscopic and robotic surgery: not only a matter of fashion. *J Hepato-Biliary-Pancreat Sci* 18(4):506–509. <https://doi.org/10.1007/s00534-011-0385-6>
  19. Krempien R, Hoppe H, Kahrs L, Daeuber S, Schorr O, Eggers G, Bischof M, Munter MW, Debus J, Harms W (2008) Projector-based augmented reality for intuitive intraoperative guidance in image-guided 3D interstitial brachytherapy. *Int J Radiat Oncol Biol Phys* 70(3):944–952. <https://doi.org/10.1016/j.ijrobp.2007.10.048>
  20. Zitová B, Flusser J (2003) Image registration methods: a survey. *Image Vis Comput* 21(11):977–1000. [https://doi.org/10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9)
  21. Souzaki R, Ieiri S, Uemura M, Ohuchida K, Tomikawa M, Kinoshita Y, Koga Y, Suminoe A, Kohashi K, Oda Y, Hara T, Hashizume M, Taguchi T (2013) An augmented reality navigation system for pediatric oncologic surgery based on preoperative CT and MRI images. *J Pediatr Surg* 48(12):2479–2483. <https://doi.org/10.1016/j.jpedsurg.2013.08.025>
  22. Puerto-Souza GA, Mariottini GL (2013) Toward long-term and accurate augmented-reality display for minimally-invasive surgery. In: 2013 IEEE international conference on robotics and automation, 6–10 May 2013, pp 5384–5389. <https://doi.org/10.1109/icra.2013.6631349>
  23. Stoyanov D, Darzi A, Yang GZ (2010) A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery. *Comput Aided Surg* 10(4):199–208. <https://doi.org/10.3109/10929080500230379>
  24. Chang P-L, Stoyanov D, Davison AJ, Edwards PE (2013) Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery. In: medical image computing and computer-assisted intervention—MICCAI 2013, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg, pp 42–49
  25. Totz J, Thompson S, Stoyanov D, Gurusamy K, Davidson BR, Hawkes DJ, Clarkson MJ (2014) Fast semi-dense surface reconstruction from stereoscopic video in laparoscopic surgery. In: Information processing in computer-assisted interventions, 2014. Springer, Cham, pp 206–215
  26. Bouchard C, Magill JC, Nikonovskiy V, Byl M, Murphy BA, Kaban LB, Troulis MJ (2012) Osteomark: a surgical navigation system for oral and maxillofacial surgery. *Int J Oral Maxillofac Surg* 41(2):265–270. <https://doi.org/10.1016/j.ijom.2011.10.017>
  27. Casap N, Nadel S, Tarazi E, Weiss EI (2011) Evaluation of a navigation system for dental implantation as a tool to train novice dental practitioners. *J Oral Maxillofac Surg* 69(10):2548–2556. <https://doi.org/10.1016/j.joms.2011.04.026>
  28. Venosta D, Sun Y, Matthews F, Kruse AL, Lanzer M, Gander T, Grätz KW, Lübbers H-T (2014) Evaluation of two dental registration-splint techniques for surgical navigation in cranio-maxillofacial surgery. *J Cranio-Maxillofac Surg* 42(5):448–453. <https://doi.org/10.1016/j.jcms.2013.05.040>
  29. Badiali G, Ferrari V, Cutolo F, Freschi C, Caramella D, Bianchi A, Marchetti C (2014) Augmented reality as an aid in maxillofacial surgery: validation of a wearable system allowing maxillary repositioning. *J Cranio-Maxillofac Surg* 42(8):1970–1976. <https://doi.org/10.1016/j.jcms.2014.09.001>
  30. Wang J, Suenaga H, Yang L, Liao H, Ando T, Kobayashi E, Sakuma I (2015) 3D surgical overlay with markerless image registration using a single camera. In: Linte CA, Yaniv Z, Fallavollita P (eds) Augmented environments for computer-assisted interventions. Springer, Cham, pp 124–133
  31. Yang J, Li H, Campbell D, Jia Y (2016) Go-ICP: a globally optimal solution to 3D ICP point-set registration. *IEEE Trans Pattern Anal Mach Intell* 38(11):2241–2254. <https://doi.org/10.1109/TPAMI.2015.2513405>
  32. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometr Intell Lab Syst* 2(1):37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
  33. Ulrich M, Wiedemann C, Steger C (2012) Combining scale-space and similarity-based aspect graphs for fast 3D object recognition. *IEEE Trans Pattern Anal Mach Intell* 34(10):1902–1914. <https://doi.org/10.1109/TPAMI.2011.266>
  34. Wang J, Kobayashi E, Sakuma I (2015) Coarse-to-fine dot array marker detection with accurate edge localization for stereo visual tracking. *Biomed Signal Process Control* 15:49–59. <https://doi.org/10.1016/j.bspc.2014.09.008>
  35. Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. *Math Program* 45(1):503–528. <https://doi.org/10.1007/bf01589116>
  36. Wang J, Ji X, Zhang X, Sun Z, Wang T (2018) Real-time robust individual X point localization for stereoscopic tracking. *Pattern Recognit Lett* 112:138–144. <https://doi.org/10.1016/j.patrec.2018.07.002>

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