IEEE Journal of Translational Engineering in Health and Medicine

Received 25 August 2023; revised 19 October 2023 and 8 November 2023; accepted 8 November 2023.
Date of publication 13 November 2023; date of current version 1 February 2024.

Digital Object Identifier 10.1109/JTEHM.2023.3332088

Registration Sanity Check for AR-guided Surgical Interventions: Experience From Head and Face Surgery

SARA CONDINO[1], FABRIZIO CUTOLO[1], MARINA CARBONE[1], LAURA CERCENELLI[2], GIOVANNI BADIALI[1], NICOLA MONTEMURRO[3], AND VINCENZO FERRARI[1], (Member, IEEE)

[1]Department of Information Engineering, University of Pisa, 56126 Pisa, Italy
[2]eDIMES Laboratory of Bioengineering, Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40138 Bologna, Italy
[3]Department of Neurosurgery, Azienda Ospedaliera Universitaria Pisana (AOUP), 56127 Pisa, Italy

CORRESPONDING AUTHOR: F. CUTOLO (fabrizio.cutolo@unipi.it)

This work was supported in part by the HORIZON2020 Project Video-Optical See-Through Augmented Reality (AR) surgical System (VOSTARS) Call: ICT-29-2016 Photonics Key Enabling Technologies (KET) 2016 under Project 731974; in part by the Next Generation European Union (EU) Project through Ecosistema dell'Innovazione Tuscany Health Ecosystem (The Piano Nazionale di Ripresa e Resilienza (PNRR), Spoke 9: Robotics and Automation for Health) under Grant ECS00000017; and in part by the Italian Ministry of Education and Research (MUR) in the Framework of the FoReLab and CrossLab Projects (Departments of Excellence).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Area Vasta Emilia Centro under Approval Nos: 3749/2019 and 22/10/2019 and notified to the Italian Ministry of Health in December 2019.

This article has supplementary downloadable material available at https://doi.org/10.1109/JTEHM.2023.3332088, provided by the authors.

ABSTRACT Achieving and maintaining proper image registration accuracy is an open challenge of image-guided surgery. This work explores and assesses the efficacy of a registration sanity check method for augmented reality-guided navigation (AR-RSC), based on the visual inspection of virtual 3D models of landmarks. We analyze the AR-RSC sensitivity and specificity by recruiting 36 subjects to assess the registration accuracy of a set of 114 AR images generated from camera images acquired during an AR-guided orthognathic intervention. Translational or rotational errors of known magnitude up to ± 1.5 mm/$\pm 15.5°$, were artificially added to the image set in order to simulate different registration errors. This study analyses the performance of AR-RSC when varying (1) the virtual models selected for misalignment evaluation (*e.g.*, the model of brackets, incisor teeth, and gingival margins in our experiment), (2) the type (translation/rotation) of registration error, and (3) the level of user experience in using AR technologies. Results show that: 1) the sensitivity and specificity of the AR-RSC depends on the virtual models (globally, a median true positive rate of up to 79.2% was reached with brackets, and a median true negative rate of up to 64.3% with incisor teeth), 2) there are error components that are more difficult to identify visually, 3) the level of user experience does not affect the method. In conclusion, the proposed AR-RSC, tested also in the operating room, could represent an efficient method to monitor and optimize the registration accuracy during the intervention, but special attention should be paid to the selection of the AR data chosen for the visual inspection of the registration accuracy.

INDEX TERMS Augmented reality, computer-assisted surgery, image-to-patient registration, sanity check.

I. INTRODUCTION

RECENTLY, there has been a growing interest in using augmented reality (AR) as a navigation tool for image-guided surgery [1], [2], [3]. In conventional image-based virtual reality (VR) navigators, the real-time guidance data are rendered in a virtual scene after it has been spatially aligned to the patient's anatomy (*i.e.*, registered). The guidance information consists of simplified 3D reconstructions of the patient's anatomical structures derived from preoperative images, and it comprises virtual navigation aids associated with the pose of the surgical tool tracked in real-time.

The biggest benefit of AR-based tools over conventional VR surgical navigators is that the surgeon can visualize the guidance information directly onto the patient's body

IEEE Journal of Translational Engineering in Health and Medicine

(*i.e.*, in situ visualization) [4], [5], and anatomy-related data are inherently not simplistic. In both cases (*i.e.*, VR and AR navigators), the accuracy and the safety of the intervention rely on the efficacy of the registration procedure. Generally speaking, registration is "*the determination of a geometrical transformation that aligns one view of an object with another, where a view can be an image [...] but it can also be the physical object itself*" [6]. In image-guided surgery, the two views comprise a radiological image (*e.g.*, a segmented CT or MRI) and some part of the patient's anatomy, with the resulting registration, referred to as "image-to-patient" registration [6].

If the registration is not performed correctly and/or is not accurate enough, the guidance information (*e.g.*, virtual anatomical models, planned cutting lines, drilling/biopsy trajectories) can be misleading and even dangerous for the patient. Indeed, it may distort and adversely affect the visuospatial targeting and the spatial reasoning ability of the surgeon throughout the procedure: for example, the surgeon may fail to recognize important anatomical structures or tumor margins surrounding the surgical target, affecting the outcomes and the morbidity of the surgery.

A common strategy adopted to check the registration accuracy in traditional VR navigators requires the surgeon to "touch" with a tracked probe some univocally identifiable anatomical landmarks, preferably close to the region of interest (ROI), while the relative positioning of the probe and the anatomical model is also checked on the navigator screen. The distance between the probe tip and the anatomical landmark displayed on the navigator screen provides an immediate estimate of the registration quality in terms of target registration error (TRE) at that point. The TRE tends to vary slowly within a specific region of interest (ROI), and a check of at least three not colinear points is an adequate validation mean to detect any rotation errors around the surgical target. If the TRE is too high, then the registration may be wrong, and it should be repeated; otherwise, the system is likely to work correctly [6], [7].

As the registration can degrade during the intervention (*e.g.*, due to involuntary movements of the patient and/or of the optical markers pinned to the anatomy) surgeons should repeat this control procedure, which we will henceforth call registration "sanity check" (RSC), several times on each landmark. However, RSC is usually performed only once (immediately after registration) and using a single target. Indeed, finding unambiguously identifiable landmarks can be challenging, and repeating the check several times can result in an excessive increase in surgical time.

Monitoring and optimizing the registration accuracy during the intervention execution is an open challenge. Albeit many attempts of using intra-operative imaging to predict anatomy displacements for rigid [8], [9], [10] and non-rigid structures [11], [12], [13] have been proposed, "*simpler methods that aim to give the surgeon control over the registration not only at the beginning of surgery but also during surgery, have not fully been explored*" [14]. However, in navigators based on AR, the RSC can be particularly intuitive and straightforward. Indeed, the virtual information is overlaid directly onto the patient's anatomy, and the surgeon can perform an immediate visual validation of the alignment between the virtual content with its physical counterpart, thus avoiding the unsafe eyesight shift away from the patient. This allows a real-time check of the registration quality, potentially on many anatomical landmarks simultaneously, with a reduced impact on the surgical workflow. If errors occur, the surgeon may employ a registration correction strategy, such as using a tracked probe to reacquire an intra-operative point cloud for refining the initial registration and correcting the mismatch between the real and virtual contents. For example, Drouin et al. [15] describe a simple method to improve AR overlay in neurosurgery by tracing curves along the surface of exposed vessels using a tracked probe. The proposed method allows for correction of the initial registration that may have degraded due to draping, attachment of skin retractors, "brain shift" etc.

Defining a fast and effective strategy for visually monitoring the registration error is a key issue in the design of a navigation system. According to Morienau et al. [16], a user-oriented approach is needed to support the design of the optimal data (*i.e.*, content type and amount) displayed by an image-guided surgical system to optimize the actual information perceived by operators according to their level of knowledge. Rather than visualizing the full virtual anatomical models, one method explored in the literature for the visual assessment of the AR overlay relies on the use of model contours.

Thompson et al. [17] propose a method for quantitative in-vivo estimation of the registration error during laparoscopic AR surgery based on the visible misalignment of exposed organ contours (*e.g.*, liver contours) to infer the misalignment of hidden anatomical targets. Another example of using contours in AR navigation is the study by Amir-Khalili et al. [18] which describes a framework helping the surgeon in localizing excision margins in robot-assisted interventions. More specifically, they create an uncertainty-encoded AR view considering shape boundary uncertainties in the segmentation of the pre-operative CT (*i.e.*, kidney and tumor surfaces in their experiment) to be overlaid on stereo-endoscopy. Moreover, they verify the registration outcome by visualizing the contour of the projected mesh on both left and right endoscopic views.

The use of object contours has also been proposed for applications outside of surgery. According to [19] "*the effects of registration errors can be mitigated through the development of adaptive user interfaces that tailor the information display as a function of registration errors*". More specifically, they propose a statistical method for estimating the registration error and show how these estimates can be used in AR interfaces. According to their study, a compelling method of highlighting a convex object in AR is to render

a convex hull that can be expanded or shrunk based on the registration error estimate. Using the expanded hull, one can get a highlight that encloses the region where the object is located, even if it lacks precision. Similarly, using a restricted hull, one can get a region that certainly contains a part of the object. Selecting the best-highlighting strategy (normal, contracted, expanded, or a combination of them) depends on the density of the targets and the specific application.

Another simple option for visually estimating the registration error is to display virtual spheres at anatomical landmarks that can be easily identified by the surgeon both in preoperative images and during surgery. In [20] for example, the authors have selected seven anatomical landmarks (*i.e.*, the canthi of the eyes, the pronasal point at the anterior apex of the nose, the subnasal point, and the trago (ear)) for the evaluation of a wearable AR platform for neurosurgery based on patient-specific templates that allow for fast, non-invasive, and fully automatic planning-to-patient registration. The rationale is to display AR spheres at the aforementioned facial anatomical landmarks allowing the surgeon to perform a straightforward visual estimate of the template placement from different viewpoints. The AR spheres can also help the surgeon in optimizing the registration as they provide a reference to adjust the positioning of the template until the AR spheres appear perfectly aligned with the corresponding anatomical landmarks. The nasion, top of the nose, and eyelids are used as anatomical landmarks in the clinical trial reported in [21] concerning a tablet-based AR system for neurosurgical guidance. In that work, the edge of the virtual scalp, together with the aforementioned landmarks were used to evaluate the AR accuracy overlay.

In conclusion, although some studies have proposed methods for verifying RSC in the operating room using AR, the literature is lacking both criteria for establishing standard methods for performing efficient RSC, and also studies on the effectiveness of the procedures proposed to date. In particular, in state-of-the-art works, the AR data are commonly selected by the authors based on their previous experiences and no studies have been carried out, to the best of the authors' knowledge, on the influence of the type of AR data on the registration results.

This paper concerns methods for qualitative visual estimation of registration in AR-guided surgery and focuses on the following fundamental and not yet answered questions:

1) Q1 - Does the AR-RSC performances depend on the type of AR data (*e.g.*, object contours, solid objects) selected for the misalignment assessment? That is, are there optimal data that allow for finer identification of registration errors (*i.e.*, smaller errors)?
2) Q2 - Do the AR-RSC performances vary with the type (translation/rotation) of the registration error? Are there components of the registration error, among the three translations and three rotations, that are more difficult to identify?
3) Q3 - Do the AR-RSC performances depend on the user's level of experience in using AR technologies?

In an attempt to answer these questions, we performed a user study based on the visualization of AR scenes created from images acquired during an experimental maxillofacial surgery procedure guided by a proof-of-concept AR Head-Mounted Display (HMD) for surgery (*i.e.*, the VOSTARS system) [22].

II. MATERIALS AND METHODS

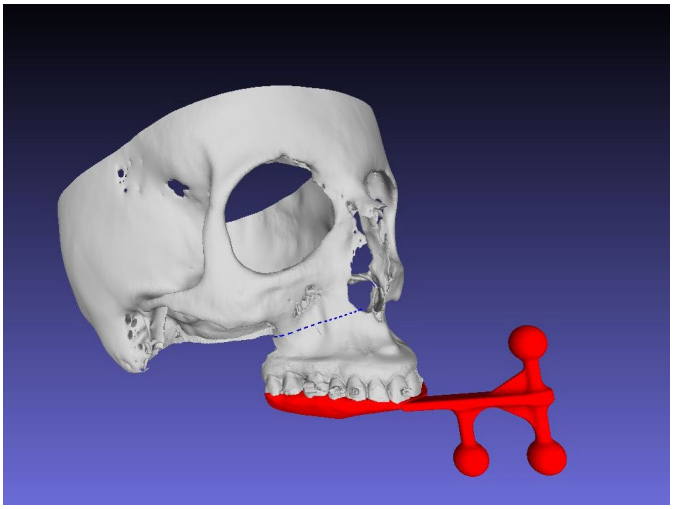
A. DESCRIPTION OF THE INTERVENTION SELECTED FOR THIS STUDY

The use case selected for this study is an experimental Le Fort1 osteotomy, guided by the VOSTARS HMD. The Le Fort1 osteotomy is one of the most common surgical procedures to correct dentofacial deformities and it involves the cutting of the maxilla along a preoperatively defined trajectory (Figure 1a). The VOSTARS system is a custom-made hybrid video and optical see-through AR HMD (Figure 1b) paired with a dedicated and non-distributed software framework, specifically conceived for running AR applications for surgical guidance by supporting in situ visualization of the surgical plan and other medical data (*e.g.*, 3D virtual models of the patient anatomy, or vital signs) [22], [23], [24].

In the video see-through (VST) modality, the VOSTARS platform offers an accurate registration between digital and real data; hence, this modality is strongly recommended for guiding high-precision surgical tasks requiring sub-millimetric accuracy, such as Le Fort1 osteotomies. The system was tested in-vitro and in-vivo, and the results of a seven-patient clinical trial show that when using VOSTARS for VST-guided Le Fort1 procedures, on average, 86% of the osteotomy length falls within ± 0.5 mm accuracy [23]. The image-patient registration strategy implemented to guide the Le Fort1 intervention is based on the use of a patient-specific occlusal splint incorporating three spherical markers (Figure 4b), which can be optically tracked by the VOSTARS system. The VOSTARS system has a dedicated inside-out optical tracking mechanism: the head-anchored RGB cameras used to implement the VST augmentation also allow the stereo localization of the spherical markers. The virtual 3D planning (*i.e.*, the osteotomy trajectory) and the occlusal splint are designed based on preoperative CT images, and the positions of the three markers are known in the CT dataset reference system. Thus, by computing the position of the three markers with respect to the HMD, the registration matrix can be derived in a closed-form fashion through the estimation of the rigid transformation that aligns the two sets of corresponding triplets of 3D points.

Therefore, the patient-specific and trackable occlusal splint acts as a registration template and allows us to skip the preoperative manual registration procedure. Nevertheless, the registration accuracy at the beginning and during the navigated procedure strongly relies on the initial placement of the splint on the patient's teeth and on its stability over time. For this reason, to monitor the effectiveness of the registration during the procedure [23], we developed an AR-RSC

IEEE Journal of Translational Engineering in Health and Medicine



(a) Example of Le Fort1 osteotomy (blue dotted line) planned for a patient recruited in the VOSTARS clinical trial. The 3D model of the designed occlusal splint is represented in red.

(b) The VOSTARS HMD prototype.

FIGURE 1. VOSTARS navigation system and surgical planning for LeFort 1 osteotomy.

modality and used it in clinical trials. In this regard, a video of the system at work during surgery is enclosed in the present paper as supplementary material.

B. GENERATION OF AR IMAGES

Three AR images are extracted from a video captured by the VOSTARS HMD in the operating room during one of the above-mentioned clinical trials. The three images were selected by a panel of three engineers experienced in AR among frames for which the estimated tracking error was less than 0.5 mm. All three engineers rated positively the alignment of the AR content with its physical counterpart. More in particular, we selected three images acquired by the left camera of the VOSTARS HMD and stored without the AR content. The following 3D models were selected as reasonable VR content for the RSC: incisor teeth, their gingival margins, and orthodontic brackets (commonly affixed on the patient's teeth before this kind of intervention) (Figure 2). The choice of the reference landmarks is dictated by both the specific surgical task, which in our case is the Le Fort1 osteotomy of the upper maxilla bone, and by the visibility of the specific anatomical structures involved in the procedure. Interesting examples of other landmarks used in the literature for the estimation of registration errors during AR navigation in maxillofacial surgery and dental implant placement are reported in [25], [26], [27], and [28].

The extrinsic parameters computed by the VOSTARS platform, and representing the estimated pose of the anatomy relative to the camera were saved for each of the three AR images. The extrinsic parameters, together with the three real images, the 3D models of the VR content, and the HMD cameras' intrinsic parameters (representing the camera

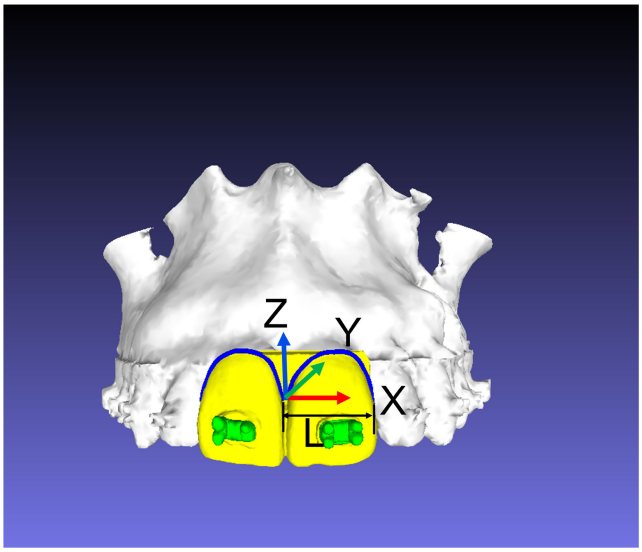


FIGURE 2. VR models selected for the RSC (incisor teeth in yellow, gingival margins in blue, and orthodontic brackets in green) and the reference system used to express the error components.

projection model), were imported by a Unity3D AR application to generate a new dedicated set of AR images for the specific purpose of this work. The Unity3D application was developed to iteratively generate AR images for each of the three frames and each of the three virtual models (incisor teeth, gingival margins, and brackets). The generation of an AR scene requires the configuration of a virtual camera using the linear part of the intrinsic parameters of the corresponding real camera to obtain the same projection model and guarantee the virtual–to–real matching. To do this in Unity, we used the "Physical Camera" component that

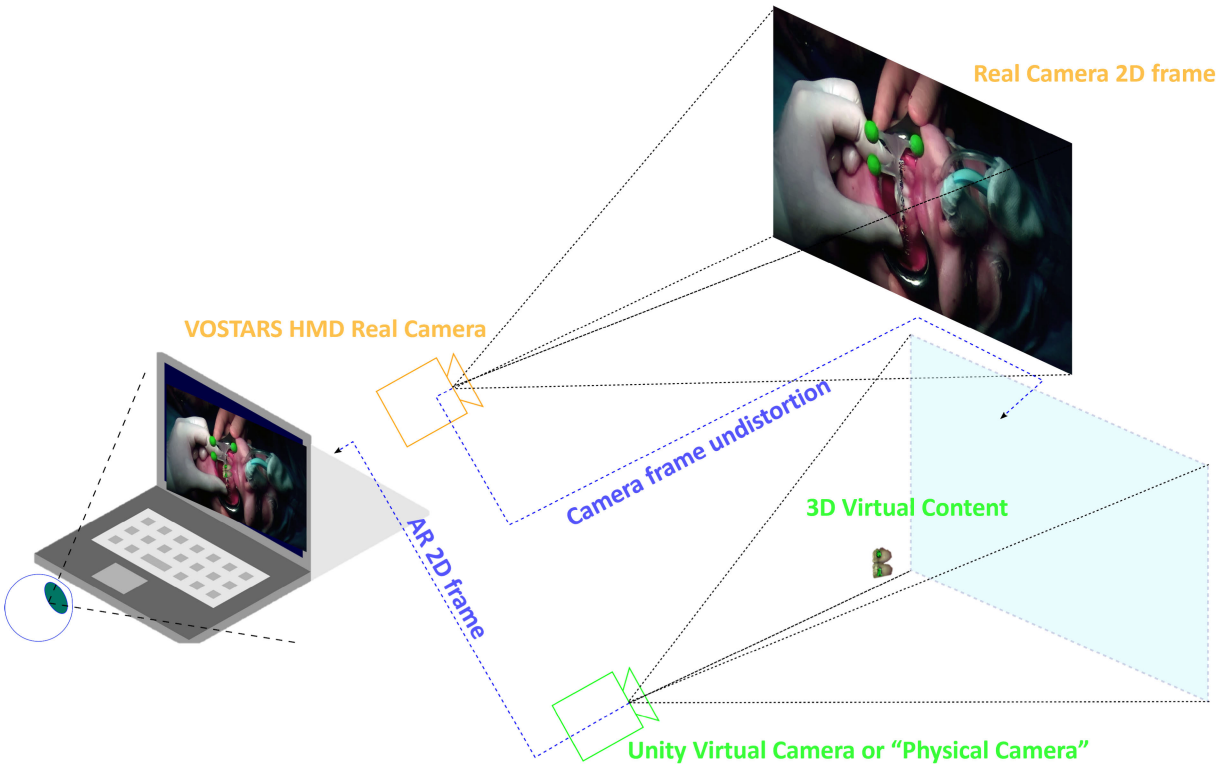


FIGURE 3. Scheme of the AR application developed in Unity3D under video see-through modality.

can simulate the linear real-world camera attributes: focal length, sensor size, and lens shift as detailed in [29], whereas the non-linearities associated with the distortion introduced by the lens are compensated by undistorting the real camera frame before rendering them on the background of the AR image (Figure 3). The extrinsic parameters of the virtual camera are modeled using the real camera pose computed frame-by-frame by the inside-out marker-based tracking of the VOSTARS platform.

For our specific application, a predefined error was iteratively added to the real camera pose to generate an intentional misplacement of the projection of the VR content onto the actual image plane, thus simulating a registration error.

A total of 342 AR images (*i.e.*, three sets of 114 images for each of the three selected frames) were generated including AR images with no intentional error, and AR images affected by pure translational or rotational error along a single axis (*e.g.*, 0.5mm translational error along x). No images with a combined translation and rotation error were produced for this study. The resulting AR images were classified into 4 groups based on the magnitude of the error:

- Grade0 error images: no intentional error added.
- Grade1 error images: intentional 0.5mm or −0.5mm translational error added along a single axis, or intentional 5.5° or −5.5° rotational error added along a single axis.
- Grade2 error images: intentional 1.0 mm or -1.0mm translational error added along a single axis, or intentional 10.5° or −10.5° rotational error added along a single axis.
- Grade3 error images: intentional 1.5mm or −1.5mm translational error added along a single axis, or intentional 15.5° or −15.5° rotational error added along a single axis.

Figure 2 depicts the reference system used to express the error components, with the origin in the centre of the frontal teeth coinciding with the centre of the splint. This point has been selected as a clear distinguishable origin and a natural centre of rotation for the splint during its manual handling.

The magnitude E_i of translational errors for each Grade i was defined at increments of 0.5mm. Moreover, for each Error Grade i, the rotational error component α_i was instead defined such that

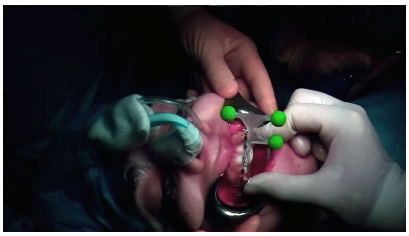
$$L \cdot sen(\alpha_i) = E_i \tag{1}$$

where L is the distance between the reference system origin and the lateral edge of the tooth model/the endpoint of the gingival margin model (see Figure 2), in order to obtain the same maximum displacement due to translation or rotation.

C. STUDY PROTOCOL

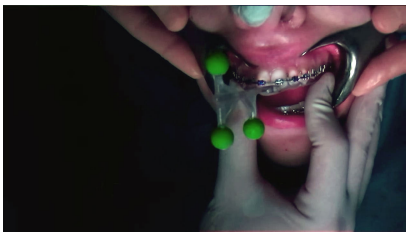
The study protocol involved administering 114 images to a selected group of users with different backgrounds (both technical and medical) and collecting users' perceptions of the correctness of AR registration through binary yes/no questions. The set of AR images generated, for one of the

IEEE Journal of Translational Engineering in Health and Medicine



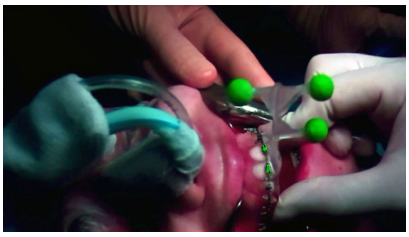
(a) Frame selected for training

(b) Frame 1 selected for testing



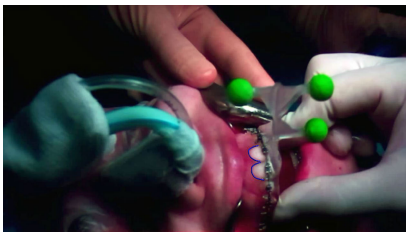
(c) Frame 2 selected for testing

FIGURE 4. **Selected video frames for the study.**



(a) Grade 0; Model=Brackets.

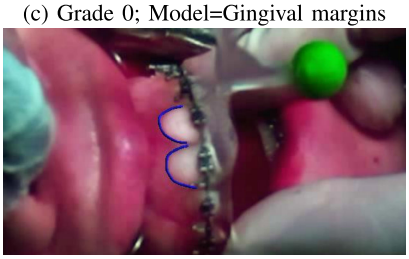
(b) Grade 0; Model=Incisor teeth



(c) Grade 0; Model=Gingival margins

(d) Grade 0; Model=Brackets; Zoomed detail;

(e) Grade 0; Model=Teeth; Zoomed detail;



(f) Grade 0; Model=Gingival margins; Zoomed detail;

FIGURE 5. **The three Grade 0 error images generated for one of the selected frame with the zoomed details.**

three selected frames, was used to develop a training session, whereas the other two were used for testing (Figure 4).

114 images were administered to subjects that included, for each of the 2 frames, 3 Grade0 error images (*i.e.*, the three AR images generated using the three different VR models -incisor teeth, their gingival margins, and orthodontic brackets- without adding any intentional error) (Figure 5), and 54 images affected by intentional errors (*i.e.*, , 18 Grade1 error images, 18 Grade2 error images, and 18 Grade3 images). The latter (Figure 6) included, for each virtual model (*i.e.*, Brackets, Incisor teeth, Gingival margins), the 3 error Grades (*i.e.*, Grade1, Grade2, Grade3) images for each different error component (*i.e.*, tx, ty, tz, Rx, Ry, Rz).

In this work, submillimetric errors (*i.e.*, Grade 1 errors) were considered compatible with the target accuracy for image-guided precision surgery. Therefore, in each experiment, the total number (P) of 'Positive' images (*i.e.*, images affected by a Grade2 error or Grade3 error, considered detrimental to the accuracy of the intervention) is 72; while the total number (N) of 'Negative' images (*i.e.*, images not affected by intentional error or affected by Grade1 intentional error) is 42. The following paragraphs furnish details on the application developed for the AR-RSC test, the demographic of participants recruited for the study, and the analysis of the results.

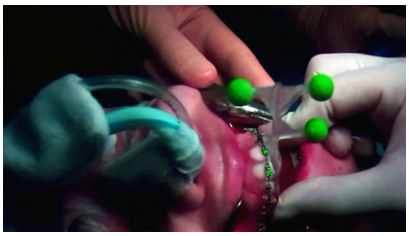
1) UNITY APPLICATION

A Windows 10 software AR application was developed with Unity3D to acquire users' judgment on the AR content registration based on a qualitative visual estimation of the virtual-to-real alignment.

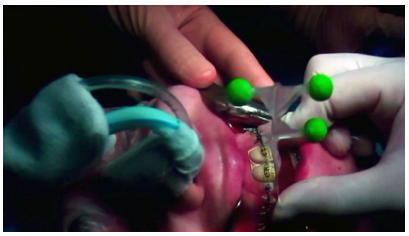
The application provides the user with all the information needed to perform the test correctly and includes a form for collecting personal data. It also includes a training module that displays examples of Grade 0 and Grade 1 error images, correctly classified as being Negative, and Grade 2 and Grade 3 error images, correctly classified as being Positive. The test module is designed to present the 114 AR images in a random order, to allow the user to re-examine an image (back button) and to collect the following data: presentation order of the AR images, time spent for the evaluation of each AR image, and user binary feedback on the AR registration correctness.

2) DEMOGRAPHICS

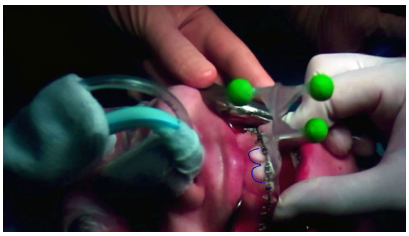
A total of 24 technicians (engineers and physicists) and 12 medical doctors (maxillofacial surgeons, neurosurgeons,



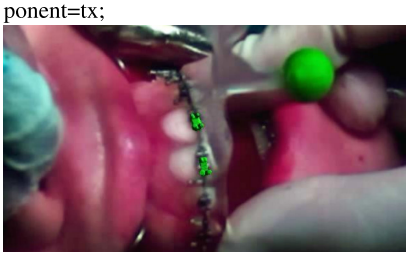
(a) Grade 2; Model=Brackets; Error Component=tx;



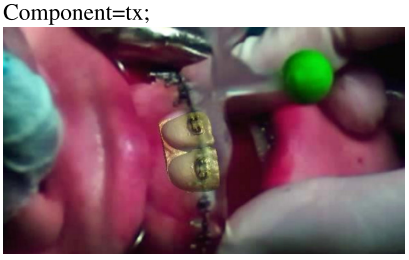
(b) Grade 2; Model=Incisor teeth; Error Component=tx;



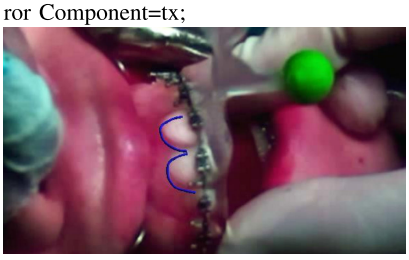
(c) Grade 2; Model=Gingival margins; Error Component=tx;



(d) Grade 2; Model=Brackets; Error Component=tx; Zoomed detail;



(e) Grade 2; Model=Teeth; Error Component=tx; Zoomed detail;



(f) Grade 2; Model=Gingival margins; Error Component=tx; Zoomed detail;

FIGURE 6. Example of Grade 2 error images generated for one of the selected frames.

TABLE 1. Demographics of the 36 participants involved in the user study.

Information	Value
Gender (female, male)	(18, 18)
Profession (technician, medical doctors)	(24,12)
Age (min, max, mean, std)	(23, 57, 33.8, 7.9)
Experience in AR (Expert, Intermediate, Novice)	(6, 12, 18)

and dentists) were recruited from the University of Pisa and the University of Bologna. Besides demographic data, we also asked the participants to rate their experience with AR methods to get the opportunity to correlate it with their performance Table 1.

3) DATA ANALYSIS

The AR-RSC test result for each AR image was categorised, based on the image type and on the opinion expressed by the user, as reported below:

- True Positive (TP): in the case of a Grade2 or Grade3 error image correctly identified as being Positive.
- False Positive (FP): in the case of a Grade0 or Grade1 error image incorrectly identified as being Positive.
- True Negative (TN): in the case of a Grade0 or Grade1 error image correctly identified as being Negative.
- False Negative (FN): in the case of a Grade2 or Grade3 error image incorrectly identified as being Negative.

For each session, the test sensitivity, also known as true positive rate (TPR), and the test specificity, also known as true negative rate (TNR), were calculated as follows:

$$TPR = {}^{TP}/_{P}$$
$$TNR = {}^{TN}/_{N} \quad (2)$$

TPR is a measure of how well the registration sanity check can identify true positives, while TNR is a measure of how well the registration sanity check can identify true negatives. The TPR and the TNR were calculated for each of the three AR models (TPRMOD and TNRMOD) selected for this study (*i.e.*, brackets models, incisor teeth models, gingival margin model) to verify Q1 (Does the AR-RSC sensibility depend on the type of AR data selected for the misalignment assessment?).

Moreover, the TPR and the TNR were calculated for each registration error component (TPRCOMP nd TNRCOMP), to verify Q2 (Does the AR-RSC sensibility vary with the type of the registration error?). In both cases, the Friedman Test was used to determine whether the observed differences are statistically significant.

Finally, the Mann-Whitney U test and the Kruskal-Wallis test were used to understand whether the users' performances in executing the AR-RSC, in terms of TPR and TNR, differ based on their profession and their experience in using AR technologies, respectively, to verify Q3 (Does the AR-RSC performances depend on the user's level of experience in using AR technologies?).

III. RESULTS

Table 2 summarises the overall TPR and TNR and answers Q3 by showing the AR-RSC performance, in terms of TPR and TNR, of participants with different levels of experience in using AR technologies. The Kruskal-Wallis test revealed no statistically significant differences in performance according to user AR experience for either TPR (p=0.631) or TNR (p=0.681).

Table 3 and Table 4 answer Q1, reporting the TPR and the TNR for each of the three AR models (Brackets, Incisor

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 2. Global TPR and TNR..

	TPR(%)		TNR(%)	
	MEDIAN	IQR	MEDIAN	IQR
ALL PARTICIPANTS	71.5	18.1	59.5	23.8
AR EXPERT	75.0	5.6	59.5	9.50
AR INTERMEDIATE	72.2	13.2	9.5	21.4
AR NOVICE	70.1	20.8	65.5	28.6

TABLE 3. TPR for each of the three AR models.

	TPR(%)					
Error Grade	Brackets		Incisor teeth		Gingival margins	
	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR
Grade 3 (± 1.5 mm, or ± 15.5°)	83.3	16.7	66.7	16.7	83.3	8.30
Grade 2 (± 1 mm, or ± 10.5°)	75	25	54.2	25	66.7	16.7
Global	79.2	16.6	60.4	18.8	75	14.6

TABLE 4. TNR for each of the three AR models.

	TNR(%)					
Error Grade	Brackets		Incisor teeth		Gingival margins	
	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR
Grade 1 (± 0.5 mm, or ± 5.5°)	50	41.7	66.7	29.2	58.3	33.3
Grade 0 (no intentional error)	75	50	100	50	100	25
Global	57.1	35.7	64.3	32.2	60.7	35.7

teeth, Gingival margins). There was a statistically significant difference in TPR depending on which type of AR model was shown: p-values were less than 0.001, for both Grade3 errors, Grade2 errors, and globally. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. Results show that the increase in TPR in Brackets vs Incisor teeth margins is statistically significant ($p < 0.001$ for Grade3, Grade2 and Globally), as well as the increase in TPR in Gingival margins vs Incisor teeth margins ($p < 0.001$ for Grade3, Grade2 and Globally), whereas no statistically significant differences were found in TPR between Brackets and Gingival margins ($p = 0.5408$ for Grade3, $p = 0.4081$ Grade2, and $p = 0.3296$ Globally).

Moreover, there was a statistically significant difference in TNR depending on which type of AR model was shown for Grade1 (p=0.002) and Globally (p=0.003), while no statistically significant difference in TNR was found for Grade0 (p=0.195). Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. Results show that the increase in TNR in Incisor teeth vs Brackets teeth margins is statistically significant for Grade1 ($p = 0.015$) and Globally ($p = 0.016$), while no statistically significant differences were found in TNR between Brackets and Gingival margins ($p = 0.017$ for Grade1, and p=0.045 Globally) and Incisor teeth and Gingival margins ($p = 0.898$ for Grade1, and p=0.708 Globally).

Table 5 and Table 6 answer Q2. Table 5 reports the global TPR and the TNR for translation and registration errors, while Table 6 reports the TPR for each translation and registration component. As highlighted in 5 there were no statistically significant differences in TPR and in TNR for translation and rotation components.

Results show that the increase in TPR in tx vs ty, and in tz vs ty is statistically significant for Grade 3, Grade 2, and Globally ($p < 0.001$ for Grade 3, Grade 2, and Globally). Moreover, a statistically significant increase in TPR in tx vs tz was found for Grade 2 ($p = 0.006$). Instead, no statistically significant differences were found in TPR between tx and tz for Grade3 ($p = 0.789$), and Globally (p=0.029). As for Rotation components, a statistically significant increase in TPR was found in Ry vs Rx, in Rz vs Rx ($p < 0.001$ for Grade 3, Grade2, and Globally), and in Ry vs Rz ($p < 0.001$ for Grade 3, $p = 0.002$ Grade 2, and $p < 0.001$ Globally).

IV. DISCUSSION

According to the literature, monitoring and optimizing the accuracy of registration during the performance of surgery is an open challenge in computer-assisted surgery: indeed, there is a need for simple methods that allow the surgeon to monitor registration not only at the beginning but also during surgery.

This study investigates a simple registration sanity check method for AR-guided surgery navigation (AR-RSC) based on the visualization of virtual 3D models of anatomical or artificial landmarks, or their contours. More particularly this research focuses on three fundamental and not jet-answered questions: (Q1) Does the AR-RSC performances depend on the type of AR data?; (Q2) Do the AR-RSC performances vary with the type of registration error? (Q3) Do the AR-RSC performances depend on the user's level of experience in using AR technologies?

The following is a discussion of the results obtained for the use case selected for this work, an experimental Le Fort1 osteotomy guided by HMD VOSTARS.

A. ANSWER to QUESTION 1

The study performed reveals the dependence of the sensitivity and specificity of the AR-RSC on the virtual models selected for misalignment assessment: in our specific application, the tooth models obtained globally a TPR of 60.4% (overall median value), whereas significantly better TPRs were obtained with brackets and tooth margins, achieving 79.2% and 75% performance, respectively. Such results were obtained for synthetic translational or rotational errors along a single axis equal to (Grade3) or less (Grade2) than ± 1.5mm, or ± 15.5°, respectively. TPRs of up to 83.3% were reached with brackets and tooth margins for Grade 3 errors.

In contrast, the tooth models were significantly better than brackets for TNR (overall median of 64.3% vs 57.1% for images not affected by an intentional error, or affected by translational or rotational errors along a single axis equal to ± 0.5 mm, or ± 5.5°, respectively). Probably, because the virtual model of the teeth occludes the real anatomy more

TABLE 5. TPR and TNR for translation and registration errors.

Error Grade	TPR(%)					TNR(%)				
	Translation		Rotation		p	Translation		Rotation		p
	MEDIAN	IQR	MEDIAN	IQR		MEDIAN	IQR	MEDIAN	IQR	
GRADE 3 (± 1.5 mm, or ± 15.5°)	83.3	50	83.3	50	0.744	/	/	/	/	/
GRADE 2 (± 1 mm, or ± 10.5°)	66.7	41.7	58.4	66.7	0.609	/	/	/	/	/
GRADE 1 (± 0.5 mm, or ± 5.5°)	/	/	/	/	/	66.7	25	66.7	50	0.212
GLOBAL	75.0	41.7	66.7	50	0.889	66.7	25	66.7	50	0.212

TABLE 6. TPR for each registration error component.

Error Grade	TPR(%)											
	tx		ty		tz		Rx		Ry		Rz	
	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR	MEDIAN	IQR
GRADE 3 (± 1.5 mm, or ± 15.5°)	100	16.7	41.7	25.1	100	16.7	50	33.4	100	0	66.7	33.3
GRADE 2 (± 1 mm, or ± 10.5°)	83.3	33.3	33.3	33.3	66.7	41.7	33.3	33.3	100	0	50	33.4
GLOBAL	91.7	16.7	41.7	25	83.3	16.7	50	25	100	0	58.3	25

than the tooth margins, the user is more likely to believe the registration is correct, and as a result, worse performance in terms of TPR and better performance in terms of TNR is obtained. In general, for both diagnostic and screening tests, there is a trade-off between sensitivity and specificity, with higher sensitivities being linked to lower specificities and vice versa [30]. The main goal of AR-RSC is to identify every potential registration error, that could compromise the accuracy of the intervention: thus, the number of false negatives should be low, which requires high sensitivity. False positives, on the other hand, do not compromise the accuracy of the intervention, so they are less dangerous from a clinical point of view, but they can still be detrimental as they can lead the user not to trust the AR guide and/or unnecessarily repeat registration-related steps (*e.g.*, , the splint positioning in the surgical case selected as an example in this paper), thus uselessly lengthening the surgical time.

B. ANSWER to QUESTION 2

According to this work, the sensitivity of the proposed AR-RSC does not significantly vary for translation and rotation errors, but among the translations, the most difficult component to detect is the one along the axis perpendicular to the image plane (*i.e.*, , the y-axis along the antero-posterior direction, in our case). As could reasonably be expected, translation errors along the depth are more difficult to estimate than transversal translations. As for the rotation components, those along the y-axis are the easiest to identify (the median of the TPR reaches 100% also for submillimetric errors), while the most difficult to identify are those along the x-axis (i.e., in the sagittal plane) for which the sensitivity is halved (overall, the median is 50%). Thus, to rule out such error components, which during a real clinical application may possibly be related to splint malpositioning, the surgeon must rely primarily on haptic feedback during the mechanical engagement of the splint into the teeth, rather than visual feedback. An alternative solution could be performing the visual sanity check by using two (or more) landmarks that lie in mutually orthogonal planes. Instead, any tracking-related error components can be identified through the quantitative estimation of fiducial registration error, which is provided in real-time by the VOSTARS system.

C. ANSWER to QUESTION 3

An important point highlighted by this study is that users' performance is not related to their level of experience with AR, suggesting that the proposed method could be successfully employed even by users who have no prior experience in using AR applications. However, caution should be taken in transferring the results of this study to real clinical applications. In fact, in this work, subjects were presented only with static images, whereas in the real application, surgeons may vary their point of view and, in uncertain cases, may verify the correspondence between virtual and real from different perspectives. The application developed to collect data for this study instead forces the user to make a negative or positive judgment for each image presented. The choice to employ static images is related to the need to have a set of AR images with a validated image-patient registration, and to synthetically add an error of known magnitude, and to present the same set of resulting images to different users.

In conclusion, results obtained for the use case selected for this work, suggest that:

1) the AR-RSC performances can significantly vary depending on the AR data selected for the visual estimation (in our case the 3D virtual models of brackets, incisor teeth, and gingival margins);
2) there may be components of the misregistration error that are more difficult to identify than others;
3) results of the AR-RSC could be not related to the users' level of experience with AR.

The results of this study, obtained for a particular type of surgery, can be used to draw an important general consideration: special attention should be paid to the selection

of AR data for visual inspection and validation of the registration in such a way as to maximize the sensitivity of the proposed AR-RSC method. If in the particular surgical procedure, it is possible to visualize landmarks placed in two planes that are orthogonal to each other (e.g., frontal and sagittal plane), it is suggested to repeat the evaluation from different perspectives to strengthen the identification of any error components in all three directions. The experimental strategy employed in this paper, based on the administration of images with synthetic errors of different magnitudes introduced by the experimenter to simulate the effect of misregistrations, can be used for the selection of optimal landmarks in the preclinical evaluation phase of AR-based navigation systems. Once the optimal AR data are selected, the AR-RSC method offers a straightforward method to monitor and optimize the registration accuracy during the intervention, allowing for a real-time visual check of the registration quality, potentially on several anatomical landmarks simultaneously, therefore with a reduced impact on the surgical workflow.

In summary, this paper describes a simple qualitative method to visually detect registration errors during AR-guided surgery , reports our experience from Head and Face Surgery, and furnishes an experimental method for optimizing the AR-RSC performance by selecting the optimal VR content.

REFERENCES

[1] P. Vávra et al., "Recent development of augmented reality in surgery: A review," *J. Healthcare Eng.*, vol. 2017, Aug. 2017, Art. no. 4574172.

[2] T. Sielhorst, M. Feuerstein, and N. Navab, "Advanced medical displays: A literature review of augmented reality," *J. Display Technol.*, vol. 4, no. 4, pp. 451–467, Dec. 2008.

[3] J. Yoon et al., "Augmented reality for the surgeon: Systematic review," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 4, p. e1914, 2018.

[4] N. Cattari, S. Condino, F. Cutolo, M. Ferrari, and V. Ferrari, "In situ visualization for 3D ultrasound-guided interventions with augmented reality headset," *Bioengineering*, vol. 8, no. 10, p. 131, Sep. 2021. [Online]. Available: https://www.mdpi.com/2306-5354/8/10/131

[5] N. Navab, S.-M. Heining, and J. Traub, "Camera augmented mobile C-arm (CAMC): Calibration, accuracy study, and clinical applications," *IEEE Trans. Med. Imag.*, vol. 29, no. 7, pp. 1412–1423, Jul. 2010.

[6] J. M. Fitzpatrick, "The role of registration in accurate surgical guidance," *Proc. Inst. Mech. Eng., H, J. Eng. Med.*, vol. 224, no. 5, pp. 607–622, May 2010.

[7] H. H. L. Chan et al., "An integrated augmented reality surgical navigation platform using multi-modality imaging for guidance," *PLoS ONE*, vol. 16, no. 4, Apr. 2021, Art. no. e0250558, doi: 10.1371/journal.pone.0250558.

[8] H. Liu and F. R. Y. Baena, "Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery," *IEEE Access*, vol. 8, pp. 42010–42020, 2020.

[9] H. Suenaga et al., "Vision-based markerless registration using stereo vision and an augmented reality surgical navigation system: A pilot study," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–11, Dec. 2015.

[10] H. Gueziri and D. Collins, "Fast registration of CT with intra-operative ultrasound images for spine surgery," in *Proc. Int. Workshop Challenge Comput. Methods Clin. Appl. Spine Imag.*, 2018, pp. 29–40.

[11] N. Golse, A. Petit, M. Lewin, E. Vibert, and S. Cotin, "Augmented reality during open liver surgery using a markerless non-rigid registration system," *J. Gastrointestinal Surg.*, vol. 25, no. 3, pp. 662–671, Mar. 2021.

[12] H. Rivaz and D. L. Collins, "Deformable registration of preoperative MR, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 7, pp. 1017–1028, Jul. 2015.

[13] L. W. Clements et al., "Deformation correction for image guided liver surgery: An intraoperative fidelity assessment," *Surgery*, vol. 162, no. 3, pp. 537–547, Sep. 2017.

[14] É. Léger, J. Reyes, S. Drouin, D. L. Collins, T. Popa, and M. Kersten-Oertel, "Gesture-based registration correction using a mobile augmented reality image-guided neurosurgery system," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 137–142, Oct. 2018.

[15] S. Drouin, M. Kersten-Oertel, and L. Collins, "Interaction-based registration correction for improved augmented reality overlay in neurosurgery," in *Proc. Workshop Augmented Environ. Comput.-Assist. Intervent.*, Oct. 2015, pp. 21–29.

[16] T. Morineau, X. Morandi, N. Le Moëllic, and P. Jannin, "A cognitive engineering framework for the specification of information requirements in medical imaging: Application in image-guided neurosurgery," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 8, no. 2, pp. 291–300, Mar. 2013.

[17] S. Thompson et al., "In vivo estimation of target registration errors during augmented reality laparoscopic surgery," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 13, no. 6, pp. 865–874, Jun. 2018.

[18] A. Amir-Khalili, M. S. Nosrati, J.-M. Peyrat, G. Hamarneh, and R. Abugharbieh, "Uncertainty-encoded augmented reality for robot-assisted partial nephrectomy: A phantom study," in *Proc. Int. Workshop Med. Imag. Virtual Reality*, Sep. 2013, pp. 182–191.

[19] B. MacIntyre, E. M. Coelho, and S. J. Julier, "Estimating and adapting to registration errors in augmented reality systems," in *Proc. IEEE Virtual Reality*, Mar. 2002, pp. 73–80.

[20] S. Condino et al., "Evaluation of a wearable AR platform for guiding complex craniotomies in neurosurgery," *Ann. Biomed. Eng.*, vol. 49, no. 9, pp. 2590–2605, Sep. 2021.

[21] E. Watanabe, M. Satoh, T. Konno, M. Hirai, and T. Yamaguchi, "The trans-visible navigator: A see-through neuronavigation system using augmented reality," *World Neurosurg.*, vol. 87, pp. 399–405, Mar. 2016.

[22] L. Cercenelli et al., "The wearable VOSTARS system for augmented reality-guided surgery: Preclinical phantom evaluation for high-precision maxillofacial tasks," *J. Clin. Med.*, vol. 9, no. 11, p. 3562, Nov. 2020.

[23] M. Carbone et al., "Architecture of a hybrid video/optical see-through head-mounted display-based augmented reality surgical navigation platform," *Information*, vol. 13, no. 2, p. 81, Feb. 2022. [Online]. Available: https://www.mdpi.com/2078-2489/13/2/81

[24] F. Cutolo, B. Fida, N. Cattari, and V. Ferrari, "Software framework for customized augmented reality headsets in medicine," *IEEE Access*, vol. 8, pp. 706–720, 2020.

[25] L. Shao et al., "Augmented reality navigation with real-time tracking for facial repair surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 6, pp. 981–991, Jun. 2022, doi: 10.1007/s11548-022-02589-0.

[26] L. Shao et al., "Augmented reality calibration using feature triangulation iteration-based registration for surgical navigation," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105826. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522005856

[27] L. Ma et al., "Augmented reality surgical navigation with accurate CBCT-patient registration for dental implant placement," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 47–57, Jan. 2019, doi: 10.1007/s11517-018-1861-9.

[28] L. Shao et al., "Robot-assisted augmented reality surgical navigation based on optical tracking for mandibular reconstruction surgery," *Med. Phys.*, vol. 51, no. 4, pp. 363–377, Jul. 2023. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16598

[29] S. Condino, G. Turini, V. Mamone, P. D. Parchi, and V. Ferrari, "Hybrid spine simulator prototype for X-ray free pedicle screws fixation training," *Appl. Sci.*, vol. 11, no. 3, p. 1038, Jan. 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/3/1038

[30] D. Carvajal and P. Rowe, "Sensitivity, specificity, predictive values, and likelihood ratios," *Pediatrics Rev./Amer. Acad. Pediatrics*, vol. 31, pp. 511–513, Dec. 2010.

Open Access funding provided by 'Università di Pisa' within the CRUI CARE Agreement