

Received December 5, 2019, accepted December 18, 2019, date of publication December 24, 2019, date of current version January 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962122

# Software Framework for Customized Augmented Reality Headsets in Medicine

FABRIZIO CUTOLO<sup>1</sup>, (Member, IEEE), BENISH FIDA<sup>1</sup>, NADIA CATTARI<sup>2</sup>,  
AND VINCENZO FERRARI<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

<sup>2</sup>EndoCAS Centre, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56124 Pisa, Italy

Corresponding author: Fabrizio Cutolo (fabrizio.cutolo@endocas.unipi.it)

This work was supported by the HORIZON2020 Project VOSTARS (Grant Number 731974). Call ICT-29-2016, Photonics KET 2016.

**ABSTRACT** The growing availability of self-contained and affordable augmented reality headsets such as the Microsoft HoloLens is encouraging the adoption of these devices also in the healthcare sector. However, technological and human-factor limitations still hinder their routine use in clinical practice. Among them, the major drawbacks are due to their general-purpose nature and to the lack of a standardized framework suited for medical applications and devoid of platform-dependent tracking techniques and/or complex calibration procedures. To overcome such limitations, in this paper we present a software framework that is designed to support the development of augmented reality applications for custom-made head-mounted displays designed to aid high-precision manual tasks. The software platform is highly configurable, computationally efficient, and it allows the deployment of augmented reality applications capable to support in situ visualization of medical imaging data. The framework can provide both optical and video see-through-based augmentations and it features a robust optical tracking algorithm. An experimental study was designed to assess the efficacy of the platform in guiding a simulated task of surgical incision. In the experiments, the user was asked to perform a digital incision task, with and without the aid of the augmented reality headset. The task accuracy was evaluated by measuring the similarity between the traced curve and the planned one. The average error in the augmented reality tests was < 1 mm. The results confirm that the proposed framework coupled with the new-concept headset may boost the integration of augmented reality headsets into routine clinical practice.

**INDEX TERMS** Augmented reality, image-guided surgery, head-mounted displays, application software, target tracking.

## I. INTRODUCTION

The healthcare sector represents one of the most promising and fascinating fields of application for visual augmented reality (AR), with potential uses in medical education and training, surgical planning, remote surgery, robot-assisted surgery, and surgical navigation [1], [2].

Particularly in image-guided surgery, the need to integrate medical imaging into the surgical workflow has encouraged the research for new visualization modalities based on AR that could act as surgical guidance or alternatively as tool for surgical planning and/or diagnosis [3]–[6]. AR interfaces have the potential to shift the paradigm of how

medical imaging is commonly deployed into the operating room (OR). This is owing to the ability of AR to allow the ubiquitous enrichment of the surgical scene with computer-generated elements derived from medical datasets: AR technology is able to provide physicians with a virtual navigation aid contextually blended with the real surgical scenario (i.e., in situ) [7], [8]. Recently, this trend has been further supported by the increasing capability of mobile graphics and computing power that has led to the development of self-contained and affordable AR headsets such as the Microsoft HoloLens, the Meta Two, and the MagicLeap One [9].

Wearable AR systems based on head-mounted displays (HMDs) are deemed as the most ergonomic and effective solutions to guide those procedures that are manually performed under the surgeon's direct vision due to their ability

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoxiang Zhang.

to preserve the user's egocentric viewpoint [6], [10]. This applies for instance to all those surgical procedures that involve the cutting/incision of exposed human body tissues (i.e., epithelial tissue, muscle tissue, connective tissue, and nervous tissue) and therefore it involves almost all the surgical sub-specialties.

From a technological standpoint, AR HMDs can be categorized according to the see-through paradigm they implement: video see-through (VST) HMDs and optical see-through (OST) HMDs [11].

In standard OST HMDs, the user's visual perception of the real world is augmented by rendering the virtual content on a two-dimensional (2D) micro display placed outside the user's field of view, and by sending the display images to the user's eye by means of an optical combiner [12]. A collimation optics (i.e., the eyepiece) is placed between the optical combiner and the micro display to focus the virtual image so that it appears magnified at a comfortable viewing distance on a semitransparent surface of projection [13].

Differently, in VST HMDs, the direct perception of the world is not preserved since it is mediated by one or two front-facing cameras mounted on the visor. In this way, the camera views of the world are first digitally blended with the virtual content and then rendered on the micro displays of the visor.

Nowadays, OST HMDs are the leading edge and the major output medium of wearable AR technology, and several consumer level headsets have been recently developed following the success of the Microsoft HoloLens. Nevertheless, even if AR technology is continuously evolving, technological and human-factor limitations still hinder the routine use of such devices in routine clinical practice [14].

The major technological limitations of commercial OST HMDs, are due to their general-purpose nature and the lack of a standardized software framework for medical applications. As regards the first limitation, AR headsets were, and still are, mostly designed for parallel viewing: the virtual content is projected at a fixed focal distance (normally between 2 m and infinity), and therefore they generate perceptual conflicts, such as vergence-accommodation conflict and focus rivalry [15], [16], when used to interact with objects closer to the viewer's eyes (i.e., at arm's reach) [17]. In addition, when embedded with tracking sensors and computing units, they are generally quite cumbersome and thus rather uncomfortable for a prolonged use with the head tilted such as during manual tasks. These aspects raise serious concerns regarding the effectiveness of such devices to aid manual tasks that require high level of hand-eye coordination (e.g., in surgery). As regards the second limitation, most commercial HMDs do not take into account the operational constraints imposed by the surgical context and none of them is provided with a device-independent software framework specifically suited to surgical guidance [18].

The basic condition for the acceptance of a new technology, such as AR HMDs, in the OR is related to its ability of being smoothly integrated into the workflow of the intervention, without affecting and disturbing the surgeon's activity during

the rest of the procedure [19]. This principle applies not only to the hardware, but also to the software architecture, which is the core of the AR applications (ARAs). This means that the software framework should be as much as possible device-independent and highly configurable in order to be tailored to different surgical scenarios and applications.

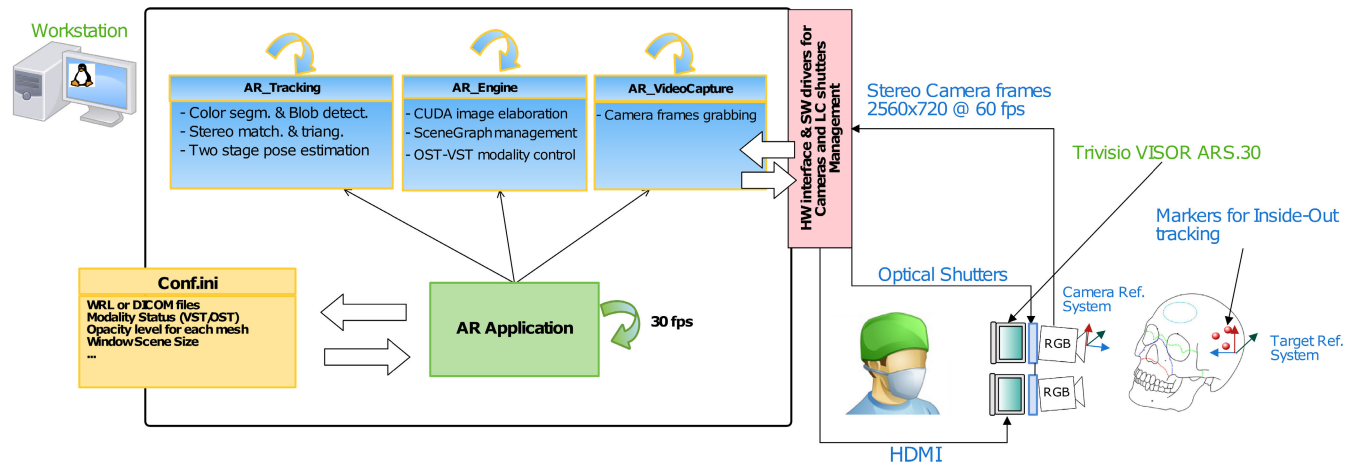
To address these challenges, since 2016, we have been coordinating the European project VOSTARS (Video and Optical See-Through Augmented Reality Surgical Systems, Project ID: 731974 [20]). The project aim is to develop a new-concept AR headset able to provide both video and optical see-through-based augmentations and to validate it as tool for surgical guidance. The main goal of the project is to combine the advantages of both the see-through paradigms towards the definition of AR visualization modalities capable to adapt to different phases of the surgical workflow.

In this paper, we describe the components and the features of the software framework that were developed in the course of the project and we also unveil the most relevant properties of an early version of the custom-made hybrid video/optical see-through HMD that was developed and used as testing platform.

In addition, we also provide a qualitative evaluation and a quantitative assessment of the AR software framework through a user study. The experimental study aims at assessing the efficacy of the proposed AR platform (custom-made HMD plus software framework) in guiding a simulated task of tissue incision. The main contributions of the work are:

- A software framework capable to support the deployment of AR applications on customized headsets for image-guided surgery and surgical simulation.
- A software framework implemented in CUDA architecture capable to deploy both optical and video see-through-based augmentations in a computationally efficient fashion (average frame period  $\sim 0.029$  s).
- A software framework that features an highly optimized inside-out optical tracking algorithm specifically suited for use in a surgical scenario (processing speed for the tracking  $\sim 0.007$  s).
- A software framework that is highly configurable in terms of rendering or tracking features.
- A custom-made stereoscopic hybrid video-optical see-through HMD that was designed to fulfil strict requirements towards the realization of a functional and reliable AR-based surgical navigator.
- An experimental study in which the software framework and the custom-made HMD were tested in terms of task performance, efficiency, and usability.

The rest of the paper is organized as follows. Section II highlights the functionality of the existing AR frameworks in medical applications. A detailed description of the hardware and software components and AR application and the shared library classes with their functionalities is provided in section III. In section IV, the experimental study protocol and its assessment are described in detail. The result of the study



**FIGURE 1.** Overview of the hardware and software components of the Augmented Reality (AR) platform for surgery. The AR framework runs on a single workstation and can implement both the optical see-through (OST) and the video see-through (VST) mechanisms.

and AR software usability are discussed in the Section V, while Section VI draws the conclusions and future work.

## II. RELATED WORKS

In literature, few fully functional AR framework for medical applications have been proposed.

In 2006, a standardized software architecture for computer-assisted surgery, named CAMPAR, was proposed [21]. The software integrates methods for image processing and visualization of medical volume data together with an efficient synchronization mechanism based on network time protocol; the protocol guarantees the correct integration of multiple tracking and visualization systems from different manufacturers. This platform was used in studies designed for the improvement of the intraoperative AR visualization and depth perception in laparoscopic surgery with an endoscope [22] and in orthopedic and trauma surgery with a custom-made VST HMD [23] or a re-engineered video augmented mobile C-arm system [24].

Another highly distributed software framework tailored for the development of projection-based ARAs in the OR, was proposed in 2016 [25]. The core of the multi-layer architecture is based on a communication module implemented using Google Protocol Buffers [26] for exchanging messages between peers over a transport layer.

In [18], an AR framework for surgical guidance in minimally invasive surgery was proposed. Here too, the framework implements a distribute architecture based on a different open-source protocol (OpenIGTLink [27]) to perform inter-processes communications ensuring high interoperability. The framework was tested for the intraoperative guidance during laparoscopic liver surgery in combination with Da Vinci's surgical robot.

Overall, all the mentioned frameworks rely on a distributed architecture and are therefore based on the appropriate selection of a dedicated protocol for synchronization and communication between different computing units and/or tracking or visualization devices.

Our software framework is architecturally simpler since it runs on a single computing platform and it features a traditional architecture with a single process running (1 executable file) and multiple shared libraries (1 library per module). Our goal was to implement a framework that could be ultimately compatible with the adoption of an embedded computing unit so to reinforce the compactness of the whole AR platform. The software framework is highly versatile and configurable thanks to its modularity, and it supports the deployment of AR applications on customized headsets for image-guided surgery and surgical simulation.

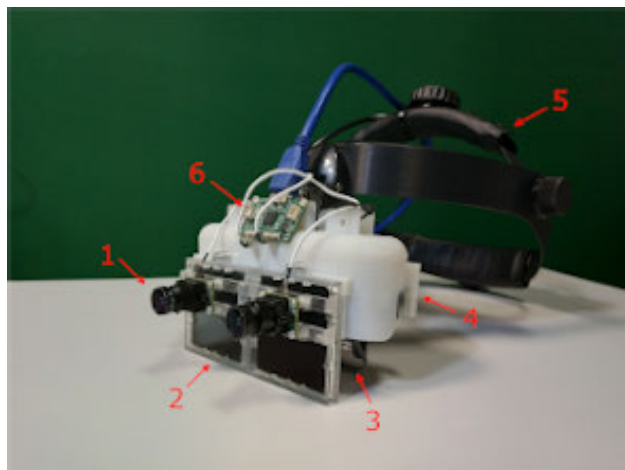
## III. MATERIALS AND METHODS

This section provides a detailed description of the hardware and software components. All components are depicted in Fig. 1.

### A. HARDWARE COMPONENTS

The AR framework runs on a standard workstation class PC with the following specifications: Intel Core i7-4770 CPU @ 3.40 GHz with 4 cores and 12 GB RAM. Graphic card processing unit (GPU) is a Nvidia GeForce GTX 1050 (2GB) with 640 CUDA Cores.

Our quasi ortho-stereoscopic HMD for AR-based surgical navigation was designed and assembled by reworking and re-engineering a commercial binocular OST visor (ARS.30 by Trivisio [28]) with a similar approach to our previous works [29], [30] (Fig. 2). Our HMD is able to yield both the see-through mechanisms (VST and OST) through the use of a pair of liquid-crystal (LC) optical shutters placed in front of the beam combiner of the see-through displays [31]. The LC panels can be electronically controlled to modify the transparency of the display. This feature allows switching between the unaided binocular view (i.e., OST mode with shutters off) and the camera-mediated view (i.e., VST mode with shutters on). Under OST mode, only the computer-generated elements are rendered onto the two microdisplays



**FIGURE 2.** The custom-made hybrid video/optical see-through head-mounted display. 1→Pair of stereo camera for the inside-out optical tracking and the camera-mediated view. 2→Pair of LC optical shutters for the video-optical switching mechanism. 3→Beam combiner of the see-through display. 4→Plastic frame that holds all the components around the optical see-through visor. 5→The head-mount. 6→The electronic board for the stereo synchronization of the camera frames.

of the visor, whereas under VST mode the real views of the world are grabbed by the external RGB cameras and the virtual elements are digitally added to them before the augmented frames are rendered on the two microdisplays.

The ARS.30 visor is provided with dual SXGA OLED panels with 1280x1024 resolution, a diagonal field-of-view (FOV) of 30° and an eye-relief of 3 cm each. The two panels are controlled by the workstation via HDMI and the angular resolution of the OST display is  $\approx 1.11$  arcmin/pixel. The collimation optics of the visor used in our experiments was entirely re-engineered to offer a focal length of about 50 cm, which represents a defining and original feature to mitigate the vergence-accommodation conflict and the focus rivalry when used for close-up works. By way of example, the Microsoft HoloLens projects the “hologram” at a fixed distance of about 2m, which makes it perceptually uncomfortable for aiding high-precision manual tasks [14].

The visor comprises also a head support with a flip-up mechanism as that of a standard magnifying glasses headset or a welding helmet. This allows reducing head tilt during manual tasks and thus it increases the overall ergonomics and usability of the HMD for close-up works.

A 3D printed plastic frame was also built to incorporate the two LC shutters and to act as support for the pair of front-facing USB 3.0 RGB cameras. The stereo camera pair is composed by two LI-OV4689 cameras by Leopard Imaging, both equipped with 1/3” OmniVision CMOS 4M pixels sensor (pixel size of  $2\mu\text{m}$ ). The cameras are stereo synchronized through a dedicated board (LI-OV580-STEREO), which also includes a USB 3.0 interface. By means of the ABS support, the cameras are mounted with an anthropometric interaxial distance ( $\sim 6.3$  cm) and with a fixed convergence angle that provides sufficient stereo overlap at about 40 cm (i.e., an average working distance for manual tasks).



**FIGURE 3.** The tablet PC used in the user study. 1→Rigid shell for the tablet PC. 2→Fluorescent markers for the optical tracking. 3→Digital pen. 4→Tablet PC.

The resulting offset/parallax between cameras and eye/display is of about 5.9 cm along the display optical axis (z-axis) and of 2.2 cm along the display vertical axis (y-axis). The stereo cameras configuration adopted in the study is: 2560x720@60 frames-per-second (fps). Both the cameras are equipped with a M12 lens support whose focal length ( $f = 8$  mm) is chosen to ensure a sufficient camera FOV able to cover the entire display FOV at 40 cm as well as to mitigate the zoom factor due to the eye-to-camera parallax along the display optical axis. To perform the user study, we used a tablet PC Asus ZenPad 3S 10 Z500M and a digital stylus (Fig. 3). More details on the experimental setting are provided in Section IV.

## B. SOFTWARE TOOLS AND LIBRARIES

The software framework was built in C++ under Linux Operating System (Ubuntu 16.04) with an object-oriented design. We took advantage of the Compute Unified Device Architecture framework (CUDA Toolkit 8.0) to develop an application able to harness the power of the GPU through parallel computing over the GPU cores.

The chosen integrated development environment (IDE) is Nsight™ Eclipse Edition by NVIDIA; Nsight provides an all-in-one integrated environment to edit, build, debug and profile CUDA C/C++ applications.

We opted for OpenCV [32] (Open Source for Computer Library, ver. 3.3.1.) for the machine vision libraries. OpenCV is a cross-platform API for computer vision created by Intel that covers several areas of applications and allows for low-level image processing methods and high-level computer vision algorithms. The OpenCV GPU routines are written using CUDA, therefore they benefit from the underlying CUDA ecosystem. The adoption of the OpenCV::CUDA routines makes the CUDA architecture more versatile and easily adaptable to different devices (embedding different CPUs and nVidia GPUs).

As regards the rendering of the scenegraph (SG), we used the 3D graphics and visualization VTK library,



```
[CONF]
WRL_FILES = ".\object.wrl"

;Opacity level: s for solid, t for transparency, d for delete
OPACITY = "s"

;Window scene size
IMG_ELAB_W = 2560
IMG_ELAB_H = 720

;Cameras ID : format : <device id> (mono/stereo)
CAM_ID_1 = "0,STEREO"
CAM_ID_2 = -1

;Pipelines: for each valid camera a pipeline is needed
;CAM_ID_1_PIPELINES = "nvcamerasrc ! 'video/x-raw(memory:NVMM), ...."

;See-through modality: OST / VST
MODALITY = VST

;Viewport mode: STEREO / MONO
MODE = STEREO

;Localization method: MONO_GRID / MONO_MARKER / STEREO_GRID / STEREO_MARKER
LOC_MODE = STEREO_MARKER

;Image warping: Magnification / Homography
WARP = Magnification

;Localizar enabled / not enabled
;ENABLED = TRUE|FALSE
ENABLED = TRUE

;Camera parameters files
CAM_CALIB_FILE_1 = ".\cam_left-1280x720.yml"
CAM_CALIB_FILE_2 = ".\cam_right-1280x720.yml"

;Camera settings files
CAM_SETTINGS_FILE_1 = ".\cam_left-settings.txt"
CAM_SETTINGS_FILE_2 = ".\cam_right-settings.txt"
```

FIGURE 4. Template of the Conf.ini file.

version 8.2.0 [33]. VTK is an open-source platform-agnostic C++ library for 3D computer graphics, modelling, and volume rendering suited to medical images. VTK offers all the methods and classes that support the visualization of reconstructed 3D images from sets of radiological 2D images (i.e., Direct volume rendering). Data from 3D ultrasound systems, computed tomography (CT) and magnetic resonance imaging (MRI) scans can be managed in VTK.

We used a dedicated OpenCV module (OpenCV::Viz) as a wrapper of methods and classes of the VTK framework into the OpenCV platform. OpenCV::Viz allows for the creation of the AR visualization windows and the management of the rendering event loop, both referring to the underlying VTK architecture.

### C. SOFTWARE FRAMEWORK

The core function of the software, under VST modality, is to process and augment the images grabbed by the pair of front-facing RGB cameras before they are sent to the two microdisplays of the visor. Most of the application directives are configurable through an apposite configuration file (i.e. Conf.ini). The Conf.ini file, whose template is shown in Fig. 4, contains the following features that characterize the ARA and the components of the SG in terms of rendering or tracking requirements:

- The path to the folder containing the 3D mesh files and/or DICOM files to be imported by the ARA.
- The possibility to select, for each mesh, the opacity level within the SG.
- The possibility to select the camera ID and format: mono/stereo.
- The possibility to select the window scene size (i.e., AR window resolution).

- The possibility to select the camera processing pipeline.
- The possibility to select the see-through modality: OST/VST.
- The possibility to select the viewport mode: stereo/mono.
- The possibility to select the localization method: monochromatic spherical markers/planar marker.
- The possibility to warp the image before rendering (magnification or homography-based transformation).
- The possibility to activate/deactivate the localization.
- The possibility to load the camera setting files.
- The possibility to load camera parameters files.

Two types of 3D files, VRML and PLY, can be imported. As an alternative, DICOM files can also be loaded for direct volume rendering. The files containing the camera parameters (e.g., intrinsic and extrinsic camera parameters) and files containing the camera settings (e.g., fps, white balance, contrast, brightness) are imported by the ARA. The software framework can switch at run-time between OST and VST modalities. The localization can also be turned on and off upon request; when off, the software allows for a direct interaction with the surgical map. In this way, the application can also be used for surgical planning purposes.

The application and the shared library classes with their functionalities are described in the next subsections. For clarity, methods are in *Italic*.

### D. APPLICATION WORKFLOW

In any AR application that aims to be real time, the critical challenge is to address the computational complexity of all the image processing tasks associated to the rendering and the tracking processes in a computationally efficient way. For this reason, while designing our ARA, we leveraged the instruction level parallelism for improving the computational capacity by concurrent execution of threads (i.e., multithreading).

The VST paradigm implemented by the software framework can be functionally and logically described as follows.

Upon startup, the ARA reads the Conf.ini file to configure its internals and to initialize the instances of the three main AR libraries:

- AR\_VideoCapture
- AR\_Engine
- AR\_Tracking

The AR\_VideoCapture library captures RGB cameras frames of the real scene and dispatches them to the AR\_Engine library; in AR\_Engine, these video frames are processed, augmented, and rendered onto the two viewports associated to the microdisplays of the visor. The machine-vision methods needed for yielding the image registration between real scene and virtual elements are performed by the AR\_Tracking library.

### E. AR\_VIDEOCAPTURE LIBRARY

The AR\_VideoCapture library exploits the OpenCV::VideoCapture class, which provides C++ API for capturing videos from cameras. A dedicated thread, defined in a specific

class named `AR_VideoCapture::VideoDeviceThread`, continuously grabs the camera frames and stores them in an area of memory. The method for retrieving the buffered frames is named `VideoDeviceThread::getLastVideoFrame()`.

The `VideoDeviceThread` class is highly configurable to access different types of video cameras by using their associated pipelines/camerasIDs. Camera parameters such as fps, brightness, saturation, contrast, are adjustable as well (through the camera settings files). This design ensures that the library can be adapted to the special needs of the ARA.

#### F. AR\_ENGINE LIBRARY

The `AR_Engine` library is the key component in the ARA; it has a simplified interaction with the other shared libraries and controls the real-time rendering of the whole AR scene through two main classes: `AR_Engine::Sceneview` and `AR_Engine::Monoview`.

The `Sceneview` class is a derived class of the `OpenCV::Viz::Viz3d` class and it represents the 3D visualizer window. The single elements of the SG, namely the widgets associated to 3D meshes and the virtual cameras, are initialized retrieving data from the main application.

In our stereo side-by-side ARA, the `Sceneview` class manages the rendering loop and handles the two instances of the `Monoview` class (one for each viewport for our stereoscopic ARA) in a side-by-side fashion. The `Sceneview` class also manages the `AR_Tracking` object and methods (see section G). In the `Sceneview` class, parallel computing is introduced by calling a single thread for each `Monoview` object for the image processing part of the `AR_Tracking` module (i.e., `AR_Tracking::image_processing`).

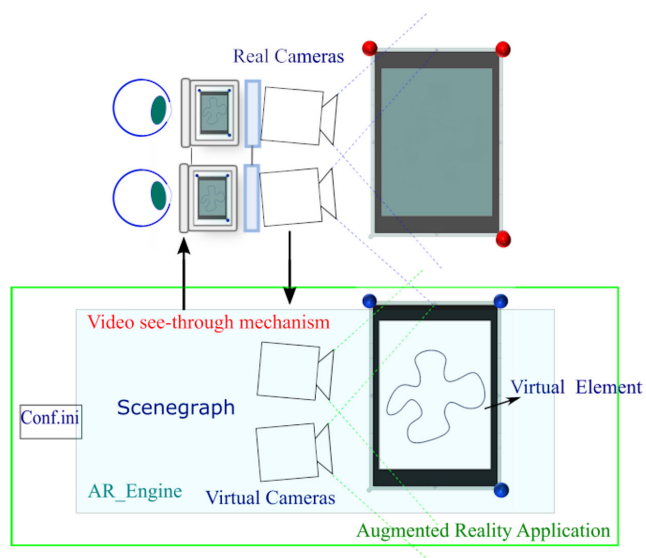
The `Sceneview` class retrieves data from the `AR_Tracking` class to control the geometrical relations between the elements of the SG. In the `Monoview` class, all functions related to the AR rendering are managed. Each of the two instances of the `Monoview` class communicates directly with the associated instance of the `AR_VideoCapture` to retrieve the buffered video frames.

#### 1) CUDA IMAGE ELABORATION (OPTIMIZED UNDISTORTION AND CUDA WARPING)

Camera images are remapped using `OpenCV::CUDA` routines. The non-linear part of the internal model of the cameras, due to lens radial distortion, is compensated by applying a non-linear remapping of the camera images. To achieve this, we created an optimized CUDA version of the standard `OpenCV::undistort` routine (`Monoview::undistort_optimized`).

The underlying rationale behind our optimized version of the `undistort` routine, is that the `OpenCV::CUDA::initUndistortRectifyMap` is performed just once and, within the render loop, the `OpenCV::CUDA::remap` is then called.

After undistortion, and if selected in the `Conf.ini` file, a linear remapping of the images can be performed by calling the `OpenCV::CUDA::warpPerspective` routine.



**FIGURE 5.** Video see-through paradigm implemented by the software framework. To accurately register a virtual element to a target object (e.g., the surface of a tablet PC) the virtual element must be observed by a couple of virtual viewpoints whose process of image formation mimic that of the real cameras in terms of intrinsic and extrinsic parameters.

#### 2) SCENEGRAPH RENDERING

To achieve an accurate alignment between real and virtual content, the virtual content (i.e., the SG) must be observed by a couple of virtual viewpoints (i.e., `Monoview::Virtual_camera`) whose processes of image formation mimic those of the real cameras in terms of intrinsic and extrinsic parameters. To this end, the conditions to be satisfied are twofold. First, the intrinsic and extrinsic parameters of the virtual stereo cameras must be initialized loading the data of a standard stereo camera calibration routine [34]. Secondly, the pose of the virtual elements with respect to the virtual cameras must be set equal to those of the real elements to be augmented. In our framework, this condition is satisfied by applying an optical marker-based tracking method. Fig. 5 shows the VST mechanism implemented by the ARA.

Notably, even though in this study we used the headset solely under VST mode, the `AR_Engine` library is capable of providing both optical and video see-through-based augmentations. This functionality is implemented by adapting the projection transformations of the virtual cameras, changing the rendering modality, and controlling the transparency of the LC shutters.

#### G. AR\_TRACKING LIBRARY

The `AR_Tracking` library computes the pose of the target with respect to the HMD by means of a dedicated inside-out optical tracking algorithm.

#### 1) INSIDE-OUT MARKER-BASED OPTICAL TRACKING METHOD

Most commercial image-guided surgery systems rely on outside-in electromagnetic and/or optical tracking methods [35]. Nowadays, optical tracking systems based on the

infrared detection of spherical markers are the state-of-the-art in surgical tracking [36] and they are preferred over electromagnetic trackers since they are not affected by the presence of ferromagnetic and/or conductive materials [37]. In addition, to achieve a tracking accuracy comparable to that obtained through standard optical trackers, the distance of the tracked body (patient or surgical tool) from the electromagnetic field generator should be limited to 30 cm [38].

The external optical trackers currently embedded in commercial surgical navigators (e.g., Claron Technology Micron-tracker [39], (Northern Digital Polaris optical [40], Medtronic StealthStation system [41]) are not sufficiently flexible for further development and customization due to proprietary techniques and libraries [42]. In addition, standard outside-in optical tracking solutions introduce unwanted line-of-sight constraints and add technical complexity to the surgical workflow [24], [43]. For this reason, we deemed it purposeful to design a dedicated inside-out optical tracking mechanism that could feature a simple installation phase and that it could be easily integrated in our AR software framework and modified according to the specific application scenario. Our optical tracking solution does not require obtrusive external trackers for localizing the spherical markers anchored to the target scene [44].

Compared to planar markers, such as the ones used in most AR applications based on commercial headsets, small spherical markers contribute reduced line-of-sight constraints and they can be conveniently placed on the patient's body and/or around the working area with a lesser logistic impact in the setup phase. The tracking method exploits the same head-anchored RGB stereo camera pair used for implementing the VST mechanism and it does not rely on additional tracking cameras (infrared and/or RGB), which would dramatically complicate the calibration process that need to be carried out before the actual procedure [36]. The usability of the tracking mechanism was further improved by using sets of three monochromatic markers, considering that three is the minimum set of markers that yields a finite number of solutions (i.e. four solutions) to the camera pose estimation problem [45]. Our method tackles the ambiguity of the perspective-3-point (P3P) problem by leveraging the stereoscopic settings of the VST headset.

## 2) OVERVIEW OF THE TRACKING ALGORITHM

Camera pose estimation problem is divided into four main stages. First, the pixel coordinates of the markers' centroids are determined through a color segmentation and blob detection phase. Next, the algorithm performs the stereo matching of the two triplets of image points and derives the 3D coordinates of them in the left camera reference system (CRS) through stereo triangulation. Then, the pose of the target reference system (TRS) with respect to the CRS is computed by solving in closed-form the absolute orientation problem (AOP) [46]. As a last step, the pose is refined through a Levenberg-Marquardt optimization algorithm. Each step is explained in more details in the following subsections.

The algorithm presented here achieves significant improvements compared to that presented in [43], both in terms of reliability and frame rate. Major differences are in the methods implemented for markers detection and in the methods used to solve the stereo and the 3D-3D correspondences.

## 3) COLOR SEGMENTATION AND BLOB DETECTION

The optical tracking algorithm features the processing of the camera images retrieved from the AR\_VideoCapture library. These methods are included in the *AR\_Tracking::img\_proc*. routine.

Spherical markers are detected through image segmentation in the Hue-Saturation-Value (HSV) color space and blob detection. Color-based image segmentation must assure a robust trade-off between illumination invariance and absence of segmentation overlaps among differently colored regions. The use of monochromatic markers ensures high robustness even with non-controllable and inconsistent lighting conditions, since incorrect labeling can be prevented. To counter the limitation of using visible light as source of information, we used red fluorescent pigmentation for our spherical markers, since fluorescent dyes peak the S channel of the HSV color space and boost the response of the camera CMOS sensor. In addition, such pigmentation has a high V range, which makes the segmentation sufficiently robust to non-uniform levels of illumination intensity, shadows and shadings.

As suggested in [47], some care must be taken for the camera color settings. With the camera white balancing ON the color channels are continuously remapped. This changes the original red channel values and thus modifies the image left by the fluorescent color. Therefore, since in an OR the quality of white colors is not high, in our application we decided to turn the camera white balancing off. This part of the image processing is implemented through OpenCV::CUDA routines to leverage the power of the GPU computing.

After image segmentation, blob detection is performed on both stereo images using an improved version of the *OpenCV::SimpleBlobDetector* class. This class performs several filtrations of returned blobs based on a set of parameters. In our algorithm we dictated the following constraints for the connected regions:

- Extracted blobs must have an area  $\geq 50$  pixels.
- Extracted blobs must have a circularity value  $\frac{4\pi \text{Area}}{\text{perimeter}^2} \geq 0.5$ .
- Extracted blobs must have a convexity value  $\frac{\text{Area}}{\text{Area of blob convex hull}} \geq 0.5$

The centroids of the selected regions are determined using the spatial moments. These image points correspond to the projections of the centroids of the markers on the image planes of the two cameras. Fig. 6 shows the results of the color segmentation and blob detection steps.

## 4) STEREO MATCHING AND STEREO TRIANGULATION

By working with a set of three indistinguishable markers, it is not possible to localize them in the CRS without



(a) Original camera frame.



(b) Binary image after color image segmentation.



(c) Result of the blob detection.

**FIGURE 6.** Results of the image processing on the camera frames.

ambiguity, since the correspondence between projected points on the left and on the right camera images (conjugate points) is unknown (i.e., stereo correspondence problem). In our algorithm, we solve the stereo correspondence problem applying standard epipolar geometry rules to the six permutations of matches between the triplets of feature points on the stereo images. Over the six permutations of matches of conjugate points ( $j = 1 : 6$ ), our method (*AR\_Tracking::find\_epipolar\_conf*) finds the correspondence  $j_{corr}^{2D}$  that minimizes a cost function  $E_j(d)$  obtained computing the sum of three absolute distances  $d_i$  between the epipolar lines and the points on the conjugate image.

$$j_{corr}^{2D} = \arg \min_j E_j(d) == \arg \min_j \sum_{i=1}^3 d_i \quad (1)$$

where  $d_i$  is computed as follows: given a point  $p_l = (x_l, y_l)$  on the left image and its corresponding on the right image  $p_r = (x_r, y_r)$ , the equation of the epipolar line in its implicit notation (on the right image) is computed starting from the stereo correspondence equation based on the fundamental

matrix  $F$ :

$$\forall(x_l, y_l) \leftrightarrow (x_r, y_r) \quad (2)$$

$$[x_r, y_r, 1] F \begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = 0 \quad (3)$$

$$[x_r, y_r, 1] \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \rightarrow ax_r + by_r + c = 0 \quad (4)$$

Therefore, for each correspondence  $i$  between  $p_l$  and  $p_r$ ,  $d_i$  is trivially computed as:

$$d_i = \frac{|ax_r + by_r + c|}{\sqrt{a^2 + b^2}} \quad (5)$$

After solving stereo correspondence, the 3D position of each marker in the CRS is computed through stereo triangulation knowing the camera intrinsic parameters and the relative pose between the two cameras (we use the *OpenCV::triangulatePoints* routine).

### 5) 3D-3D CORRESPONDENCE AND ABSOLUTE ORIENTATION PROBLEM

In *AR\_Tracking*, we simultaneously solve the 3D-3D correspondence problem and the AOP (i.e. we compute  $[R_{corr}|t_{corr}]$ ) through a method named *AR\_Tracking::estimate\_pose\_AOP*.

The 3D-3D correspondence between the two triplets of 3D points with the relative pose between their associated coordinate system is determined by picking, over the six possible permutations  $j$ , the configuration  $j_{corr}^{3D}$  that yields the lowest root mean square of fiducial registration error ( $FRE^2$ ) computed through a closed-form fitting method.

A key factor in performing accurate measurements with stereo cameras is to know with extreme confidence the relative pose between them. The major drawbacks of using head-anchored stereo cameras with an anthropometric interaxial distance are the non-ideal stability in the constraints between the two cameras and the presence of a reduced baseline length ( $b$ ). Both these features lead to inaccuracies in measuring the 3D position of the markers and therefore in estimating the pose of the TRS [48].

The major error contribution is measured along the axis orthogonal to the baseline ( $z$ -axis) and it increases with the square of the distance. The depth resolution at the distance  $Z$  can be computed as follows:

$$\Delta Z \approx \frac{Z^2}{fb} \Delta d \quad (6)$$

By way of example, let us consider fixed and ideally error-free estimates of the baseline length  $b = 65$  mm, and of the left camera focal length  $f = 8$  mm. Given a stereo disparity accuracy  $\Delta d$  of  $\pm 1$  pixel (corresponding for our camera to  $\pm 4 \mu m$ ), the associated depth resolution  $\Delta Z$  is approximately  $\pm 1.9$  mm for  $Z = 50$  cm. In this case, the closed-form solution of the AOP cannot yield a sufficiently accurate result.



To counter this problem, we added a second stage of pose estimation aimed at refining separately the pose of the two cameras.

#### 6) EXTERIOR ORIENTATION PROBLEM OR POSE REFINEMENT

This second stage, performed by a method named *AR\_Tracking::refine\_pose*, minimizes a cost function formulated as the sum of the square measurement error (reprojection residuals  $d_i$ ) between measured image points  $p_i$  and calculated projections  $\hat{p}_i$  of the corresponding world points ( $P_i$ ):

$$\begin{aligned} [R_{\text{refined}} | T_{\text{refined}}] &= \arg \min_j \sum_{i=1}^3 d(p_i, \hat{p}_i)^2 \\ &= \arg \min_j \sum_{i=1}^3 \left\| p_i, \hat{p}_i(K, \hat{R}, \hat{t}, P_i) \right\|^2 \quad (7) \end{aligned}$$

where  $K$  is the matrix of intrinsic parameters (fixed) and  $\hat{R}, \hat{t}$  are the rotation matrix and translation vector to be optimized. The refinement method exploits the *OpenCV::solvePnP* routine, which runs an iterative Levenberg-Marquardt optimization algorithm and yields subpixel accuracy in the image plane.

#### H. COMPUTING EFFICIENCY AND ACCURACY OF THE AUGMENTED REALITY APPLICATION

The average frame rate of the AR application is 30 fps: the application takes on average  $\sim 0.029$ s (with a standard deviation of  $\sim 0.007$  s) to complete the entire AR mechanism from camera frame recording to AR rendering. The CUDA-based GPU implementations of the machine vision routines helps improving the computing efficiency of the tracking loop, which takes on average  $\sim 0.007$  s (std  $\sim 0.002$  s).

As regards the AR registration accuracy, the average FRE measured in static conditions over 1000 frames is  $\sim 0.24$  mm (std  $\sim 0.05$  mm), whereas the root mean square of the reprojection residuals (overlay error) onto the image plane is  $< 1$  pixel.

However, the evaluation of a wearable AR platform in terms of AR accuracy is affected by many factors other than the tracking accuracy. Just to mention a few: the errors induced by the display in terms of low angular resolution; the low contrast ratio; the optical distortion of the display; the zoom factor; the errors induced by the image-to-target registration; the optical aberrations typical of HMDs that alter the user's perception of depth [49]. To provide an estimation of the impact of all these factors on user's performance, we conducted an experimental study. The study was designed to assess the efficacy and the reliability of the AR platform (custom-made HMD plus software framework) in guiding a simulated task of tissue incision.

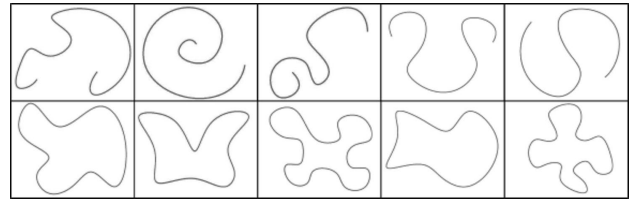


FIGURE 7. B-splines curves used for the experimental study.

#### IV. USER STUDY

The definition of an accurate line of incision is paramount in many surgical procedures. In brain surgery, the drawing of the skull incision lines before the craniotomy and the dural opening are essential tasks. This is because possible deviations from the ideal trajectory may impact the working area of the surgeons in relation to deep structures to be avoided during lesion targeting [50], [51]. In craniomaxillofacial surgery, several corrective surgeries involve the precise location of the osteotomy lines as for the Le Fort fractures [52]. In plastic surgery, the surgeon is often asked to follow well established incision lines on the skin in an effort to achieve inconspicuous scars [53].

In the user study, the simulated task had to be repeatable within participants and with a clearly measurable metrics that could capture the benefits of AR. For these objectives, we used the display of a tablet notebook (Asus ZenPad 3S 10 Z500M) as a digital surface of incision and a digital pen as a digital scalpel. The display resolution is of 2048x1536 with a pixel density (ppi) of 264 (i.e.,  $\sim 10.4$  pixel/mm). AutoDesk Sketchbook application was used as drawing software. As drawing tool, we used a technical pen with tip size of 4 pixels (i.e.,  $\sim 0.4$  mm). Each subject was asked to trace a line with a digital stylus onto the display of the tablet under two conditions:

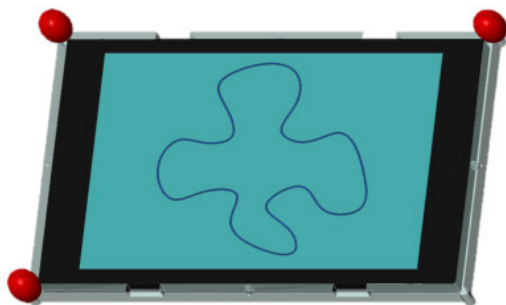
- Digital incision with naked eye (NK\_Dinc).
- Digital incision with AR guidance (AR\_Dinc).

In the NK\_Dinc tests, the user had to trace a line with the digital pen trying to follow the pre-loaded curve added as a secondary layer on the digital canvas. These naked-eye tests were useful to evaluate the efficacy of the experimental setting that digitally simulates surgical incision.

In AR\_Dinc tests, the guiding line to be traced was provided to the user through the AR HMD. Here the user had to trace the curve on a blank digital canvas. During these tests, the contrast of the tablet display was lowered to increase the real-to-virtual contrast ratio.

A total of ten B-spline curves (BSPs) were designed: five closed curves and five open curves (Fig. 7). The BSPs were CAD designed using the PTC Creo Parametric 3D Modelling software (ver. 3.0). Each BSP has a width of 1 mm.

A rigid shell for the tablet was 3D printed with a rapid prototyping machine (Stratasys Objet30 Prime) to support the markers for the optical tracking algorithm. The case was designed with the same CAD software used for the BSPs and it was provided with three spheres of 1.8 cm diameter (Fig. 8). The spheres were colored using a red fluorescent dye.



(a) CAD design of the rigid shell for the tablet PC with the optical markers and the B-spline curve.



(b) Augmented reality of the tablet with the virtual spline registered onto the surface of the tablet display.

**FIGURE 8.** Conceptual design (CAD) and real embodiment of the digital incision testing platform.

The virtual BSPs were projected on the display of the tablet using the pose computed by tracking the markers embedded into the shell. In order to perfectly fit the tablet into the case, thus reducing possible inaccuracies in the estimation of the position of the physical display during tracking, the tablet shell was provided with four screw holes to fix tablet and case in a stable position.

The markers, the physical display, and the BSPs were all referred to the same reference system (TRS) dictated in the CAD software. Each BSP was exported as a single VRML file. Fig. 9 shows the experimental setting with a user performing one of the AR\_Dinc.

**A. PARTICIPANTS AND STUDY PROTOCOL**

Twelve participants were recruited from university students, staff, and faculty members. The demographic information about the users are listed in Table 1. All participants had normal vision acuity or corrected visual acuity with the aid of prescription glasses or contact lenses. The participants had to rate their experience with AR, with digital pens, and with HMDs to have a baseline and assess the familiarity of users with the procedures/tools and AR.

The twelve subjects each performed the two group of digital incision tests, resulting in  $10 \times 2 \times 12 = 240$  trials overall. Before the test session, each participant read and signed an informed consent. In the experiment procedure with the HMD, each participant was instructed about the test and was involved in two training sessions with a specific



**FIGURE 9.** Experimental setting during an augmented-reality trial with the subject wearing the AR headset.

**TABLE 1.** Demographics of the ten participants to the user study.

General info	Value
Gender (male; female; non-binary)	(5; 7; 0)
Age (min; max; mean; STD)	(26; 43; 30.75; 5.75)
IPD (mm) (min; max; mean; STD)	(52.4; 67.3; 57.3; 4.2)
Handedness (left:right;ambidextrous)	(1;11;0)
Uncorrected poor vision (no, yes)	(7, 5)
AR experience (none, limited, familiar, experienced)	(1, 3, 3, 5)
HMD experience (none, limited, familiar, experienced)	(1, 3, 3, 5)
Digital pen experience (none, limited, familiar, experienced)	(4, 3, 4, 1)

BSP different from the ones used in the assessment sessions. No further assistance was provided to the participants. For each subject, the sequence of tests (NK\_Dinc or AR\_Dinc) was randomly assigned.

During the assessment sessions, each participant was asked to report any spatial jitter or drift of the virtual content and stop the task if any. To improve the overall ergonomics and allow the user to work while seated comfortably with the screen at eye level, the tablet and the case were placed over a wooden holder with a tilted surface (45° of inclination). In all the experiments, we used a professional studio illuminator (DynaSun 3X CY25WT) with three spots to imitate the OR lighting conditions.

**B. QUALITATIVE ASSESSMENT OF THE AR PLATFORM**

After completing both groups of tests, each participant was asked to fill in the demographic survey and a Likert

**TABLE 2. Results of the seven-point monotone Likert questionnaires (1: Strongly Disagree; 7: Strongly Agree), with calculated p - values according to Mann-Whitney U test.**

Item	Questions	Median (IQR)	p-value HMD exp.
			p-value AR exp. p-value dig. pen exp.
1	The task is interesting and engaging	6 5 ~ 7	0.828 0.828 0.078
2	Stereoscopic vision is good and comfortable	5.5 5 ~ 6.5	0.424 0.424 0.205
3	No significant distortion of the scene is perceived through the visor (double vision, blurred vision, ...)	5 4 ~ 7	0.731 0.731 0.748
4	No discomfort is felt when performing the task (visual fatigue, dizziness, nausea, ...)	5 5 ~ 6	0.913 0.913 1
5	The trajectory is easy to follow and draw	6 5 ~ 6	0.913 0.913 0.470
6	Virtual trajectory occluding the tip of the pen does not affect significantly the performance of the task	5 4 ~ 6	0.420 0.420 0.359
7	Been successful in completing the task	5 5 ~ 6	0.283 0.283 0.495
8	Not been stressed, irritated, frustrated, annoyed in completing the task	6 5.5 ~ 7	0.849 0.849 0.255
9	Likely to try again	6 5 ~ 6	1 1 1

questionnaire in terms of usability, functionality, and technology acceptance. The Likert questionnaire, shown in Table 2, comprises 9 items, each evaluated using a seven-point monotone Likert scale (from 1 = strongly disagree, to 7 = strongly agree) as previously done in [54] and in [55].

### C. QUANTITATIVE EVALUATION

The goal of the quantitative evaluation of the AR application was to measure the similarity between the two trajectories: the virtual BSP associated to the planned path of incision, and the actual curve traced by the user, both with and without the AR guidance. The time for completing the tasks under the two conditions was also measured. The Hausdorff distance ( $H_{dis}$ ) provides a metric for addressing the problem of measuring the distance between two sets of points that form the two curves. By computing the Hausdorff distance, we measure the “closeness” between the two trajectories. The  $H_{dis}$  measures how far two subsets of a metric space (for us  $\mathbb{R}^2$ ) are from each other and it is defined as:

$$H_{dis} = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \quad (8)$$

where sup is the supremum, inf the infimum, and  $d(x, y)$  the denotes the Euclidean distance in  $\mathbb{R}^2$  between points of the two curves.

The results and the statistical analysis were both processed in MATLAB® (R2018b MathWorks, Inc., Natick, Massachusetts, US). The quantitative evaluation for each trial was broken down into the following steps:

- The image containing the traced curve is exported from the drawing software as a .png file (image resolution/display resolution = 2048x1536).
- The associated planned BSP is exported from the CAD software as a .png file with the same image resolution.
- The two images are binarized.
- $H_{dis}$  between the two curves is computed to assign a score to the similarity between the two trajectories. The distance values are converted from pixels to mm using the ppi of the tablet display.

### D. STATISTICAL ANALYSIS

Responses to the Likert questionnaire were summarized using median with dispersion measured by interquartile range. We carried out nonparametric Mann-Whitney U test to assess whether the answer tendencies differ based on the user’s experience with AR, HMD, and digital pen. A p-value < 0.05 was considered statistically significant.

Quantitative results were presented for each user in terms of average value, standard deviation and max value of the similarity between the traced curve and the planned one ( $H_{dis}$ ). Time to completion ( $T_{compl}$ ) was equally measured. A post-hoc analysis with Wilcoxon signed-rank test was conducted for the  $H_{dis}$  and  $T_{compl}$  between tests with and without AR. A p-value < 0.05 was considered statistically significant. For each condition, a Mann-Whitney U test was also conducted to evaluate whether the subject performance differ based on his/her previous experience with AR, with HMD, or with digital pen. A p-value < 0.05 was considered statistically significant.

## V. RESULTS

### A. RESULTS OF THE QUALITATIVE EVALUATION

Table 2 shows the results of the Likert questionnaire. Subjects expressed an overall positive opinion regarding the user study. We obtained almost top score for the enjoyability (items 1 and 9), the ease-to-use (item 5) and for the comfort (item 7) of the task. The nonparametric Mann-Whitney U tests revealed that there was not any statistically significant difference in the answers among participants with at least some level of experience with AR, HMDs, and digital pens.

### B. RESULTS OF THE QUANTITATIVE EVALUATION AND DISCUSSION

All the twelve participants completed the 10 AR\_Dinc tasks without perceiving any spatial jitter or drift of the virtual content, thus confirming that the optical tracking proved to be sufficiently robust. As reported in Table 3 and Table 4, for each user, the results under the treatment (AR\_Dinc) and control condition (NK\_Dinc) were summarised in means, standard deviations and max values of the Hausdorff distances  $H_{dis}$  and completion times  $T_{compl}$ . The overall mean, standard deviation and max values of  $H_{dis}$  were 0.91 mm, 0.14 mm, and 1.31 mm for the NK\_Dinc tests and 0.98 mm, 0.17 mm, and 1.63 mm for the AR\_Dinc tests. As for the completion times, the mean, standard deviation, and max

**TABLE 3.** Quantitative evaluation results (Hausdorff distances).

User	HMD			Naked eye		
	Mean	Std Dev	Max	Mean	Std Dev	Max
1	0.98	0.14	1.19	0.97	0.15	1.24
2	0.96	0.22	1.54	0.95	0.16	1.31
3	0.96	0.19	1.31	0.88	0.12	1.20
4	1.00	0.13	1.21	0.89	0.12	1.02
5	0.94	0.12	1.16	0.90	0.12	1.10
6	0.99	0.23	1.54	0.92	0.17	1.27
7	0.95	0.10	1.08	0.87	0.15	1.17
8	0.96	0.17	1.21	0.91	0.12	1.11
9	1.08	0.28	1.63	0.90	0.15	1.25
10	0.99	0.16	1.24	0.94	0.17	1.30
11	0.94	0.12	1.09	0.90	0.14	1.16
12	0.91	0.14	1.21	0.87	0.11	1.12

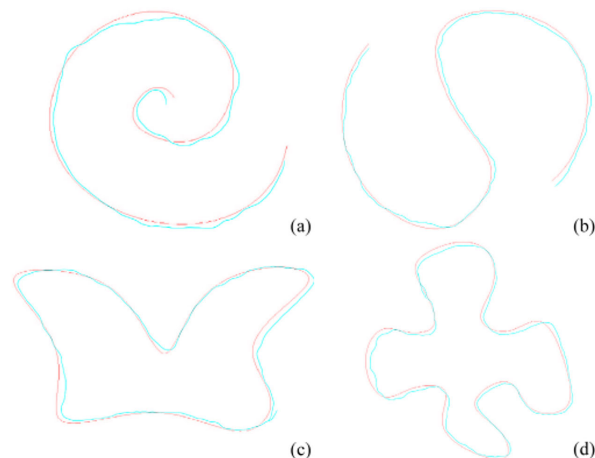
**TABLE 4.** Quantitative evaluation results (time to completion).

User	HMD			Naked eye		
	Mean	Std Dev	Max	Mean	Std Dev	Max
1	24.1	4.8	31.5	32.1	7.9	43.3
2	28.8	3.2	35.5	41.2	14.9	64.3
3	42.5	12.7	67.5	38.6	14.7	70.8
4	34.9	11.7	56.2	42.4	6.5	54.2
5	26.4	5.1	33.1	31.0	6.0	41.3
6	14.4	4.7	23.3	11.9	2.5	16.6
7	31.3	4.4	37.3	32.5	7.6	50.6
8	34.5	11.4	52.4	49.7	13.2	73.3
9	23.9	5.3	35.1	29.3	7.1	43.7
10	26.5	6.8	35.6	24.8	7.9	42.4
11	53.3	10.0	66.5	34.6	8.0	49.4
12	28.3	7.4	42.9	39.0	8.1	57.8

values were 33.9 s, 12.9 s, and 73.3 s for the NK\_Dinc tests and 30.7 s, 12.2 s, and 67.5 s for the AR\_Dinc tests. Therefore, on average, the users performed slightly better with the naked-eye than with the HMD (0.07 mm of spatial difference on average) but taking more time (3.2 s of time difference on average that is approximately ~ 10% of the overall task duration).

The post-hoc analysis with Wilcoxon signed-rank confirmed that there was globally a statistically significant tendency ( $p = 0.0011$ ) in achieving slightly better performance accuracy with the tests performed with the naked eye compared to the AR\_Dinc tests. On the other hand, the same analysis on the completion times also revealed there was a statistically significant difference ( $p = 0.0022$ ) between the AR\_Dinc tests and the NK\_Dinc tests.

The major cause for such tendencies lies in the particular experimental setting adopted that simulates a digital incision task. With a visual check, during the NK\_Dinc tests, the user was able to immediately correct the incision direction whenever he/she perceived that the line drawn was deviating from the planned one. This real-time correction generally



**FIGURE 10.** Example of fused images from planned b-spline curves and traced curves. (a) spline 2, (b) spline 5, (c) spline 7, (d) spline 10.

increased the NK\_Dinc tests duration, and it was not possible in the AR\_Dinc tests since the contrast of the tablet display was kept low for maintaining the real-to-virtual contrast ratio sufficiently high.

It should be also noted that, in our experiments, the control condition represents an ideal scenario per se: a digital path of incision directly superimposed over the surface of incision. Despite that, the results obtained with the AR guidance proved to be statistically not too far from those obtained without the AR guidance. As additional consideration, we can hypothesize that the benefits of an AR guidance would be even more evident if compared with a standard surgical navigation approach with the user (i.e., the surgeon) having to mentally try to map the planning information provided on an external screen onto the surgical field. We also hypothesize that, in a real surgical scenario, AR guidance would reduce the stress level during the incision and it would help reducing task duration whilst increasing task precision.

By way of illustration, Fig. 10 shows four composite images containing the fused versions of traced trajectories (i.e., the cyan line) with planned ones (i.e., the pink line) for AR\_Dinc tests.

The Mann-Whitney U test revealed that there were no significant differences in accuracy performance between users with at least some level of experience with AR ( $p = 0.315$ ), HMDs ( $p = 0.315$ ), and digital pens ( $p = 0.989$ ). As regards the differences in  $T_{compl}$ , no statistically significant difference occurred between users with AR experience ( $p = 0.058$ ), and HMD experience ( $p = 0.058$ ), whereas a statistically significant difference was measured between users with at least some level of experience with digital pens ( $p < 0.001$ ).

**C. SOFTWARE USABILITY: RESULTS FROM EARLY APPLICATIONS**

An early version of the software framework was used in a study published in 2018 [30]. The study presented an automatic calibration procedure suited for OST HMDs with



infinity focus. The goal of any OST display calibration is to estimate the projection parameters of the virtual rendering camera that models the combined eye-display system and whose values vary according to the position of the user's eye with respect to the display. Unfortunately, depending on the proprietary platform associated to most consumer level OST HMDs, the control of the low-level rendering camera is often restricted by compatible interfaces [56]. For this reason, our framework is particularly useful in tuning the extrinsic and the intrinsic projection parameters of the virtual rendering camera computed during a calibration stage.

More recently, the software framework was used in a study aimed at evaluating the effect of a perspective conversion of the camera frames in restoring the natural perception of the three-dimensional space in non-orthostereoscopic VST HMDs [49]. In both these studies, the LUMUS OE-33 [57] was used as OST HMD, proving that the software can be appropriately used with different types of commercial AR HMDs.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a novel software framework for the deployment of AR applications able to support in situ visualization of medical imaging data. The software is suited for customized AR headsets specifically conceived for guiding high-precision manual tasks such as surgical incisions.

The software framework leverages the instruction level parallelism provided by CUDA architecture and it is capable of providing both optical and video see-through-based augmentations and it is computationally efficient (average frame period  $\sim 0.029$  s). The framework also features highly optimized stereoscopic optical marker-based tracking routines, which allow achieving a processing speed for the tracking  $\sim 0.007$  s per stereo frame.

We designed an experimental study to evaluate qualitatively and quantitatively the efficacy and the reliability of the entire AR platform (custom-made HMD plus software framework) in guiding a simulated task of tissue incision. The results were given in terms of perceived workload and comfort, performance accuracy, and completion time.

The qualitative results of our experiments show that the AR platform is generally regarded as engaging, ergonomic, and beneficial to the achievement of the task. The level of discomfort and frustration experienced by the participants during the tests were generally low.

The quantitative results suggest that the AR platform could be used to guide high-precision tasks: the average difference between traced and planned lines was of  $\sim 0.98$  mm for the AR tests, only 0.07 mm higher than the average incision accuracy achieved by performing the same task with naked eyes. On the other hand, the completion times were generally higher for the naked eye tests than for the AR tests. To the best of our knowledge, there is no AR software framework capable today of deploying AR applications with different types of head-mounted displays for medical applications and devoid of platform-dependent tracking techniques and/or complex

calibration procedures. The obtained results strongly encourage us to speed-up the clinical assessment of the entire AR platform.

It is envisioned that the proposed software platform coupled with our new-concept AR headset will boost the transfer of AR technology into routine clinical practice.

## ACKNOWLEDGMENT

The authors would like to thank S. Mascioli and C. Freschi for their support on the preliminary selection of the most appropriate software libraries and tools and R. D'Amato for his support in designing and assembling the AR visor.

## REFERENCES

- [1] L. Chen, T. W. Day, W. Tang, and N. W. John, "Recent developments and future challenges in medical mixed reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2017, pp. 123–135.
- [2] M. Eckert, J. S. Volmerg, and C. M. Friedrich, "Augmented reality in medicine: Systematic and bibliographic review," *JMIR Mhealth Uhealth*, vol. 7, no. 4, Apr. 2019, Art. no. e10967.
- [3] T. M. Peters, "Image-guidance for surgical procedures," *Phys. Med. Biol.*, vol. 51, no. 14, pp. R505–R540, Jun. 2006.
- [4] 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.
- [5] F. Cutolo, M. Carbone, P. D. Parchi, V. Ferrari, M. Lisanti, and M. Ferrari, "Application of a new wearable augmented reality video see-through display to aid percutaneous procedures in spine surgery," in *Proc. Augmented Reality, Virtual Reality, Comput. Graph. (AVR)*, L. T. De Paolis and A. Mongelli, Eds. Cham, Switzerland: Springer, 2016, pp. 43–54.
- [6] P. Vávra, J. Roman, P. Zonca, P. Ihnát, M. Nemeč, J. Kumar, N. Habib, and A. El-Gendi, "Recent development of augmented reality in surgery: A review," *J. Healthcare Eng.*, vol. 2017, Sep. 2017, Art. no. 4574172.
- [7] M. Kersten-Oertel, P. Jannin, and D. L. Collins, "The state of the art of visualization in mixed reality image guided surgery," *Comput. Med. Imag. Graph.*, vol. 37, no. 2, pp. 98–112, Mar. 2013.
- [8] B. Fida, F. Cutolo, G. di Franco, M. Ferrari, and V. Ferrari, "Augmented reality in open surgery," *Updates Surg.*, vol. 70, no. 3, pp. 389–400, Sep. 2018.
- [9] H. Liu, E. Auvinet, J. Giles, and F. Y. Rodriguez Baena, "Augmented reality based navigation for computer assisted hip resurfacing: A proof of concept study," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1595–1605, Oct. 2018, doi: [10.1007/s10439-018-2055-1](https://doi.org/10.1007/s10439-018-2055-1).
- [10] F. Cutolo, P. D. Parchi, and V. Ferrari, "Video see through AR head-mounted display for medical procedures," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 393–396.
- [11] J. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence*, vol. 9, no. 3, pp. 287–309, Jun. 2000.
- [12] S. A. Benton, *Selected Papers on Three-Dimensional Displays*, vol. MS162. Washington, DC, USA: SPIE Press, 2001.
- [13] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 362–371, Jun. 2011.
- [14] 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](https://doi.org/10.1007/s10439-019-02299-w).
- [15] 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 (IEEEVR)*, Mar. 2016, pp. 253–254.
- [16] H. Hua, "Enabling focus cues in head-mounted displays," *Proc. IEEE*, vol. 105, no. 5, pp. 805–824, May 2017.
- [17] 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.*, to be published.

- [18] B. Marques, R. Plantefève, F. Roy, N. Haouchine, E. Jeanvoine, I. Peterlik, and S. Cotin, "Framework for augmented reality in Minimally Invasive laparoscopic surgery," in *Proc. 17th Int. Conf. E-health Netw., Appl. Services (HealthCom)*, Oct. 2015, pp. 22–27.
- [19] N. Navab, J. Traub, T. Sielhorst, M. Feuerstein, and C. Bichlmeier, "Action- and workflow-driven augmented reality for computer-aided medical procedures," *IEEE Comput. Graph. Appl.*, vol. 27, no. 5, pp. 10–14, Sep. 2007.
- [20] *Vostars h2020 Project, g.a. 731974*. Accessed: Dec. 26, 2019. [Online]. Available: <http://www.vostars.eu/>
- [21] T. Sielhorst, M. Feuerstein, J. Traub, O. Kutter, and N. Navab, "Campar: A software framework guaranteeing quality for medical augmented reality," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 1, pp. 29–30, Jun. 2006.
- [22] M. Feuerstein, T. Mussack, S. M. Heining, and N. Navab, "Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection," *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 355–369, Mar. 2008.
- [23] J. Traub, P. Stefan, S. M. Heining, T. Sielhorst, C. Riquarts, E. Euler, and N. Navab, "Hybrid navigation interface for orthopedic and trauma surgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, R. Larsen, M. Nielsen, and J. Sporring, Eds. Berlin, Germany: Springer, 2006, pp. 373–380.
- [24] 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.
- [25] M. S. Pérez, M. C. Trinidad, D. Karatzas, A. C. Calaf, and P. P. V. Alcocer, "Development of general-purpose projection-based augmented reality systems," *IADIS Int. J. Comput. Sci. Inf. Syst.*, vol. 11, no. 2, pp. 1–18, 2016.
- [26] *Google Developers. Protocol Buffers*. Accessed: Dec. 26, 2019. [Online]. Available: <http://developers.google.com/protocol-buffers/>
- [27] *Openiglink Open Network Interface for Image-Guided Therapy*. Accessed: Dec. 26, 2019. [Online]. Available: <http://openiglink.org/>
- [28] (2019). *Trivisio, Lux Prototyping*. [Online]. Available: <https://www.trivisio.com/>
- [29] F. Cutolo, U. Fontana, and V. Ferrari, "Perspective preserving solution for quasi-orthoscopic video see-through HMDs," *Technologies*, vol. 6, no. 1, p. 9, Jan. 2018.
- [30] U. Fontana, F. Cutolo, N. Cattari, and V. Ferrari, "Closed-loop calibration for optical see-through near eye display with infinity focus," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2018, pp. 51–56.
- [31] F. Cutolo, U. Fontana, M. Carbone, R. D'Amato, and V. Ferrari, "[POSTER] hybrid video/optical see-through HMD," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Oct. 2017, pp. 52–57.
- [32] (Nov. 2019). *OpenCV*. [Online]. Available: <http://opencv.org/>
- [33] (Jul. 2019). *Vtk, the Visualization Toolkit*. [Online]. Available: <http://vtk.org/>
- [34] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [35] M. Maybody, C. Stevenson, and S. B. Solomon, "Overview of navigation systems in image-guided interventions," *Techn. Vascular Intervent. Radiol.*, vol. 16, no. 3, pp. 136–143, Sep. 2013.
- [36] R. Hussain, A. Lalande, C. Guigou, and A. B. Grayeli, "Contribution of augmented reality to minimally invasive computer-assisted cranial base surgery," *IEEE J. Biomed. Health Inform.*, to be published.
- [37] 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.
- [38] 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](https://doi.org/10.1007/s11548-015-1348-1).
- [39] (Nov. 2019). *Claron Technology Microntracker*. [Online]. Available: <https://www.claronav.com/microntracker/>
- [40] (Nov. 2019). *Northern Digital inc., Polaris Optical Tracking Systems*. [Online]. Available: <https://www.ndigital.com/medical/products/polaris-family/>
- [41] (Nov. 2019). *Medtronic inc., Stealth Station s7 Surgical Navigation System*. [Online]. Available: <https://www.medtronic.com/us-en/healthcare-professionals/products/neurological/surgical-navigation-systems/stealthstation.html>
- [42] Z. Zhou, B. Wu, J. Duan, X. Zhang, N. Zhang, and Z. Liang, "Optical surgical instrument tracking system based on the principle of stereo vision," *J. Biomed. Opt.*, vol. 22, no. 6, pp. 1–14, 2017, doi: [10.1117/1.JBO.22.6.065005](https://doi.org/10.1117/1.JBO.22.6.065005).
- [43] F. Cutolo, C. Freschi, S. Mascioli, P. D. Parchi, M. Ferrari, and V. Ferrari, "Robust and accurate algorithm for wearable stereoscopic augmented reality with three indistinguishable markers," *Electronics*, vol. 5, no. 3, p. 59, Sep. 2016.
- [44] M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya, "A stereoscopic video see-through augmented reality system based on real-time vision-based registration," in *Proc. IEEE Virtual Reality*, New Brunswick, NJ, USA, Mar. 2000, pp. 255–262.
- [45] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003.
- [46] 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, Sep. 1987.
- [47] 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.
- [48] M. Kytö, M. Nuutinen, and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," *Proc. SPIE*, vol. 7864, Jan. 2011, pp. 168–176.
- [49] N. Cattari, F. Cutolo, and 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.
- [50] Z. Dai, R. Yang, F. Hang, J. Zhuang, Q. Lin, Z. Wang, and Y. Lao, "Neurosurgical craniotomy localization using interactive 3D lesion mapping for image-guided neurosurgery," *IEEE Access*, vol. 7, pp. 10606–10616, 2019.
- [51] F. L. Cohen, D. Mendelsohn, and M. Bernstein, "Wrong-site craniotomy: Analysis of 35 cases and systems for prevention," *J. Neurosurg.*, vol. 113, no. 3, pp. 461–473, 2010.
- [52] S. Abela, D. Tewson, S. Prince, M. Sherriff, and D. Bister, "Precision of le fort I osteotomy: Planning versus outcome. a multi-centre retrospective study," *J. Orthodontics*, vol. 44, no. 1, pp. 8–13, 2017.
- [53] G. Lemperle, M. Tenenhaus, D. Knapp, and S. M. Lemperle, "The direction of optimal skin incisions derived from striae distensae," *Plastic Reconstructive Surg.*, vol. 134, no. 6, pp. 1424–1434, 2014.
- [54] H. Xue, P. Sharma, and F. Wild, "User Satisfaction in Augmented Reality-Based Training Using Microsoft HoloLens." *Computers*, vol. 8, no. 1, p. 9, 2019.
- [55] P. Morillo, J. M. Orduña, S. Casas, and M. Fernández, "A comparison study of AR applications versus pseudo-holographic systems as virtual exhibitors for luxury watch retail stores," *Multimedia Syst.*, vol. 25, no. 4, pp. 307–321, Aug. 2019.
- [56] E. Azimi, L. Qian, N. Navab, and P. Kazanzides, "Alignment of the virtual scene to the tracking space of a mixed reality head-mounted display," 2018, *arXiv:1703.05834*. [Online]. Available: <https://arxiv.org/abs/1703.05834>
- [57] (Nov. 2019). *Lumus Optical*. [Online]. Available: <https://lumusvision.com/products/>



**FABRIZIO CUTOLO** (Member, IEEE) received the B.Sc. and M.Sc. degrees 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 Post-Graduate Research Associate with the Department of Information Engineering, University of Pisa. His research interests are in developing and evaluating new mixed reality solutions for image-guided surgery and surgical simulation, machine-vision applications, visual perception, ubiquitous tracking, and human-machine interfaces for rehabilitation. He has been involved in several national and international research projects. He is also a WP Leader of the HORIZON2020 Project VOSTARS, Call ICT-29-2016.



**BENISH FIDA** received the M.Sc. degree in computer science from the University Institute of Information Technology, PMAS Rawalpindi, Pakistan, in 2012, and the Ph.D. degree in biomedical engineering from the University of Roma Tre, Rome, Italy, in 2016. She is currently a Post-Graduate Researcher with the Department of Information Engineering, University of Pisa, Pisa, Italy. Her research interests are focused on the development of algorithms and solutions in the field of biomedical signal processing and computer vision. She is also involved in the HORIZON2020 Project VOSTARS as a Software Developer.



**NADIA CATTARI** received the M.Sc. degree in biomedical engineering from the University of Pisa, where she is currently pursuing the Ph.D. degree in clinical and translational science with the EndoCAS Centre. Her research activities span vision augmentation, stereoscopic 3D displays, computer graphics, analysis of perceptual issues in the interaction with augmented environments, and calibration in augmented reality. She is involved in research regarding the efficacy of wearable augmented reality head-mounted displays as tools for surgical guidance and/or surgical training in various types of procedures.



**VINCENZO FERRARI** 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. He is the author of more than 80 peer-reviewed publications and has five patents. He is also the Coordinator of the EndoCAS Centre, University of Pisa. His research interests involve image-guided surgery and simulation, computer vision and augmented reality devices and applications in medicine. He is involved in several national and international research projects. He is also the Coordinator of the HORIZON2020 Project VOSTARS, Call ICT-29-2016.

...