Projected Augmented Reality to Drive Osteotomy Surgery: Implementation and Comparison With Video See-Through Technology

VIRGINIA MAMONE[ID]**, VINCENZO FERRARI**[ID]**, (Member, IEEE),
SARA CONDINO**[ID]**, AND FABRIZIO CUTOLO**[ID]**, (Member, IEEE)**
Dipartimento di Ingegneria dell'Informazione, University of Pisa, 56122 Pisa, Italy
EndoCAS Center for Computer-Assisted Surgery, 56124 Pisa, Italy

Corresponding author: Virginia Mamone (virginia.mamone@endocas.unipi.it)

ABSTRACT In recent years, the spreading of visual augmented reality as an effective tool in image-guided surgery, has stimulated the research community to investigate the use of commercial augmented reality headsets a broad range of potential applications. This aroused enthusiasm among clinicians for the potential of augmented reality, but also revealed some technological and human-factor limitations that still hinder its routine adoption in the operating room. In this work, we propose an alternative to head-mounted displays, based on projected augmented reality. Projected augmented reality completely preserves the surgeon's natural view of the operating field, because it requires no perspective conversion and/or optical mediation. We selected a cranio-maxillofacial surgery application as a benchmark to test the proposed system and compare its accuracy with the one obtained with a video see-through system. The augmented reality overlay accuracy was evaluated by measuring the distance between a virtual osteotomy line and its real counterpart. The experimental tests showed that the accuracy of the two augmented reality modes is similar, with a median error discrepancy of about 0.3 mm for the projected augmented reality mode. Results suggest that projected augmented reality can be a valuable alternative to standard see-through head-mounted displays to support in-situ visualization of medical imaging data as surgical guidance.

INDEX TERMS Augmented reality, projected augmented reality, video see-through, head-mounted display, computer-assisted surgery, cranio-maxillofacial surgery, augmented reality accuracy.

I. INTRODUCTION

The surgeon bears a great responsibility: a cut into the patient's body can either save the patient's life or cause irretrievable damage. When executing a surgical procedure, the surgeon has both to deal with the resulting tension and also to keep in mind a great deal of information about the patient's pathology and anatomy. Visual augmented reality (AR) technologies assist the surgeon by merging patient-specific information from medical imaging with the patient's anatomy. Since the first attempts in the 90's [1], the interest for AR has grown, and recent progress in miniaturized computational units, optics, and photonics devices have led to the development and then commercialization of AR head mounted

displays (HMD). These devices are capable to support the in-situ visualization of preoperative and intraoperative medical imaging data from an egocentric viewpoint. An increasing number of studies propose the use of commercial AR HMDs to guide surgical procedures. In a recent work, Pratt *et al.* used Microsoft HoloLens to assist reconstructive surgery in the identification, dissection and execution of vascular flaps [2]. In [3], the authors adopted the Google Glass head-up display to perform intra-operative neuronavigation for tumor resection. However, recent works have shown that most consumer-level devices have technological and human-factor limitations that hinder their routine use in healthcare [4]. These limitations differ depending on the AR paradigm chosen, which can be either video see-through (VST) or optical see-through (OST). VST HMDs record and reproduce camera frames of the real world using one or two cameras anchored

The associate editor coordinating the review of this manuscript and approving it for publication was Chua Chin Heng Matthew[ID].

to the HMD and then present the resulting images to the user with merged virtual content superimposed on them. On the other hand, OST devices are based on optical combiners (e.g., waveguides or semi-reflective mirrors) that preserve the direct view of the world and simultaneously reflect the computer-generated images into the user's eye. In both cases, the virtual content is commonly projected at a fixed focal distance which corresponds for VST displays to the visualization distance of the augmented camera images and, for the OST displays, to the visualization distance of the virtual information only. General-purpose HMDs feature a distance of the focal plane usually between 2 meters and infinity. When dealing with closer objects, as in the execution of manual tasks, this produces perceptual problems such as vergence-accommodation conflict and focus rivalry [5], [6]. As a result, natural eye accommodation and retinal blur effects are not stimulated, thus producing visual fatigue. Recent research studies proved that these human-factor limitations compromise the user's performance in tasks that require to focus simultaneously on real and virtual content [7]. VST devices also present problems related to the impaired and distorted perception of the real scene produced by the perspective conversion of the camera frames [8]. To this day, OST HMDs are the leading edge of AR research: as opposed to VST solutions, they preserve the direct view of the world, allowing the user to maintain his or her own natural visual experience. However, perceptual limitations still hinder their use to support high precision manual tasks: in [9], the authors proved that using Microsoft HoloLens to guide manual tasks reduces the user's performances because of incorrect focus cues and visual discomfort. Based on these considerations, the use of commercial HMDs is not recommended in surgery, where a high level of precision is required to maximize clinical outcome [10]. In this regard, in the review study on AR in Oral and cranio-maxillofacial Surgery, authors claimed that a 1-2 mm accuracy is an acceptable range [11]. In their work Carbone *et al.* suggest to develop an AR HMD specifically designed for surgical guidance to meet these requirements, by taking into account the surgical working distance around a few tens of centimetres, and correctly focusing both real and augmented information [12].

In recent years, a different technology for AR has started to make its way into the current medical literature: it is based on video projectors that, when properly calibrated with a tracking system, can provide AR information directly on the structure of interest, overcoming the perceptual limitations introduced above owing to focal rivalry and vergence-accommodation conflict. This is the case, for example, of Mewes *et al.* [13], who used an ultra-long-throw projector to guide the radiologist during interventional magnetic resonance imaging; or Gierwialo *et al.* [14], who applied projected AR in surgical training and education, allowing medical students to visualize the internal structures of a liver mannequin. Both of these works rely on the use of heavy and cumbersome projectors, and for their successful integration into the operating room, they require a review and reorga-

nization of the equipment layout. This is due to the specific application requirements, which impose, for example, the use of an ultra-long-throw projector in [13] rather than a smaller consumer-grade projector. These constraints, however, are not inherent in the use of projected AR in medicine and surgery. Recently, technological development led to a high brightness even from small LED sources and to increasing miniaturization of micro-lenses, thus resulting in lightweight and handy projectors. This paved the way for new biomedical applications: a mini projector in combination with a laparoscopic ultrasound has been proposed to improve efficacy and safety in laparoscopic partial nephrectomy [15]; in [16], the authors developed a portable system for monitoring skin ulcers through projected AR; in [17], the authors presented a mini projector for guided neurosurgery that allows the regions of interest to be visualized directly on the patient's head, skull, or brain. Projected AR relies on a different paradigm than egocentric AR based on HMDs and has the potential to overcome the limitations that prevent AR from routinely entering the operating room. However, projected AR has also some limitations. Some of these will be discussed later in this work, while others can be avoided by acting upstream. This is the case with the parallax error, that causes the location of a projected internal structure to be perceptually consistent only from a single viewpoint [18]. If the surgeon head is not tracked, as in most cases, this viewpoint is coincident to that of the projector. Observing the projected AR structure from any other viewpoint will produce an error in its perceived 3D position. The parallax error is completely resolved when projecting structures that reside on the surface as it is done in this study.

In this work, we introduce a lightweight and portable device for projected AR specifically designed for surgery. The system consists of two main components: a mini projector and a stereo camera rig. A cranio-maxillofacial surgery application was chosen as a benchmark to test the proposed system. Cranio-maxillofacial surgery treats various types of diseases and aesthetic-functional problems of the face and facial skeleton. As detailed in the next section, the selected application is a Le Fort I osteotomy, which addresses deformities of the lower midface, such as gowth abnormalities, congenital asymmetries, and relative disproportion to position of the mandible. Cranio-maxillofacial surgery has undergone a significant evolution over the past 20 years due to the extensive adoption of virtual surgical planning, surgical navigation tools, and computer-aided design/manufacturing techniques. The usefulness of virtual planning, as well as of 3D stereolithographic models and osteotomy guides, has been widely documented in the plastic surgery and cranio-maxillofacial surgical literature [19], [20]. Moreover surgical navigators have been shown to optimize the functional and aesthetic outcomes in patients with dentofacial deformities by assisting the surgeon in transferring the preoperative virtual plan accurately to the operating theatre, avoiding critical anatomical structures, and verifying surgical outcomes intraoperatively [21].

We tested our portable AR projector in-vitro to guide the surgeon in performing osteotomies. An osteotomy is any surgery that cuts and reshapes the bones. The precise location of an accurate osteotomy line is a fundamental task in many surgical procedures. In orthognathic surgery, the accurate localization of the osteotomy lines is essential to properly manipulate and reposition the bone fragment. In this work we decided to use one of Le Fort's lines as a reference. In the early 19th century, Le Fort identified three osteotomy pathways on the skull that were later used to correct specific deformities [22]. As a reference procedure for this work we selected the Le Fort I osteotomy that is performed by surgeons to correct dentofacial deformities [23]. The need to accurally follow the osteotomy lines, as already mentioned, is not limited to cranio-maxillofacial surgery. In a recent work, the authors tested AR tumor resection on animal femurs and found that AR navigation allows both experienced and inexperienced surgeons to follow the surgical margins within the required tolerances with better accuracy [24]. In neuro-oncological surgery, AR has already been successfully applied mainly for tumor resection. A study involving 17 real clinical cases demonstrated the usefulness of AR in skin incision, craniotomy, and dural opening [25]. In this work, we describe the results of a direct comparison in the AR visualization accuracy between a custom-made VST HMD and a projected AR device. We paid attention to shield the test results from any additional sources of variability other than the different AR methodology, such as lighting conditions or tracking accuracy. The aim of this work is to objectively evaluate the accuracy of the projected AR in terms of overlay of the AR features with ground truth to demonstrate that projected AR can be a valuable alternative to HMDs in surgery.

II. MATERIALS AND METHODS: HARDWARE AND SOFTWARE

This section provides a detailed description of the device hardware and software components. The projected AR device is depicted in Figure 1.

The next subsection gives details of the hardware, covering both the cameras, the projector and the computer platform (A. "Hardware Components"). Then, we provide a description of the rigid transformations involved and the relative notation (B. "Transformation Formulation"). In C. "Stereo Rig Calibration", we discuss the calibration of the stereo camera pair, preliminary step to the optical tracking and to the implementation of VST AR. In D. "Marker-based Optical Tracking", we then discuss the tracking method. Finally we provide information about the projector calibration (E. "Projection Module Calibration") and the AR software framework used in the experimental tests (F. "AR Software Framework").

A. HARDWARE COMPONENTS

The portable device for projected AR is composed of two functional modules: the tracking module, and the projection module. The tracking module consists of a stereo rig. The
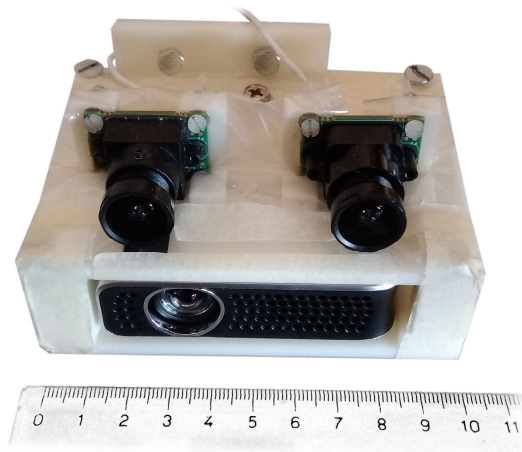

FIGURE 1. Projected AR device, featuring the OV580 stereo camera pair by Leopard Imaging and the Philips PicoPix 4010 projector.

stereo camera pair used for optical tracking is composed of two LI-OV4689 cameras by Leopard Imaging, both equipped with a 1/3'' OmniVision CMOS 4M pixels sensor (pixel size: $2\mu m$). The cameras have a horizontal field of view of 68° that corresponds to an angular resolution, $R \approx 3.6$ arcmin/pixel. The cameras are stereo synchronized through a dedicated board (LI-OV580-STEREO), which also includes a USB 3.0 interface. The stereo cameras configuration adopted in the study is: $2560 \times 720@60$ frames-per-second (fps).

The projection module includes a reference camera, i.e. the left camera of the stereo rig, and the projector. The projector is a Philips PicoPix 4010 mini projector, which features a resolution of 854×480 pixels and a brightness of 100 lumens. An ABS support, produced through rapid prototyping, constraints the stereo camera pair and the projector. The weight of the device is 200 g and the overall dimensions are $9.5 \times 7 \times 6\ cm^3$.

B. TRANSFORMATION FORMULATION

The modules, together with the related transformations and reference systems, are shown in Figure 2. Rigid transformations from reference system {A} to {B} are denoted by uppercase bold letters: $_A^B\mathbf{T}$. Reference systems are denoted by uppercase letters, enclosed in curved brackets: {C} is the left camera reference system; {O} is the optical frame reference system where the coordinates of the target object are known; {P} is the projector reference system. Both 3D and 2D points are represented by homogeneous vectors. 3D points are denoted by uppercase letters with a left superscript denoting the associated reference coordinate system (e.g., a 3D point in camera coordinates is ^{C}L. 2D image points are denoted by lowercase letters (e.g., an image point in virtual screen coordinates is denoted by l). The 6 DoF transformation representing the change of reference system of target points from {O} to {P}, and associated to the two functional blocks

is:

$$^{P}L = {}_{C}^{P}\mathbf{T} \, {}_{O}^{C}\mathbf{T} \, {}^{O}L \qquad (1)$$

where ${}_{O}^{C}\mathbf{T}$ is the tracked pose of the optical frame with respect to the left camera, and ${}_{C}^{P}\mathbf{T}$ is the calibrated pose between the left camera coordinate system and the projector coordinate system.

C. STEREO RIG CALIBRATION

Stereo rig calibration provides the intrinsic camera parameters and the extrinsic stereo parameters (i.e., the relative pose between the two cameras) required to track the target object. Initially, the intrinsic calibration of the stereo cameras is performed. This provides the projection transformations of both cameras, I_L and I_R, that map the points in space with the image plane of the two cameras according to formula 2. The intrinsic calibration also estimates the lens distortion parameters, k_L and k_R. A set of 20 images of an 8×7 chessboard pattern with 10 mm side were captured for each camera and processed using the MATLAB Single Camera Calibration App. The calibration routine implementing Zhang's calibration method, was performed for each camera individually to obtain $I_{(L/R)}$ and $k_{(L/R)}$ [26], [27].

Subsequently, stereo calibration is performed through the same chessboard pattern, but images are now simultaneously captured by both cameras. Stereo calibration routine aims to compute the geometrical relationship between the two cameras in terms of rotation matrix and translation vector (i.e., extrinsic parameters). The images from the left and right camera are processed by the MATLAB Stereo Camera Calibration App. The intrinsic parameters of the right and left camera are provided as input to the procedure and they are kept fixed, so that the calibration routine only estimates the extrinsic parameters. This approach is preferred over an all-in-one calibration process that estimates the intrinsic and stereo parameters, because the acquisition of the intrinsic calibration images is not constrained by the need to simultaneously visualize the pattern on both cameras. With a two-step stereo camera calibration process, it is possible to cover a wider area of the field of view and we obtain a more accurate and reliable estimation of the intrinsic calibration parameters.

D. MARKER-BASED OPTICAL TRACKING

The optical tracking algorithm that provides the pose of the target object (${}_{O}^{C}\mathbf{T}$) is based on the stereo localization of a set of three spherical markers. Compared to planar markers, such as the ones used in most AR applications, small spherical markers reduce the obstruction of the surgeon's line-of-sight on the operating field and they are more suitable for use in a surgical scenario as they reduce the logistic burden in the setup phase. When the structures of interest are visible, optical tracking solutions are generally preferable in image-guided surgery applications over electromagnetic tracking for the accuracy they can provide [28], and because they generally do not require wires for connecting the tracked body to the tracker. Electromagnetic trackers are particularly suited

for tracking hidden structures [29], [30], but their accuracy and their reliability are severely degraded by the presence of ferromagnetic and/or conductive materials [31] and by the distance of the tracking body (patient or surgical tool) from the field generator (that should be at most 30 cm [32]).

The tracking algorithm for the estimation of the pose of the optical frame is described in more detail in [33], [34]. A short description of its main features is given below. The algorithm can be broken down into four main steps:

- Marker detection. The centroids of the markers are detected on both camera images through Hue-Saturation-Value (HSV) color space segmentation and blob detection. To increase their detection efficiency, markers are colored with a fluorescent dye, which peaks the S channel and boosts the response of the camera CMOS sensor, thus making segmentation more robust against variations in lighting, shadows and shades [35].
- Stereo matching and stereo triangulation. Epipolar geometry constraints are used to determine the correspondence between points on the left and right camera image. After solving the stereo correspondence, the 3D position of the three markers in the camera coordinate system ${}^{C}L_i$ is computed through stereo triangulation knowing the camera intrinsic parameters and the relative pose between the two cameras.
- First pose estimation step. The 3D-3D correspondence between the two triplets of 3D points, together with the pose between their associated coordinate systems (${}_{O}^{C}\mathbf{T}$) are determined by picking, from among the six possible permutations, the configuration that has the lowest root mean square of the fiducial registration error computed through an unambiguous closed-form fitting method [36].
- Second camera pose estimation step. This final stage refines the pose of the target object with respect to each of the two cameras separately. The method, based on a non-linear iterative Levenberg–Marquardt optimization [37], minimizes a cost function formulated as the sum of the square measurement error (re-projection residuals) between detected image points and calculated projections of the corresponding world points.

E. PROJECTION MODULE CALIBRATION

As with the stereo rig, calibration of the projection module is essential for accurately determining the projection equation that yields a good spatial alignment of the projected virtual content over the real target. The projector is modeled as a "reverse" camera, which projects rather than acquires. Therefore, the considerations about intrinsic and stereo parameters remains valid. The calibration is performed by acquiring a set of images, in which a digital 15×8 chessboard with 80 px side is projected on a plane identified by a 10×5 printed chessboard with a 10 mm side. For example, Figure 3 shows the acquisition of one of the images from the calibration set. The projector is calibrated with respect to the
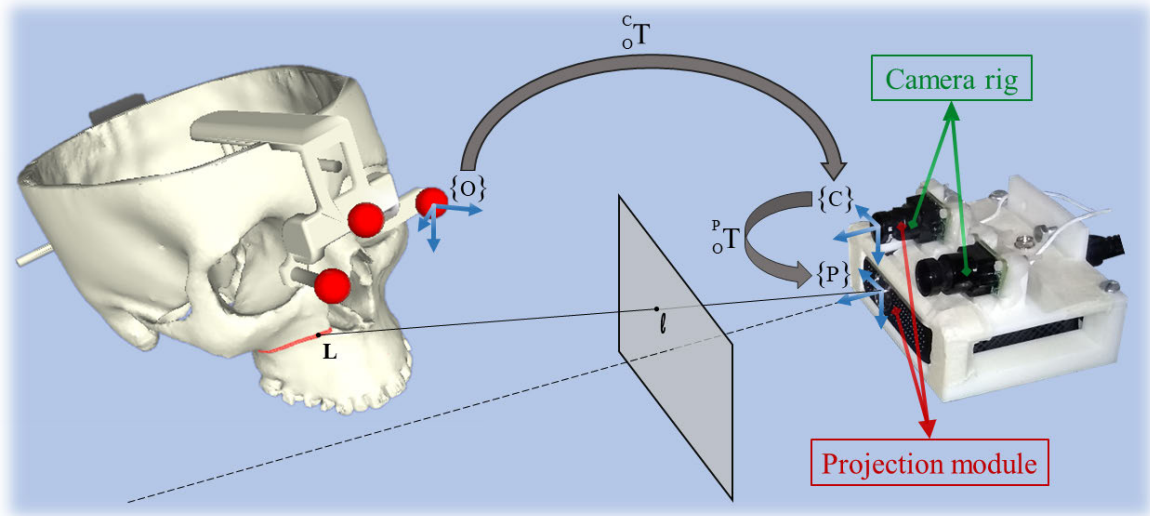
FIGURE 2. Transformations involved in the projected AR image formation. On the left, the skull replica with the optical tracking markers and the corresponding reference system, {O}. On the right, the two functional modules that make up the device, the camera rig and the projection module.
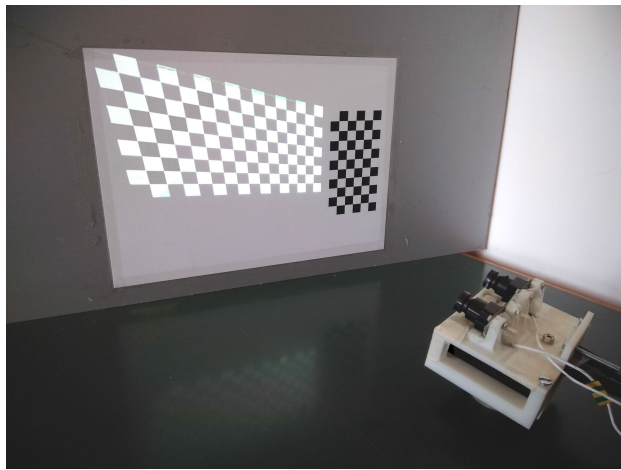

FIGURE 3. The stereo rig captures an image for the calibration set. On the left, the digital 15 × 8 chessboard. On the right is the 10 × 5 printed chessboard.

reference camera by using the method proposed in [38] and it provides us with ${}_C^P\mathbf{T}$ and I_P.

Having determined both transformations in equation 1, it is now possible to change the target pose coordinates from {O} to {P}. We can thus obtain the AR image by mapping the target coordinates onto the projector image plane as follows:

$$\ell = I_P \,{}^P L \qquad (2)$$

where I_P is the projector camera intrinsic matrix.

F. AR SOFTWARE FRAMEWORK

The software framework, whose structure and functionalities were presented in [34], is conceived for the deployment of VST and OST applications with custom-made HMDs and it supports in situ visualization of medical imaging data.

In our work, we adapted the original software to deploy also projected AR.

The AR software framework runs on a laptop with Intel Core i7-9750H CPU@ 2.60 GHz, 6 cores and 12 GB RAM. Graphic card processing unit (GPU) is a Nvidia GeForce RTX 2060.

The software framework is developed in C++ under Linux operating system with an object-oriented design and a CUDA-based architecture to leverage the power of GPU computing. The software relies on OpenCV API (Open Source for Computer Library, version 3.3.1.) for computer vision algorithms and low-level image operations [39]. To render the virtual scene, we used a 3D graphics and visualization VTK library [40]. The software framework allows for switching between three rendering modes. In the experimental tests, we considered only the left camera image for AR accuracy evaluation.

- The VST AR mode produces a $2 \times 1280 \times 720$ stereo image. In this case, the key function of the software is to process, undistort, and augment the images grabbed by the pair of RGB stereo cameras, before they are rendered on the display. Image undistortion is a fundamental step in machine vision because it corrects images from distortions introduced by the lenses. The camera frames are also processed to perform the marker-based optical tracking (see Section II-D. The virtual content is a 3D mesh representing, in our tests, a blue-colored virtual line (Figure 4-A). Details about the generation of the AR images are deepened in [34].

- The projected AR mode works as follows. The projector projects the line directly onto the target object, as described in section II. In this case, the camera frames grabbed for optical tracking are undistorted and rendered without any merged virtual content, as in Figure 4-B.
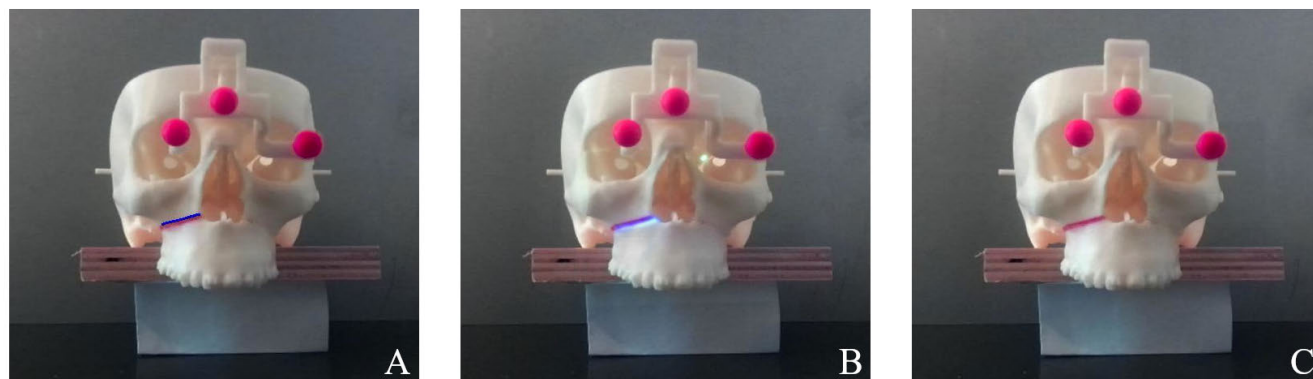
FIGURE 4. Rendering modes A) VST AR mode. B) Projected AR mode. C) Real view mode.

This is not necessary for the user's appreciation of AR, but it is essential for accuracy comparison.

- The real view mode does not generate any AR content and it solely consists in the rendering of the undistorted camera view of the target object. The resulting image serves as a ground truth, as it shows the exact position of the associated real line (shown in red in Figure 4-C).

III. MATERIALS AND METHODS: EXPERIMENTAL PROTOCOL

Projected AR technology and VST AR technologies are compared with one another on the overlay accuracy of the AR features with respect to their exact correct position. This information is obtained through the real view mode that provides the position of the cutting line, colored red on the skull replica. The experimental protocol involves the collection of one hundred image sets, each consisting of three images. The flowchar in Figure 5 shows the location of the various software blocks and their inputs and outputs. The peculiarity of this study is that the camera pair provides the tracking system for both the VST and projected modes. In our experimental tests, the two devices produced AR starting from the same tracking data. This allowed us to directly compare the two AR modes, by removing the tracking accuracy from the variability sources when evaluating AR accuracy. As shown in Figure 5,the acquisition of each individual set begins by capturing a stereo image pair. This is then processed by the marker-based optical tracking that produces the cutting line pose. Following, this information is passed to both the software for the projected AR mode and for the VST AR mode, producing the first two images of the set, the virtual image and the projected image. The third and last image of each set is output from the real view mode. The three images so obtained undergo steps 1, 2, and 3 for the extraction of the centerline of the cutting line, which will be discussed in detail in the next sub-sections. Once identified the centerline of the virtual cutting line and the centerline of the projected cutting line had been identified, we determined their distance from the real cutting line. This provides a pixel measure of the
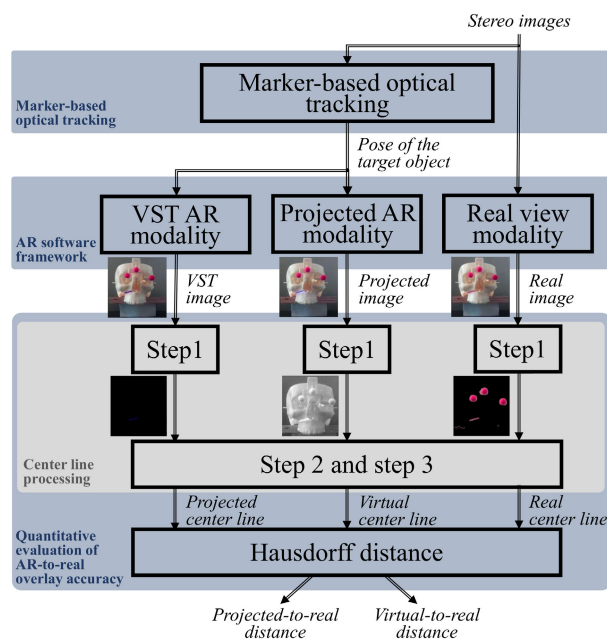

FIGURE 5. Flowchart representing the acquisition and processing for each of the one hundred image sets. Each set provides a measure of AR-to-real overlay accuracy for each of the two modes: projeted-to-real distance and VST-to-real distance.

overlay accuracy of the two modes. The procedure is repeated one hundred times. In the end, one hundred projected-to-real distance values and one hundred VST-to-real values are obtained.

A. EXPERIMENTAL SET-UP

Tests are performed on a 3D-printed replica of a patient-specific skull. As reference feature, we considered a Le Fort 1 osteotomy line. The skull replica presents a real osteotomy line 3D printed on the surface of the skull, according to a simulated virtual plan. In order to facilitate the detection of the correct osteotomy, the real osteotomy was colored using red fluorescent dye. This artificial fracture provides
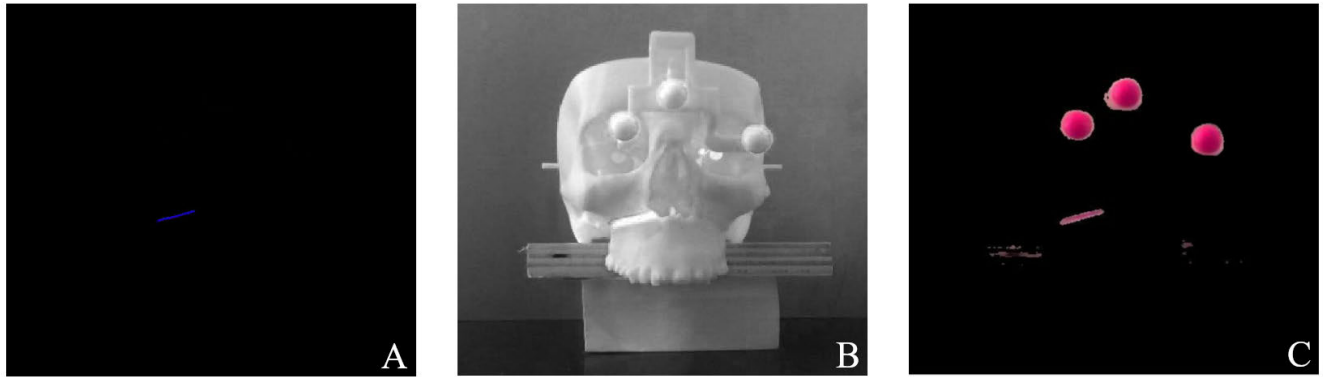
FIGURE 6. Step 1 - image preparation: A) VST image processing to emphasize the virtual osteotomy line; B) Projected AR image processing to emphasize the projected osteotomy line; C) Real view image processing to emphasize the real osteotomy line.
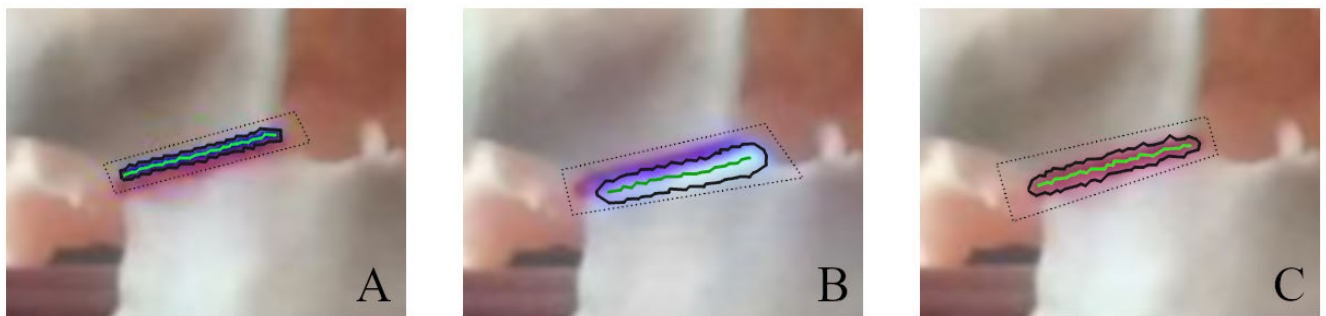

FIGURE 7. Step 2 and 3 - center line formation: A) VST mode; B) Projected AR mode; C) Real view mode. The black dotted line indicates manual selection, the continuous black line indicates the active contours output, and the green line denotes the centerline after skeletonization.

a reference to assess the misalignment with respect to the augmented osteotomy line generated under both VST and projected AR. The patient-specific replica of the skull also embeds the three markers defining {O}.

We acquired a static image (i.e., left camera frames) for each of the three AR modes according to the following steps:

1) The 3D printed replica of the human skull was placed within the field of view of the stereo cameras.
2) The augmented view of the left camera was acquired under VST mode in the AR application.
3) The AR application, switched to projected AR mode, recorded the real view of the left camera with projected AR.
4) The application, switched to real-view mode, recorded the real view of the left camera with no augmentation.

This process was repeated for one hundred different positions of the skull with respect to the projector. The three images of each set were acquired subsequently within a short period of time, taking care that the lighting conditions remained unchanged.

B. QUANTITATIVE EVALUATION OF AR-TO-REAL OVERLAY ACCURACY

In our quantitative evaluation, we compared the overlay accuracy between real and augmented features of VST and projected AR. To this aim, we considered the Hausdorff distance between real and augmented osteotomy lines as a metric for measuring AR overlay accuracy. The Hausdorff distance is defined as [41]:

$$d_H = max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y), \} \quad (3)$$

where X and Y are two subsets of a metric space, in this case the Euclidean space, and d(x, y) is the distance between two generic points in subset X and Y. The Hausdorff distance is the longest of all the distances from a point in one set to the closest point in the other set.

The procedure to extract the pixel coordinates of the osteotomy lines from the images is performed in MATLAB and is broken down into the following steps:

- *Step 1:* The images are converted into the Hue-Saturation-Value (HSV) color scale and processed to emphasize the feature sought within the image. More specifically:
  - For images acquired in VST mode, we looked for the virtual osteotomy line, that has been artificially added into the image through the VTK rendering libraries. Secondly, a band-pass filter centered on the H channel of the virtual line (i.e., blue) and a high pass filter on the S channel were applied.
  - The feature to be found in projected mode is the projected osteotomy line. In this case, only the V channel was isolated, as it contains the information about the brightness of the image. The

pixels representing the projected osteotomy line have intrinsically higher brightness values than the other parts of the image.

- For images acquired in real view mode, the objective was to isolate the real osteotomy line, which is marked in red on the skull dummy. The image is therefore band-pass filtered on H channel to isolate the red and high-pass filtered on the S channel.

The images resulting from these processes are shown in Figure 6.

- *Step 2:* The pixels of the osteotomy line are detected using Chan-Vese active contours region growing technique [42]. This technique begins to iterate from an initial curve, which is manually selected. The evolving curve stops on the detected boundary via Mumford-Shah segmentation and provides positive results even when boundaries are smooth. A particular of the boundaries before and after the region growing technique is depicted in Figure 7: the coarse boundaries are dashed in black, the boundaries detected by the Chan-Vese algorithm are highlighted with a black continuous line.
- *Step 3:* Finally, the contours were collapsed to a one-pixel wide curve, as shown in green in Figure 7. This process is called skeletonization and it extracts the centerline while preserving the topology of the objects [43].

The proximity between the virtual and projected centerlines and the real centerlines can now be evaluated by measuring the Hausdorff distance between them. In Figure 8, Hausdorff distances are respectively 5 and 4.5 pixels. These values can be converted into visual angles using the conversion factor R, which expresses the visual angle associated to a single camera pixel. A comparison in pixels could be distorted by the different characteristics of the camera sensors or optics used in the experiments. By means of example, the same error in pixels could be associated with different visual angle values if the cameras being compared are different. Hereafter, the overlay errors are always expressed in arcmin rather than pixels, because they do not depend on the measurement tool used (i.e., the camera).

The two overlay errors, real-virtual and real-projected, are first analyzed individually using Shapiro-Wilk's tests to determine whether they come from normal distributions [44]. Their correlation is also evaluated. Finally, we select the best method to determine the presence of a significant difference between the samples.

IV. RESULTS

The key prerequisite for ensuring a good overlay accuracy between real and AR features is a proper calibration of the AR system. Below are the calibration accuracy indicators, from the camera intrinsic calibration to the camera-projector calibration. The following sub-section provides an in-depth analysis of the experimental test results.

The intrinsic camera calibration showed a standard percentage error on focal lengths and principal points of less than
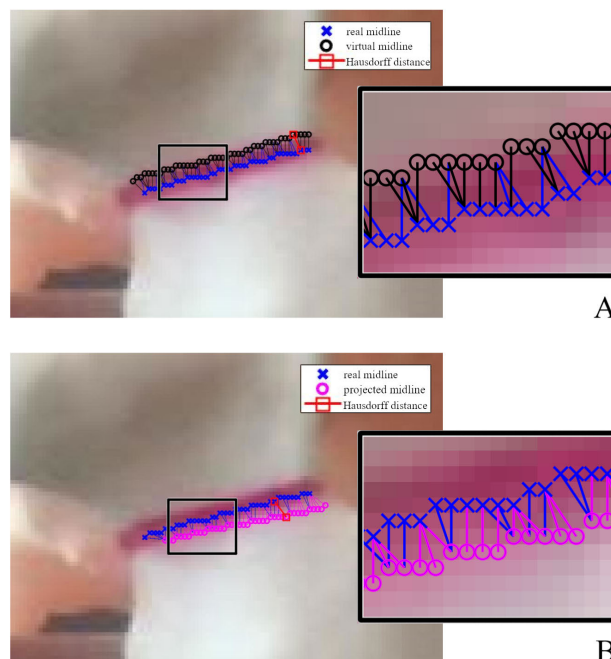

FIGURE 8. A) Calculation of the real-to-virtual Hausdorff distance. Points of the virtual osteotomy line are represented by black circles. B) Calculation of the real-to-projected Hausdorff distance. Points of the projected osteotomy line are represented by pink circles. Points of the real osteotomy line are marked with blue crosses. The points resulting from the Hausdorff algorithm are highlighted in red and the line connecting them is the Hausdorff distance between the two curves.

0.3% of the nominal value for both of them. Another indicator of intrinsic calibration accuracy is the re-projection error, which is related to the distance between a real chessboard corner detected in a calibration image, and its respective 3D point projected by using the estimated projection parameters. The right camera has an average re-projection error of 0.17 pixels, the left camera 0.19 pixels.

As for the calibration error associated with the stereo calibration, we estimated a translation error of 0.04 mm and a rotation error of 0.01° which are equivalent to 0.09% and 1.39% of the nominal values respectively. The mean epipolar error of the stereo calibration was 0.26 pixels. In a stereo camera configuration, a point in one camera view must fall along a unique line in the other camera view, named epipolar line. The epipolar error is the distance from this line to its corresponding point in the other camera image, and it is used to assess extrinsic calibration quality.

The standard percentage error on the projector intrinsic parameters is less than 0.5% for the focal length and less than 1% for the principal point. The re-projection error is 0.74 pixels, and it refers, in this case, to the pixel difference between points of the AR chessboard observed over the camera image and the respective projected 3D points.

A. EXPERIMENTAL TEST RESULTS
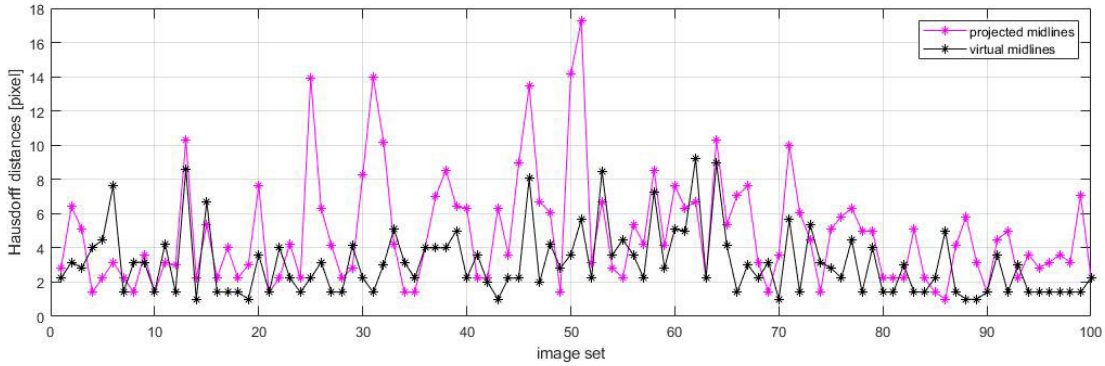The overlay errors are shown in Figure 9, whereas Figure 10 shows their histograms. The Shapiro-Wilk test rejected the

FIGURE 9. Hausdorff distances. Hausdorff distances between the real osteotomy line and the AR osteotomy line under VST mode are black. Hausdorff distances between the real osteotomy line and the AR osteotomy line under projected mode are pink.

TABLE 1. Statistics and p-values of the statistical tests.

test	statistic value	p-value
Shapiro-Wilk (real-to-virtual sample)	0.851	$1.260*10^{-7}$
Shapiro-Wilk (real-to-projected sample)	0.859	$2.353*10^{-7}$
Spearman's rank correlation	0.297	$2.736*10^{-3}$
Wilcoxon signed-rank	$3.599*10^{3}$	$7.256*10^{-7}$

TABLE 2. Median, M, and interquartile interval, IQR, of real-projected and real-virtual overlay errors expressed in image plane distance (pixels), angular difference (arcmin), and 3D space distance (mm).

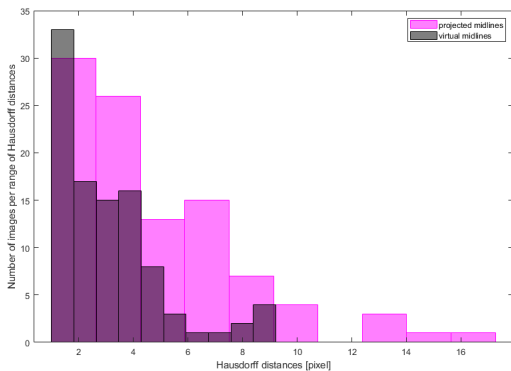| overlay error	pixels		arcmin		mm	
	M	IQR	M	IQR	M	IQR
real-to-projected	4.06	4.09	14.76	14.86	0.96	0.93
real-to-virtual	2.53	2.59	9.21	9.40	0.65	0.61

FIGURE 10. Histograms of the Hausdorff distances in VST mode (shown in black), and of the Hausdorff distances in projected mode (shown in pink).

null hypothesis of normal distribution with a 5% significance level for both samples. The tests statistics are listed in Table 1. The overlay errors, given in terms of median, M, and interquartile deviation, IQR, are M = 14.8 arcmin and IQR = 14.9 arcmin for real-projected errors, and M = 9.2 arcmin and IQR = 9.4 arcmin for real-virtual errors. These quantities are translated into spatial distances using the

information about the osteotomy line depth from the stereo rig:

$$err_{mm} \simeq Z * \tan(err_{arcmin}/60) \qquad (4)$$

The statistical indicators of the samples are shown in Table 2 in terms of arcmin, pixels, and spatial distance.

We chose the Spearman's rank correlation test to verify whether the relationship between the two samples can be described using a monotonous function [45]. The test verified the null hypothesis of ρ being significantly different from zero with a 5% significance level. Given the statistical characteristics of the samples, non-normal distributions and correlated data, we chose the non-parametric Wilcoxon signed-rank test to verify whether the two samples result from different distributions [46]. Given a 5% significance level, the test rejected the null hypothesis that the difference between the two samples came from a distribution with null median. Both Spearman and Wilcoxon statistics are shown in Table 1.

V. DISCUSSION AND CONCLUSION

In this work, we presented a lightweight and portable device for guiding surgery through projected AR. The AR features we project are surface information, as in the case of osteotomy lines. For this reason our study is not affected by the parallax error, that is typical of projected AR systems displaying structures beyond the projection surface. We compared our device to a VST HMD using the same stereo rig that is normally used for the wearable viewer.

When capturing each set of images, attention was paid to ensure that lighting conditions remained unchanged. Lighting has a significant impact both on tracking quality and on the visualization of the augmented content. As to tracking quality, the two modes receive the same tracking data, so lighting is not a source of variability, in this respect. As for visualization, the images used to perform the measurements for each set were captured under the same lighting conditions. However, a real use case could lead to complications due to the scialytic lamp, and future work will involve testing the device under the same lighting conditions as in the operating room.

The experimental tests were performed using the same tracking data for the AR generation in the two modes. The difference in the overlay error between the AR features and the ground truth is therefore only due to the difference in the AR modes.

A weak correlation in the trend of the real-to-virtual and real-to-projected overlay errors was found. This was expected because both AR modes rely on the same tracking data. However, the real-to-projected overlay error presented slightly higher values than the real-to-virtual error. This suggests that, whereas the tracking error is the same for both modes, the outcome of its propagation leads to a less accurate AR overlay in the projected mode. The tendency of the projected AR error to be greater with respect the VST AR error is confirmed by the Wilcoxon test. This may be due to the lower accuracy in the projector-camera calibration and to the extra error source introduced by the projection over a 3D surface. With regard to calibration steps, calibration results were quite accurate for the stereo cameras with very low re-projection and epipolar errors ranging in the tenth of pixel (0.17 and 0.19 pixels for right and left cameras re-projection error and 0.26 pixels for the epipolar error). On the other hand, the calibration of the projection module has a re-projection error of 0.74 pixels. Moreover, projected AR has an additional source of variability, because it is not limited to generating AR on an image (i.e., the camera image), but it provides AR directly on the target 3D surface, which makes projection accuracy sensitive to the target surface. In spite of that, the overlay accuracy of both AR modes falls within the accepted range for maxillofacial surgery of 1-2 mm ($M = 0.65$ mm for real-to-virtual error and $M = 0.96$ mm for real-to-projected error). Instead, when moving from mere visualization to the execution of a manual task, VST AR presents an additional variability term caused by the perspective conversion and the optical mediation of the display and camera lenses which both affect the visual perception of the augmented scene. Projected AR allows us to prevent the perceptual problems typically associated with AR HMDs, as it provides information exactly where it is needed, without placing screens or lenses between the surgeon and the patient.

Although the overlay error is greater for projected AR, it falls within the accepted range for maxillofacial surgery. Also, projected AR prevents the double vision, visual discomfort, and visual fatigue typical of display-mediated views under VST or OST HMDs. These perceptual limitations have been shown to hinder how manual tasks are performed [9]. Therefore, we can expect that the increased error in the projected mode may not necessarily result in a reduced performance accuracy when the user interacts with this information to follow, for instance, a projected osteotomy line.

In conclusion, this study found that the difference in the accuracy of the projected mode compared to the VST mode does not compromise its use in osteotomy surgery. Given the perceptual advantages of the projected mode, future work will include a user study based on the execution of incision-like tasks, also involving maxillofacial surgeons. The study will aim at comparing the efficacy of the two AR modes in aiding an osteotomy task for AR-guided cranio-maxillofacial surgery. Future work will also focus on improving the accuracy of the camera-projector calibration to remove its influence on the real-projected overlay error.

REFERENCES

[1] S. C. Gupta, S. A. Klein, J. H. Barker, R. J. P. M. Franken, and J. C. Banis, Jr., "Introduction of new technology to clinical practice: A guide for assessment of new VR applications," *J. Med. Virtual Reality*, vol. 1, no. 1, pp. 16–20, 1995.

[2] P. Pratt, M. Ives, G. Lawton, J. Simmons, N. Radev, L. Spyropoulou, and D. Amiras, "Through the HoloLens looking glass: Augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels," *Eur. Radiol. Exp.*, vol. 2, no. 1, p. 2, Dec. 2018, doi: 10.1186/s41747-017-0033-2.

[3] R. Diaz, J. Yoon, R. Chen, A. Quinones-Hinojosa, R. Wharen, and R. Komotar, "Real-time video-streaming to surgical loupe mounted head-up display for navigated meningioma resection," *Turkish Neurosurg.*, vol. 28, no. 4, pp. 682–688, Jan. 2018.

[4] F. Cutolo, "Letter to the editor on 'augmented reality based navigation for computer assisted hip resurfacing: A proof of concept study,'" *Ann. Biomed. Eng.*, vol. 47, no. 11, pp. 2151–2153, Nov. 2019, doi: 10.1007/s10439-019-02299-w.

[5] K. Oshima, K. R. Moser, D. C. Rompapas, J. E. Swan, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato, "SharpView: Improved clarity of defocussed content on optical see-through head-mounted displays," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2016, pp. 253–254.

[6] H. Hua, "Enabling focus cues in head-mounted displays," in *Proc. Imag. Appl. Opt., 3D, AIO, COSI, IS, MATH, pcAOP*, 2017, p. 1, Paper JTu1F.2. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=AIO-2017-JTu1F.2

[7] J. L. Gabbard, D. G. Mehra, and J. E. Swan, "Effects of AR display context switching and focal distance switching on human performance," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 6, pp. 2228–2241, Jun. 2019.

[8] N. Cattari, F. Cutolo, R. D'amato, U. Fontana, and V. Ferrari, "Toed-in vs parallel displays in video see-through head-mounted displays for close-up view," *IEEE Access*, vol. 7, pp. 159698–159711, 2019.

[9] S. Condino, M. Carbone, R. Piazza, M. Ferrari, and V. Ferrari, "Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 411–419, Feb. 2020.

[10] V. Ferrari, M. Carbone, S. Condino, and F. Cutolo, "Are augmented reality headsets in surgery a dead end?" *Expert Rev. Med. Devices*, vol. 16, no. 12, pp. 999–1001, Dec. 2019, doi: 10.1080/17434440.2019.1693891.

[11] G. Badiali, L. Cercenelli, S. Battaglia, E. Marcelli, C. Marchetti, V. Ferrari, and F. Cutolo, "Review on augmented reality in oral and cranio-maxillofacial surgery: Toward 'surgery-specific' head-up displays," *IEEE Access*, vol. 8, pp. 59015–59028, 2020.

[12] M. Carbone, R. Piazza, and S. Condino, "Commercially available head-mounted displays are unsuitable for augmented reality surgical guidance: A call for focused research for surgical applications," *Surgical Innov.*, vol. 27, no. 3, pp. 254–255, 2020, doi: 10.1177/1553350620903197.

[13] A. Mewes, F. Heinrich, U. Kägebein, B. Hensen, F. Wacker, and C. Hansen, "Projector-based augmented reality system for interventional visualization inside MRI scanners," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 15, no. 1, p. e1950, Feb. 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1950

[14] R. Gierwiało, M. Witkowski, M. Kosieradzki, W. Lisik, Ł. Groszkowski, and R. Sitnik, "Medical augmented-reality visualizer for surgical training and education in medicine," *Appl. Sci.*, vol. 9, no. 13, p. 2732, Jul. 2019.

[15] P. Edgcumbe, R. Singla, P. Pratt, C. Schneider, C. Nguan, and R. Rohling, "Follow the light: Projector-based augmented reality intracorporeal system for laparoscopic surgery," *J. Med. Imag.*, vol. 5, no. 2, p. 1, Feb. 2018.

[16] V. Mamone, M. D. Fonzo, N. Esposito, M. Ferrari, and V. Ferrari, "Monitoring wound healing with contactless measurements and augmented reality," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–12, 2020.

[17] L. Tabrizi and M. Mahvash, "Augmented reality–guided neurosurgery: Accuracy and intraoperative application of an image projection technique," *J. Neurosurg.*, vol. 123, pp. 1–6, Mar. 2015.

[18] V. Ferrari and F. Cutolo, "Letter to the editor: Augmented reality–guided neurosurgery," *J. Neurosurg.*, vol. 125, no. 1, pp. 235–237, Jul. 2016. [Online]. Available: https://thejns.org/view/journals/j-neurosurg/125/1/article-p235.xml

[19] J. I. Efanov, A. A. Roy, K. N. Huang, and D. E. Borsuk, "Virtual surgical planning: The pearls and pitfalls," *Plastic Reconstructive Surg.-Global Open*, vol. 6, no. 1, p. e1443, 2018, doi: 10.1097/GOX.0000000000001443.

[20] E. Mussi, F. Mussa, C. Santarelli, M. Scagnet, F. Uccheddu, R. Furferi, Y. Volpe, and L. Genitori, "Current practice in preoperative virtual and physical simulation in neurosurgery," *Bioengineering*, vol. 7, no. 1, p. 7, Jan. 2020, doi: 10.3390/bioengineering7010007.

[21] S. Sukegawa, T. Kanno, and Y. Furuki, "Application of computer-assisted navigation systems in oral and maxillofacial surgery," *Jpn. Dental Sci. Rev.*, vol. 54, no. 3, pp. 139–149, Aug. 2018.

[22] P. Tessier, "The classic reprint: Experimental study of fractures of the upper Jaw. 3. René le fort, M.D., Lille, France," *Plastic Reconstructive Surg.*, vol. 50, no. 6, pp. 600–607, Dec. 1972. [Online]. Available: http://europepmc.org/abstract/MED/4564611

[23] C. Hyman and E. Buchanan, "LeFort i osteotomy," *Seminars Plastic Surg.*, vol. 27, no. 3, pp. 149–154, Oct. 2013.

[24] H. S. Cho, Y. K. Park, S. Gupta, C. Yoon, I. Han, H.-S. Kim, H. Choi, and J. Hong, "Augmented reality in bone tumour resection: An experimental study," *Bone Joint Res.*, vol. 6, no. 3, pp. 137–143, Mar. 2017.

[25] P. J. Edwards, L. G. Johnson, D. J. Hawkes, M. R. Fenlon, A. J. Strong, and M. J. Gleeson, "Clinical experience and perception in stereo augmented reality surgical navigation," in *Medical Imaging and Augmented Reality*, G.-Z. Yang and T.-Z. Jiang, Eds. Berlin, Germany: Springer, 2004, pp. 369–376.

[26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[27] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 1106–1112.

[28] R. Hussain, A. Lalande, C. Guigou, and A. Bozorg-Grayeli, "Contribution of augmented reality to minimally invasive computer-assisted cranial base surgery," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2093–2106, Jul. 2020.

[29] V. Ferrari, R. M. Viglialoro, P. Nicoli, F. Cutolo, S. Condino, M. Carbone, M. Siesto, and M. Ferrari, "Augmented reality visualization of deformable tubular structures for surgical simulation," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 12, no. 2, pp. 231–240, Jun. 2016, doi: 10.1002/rcs.1681.

[30] R. Viglialoro, S. Condino, M. Gesi, M. Ferrari, and V. Ferrari, "Augmented reality simulator for laparoscopic cholecystectomy training," in *Proc. Int. Conf. Augmented Virtual Reality*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 8853, 2014, pp. 428–433.

[31] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein, "Electromagnetic tracking in medicine—A review of technology, validation, and applications," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1702–1725, Aug. 2014.

[32] J. Nijkamp, B. Schermers, S. Schmitz, S. de Jonge, K. Kuhlmann, F. van der Heijden, J.-J. Sonke, and T. Ruers, "Comparing position and orientation accuracy of different electromagnetic sensors for tracking during interventions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 8, pp. 1487–1498, Aug. 2016, doi: 10.1007/s11548-015-1348-1.

[33] F. Cutolo, V. Mamone, N. Carbonaro, V. Ferrari, and A. Tognetti, "Ambiguity-free optical–inertial tracking for augmented reality headsets," *Sensors*, vol. 20, no. 5, p. 1444, Mar. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/5/1444

[34] 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.

[35] B. Diotte, P. Fallavollita, L. Wang, S. Weidert, E. Euler, P. Thaller, and N. Navab, "Multi-modal intra-operative navigation during distal locking of intramedullary nails," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 487–495, Feb. 2015.

[36] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, May 1987, doi: 10.1109/TPAMI.1987.4767965.

[37] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, Jun. 1963. [Online]. Available: http://www.jstor.org/stable/2098941

[38] W.-C. Chang and C.-H. Wu, "Plane-based geometric calibration of a projector-camera reconstruction system," in *Proc. 10th France-Japan/8th Europe-Asia Congr. Mecatronics (MECATRONICS-Tokyo)*, Nov. 2014, pp. 219–223.

[39] (Jul. 2020). *OpenCV*. [Online]. Available: http://opencv.org/

[40] (Jul. 2020). *VTK, the Visualization Toolkit*. [Online]. Available: http://vtk.org/

[41] R. Rockafellar and R. Wets, *Variational Analysis*, 2nd ed. Springer, 2004, doi: 10.1007/978-3-642-02431-3.

[42] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[43] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-D medial surface/axis thinning algorithms," *CVGIP Graph. Model. Image Process.*, vol. 56, no. 6, pp. 462–478, 1994.

[44] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965. [Online]. Available: http://www.jstor.org/stable/2333709

[45] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904. [Online]. Available: http://www.jstor.org/stable/1412159

[46] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945. [Online]. Available: http://www.jstor.org/stable/3001968
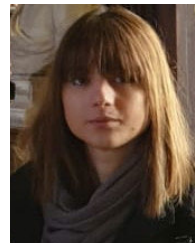
VIRGINIA MAMONE received the B.Sc. and M.Sc. degrees in biomedical engineering from the University of Pisa, Pisa, Italy, in 2016. She is currently pursuing the Ph.D. degree in information engineering with the EndoCAS Center for Computer Assisted Surgery. She has managed a Project of monitoring skin ulcers and developed an innovative device that gives both the clinician and the patient a visual feedback on the wound healing through projected AR. She also developed software for correct and automatic replacement of the aortic valve by means of an articulated robot. Her research question involves providing the surgeon with AR support for the guidance of minimally invasive surgeries in order to facilitate and improve the treatment quality.

VINCENZO FERRARI (Member, IEEE) received the Ph.D. degree from the University of Pisa. He is currently an Assistant Professor of biomedical engineering with the Department of Information Engineering, University of Pisa, where he is also a Coordinator with the EndoCAS Centre. He is the author of more than 80 peer-reviewed publications. He holds five patents. He was involved in several national and international research projects. He is a Coordinator with the Horizon 2020 Project VOSTARS (Call ICT-29-2016). His research interests include image-guided surgery and simulation, computer vision, and AR devices and applications in medicine.

SARA CONDINO received the M.Sc. degree (Hons.) in biomedical engineering and the Ph.D. degree in translational medicine from the University of Pisa, Pisa, Italy, in 2008 and 2012, respectively. She is currently a Techologist with the Department of Information Engineering, University of Pisa. She was involved in several national and international research projects. Her research interests include computer-assisted surgery, design of innovative instrumentation for surgical navigation, design of virtual and augmented reality applications for surgical navigation and simulation, design of innovative technologies based on virtual reality, and physical and hybrid simulation for rehabilitation specialists.

FABRIZIO CUTOLO (Member, IEEE) received the M.Sc. degree in electrical and computer engineering and the Ph.D. degree in translational medicine from the University of Pisa, Pisa, Italy, in 2006 and 2015, respectively. He is currently a Postgraduate Research Associate with the Department of Information Engineering, University of Pisa. He was involved in several national and international research projects. He is also a WP Leader with the Horizon 2020 Project VOSTARS (Call ICT-29-2016). His research interests include developing and evaluating new AR solutions for image-guided surgery and surgical simulation, machine-vision applications, visual perception, ubiquitous tracking, and human-machine interfaces for rehabilitation.

● ● ●