Real-time 3D scans of cardiac surgery using a single optical-see-through head-mounted display in a mobile setup

Bastian Dewitz[1,2]*, Roman Bibo[1], Sobhan Moazemi[1], Sebastian Kalkhoff[1], Stephan Recker[3], Artur Liebrecht[1], Artur Lichtenberg[1], Christian Geiger[4], Frank Steinicke[2], Hug Aubin[1] and Falko Schmid[1]

[1]Digital Health Lab Düsseldorf, Department of Cardiac Surgery, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany, [2]Institute for Human-Computer Interaction, Department of Informatics, Universität Hamburg, Hamburg, Germany, [3]Computer Science Faculty, Dortmund University of Applied Science and Arts, Dortmund, Germany, [4]Faculty of Media, University of Applied Sciences Düsseldorf, Düsseldorf, Germany

Microsoft HoloLens 2 (HL2) is often found in research and products as a cutting-edge device in Mixed Reality medical applications. One application is surgical telementoring, that allows a remote expert to support surgeries in real-time from afar. However, in this field of research two problems are encountered: First, many systems rely on additional sensors to record the surgery in 3D which makes the deployment cumbersome. Second, clinical testing under real-world surgery conditions is only performed in a small number of research works. In this article, we present a standalone system that allows the capturing of 3D recordings of open cardiac surgeries under clinical conditions using only the existing sensors of HL2. We show difficulties that arose during development, especially related to the optical system of the HL2, and present how they can be solved. The system has successfully been used to record surgeries from the surgeons point of view and the obtained material can be used to reconstruct a 3D view for evaluation by an expert. In a preliminary study, we present a recording of a captured surgery under real-world clinical conditions to expert surgeons which estimate the quality of the recordings and their overall applicability for diagnosis and support. The study shows benefits from a 3D reconstruction compared to video-only transmission regarding perceived quality and feeling of immersion.

KEYWORDS

HoloLens 2, telemedicine, telementoring, cardiac surgery, augmented reality, virtual reality

1 Introduction

The recent technological advances in the field of Mixed Reality (MR) research show great potential to support various tasks in our communication and interaction. One of these fields is medical telementoring, which describes the support of novice surgeons, the *mentees*, by expert surgeons from afar, the *mentors*, with the aim to improve operation results. Since the late 1980s, telementoring has been used in various scenarios, often to allow video-based real-time collaboration between two hospitals or between one hospital and a remote location, such as battle ships or mobile surgery units, to support real-time preparation and conduction of surgeries (Rosser et al., 2003, 2007). In many cases, telementoring performs comparable to traditional on-site mentoring and especially trainee surgeons benefit vastly considering operative times and surgery success, which implies an application of telementoring whenever on-site support is not possible (Erridge et al., 2019).

MR technology allows an elaborated form of telementoring that exceeds limits of video-based communication. The various means of communication provided by interactive Augmented Reality (AR) environments, such as interactive virtual objects, avatars, hand gestures, gaze tracking, video streaming and playback and audio communication (Galati et al., 2020), allow diverse ways of support of novice surgeons. Various systems have been proposed that suggest benefits in this area or demonstrate the technical feasibility. However, elaborate systems often use additional hardware such as depth sensors, cameras or tracking devices in a specially equipped operating room (OR) to provide an interactive communication platform. Further, in many cases, systems were not evaluated under clinical conditions (Birlo et al., 2022) which limits the findings to simulated environments.

In this research work, we present a standalone system using a single Microsoft HL2 as optical see-through head-mounted display (OST-HMD) to obtain a point cloud of a surgery utilizing only the built-in sensors. This point cloud can be streamed to a remote expert and viewed in virtual reality (VR) as immersive experience. It is specifically developed for use in clinical environments and tested with expert surgeons to evaluate the quality of the obtained data and possible applications. While we focus on capturing and evaluating 3D recordings of surgeries in this work, other important features for real-world deployment, such as bidirectional video and audio communication, are implemented as proof-of-concept to allow further investigations in the future.

The remainder of this paper structures as follows: First, we give an overview of related work regarding telementoring, 3D object reconstruction and extending the sensor capabilities of HL2. Second, we present our system with a special focus on unanticipated challenges that arose during development. Material and Methods of a preliminary study with surgeons that examined the quality of captured 3D recordings are presented in the next section, followed by the results of the

respective study. Finally, we discuss our findings and conclude on the application of HL2 as a single device for real-time 3D capturing of surgeries.

2 Related work

2.1 Optical see-through head-mounted display in telementoring

The application of MR technology in surgical context has been a popular interdisciplinary research topic in medicine, computer science, human-computer interaction and computer graphics for many years. In 1986, Roberts et al. introduced one of the first medical AR imagery devices that superimposed information from computer tomography scans upon the operation field as guidance (Roberts et al., 1986). Research-based and increasingly commercial deployments of AR technology in surgery have since then continuously risen. Besides OST-HMDs, various ways of displaying virtual content have been explored, including augmented optical systems and medical imaging devices and projection on patients (Sielhorst et al., 2008). In some cases, these early systems have been successfully deployed and evaluated in research projects under clinical conditions (Navab et al., 2012).

In the simplest case, OST-HMDs can be used to display important physiological parameters during operations by overlaying virtual content Schlosser et al. (2019). In more advanced cases, wearable spatially self-registering OST-HMDs such as Microsoft Hololens and Magic Leap can be used to display registered 3D content on the operation situs. One of the main research fields in the past 5 years for Hololens 1 and 2 as remote assistance devices are medical aids and systems (Park et al., 2021). Examples for application are, minimally invasive surgery training (Jiang et al., 2019), instrument navigation (Liebmann et al., 2019; Park et al., 2020) and overlays that contain medical data (García-Vázquez et al., 2018; Pratt et al., 2018). The most frequent experimental settings are phantom, system setup and simulator experiments whereas the application under real-world circumstances, especially for surgical procedures, has only been investigated in few research works (Park et al., 2021).

Only a few explicit telementoring systems utilizing OST-HMDs have been realized in the past and using this technology in hospitals is still not a common practice. The System for Telementoring with Augmented Reality (STAR) (Rojas-Muñoz et al., 2020a,b) transmits live RGB videos from a remote location to expert surgeons. In a similar approach, ARTEMIS (Gasques et al., 2021) equips the room with several sensors to allow a real-time 3D interaction with a generated point cloud. Similar to this approach, Long et al. (2022) transmit a 3D model for minimal invasive surgery to a remote expert for annotation. Research suggests that such telementoring systems have benefits regarding

task time and performance quality compared to audio only communication (Rojas-Muñoz et al., 2020b). To allow a real-world evaluation, some challenges still have to be overcome before such OST-HMDs can be used under surgical circumstances (Yoon, 2021). For example, the focal point of displays needs to be placed within the peri-personal space (a distance below 1 m) to avoid vergence-accomodation and focal rivalry (Carbone et al., 2020). Human factors play also a crucial part in the acceptance of new technology. Research work targeting these aims mostly at investigating spatial perception and mapping of 2D and 3D, attention shifts and surgical error caused by virtual content, and preferred visualization (Birlo et al., 2022). Research has shown that AR applications are appropriate for both experts and students in telementoring applications and user experience is mainly dependent on computer knowledge and on the role of participant (mentee vs. mentor) (Xue et al., 2019). Convenience and comfort are very important to allow an acceptance of OST-HMDs by surgeons under clinical circumstances. Complex surgeries can easily last several hours and surgeons are prevalent to neck and shoulder pain due to head-tilted and back-bend static postures during surgery (Berguer, 1999), even without additional head-worn devices. The weight of OST-HMDs must therefore be very balanced and low, so that the head can be tilted downwards for the time of surgery Carbone et al. (2020). In the future, with new generations of OST-HMDs which improve processing power, display quality and means of interaction, these issues will likely be resolved (Desselle et al., 2020) and this technology may prove to be beneficial in the practice of surgery.

2.2 Real-time 3D object reconstruction using RGBD sensors

Although equipping single operating rooms with RGB cameras is a widespread practice, for example by attaching cameras to surgical lights, real-time 3D scans of operations are not yet very common. Providing real-time 3D scans of surgeries is still a complex matter with no standardized way of capturing, sending and displaying data. Further, the utilization of additional devices is often necessary to perform tasks such as hand tracking, avatar tracking or tool tracking, or improve the quality of a 3D reconstruction (Rojas-Muñoz et al., 2020b; Gasques et al., 2021; Lawrence et al., 2021).

Reconstruction of 3D objects and environments has, however, been a main research area of computer vision and computer graphics and correspondingly the number of publications is large and diverse. Typical applications are RGBD-based odometry systems and autonomous robots, object and environmental scans using multiple frames of sensor streams, point cloud to object matching and related research questions (Zollhöfer et al., 2018). The sub-field of real-time 3D reconstruction of objects made great advances in

recent years with the emerge of commercially available and easily affordable RGBD-depth sensors and further, advances in calculation power and new algorithms.

Generally, a RGBD-sensor provides three-channel color video frames and a synchronized one-channel depth map for each pixel. To obtain a 3D representation from the RGBD data streams, the RGB frames and depth information have to be matched. Depending on the manufacturer, RGB and depth image pixels can already be associated and SDKs offer an easy method to access a colored point cloud. In other cases, the RGB image and depth image need to be aligned using the extrinsic and intrinsic camera parameters to transform pixel information from one sensor to the other. RGB and depth sensors are usually located close to each other to minimize the offset between RGB and depth pixels and avoid occlusions and missing pixels. For this task, various methods for this calibration exist, for example using point correspondences Nowak et al. (2021) or checkerboards Sylos Labini et al. (2019). After the colored point cloud is obtained, it can be rendered in different ways depending on the actual use case and requirements Zollhöfer et al. (2018). A simple way is displaying colored points in cartesian coordinates for each pixel or, more advanced, surfels (**surf**ace **el**ements) which render a lighting-affected disc at the point's position (Pfister et al., 2000). Many algorithms exist that allow a 3D mesh reconstruction from these points (Berger et al., 2014), for example, Delaunay triangulation (Lee and Schachter, 1980) and ball pivoting (Bernardini et al., 1999). Other approaches convert points to a volumetric representation and implicitly render a surface using signed distance functions and cube marching cubes (Lorensen and Cline, 1987).

Most low-cost RGBD-depth sensors can be divided into three categories: passive stereo camera systems, active infra-red stereo camera systems and active sensors. Passive stereo camera sensors use a pair of calibrated stereo cameras to match recognizable features of the environment. Examples for this type of sensors are Stereolabs ZED-mini (Stereolabs, 2022) and ORBBEC Astra (ORBBEC, 2022). Active stereoscopic sensors on the other hand, illuminate their surrounding with infrared light to reconstruct depth information. Examples for this category are Leap Motion and Intel Realsense. Sensors of the last category usually utilize either Structured Light (SL) or Time-of-Flight (ToF) approaches to estimate depth. In research, popular sensors of this category are Kinect v1 (SL), Kinect v2 (ToF), Azure Kinect (ToF) and the depth sensor of the HL2 (ToF). Most currently available sensors perform comparably well depending on the test scenario (Bajzik et al., 2020; Da Silva Neto et al., 2020).

2.3 Extending sensor capabilities of HoloLens

Only a small number of research articles exist that present the technical feasability of providing extra features for HoloLens

using additional sensors. Choudhary et al. (2021b) utilize an additional 4K camera mounted to the front of HL2 using a custom top-plate to enhance the resolution of the front camera. The systems relies on a backpack computer to capture the images and process them for a big-head magnification technique (Choudhary et al., 2021a) which can be displayed in the AR device. Leuze et al. (Leuze and Yang, 2018) use an extra depth sensor (Intel RealSense SR300) which is attached to the HoloLens 1 to allow marker-less tracking of facial features. It is, however, not connected to the HoloLens 1 but instead to an additional PC. Garon et al. (2017) mount a RealSense to the top of HoloLens 1 to enhance the quality of the depth sensor. This approach is more mobile as it uses an Intel Compute Stick PC which is attached to the back of the HMD which is only connected to a powerbank. The depth data is then streamed to the HoloLens using TCP/UDP over WiFi. Mohr et al. (2019) extended the HoloLens 1 with a planar tracking target, that allows them to use a smartphone as in input device, even outside of the field of view of the HoloLens's sensors.

Additionally to redundantly expanding the RGB and depth sensor capabilities, extending the HL2 with additional sensors has also been investigated. Erickson et al. (2019) attache thermal sensors to the HoloLens 1 and HL2 to extend human perception to see thermal radiation that is normally invisible to the human eye. The thermal sensors are also connected to a host PC *via* USB, that transmits the thermal data *via* WiFi to the HoloLens. Stearns et al. use external cameras, such as a finger-worn camera connected to a laptop (Stearns et al., 2017) or the camera of a handheld smartphone (Stearns et al., 2018), to enable users to see magnified content like text or images in the HoloLens. Besides scientific publications, a commercially available input pen that uses a 2-camera mount attached to the top of HL2 (XRGO, 2021) can be purchased as asset for HL2. Further, a Creative Commons-licensed 3D printing model exists, that can be used to mount a ZED mini stereoscopic depth camera to the top of HL2 (see (Clarke, 2020a)) and another to attach a Vive Puck to the back of HL2 (see (Clarke, 2020b)).

3 System description

3.1 Overview

While the utilization of OST-HMDs has been propesed in various medical areas, a limitation of this research field is the lack of studies that demonstrate the application in clinical situations (Birlo et al., 2022). From reviewing the literature and discussing existing systems with surgeons, we developed five key requirements, that need to be met to allow tests in real-world environments under clinical conditions and to make the system acceptable by surgeons:

- **Safety**: The system must neither increase existing, nor introduce new hazards to the patient during operation.

The system must not increase the risk of surgical error due to distractions or malfunctions.
- **Non-interference**: The setup should interfere as minimally as possible with the workflow of surgeries and tools of the surgeons, for example, individually fitted loupes. Surgeons must not need to adjust their postures to the device's requirements during operation.
- **Ease of Use**: The startup and operation needs to be as easy as possible. In this case, only the head straps of HL2 needs to be adjusted, which can easily be integrated into the *scrubbing*, the preparation process for surgeons before operation.
- **Mobility**: The setup is highly mobile and standalone to allow a fast deployment at various locations (within the hospital or outside). The HL2 and a mobile internet hotspot are the only devices for data acquisition and no additional computer is used for further processing in the OR.
- **Sterility**: Interaction with the HL2-system during surgery needs to be fully sterile. The device must not be touched during surgeries and interaction with the HL2 is therefore limited to in-air hand gestures and speech commands.

In contrast to other elaborated mutli-sensoric medical telementoring system, such as ARTEMIS (Gasques et al., 2021) and STAR (Rojas-Muñoz et al., 2020a) the environment has not to be euqipped with additional sensors, which makes deployment faster and simpler. This allows the use of the proposed system under clincal conditions. The transmitted data is not limited to RGB videos (as in (Rojas-Muñoz et al., 2020b)) and a 3D reconstruction of the operation can be provided to a remote expert.

During the development of the system, several difficulties arose that could negatively affect the operational capability of the system. In the following paragraphs, we show problems that were observed and describe approaches how they can be solved.

3.2 Software

The focus of development for HL2 lies in the easy deployment of AR Content in the real space using provided features such as hand and gesture tracking, voice commands, and 3D renderings. To enable sensor access of the HL2 for research-focused applications the HL2 has to be set to research mode and the Research Mode API (RM) by Ungureanu et al. (2020) has to be implemented. For this API, some publicly available wrappers exist that allow the utilization in Unity 3D (for example Wenhao (2021) and Gsaxner (2021)). In our approach, we developed a specialized custom DLL based on these works, that allows fast access to the depth sensor and other data. The software for HL2 was built using Unity 2019.4.34 as Universal Windows Platform (version 10.0.18362.0) application and ARM as target

FIGURE 1
Mirror Module attached to HL2. The mirror is attached with an angle of 30°. The upper part of the AHAT sensor is covered to eliminate reflections in the mirror. The spacing between mount and HL2 allows heat dissipation.

FIGURE 2
A surgeon wearing the HL2 with attached MirrorMount in the OR capturing scenes of an open-heart surgery in a typical posture during surgery. **(A)** shows the viewing axis of the surgeon, **(B)** the RGB camera axis and **(C)** the camera axis reflected by the mirror mount.

processor to allow the integration of external ARM-based plugins. The HL2 runs on version "20348.1447," as we experienced problems using the research mode with more recent versions.

3.3 Changing view direction of the color camera

The HL2 is intended to be used as communication device, especially for simulating face-to-face situations. The built-in color camera is therefore constructed to record objects in front of a forward-looking user. In the intended surgery scenario, this is not applicable as surgeons need to take a ergonomic posture in which they look downwards with a slightly forward-tilted head. Further, surgeons often use individually-fitted loupes that require a specifc head posture. As a result, recording the operation site without forcing a surgeon to take a unnatural posture is not possible using the HL2.

To overcome this limitation, we constructed a mountable mirror for the HL2 that changes the camera's viewing angle. After several prototyping and test phases, placing a mirror with a diameter of 5 cm directly above the color camera and tilting it downwards by 30°(see Figure 1) was identified as suitable arrangement. This shifts the center point of the color camera to the hands of the surgeon during operation while he or she is allowed to operate in his or her typical posture (see Figure 2).

Unfortunately, the mirror also extends into the field of view of the depth sensor of the HL2 and reflects infrared light emitted from the Time-of-Flight sensor for depth estimation. These reflections produce wrong depth information which negatively affects the 3D-reconstruction. To eliminate this possible source of error, an additional cover was placed on the upper half of the depth sensor to block the emitting IR light (visible in Figure 1). The resolution of the depth image is therefore reduced from 512×512 pixels to 512×320 pixels. The upper half is not needed to reconstruct the situs as the whole situs is visible (see Figure 5) and can therefore be discarded.

3.4 Adjusting exposure of the color camera

Lighting conditions in ORs in hospitals are typically very bright to allow a good visual perception of the situs. Usually, focused lights are pointed to the situs which produces a steep slope of brightness. This can easily be handled by the human
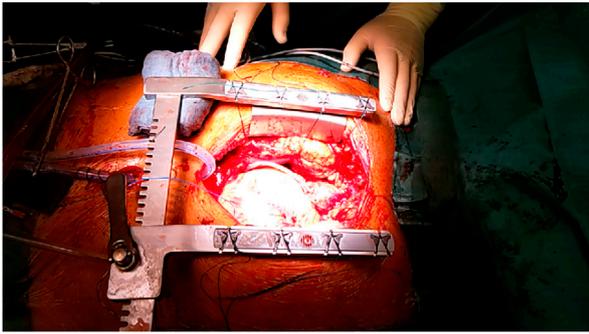
FIGURE 3
Overexposed RGB-image of open-heart surgery using automatic exposure control of HL2.

FIGURE 5
Cropped depth image from the HL2 AHAT sensor with depiction of the transmitted full-resolution region of interest (ROI) (yellow border) and the actual situs (blue border).

FIGURE 4
Visible rolling shutter artifacts of correctly exposed situs.

visual system but it can be problematic for photoelectric sensors. The high dynamic range can be seen in Figure 6, where the almost black surroundings of the situs are in fact bright cyan as seen in Figure 2. Using the default automatic exposure and ISO values from the HL2 leads to heavily overexposed images of the situs (see Figure 3), which results in the loss of important features. To obtain correctly exposed images of the situs, the HL2 needs to be modified. First, the HL2 has to be equipped with a neutral density (ND) that reduces the incoming light. In case of the ORs examined in this project, a ND-filter with ND = 4 (which equals a transmission of 6.25% of incoming light) allows a clear view on the situs while leaving enough headroom for corrections. The exposure and ISO values can be manually adjusted using the UWP-based *Windows.Media.Devices.VideoDeviceController*-methods *ExposureControl.SetValueAsync(TimeSpan)* and *IsoSpeedControl.SetValueAsync(UInt32)* as the automatic exposure with ND filter still fails to correctly expose the situs under OR lighting conditions. Only adjusting the exposure and ISO values in the software, on the other hand, leads to rolling

shutter artifacts (see Figure 4) which heavily reduce the image quality.

3.5 Data transmission

The HL2 can establish a connection to a network using the integrated WLAN or via USB-C cable. To transmit the data MixedReality-WebRTC 2.0.2 (Microsoft, 2022) (MRRTC) is used with RGB and depth video streams as video tracks plus an additional data channel. WebRTC has already been successfully utilized in comparable telementoring systems for real-time communication (See Rojas-Muñoz et al. (2020b); Gasques et al. (2021)). The video streams are encoded in YUV 4:2:0 format which are converted into RGB-values at the receiving workstation for further processing. The resolution of the color video stream is 960 × 540 pixels (the default video conferencing configuration in MRRTC). With higher resolutions, the frame rate dropped significantly and was estimated to be not usable in telementoring scenarios. The AHAT depth sensor of the HL2 has a resolution of 512 × 512 pixels (in our case, cropped to 512 × 320 pixels) with a depth resolution of 10 bit in the range between 0 and 1 m. The video encoding produces artifacts and limits the depth values to a resolution of 8bit (see Figure 7 for a comparison between full-resolution and encoded reconstruction quality).

As seen in Figures 5, 6, the region of interest (ROI) is only a small portion of the full depth and the full RGB image. To avoid loss of information or faulty information at the reconstuction end, an additional depth data stream is implemented to transmit a full-resolution (10 bit) depth stream window with an area of 200 × 140 pixels located at the lower area of the full depth image with 20 fps. To achieve an acceptable data rate, the depth information is compressed using the lossless RVL
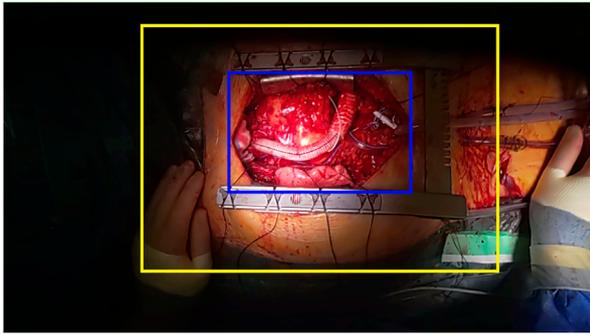
FIGURE 6
Corresponding RGB frame. The high dynamic range for correctly exposing bright parts of the situs results in partly almost black surroundings. The situs is highlighted in blue border.

compression (Wilson, 2017). The algorithm produces a byte stream with between 16.5 k and 17.5 k Byte per frame which is a compression rate of 65–70% compared to transmitting the whole data as unsigned 16 bit values (56 k Byte per frame) or around 50% respectively for only sending 10 bit-encoded frames (35 k Byte per frame). The three different streams (RGB, depth and ROI) are manually synchronized by adjusting individual delays per stream for incoming frames.

In case of a high bandwidth, the positional difference of single points of the point cloud is visible to the human eye when both point clouds are rendered side-by-side (2nd and 3rd image in Figure 7). By directly comparing the full-depth and encoded images, plateaus emerge for points with a similar distance to the sensor and additional structures (so called 'mosquito noise') appears around sharp edges. The difference between a high-bandwidth depth map and a full-resolution can also be calculated as mean absolute difference as 4.9 mm on average where 90% of values lie between 0.5 mm (5th percentile) and 13 mm (95th percentile) with a median of

3.9 mm (first quartile: 1.6 mm, third quartile: 6.9 mm). Depending on the connection speed and quality, the difference due to reduced resolution, artifacts and mosquito noise may increase drastically to a point where actual structures are not recognizable anymore, as seen in the first image of Figure 7.

In many use cases, for example when only a coarse 3D reconstruction needs to be acquired, the resolution and artifacts can be acceptable. In the medical domain this is however problematic as the judgment of a surgeon may be influenced. For example in heart surgery, the evaluation of myocardial contractions is important and artificially added or suppressed movements could lead to wrong decisions. In a 3D view, the full 10 bit resolution reconstruction reproduces a better volume of the heart and slanted surfaces are depicted more smoothly. The combination of both, a low-quality depth stream via MRRTC and a full-resolution depth stream of the ROI, allow examining the situs with the full resolution which is provided by the HL2 and perceiving the surroundings with a reduces resolution.

3.6 3D-reconstruction and Color Registration

HL2 provides two sensors that can be utilized to reconstruct a 3D view of a surgery. The locatable front color camera C_{RGB} for RGB images and the optical Time-Of-Flight (ToF) AHAT sensor C_{AHAT} for depth images. The raw AHAT image data contains the distance $d(x,y)_{AHAT}$ between sensor and environment up to 1 m mapped between 0 and 1000 for each pixel $p(x,y)_{AHAT}$. For both cameras, a pinhole camera model can be assumed to describe a mapping from 3D world coordinates to 2D image coordinates. OpenCV (Bradski, 2000) provides all necessary algorithms for this task.

To obtain a colored point cloud from both sensors, a camera calibration has to be performed which can be divided into three main steps:
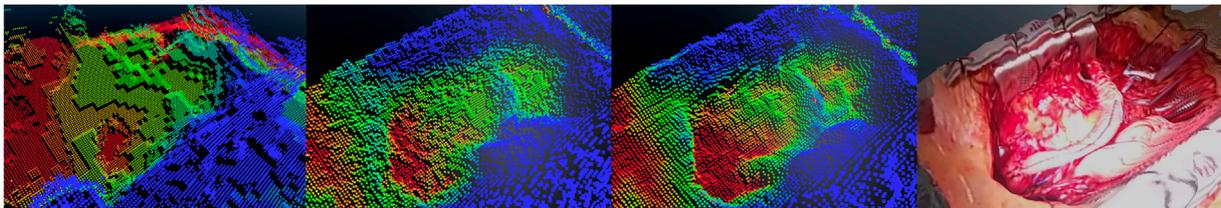
FIGURE 7
From left to right: Rendering of depth maps of the MRRTC-encoded depth map with low bandwidth (1), with high bandwidth (2) and full-resolution ROI (3). The final reconstruction (4) is provided for reference.

- **Calibration** of both sensors: Determination of camera intrinsics, extrinsics and distortion coefficients.
- **3D Reconstruction**: Calculation of 3D point cloud from depth data via pinhole model.
- **Color Registration**: Sampling of color information from the color camera to the point cloud.

First, calibration for both sensors is performed. It is a straightforward standard camera calibration with a checkerboard pattern as it was first described by Zhang (2000) to obtain intrinsic parameters K and distortion coefficients D for both sensors. Good results are obtained using a 4 × 5 checkerboard pattern with a square width of 50 mm and 30 to 40 images per sensor. The AHAT sensor is calibrated without the mirror mount to enable the use of the whole field of view of the sensor to calculate the intrinsic parameters and distortion coefficients, whereas The RGB camera is calibrated with the mirror mount attached to it. Fortunately, HL2 provides functionality that make the calibration process more convenient, compared to its predecessor HoloLens 1. For example, it is possible to utilize the depth sensor's active brightness frames for calibration, instead of the depth images. This allows the usage of a standard calibration checkerboard target rather than customized targets, as presented by Jung et al. (2015). The first steps yields two camera matrices K_{RGB} and K_{AHAT}, and two sets of sensor-specific distortion coefficients D_{RGB} and D_{AHAT}.

The second step in calibration is the calculation of the relative transformation which describes the transformation from the space of C_{AHAT} to the space of C_{RGB}. Assuming a perfectly placed mirror and a simplified model, the image of the color camera C_{RGB} is flipped vertically to allow an estimation of the camera pose with a correct orientation of the calibration pattern. OpenCV's provides the method *solvePNP* to allow the calculation of the posts of the cameras PC_{RGB} and PC_{AHAT}. With a calibration pattern visible in both sensors, the relative transformation RT from PC_{AHAT} to PC_{RGB} can be calculated using $RT = PC_{AHAT}^{-1} \cdot PC_{RGB}$. Placing the pattern in the center of the RGB sensor's image with an angle of 45° produced good results.

Finally, to obtain a colored point cloud, we adapted the method described by Sylos Labini et al. (2019) for HoloLens 1. The pixels of the AHAT image are deprojected using the distance $d(x, y)$ of each pixel $p(x,y)_{AHAT}$ in the AHAT depth image. The research mode of HL2 provides access to a look-up-table that describes the distortion of the depth image using the method *MapImagePointToCameraUnitPlane*. This method can be called in advance to obtain an float32 array that describes a 2D translation (dx, dy) for each pixel in the depth image which can be stored at the reconstruction end. A z-component of 1 is added to the described 2D translation and the resulting vector $(dx, dy, 1)$ is normalized to obtain homogeneous coordinates. The homogeneous coordinates are multiplied with the related

depth value $d(x, y)$ to obtain a point in 3D space in the camera space of C_{AHAT} with $P(x,y)_{AHAT} = d(x, y) \cdot norm(dx, dy, 1)$. To transform a 3D point to the camera space of C_{RGB}, the point is multiplied with the previously obtained matrix RT with $P(x,y)_{RGB} = RT \cdot P(x,y)_{AHAT}$. As OpenCV's coordinate system is different from Unity's, a conversion is necessary in which the y-component is flipped and the rotation around the x-axis and z-axis is inverted. Using K_{RGB} and D_{RGB}, $P(x,y)_{RGB}$ can be projected to the image coordinates $p'(u,v)_{RGB}$. If u and v of $p'(u, v)$ lie within the bounds of the RGB image, they can be used for a color lookup in the image to assign a color to the 3D point. As the RGB image is reflected in the mirror, it is again flipped vertically in this step. Points that do not lie in the RGB-image can be colored arbitrarily, for example, by mapping the distance to a color gradient.

3.7 Rendering the 3D point cloud

For each pixel in the depth image, a point can be rendered in real-time VR using Unity's ComputeBuffer calculations and geometry shaders that display a colored quad with a length of 1.5 mm (1st column in Figure 8). This reconstruction runs on the GPU and the added latency to the setup is negligible and mostly depended on the network infrastructure. For points that do not lie within the colored area, the normalized distance from a point to the depth sensor d is mapped to a gradient from teal for values close to the sensor to dark green for far-away values (visible on the right-hand side in Figure 8). These colors were chosen as they are easy to distinguish form the predominant red, violet and orange colors of the situs (3rd column in Figure 8). The video material resolution (around 300 × 200 pixels for the area within the sternum retractor (see Figure 6) is much higher than the depth image resolution (100 × 60 pixels for the area within sternum retractor, see Figure 5). To avoid loosing visual information of the RGB image, a bilateral interpolation is used to calculate sub-points between depth image points as support vectors. Each space between two adjacent depth image pixels is therefore represented by five points which results in an almost continuous surface (4th column in Figure 8).

At a distance of 50 cm, which is the typical working distance between the HL2 on the surgeon's head and the situs, the input signal of the Time-of-Flight sensor contains visible noise. The noise was measured as approximately normally distributed with a standard deviation of 2.28 mm (90% of noise lies between -4 and +4 mm for individual pixels). This results in a high-frequency bumpy surface of the 3D reconstruction with visually disturbing spikes and cavities. To minimize this noise, the 3D position of each point is smoothed by using a 5 × 5 discrete approximation of a Gaussian filter as spatial filtering for neighboring pixels (the results is shown in the 2nd column in Figure 8). After spatial smoothing, a temporal filtering that calculates the final position of a point $\dot{p}$ as $\dot{p}_t = d \cdot p_t + (1 - d) \cdot p_{t-1}$ with $d =$
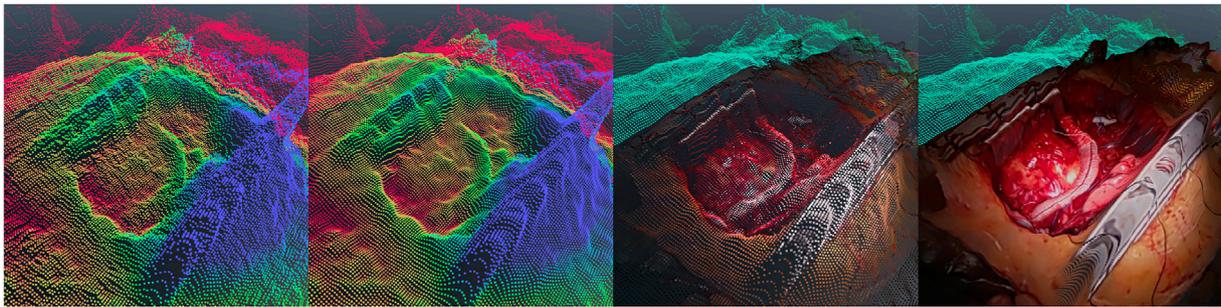
FIGURE 8
Visualization of different steps of the point cloud reconstruction. From left to right: (1) depth map (the distance to sensor is mapped to point color with blue for close point to red for far points), (2) reconstruction after Gaussian filtering, (3) RGB video stream mapped to depth pixels and (4) the result of sub pixel bilinear interpolation.

$\min\left(\frac{|p_t - p_{t-1}|_3}{.05}, .99\right)$ is used to reduce remaining jittering around the final position of a single 3D-point. The encoding of video frames and reducing the depth resolution to 8 bit also causes artifacts and removes information.

3.8 Remarks on augmented reality content

The presentation of AR content to surgeons was not a main focus in this research work. The transmission of data via MRRTC to synchronize the virtual content in the VR-control room and in the virtual scene of the AR-device has, however, been successfully implemented to allow future evaluations in this domain. Examples for future 3D content are pre- and intraoperative models and images, video and audio communication and sharing of arbitrary media content (Birlo et al., 2022). Research shows that placement and design of virtual objects under clinical conditions is still subject to current research regarding hologram accuracy (Vassallo et al., 2017; Gsaxner et al., 2019; Mitsuno et al., 2019; Galati et al., 2020; Haxthausen et al., 2021) and limitations due to the decreased field of view Galati et al. (2020). With a perfect 3D reconstruction which is mainly dependent on the sensor quality and calibration, it would be possible to attach virtual objects such as labels or draw lines on specific locations of the situs to support communication between remote expert and mentee.

While it has been demonstrated that magnifying AR-loupes can be attached to OST-HMDs (Qian et al., 2020), this approach was not considered in order to avoid hazards to patients due to falsely displayed virtual content. Instead, wearing the OST-HMD above the loupes allows surgeons to perform surgeries as usual by viewing the situs through magnification loupes, perceiving an AR-view by looking above the loupes through the OST-HMD which has been tested in a similar way with Google Glass (Yoon et al., 2017) and also, seeing the situs without magnifications looking below the loupes. By tilting their head, virtual content could also be aligned to the situs but, from a ergonomical point of view, the usability of this head posture is questionable and has to be further investigated in the future. This introduces new research questions that will be addressed in the future.

4 Materials and methods

4.1 Research question

The presented system allows capturing authentic operations under real-world conditions. The resolution is however limited by the built-in sensors of the HL2, so the question emerges whether the quality is sufficient for medical decision making. Furthermore, the mode of presentation is not limited to a 2D-RGB video stream but also allows for a 3D reconstruction. Therefore, three variables were analyzed as important factors for telemedicine applications: the feeling of presence with the two dimensions self-location (SL) and possible actions (PA), the subjective impression of media quality (MQ) and the medical appropriateness (MA). The mode of "presentation" was analyzed as a factor on three levels: a 2D video display ('2D'), a 3D reconstruction viewed on a 2D screen ("3D") and a 3D reconstruction viewed in an immersive virtual environment using a head-mounted VR-display. In order to get a clear insight in this question, a preliminary user study is performed with experienced surgeons that are able to compare the viewed material to their everyday work experience.

4.2 Data

The material showed a successful implantation of a LVAD (left ventricular assistance device) at the final stage of the surgery before stitching up of the situs began. The HL2 was worn by a surgeon in his typical posture. Visible were the beating right

A 2D video frame.

B 3D reconstruction.
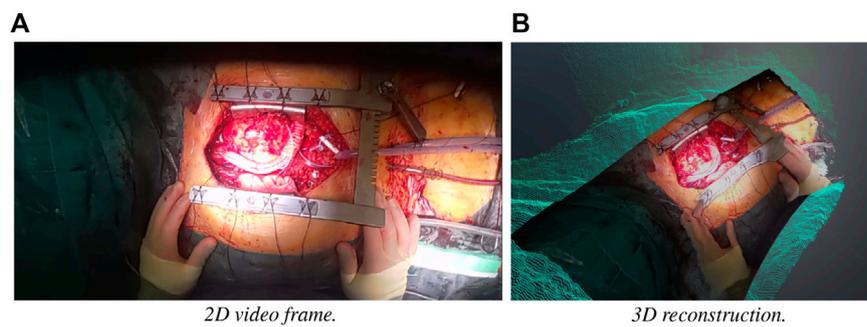
FIGURE 9
One frame of the recording used in the study. **(A)** RGB image in the video stream, **(B)** 3D reconstruction as point cloud.

ventricle, a small part of the right lung, the diaphragm, transparent surgical drains and the outflow track and driveline of the LVAD (see Figure 9A). For this stage of the prototype, we decided to show a looped 5 s clip of the beating heart without further instruments or actions as a baseline as the most basic case of material that could be used in a telementoring process for evaluation of the situs from afar. The exposure and iso were manually set to IS0 = 100 and EXP = 165750 to allow an examination of all areas of the situs (especially the less bright parts of the diaphragm on the right hand side). The material was recorded and played back using a dedicated software in Unity. The received video streams (rgb and depth) were stored with a frame rate of 30 fps as lossless '.tga' files and the ROI as byte array. For recording, the HL2 was directly connected to a Microsoft Surface Book 3 via USB NCM connection using a 2 m high-quality USB 3 cable to avoid transmission error.

4.3 Setup

For both the 2D view and the 3D model a LG 49WL95C-WE monitor was used to display the content in a window with 1920 × 1200 pixel (WUXGA) which produced a view area with a width of 44 cm and height of 28 cm. The participants were seated in a distance of 60 cm. For the VR presentation, a VIVE pro headset was used to display a simple virtual scene (a simple floor and only the colored pointcloud as object). All presentation were played on a Win 10 machine with an Intel i7 CPU, 64GB RAM and a Nvidia RTX 3090.

In all three cases, a comparable basic interaction is implemented to allow examination of the record. In case of 2D, the video material can be moved on the screen by pressing the left mouse button. It is also possible to perform zooming in and out using the mouse wheel. In case of 3D, a similar interaction is used: The mouse wheel controls the distance to the situs and the mouse can be dragged while pressing the left mouse button to rotate the camera around a point right in the

middle of the situs. Although this interaction limits the possible viewing angles it was preferred as no complex 3D navigation technique had to be learned by the participants that could negatively impact the impression of the virtual reconstruction. In VR, the situs can be explored in a natural way by walking around the 3D reconstruction and moving the head closer.

4.4 Procedure

The experiment is conducted as a within-subject (every participant experiences every presentation) design with the factor "presentation" at three levels ("2D," "3D," "VR"). The order of presentations was randomly assigned using a latin square in "Williams design." Before the experiment started, the use case of telementoring was briefly presented and important aspects of the presentation that had to be rated after each test run, such as media quality and medical appropriateness, were discussed. For each presentation, the participants were asked to freely explore the situs while having the telementoring application in mind. They were ordered to pay attention to medical details and image details that would allow decision making from afar and to state what came to their mind ("thinking-aloud"). Whenever they felt confident to rate the presentation they filled out a questionnaire. One experiment lasted typically around 10 min.

To measure feeling of presence, we used the spatial experience scale (SPES) (Hartmann et al., 2015) questionnaire. It consists of two sub scales with four items each, *self-location* (SL) which aims at the feeling of being present in a presentation and *possible actions* (PA) which aims at the feeling of being able to perform actions within a presentation. For perceived media quality (MQ), no standardized questionnaire exist, so four items were added to account for different important aspects of media material in surgery:

- **MQ1**: "The resolution and sharpness of the presentation was adequate."

- **MQ2**: "I experienced visually disturbing noise and artifacts."
- **MQ3**: 'I was able to follow the motion in the presentation'.
- **MQ4**: "The overall visual quality of the presentation was good,"

For medical appropriateness (MA), we integrated four statements that are relevant for telementoring in heart surgery:

- **MA1**: "I was able to clearly distinguish between anatomical landmarks."
- **MA2**: "I was able to evaluate myocardial contraction and cardiac function."
- **MA3**: "The presentation allowed a clear impression of the situation in the operation room."
- **MA4**: "I would be able to support medical decision making from afar using this presentation."

MQ and MA were developed with a senior surgeon at the hospital to incorporate medical expertise. All 16 questions were rated on a 7-item likert scale labeled with "fully disagree" (-3) at the left end, "neutral" (0) in the middle and "fully agree" at the right (+3).

4.5 Participants

We recruited 11 surgeons specialized in heart surgery (age: M = 36.9 SD = 5.5, 2 female, 9 male) from our hospital that have been performing surgeries for between one and 22 years (M = 11.4, SD = 6.0). One of these 11 participants had to be excluded from the study as he stated to have troubles "seeing sharp" using the HMD which may be related to him wearing glasses or a wrongly adjusted or slipped device on his head. From the remaining 10 surgeons, five stated to have "no" VR experience, three "little" and two "some."

5 Results

5.1 Analysis of measurements

The minimal, median and maximal response values from the questionnaires are displayed in Table 1 (2D, 3D, VR) alongside with the individual difference of ordinal categories for within-subject response between presentation (3D-2D, VR-2D, VR-3D). Kolmogorov-Smirnov test showed that in almost all cases answers on SL, PA, MQ and MA scales were not normally distributed. Friedman test was therefore selected as non-parametric test to analyze differences in rating of SL, PA, MQ and MA between presentation with repeated measurements. As post-hoc test, Wilcoxon signed-rank test was used for pairwise comparison. Benjamini-Hochberg procedure (Benjamini and

Hochberg, 1995) with a false discovery rate (FDR) of $\alpha = 0.1$ was used to account for multiple comparisons.

5.1.1 Spatial experience scale

The item responses on the SL and PA sub scales were averaged to obtain a questionnaire value for PA and SL. Descriptive Ratings for SL and PA are displayed in Table 1. Further, the difference in rating between presentations is displayed in Figure 10; On the SL sub scale, Friedman test showed a significant difference between groups: $\chi^2(2) = 18.05$, $p < 0.0002$. Post-hoc test showed a significant difference for 3D-2D ($p < 0.017$), VR-3D ($p < 0.017$) and VR-2D ($W = 0$, $p < 0.012$). On the PA sub scale, Friedman test showed a significant difference between groups: $\chi^2(2) = 19.1$, $p < 0.0001$. Post-hoc test showed a significant difference for 3D-2D ($p < 0.011$), VR-3D ($W = 0$, $p < 0.017$) and VR-2D ($W = 0$, $p < 0.011$). Comparing the median of ratings on both scales measured in SPES, it follows that in this setting regarding the feeling of presence a VR presentation provides the highest level of presence, followed by a 3D presentation and last a 2D presentation.

5.1.2 Media quality

Items of the MQ-scale were individually compared as absolute category response (ACR). Friedman test showed no significant difference for MQ1 ($\chi^2(2) = 0.95$, $p = 0.62$), MQ2 ($\chi^2(2) = 4.2$, $p = 0.12$) and MQ4 ($\chi^2(2) = 3.95$, $p = 0.14$). For MQ3, Friedman showed a significant difference ($\chi^2(2) = 10.05$, $p < 0.007$). Pairwise comparison showed a significant difference for VR-3D ($W = 5$, $p < 0.05$) and VR-2D ($W = 0$, $p < 0.05$). By comparing the medians of MQ3, the data suggests that the evaluation of motion does benefit from a virtual environment while 3D view and 2D video do not differ significantly.

5.1.3 Medical appropriateness

ACRs of the MA-scale were individually compared. Differences between groups are shown in Figure 12. Friedman tests showed no significant difference for MA1 ($\chi^2(2) = 5.85$, $p = 0.054$), MA3 ($\chi^2(2) = 2.85$, $p = 0.24$) and MA4 ($\chi^2(2) = 5.15$, $p = 0.076$). For MA2, the statistics showed a significant difference ($\chi^2(2) = 8.55$, $p = 0.01$). Post-hoc test showed a significant difference for VR-2D ($W = 0$, $p < 0.05$) and VR-3D ($W = 0$, $p < 0.05$). As the median of VR is higher than 2D and 3D it follows that surgeons felt significantly more able to evaluate heart function in VR.

5.2 Additional statements and observations

Six of the 10 participating surgeons showed a positive reaction (e.g., exclaiming "cool" or "awesome") when first experiencing the VR presentation. When VR was the first

TABLE 1 Median (bold) and range (min, max) for obtained data from the SPES questionnaire, MA and MQ. 2D, 3D and VR are the three levels of the factor "presentation". 3D-2D, VR-2D and VR-3D display the individual difference in rating.

	2D	3D	VR	3D-2D	VR-2D	VR-3D
SL	−**1.75** (−3, +1)	**+0.75** (−3, +2)	**+2** (+1.25, +3)	**+1** (0, +3)	**+3.25** (+1, +5.5)	**+1** (0, +5)
PA	−**1.75** (−3, 0)	**+0.25** (−2.75, +2)	**+2** (+0.25, +3)	**+1.75** (+0.25, +3.25)	**+3.25** (+1.75, +6)	**+1.25** (0, +4.25)
MQ1	**+1** (−3, +2)	**0** (−2, +2)	**+1** (−2, +2)	**0** (−4, +4)	**0** (−2, +5)	**0** (−1, +3)
MQ2	**0** (−3, +1)	**+1** (−3, +3)	**+1** (−3, +3)	**+1** (−1, +6)	**+1** (−1, +6)	**0** (−1, +1)
MQ3	**+1** (−2, +2)	**+2** (+1, +3)	**+2** (+2, +3)	**0** (−1, +3)	**+1** (0, +4)	**+1** (−1, +1)
MQ4	**+1** (−3, +2)	**0** (−2, +2)	**+2** (−1, +2)	**0** (−4, +3)	**+1** (0, +4)	**+1** (0, +4)
MA1	**+1** (−3, +3)	**+2** (0, +3)	**+2** (0, +3)	**+1** (−1, +3)	**+1** (−1, +4)	**0** (0, +1)
MA2	**+1** (−3, +3)	**+2** (0, +3)	**+2** (+2, +3)	**0** (−1, +4)	**+1** (0, +5)	**+1** (0, +2)
MA3	**0** (−3, +2)	**+1** (−1, +2)	**+1** (−2, +3)	**0** (−1, +4)	**+1** (−1, +5)	**+0** (−1, +2)
MA4	**+1** (−3, +3)	**+1** (−2, +3)	**+2** (−1, +3)	**+1** (−1, +4)	**+1** (−1, +5)	**+1** (−2, +2)

FIGURE 10
Distributions of difference in within-subjects ratings of the SL (A) and PA scale (B).
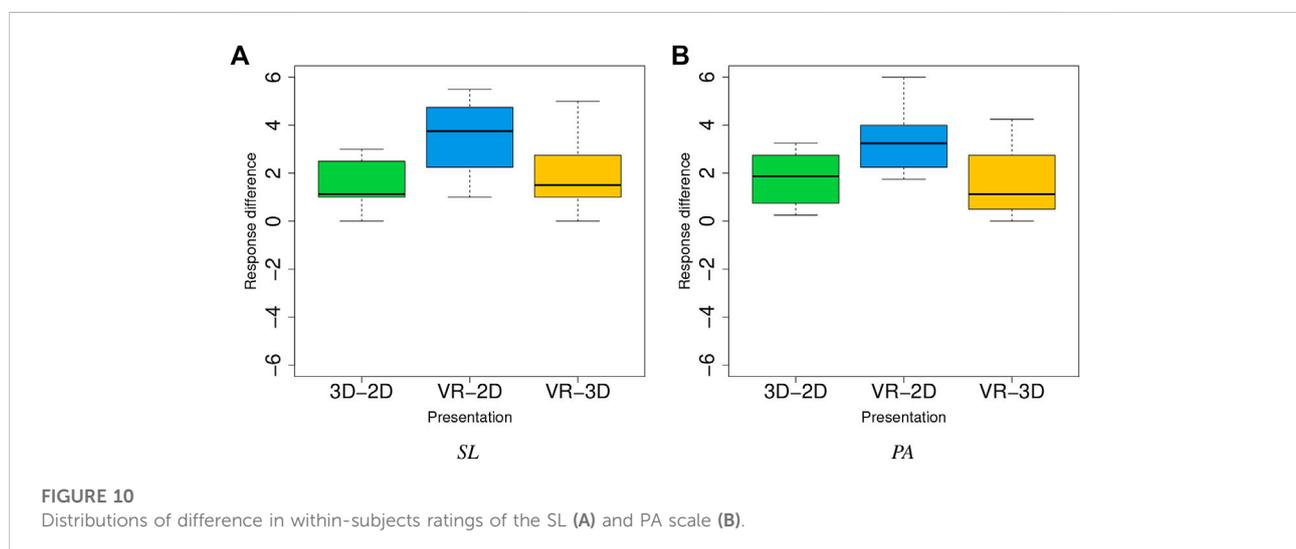
condition that was experienced, the subsequent material was in some cases considered as "boring" or it was stated that "it was better in VR." The time of being exposed to the presentations varied between subjects with a tendency of spending more time in the virtual environment, but the exact time for each condition was not measured as we did not anticipate this behavior.

Several surgeons complained about the reflex points on the surfaces and complained about the low resolution (independently from presentation). Two surgeons suggested to add parameters to make the situation in the OR more perceivable (e.g., electrocardiogram and infusion pump information).

All surgeons were able to intentionally examine the situs (e.g., zooming into specific structures) and no one reported any problems with the implemented controls or natural interaction in VR respectively, which implies that the implemented interaction techniques did not affect the ratings negatively.

6 Discussion

The proposed mounted mirror modification for the HL2 has successfully been tested under real-world conditions and demonstrates the ability to use HL2 for recording and streaming data in telementoring applications. The analysis of the data of our study shows, that the feeling of presence as measured with SPES is significantly increased and surgeons feel more like attending a surgery from afar using a VR presentation in contrast to 2D-bound presentations.

A 3D view on a screen does also increase the feeling of presence compared to a 2D video, but less than the VR presentation. This suggests that the process of telementoring might be improved to some extend by using 3D reconstructions as expert surgeons from afar could feel like being more involved in the activity around the operational situs and the OR. Further, the evaluation of motion (MQ3) and heart function (MA2) was

FIGURE 11
Distributions of difference in within-subjects ratings of MQ1 **(A)**, MQ2 **(B)**, MQ3 **(C)** and MQ4 **(D)**.
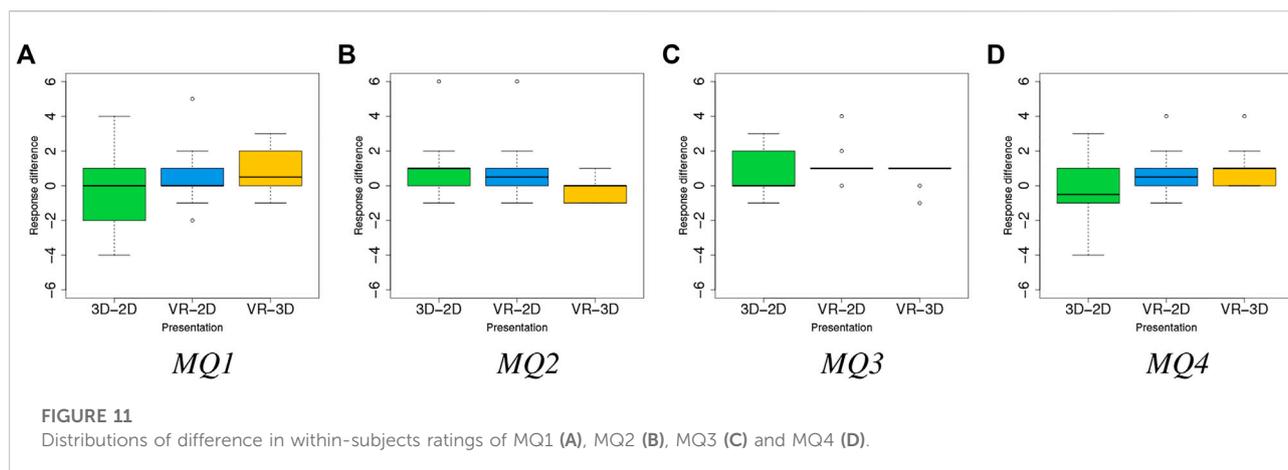
rated best in the VR presentation. The increased perception of motion and inherently better evaluation of cardiac function can probably be linked to the improved spatial perception in stereo vision using a head-mounted display. This finding stresses that task-specific immersive assistance systems might be beneficial if not even crucial for future telementoring applications.

The results of the presented study are, however, quite limited as we only showed a short clip without much specific surgical content (such as anastomosis techniques or tissue assessment). The time spend in the presentation was therefore considerably low. While the content seems a reasonable choice as baseline, more advanced scenes that show the use of tools, surgical techniques, or moments withing a surgery where a decision has to be taken could allow for more insight in more specific research questions. Considering that it was the first time of experiencing virtual reality for five of the surgeons the ratings may also be slightly biased. While SPES is a standardized tool to measure presence, the MQ and MA scale were developed by us to systematically capture the impression of surgeons. The high range of categorical responses may suggest that the questions were not precise enough or subjective biases played an important role. Further the within-subject differences showed a high variation for some items (especially visible in Figures 11A,D, 12A,B,D), which also implies individual differences that need further investigation. The difference in rating between presentation were borderline not significant for MQ2 ($p = 0.12$), MQ4 ($p = 0.14$) and especially MA1 ($p = 0.054$) and MA4 ($p = 0.076$). With further improvements of the system and more elaborated studies some findings may be made in the future considering these aspects.

The material used in our study itself was also not perfect, however, still significantly optimized in comparison with "of-the-shelf-technology." The decision to select an exposure setting that is intended to correctly exposing dark parts of the situs, seems to be not an ideal choice as reflections become more visible which was negatively commented by several surgeons. This may limit

the perception of details which should have presumably an uniform effect on all presentations. It would have been possible to select a lower exposure to reduce reflections and overexposure of bright parts in the image, although this would in return cause some loss of color information in the less exposed areas. It would probably be possible to enhance these dark parts using color correction curves which, on the other hand, may negatively impact the color fidelity. The correct settings need to be determined with surgeons, and further, in real-life use, the exposure settings need to be controllable to allow focusing on different lighting conditions and points of interest. In our recording, we did not experience any negative effect of reflections on tissue on the depth sensor, whereas the reflective material of the medical instruments and the sternum retractor causes missing information in certain areas of the reconstruction.

With the current system as presented in this work, various applications exist. While the system is not able to reconstruct fine details which would be necessary for a precise diagnosis from afar, it can produce an immersive experience of medical procedures. Especially major structures such as moving organs and tissue (for example, a beating heart) can be examined. The additional low-resolution point cloud allows an overview what is happening in an OR. A use as immersive educational tool that visualizes operations in 3D is imaginable. This could be used by students or doctors in training to learn, e.g., about specific pathologies, surgery techniques, and rare cases. In case of heart explantation, the battery life time of approximately 2–3 h of the HL2 is sufficient as this procedure usually lasts between 30 and 60 min. In other applications, the HL2 can be connected to a power supply while running the application using the USB port to expand the usage time.

Some improvements and modifications need to be made to allow a use in everyday clinical work. Other presentation that can easily be accessed such as 3D-displays and streaming of the obtained point cloud to AR-devices also seem worth exploring to

FIGURE 12
Distributions of difference in within-subjects ratings of MA1 **(A)**, MA2 **(B)**, MA3 **(C)** and MA4 **(D)**.
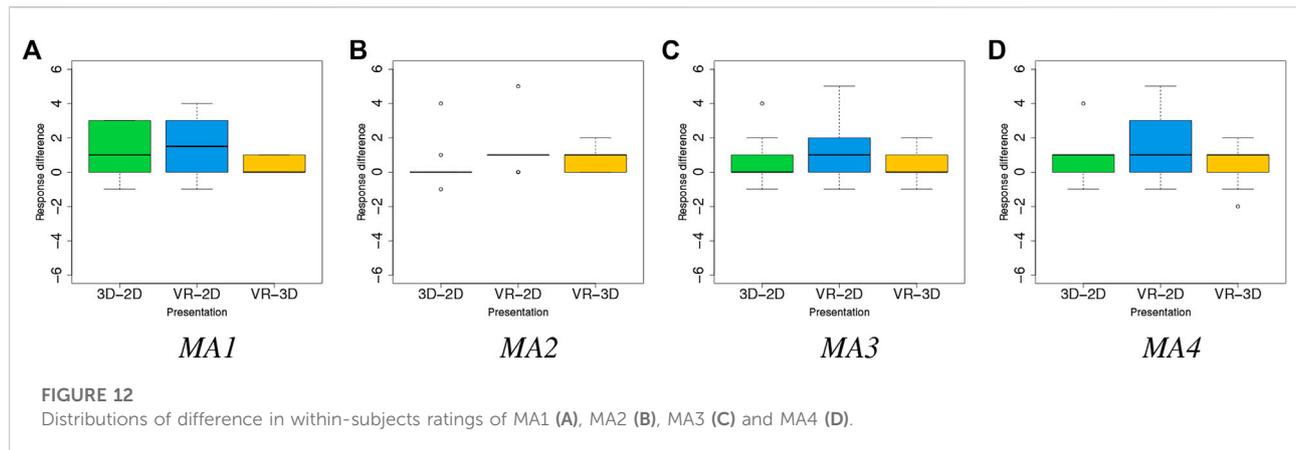
make this approach for telementoring more accessible. The RGB data quality may be improved by recording with a higher resolution on the HL2 (for video, HL2 supports up to 1920 × 1080 pixels at 30 fps) and cutting out the region of interest. The reconstruction with data of a single depth sensor becomes much more difficult when hands or tools are involved as parts of the situs are constantly occluded. Multiple combined sensors recording the situs from different angles will help to eliminate these movement induced local occlusions to a high degree. The color mapping uses a simple constrained approach in this prototype and advanced methods could be implemented to improve quality. To further improve the coherence of the point cloud, the utilization of two or more HL2, as well as potentially stationary sensors (e.g., built in OR-lights) which are registered in the same world space seems promising. By combining the different views on the situs occlusions and missing pixels could be avoided, interpolated or reconstructed with AI-based methods. At this point, the quality of depth sensor of a single HL2 seems barely sufficient to support the intended telementoring setting as the resolution (an area of around 100 × 60 pixels at he situs) and noise do not allow a precise reconstruction to make fine details visible.

Still, the use of machine learning (ML) using the data provided by the HL2 and subsequent devices of this family is interesting for advanced telementoring systems. ML-based models have already been applied to 3D images captured or visualized by HoloLens 1 and 2 for a variety of tasks such as face alignment (Kowalski et al., 2018) and object detection (Moezzi et al., 2022). Considering cardiovascular surgeons as the target users of our projects, advanced image processing and ML approaches have shown their potential for a variety of use-cases such as catheter and vessel segmentation as well as the ability to track object annotations as the organs move (Fazlali et al., 2018; Yi et al., 2020). However, under the current specifications, HL2 has several drawbacks limiting its performance for certain use-cases in cardiovascular surgery. First of all, its limited processing power makes it almost impossible to accommodate the integration of state-of-the-art ML models directly on the device. Second, the depth images suffer from relatively low resolution and signal to noise ratio. The RGB images also feature a relatively low resolution (300 × 200 pixels at the situs as displayed in Figure 6). Furthermore, differences in patients clinical characteristics and conditions might arise inconsistencies in the annotated datasets which can have negative influence on the performance of the ML models (Gudivada et al., 2017).

To mitigate the above-mentioned drawbacks in the follow-up studies, we consider collecting annotated data from multiple prospective operations as well as integrating new sensors with better precision leading to images with higher resolutions. This will facilitate incorporation of deep ML architectures such as convolutional neural networks (CNNs) and generative adversarial networks (GANs) which usually need bigger data cohorts with higher resolutions to converge. This should be emphasized that the integration of such neural networks would require higher processing power than what HoloLens is capable of. Therefore in real-time applications, the integration of advanced ML based models would need persistent wireless connections between the HL2 and a running computing server nearby. Although this would require to apply bandwidth adjustments to the running CNN models on the server to fit the maximum bandwidth supported by HL2, it has been considered feasible in a related work (Kowalski et al., 2018) Edge computing has become the method of choice for overcoming limitations of stationary and mobile devices which exhibit limitations in computation power (Mach and Becvar, 2017). Admittedly, to support an edge computing scenario, not only HL2 would need to be amended with high bandwidth network connectivity capabilities, also operation theaters would need to be equipped with appropriate radio technology. Taking into account typically low thresholds on emitted radio power, indoor 5G radio access cells can be considered a suitable

candidate to provide connectivity between edge compute nodes and HL2 kit in operation theaters (Bornkessel et al., 2021). Use cases with an interactive nature, such as telementoring, usually require a ultra low latency connection. Ultra low latency is one of the key features supported by 5G-based edge computing environments and consequently such scenarios can be investigated as one of the next steps of our research.

7 Conclusion and future work

In this paper we presented and analyzed the requirements of MR-based telementoring systems in real-world surgery settings. We identified, that for many surgeries, the usage of a HL2 and similar devices would result in an non-ergonomic posture, which especially is not possible during long duration of, e.g., cardiac surgeries. We developed the presented MirrorMount, that allows using the HL2 as telementoring device within surgeries under real-world conditions without negatively effecting the posture of surgeons—a crucial requirement for its general usability and acceptance. The device has successfully been tested to generate 3D scans that were suitable for evaluation by experienced surgeons.

The presented study shows that the obtained 3D reconstruction from a single HL2 allows for an immersive experience of operations from afar in VR. Surgeons felt more able to evaluate cardiac function which implies a possible usefulness in the medical domain. But, as our study is limited, we have to confirm these findings and their application in the real world in subsequent research work. Further, future research has to show how accurately the reconstruction is and if an application in potentially hazardous situations can be justified.

In the future, we are especially interested in evaluating advanced records of surgeries that show actual procedures and find ways for bidirectional communication that exploit the 3D reconstruction such as labels and drawn lines but also the transmission of video and possibly avatars. We also plan to use our system for live-streaming surgeries *via* a 5G network as soon as the technology is fully available at our hospital and measure the latencies of individual steps in the pipeline.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bajzik, J., Koniar, D., Hargas, L., Volak, J., and Janisova, S. (2020). "Depth sensor selection for specific application," in 2020 ELEKTRO, Taormina, Italy, May 25–28, 2020 (IEEE), 1–6.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Berger, M., Tagliasacchi, A., Seversky, L. M., Alliez, P., Levine, J. a., Sharf, A., et al. (2014). "State of the art in surface reconstruction from point clouds," in Proceedings of the Eurographics 2014, Eurographics STARs, Strasbourg, France, April 7–11, 2014, 161–185.1

Berguer, R. (1999). Surgery and ergonomics. *Arch. Surg.* 134, 1011–1016. doi:10.1001/archsurg.134.9.1011

Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., and Taubin, G. (1999). The ball-pivoting algorithm for surface reconstruction. *IEEE Trans. Vis. Comput. Graph.* 5, 349–359. doi:10.1109/2945.817351

Birlo, M., Edwards, P. J. E., Clarkson, M., and Stoyanov, D. (2022). Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. *Med. Image Anal.* 77. doi:10.1016/j.media.2022.102361

Bornkessel, C., Kopacz, T. J., Schiffarth, A.-M., Heberling, D., and Hein, M. A. (2021). "Determination of instantaneous and maximal human exposure to 5g massive-mimo base stations," in 2021 15th European Conference on Antennas and Propagation (EuCAP), Dusseldorf, Germany, March 22–26, 2021, 1–5.

Bradski, G. (2000). The OpenCV library. Dr. Dobb's Journal of Software Tools

Carbone, M., Piazza, R., and Condino, S. (2020). Commercially available head-mounted displays are unsuitable for augmented reality surgical guidance: A call for focused research for surgical applications. *Surg. Innov.* 27, 254–255. doi:10.1177/1553350620903197

Choudhary, Z., Bruder, G., and Welch, G. (2021a). *Scaled user embodied representations in virtual and augmented reality*, 1. Association for Computing Machinery. doi:10.18420/muc2021-mci-ws16-361

Choudhary, Z., Ugarte, J., Bruder, G., and Welch, G. (2021b). "Real-time magnification in augmented reality," in Proceedings - SUI 2021: ACM Spatial User Interaction 2021, Ingolstadt, Germany, September 5–8, 2021, Virtual Event, United States, November 9–10, 2021 (Association for Computing Machinery, Inc). doi:10.1145/3485279.3488286

Clarke, T. (2020a). *Hololens 2 mount for ZED mini*. [Dataset]. Available at: Https://www.thingiverse.com/thing:4561113 (Accessed Feburary 24, 2020).

Clarke, T. (2020b). *Hololens 2 vive Puck back mount*. [Dataset]. Available at: Https://www.thingiverse.com/thing:4657299 (Accessed Feburary 24, 2020).

Da Silva Neto, J. G., Da Lima Silva, P. J., Figueredo, F., Teixeira, J. M. X. N., and Teichrieb, V. (2020). "Comparison of RGB-D sensors for 3D reconstruction," in Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020, Porto de Galinhas, Brazil, November 07–10, 2020, 252–261. doi:10.1109/SVR51698.2020.00046

Desselle, M. R., Brown, R. A., James, A. R., Midwinter, M. J., Powell, S. K., and Woodruff, M. A. (2020). Augmented and virtual reality in surgery. *Comput. Sci. Eng.* 22, 18–26. doi:10.1109/mcse.2020.2972822

Erickson, A., Kim, K., Schubert, R., Bruder, G., and Welch, G. (2019). "Is it cold in here or is it just me? analysis of augmented reality temperature visualization for computer-mediated thermoception," in 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, October 14–18, 2019, 202–211.

Erridge, S., Yeung, D. K., Patel, H. R., and Purkayastha, S. (2019). Telementoring of surgeons: A systematic review. *Surg. Innov.* 26, 95–111. doi:10.1177/1553350618813250

Fazlali, H. R., Karimi, N., Soroushmehr, S. M. R., Shirani, S., Nallamothu, B. K., Ward, K. R., et al. (2018). Vessel segmentation and catheter detection in X-ray angiograms using superpixels. *Med. Biol. Eng. Comput.* 56, 1515–1530. doi:10.1007/s11517-018-1793-4

Galati, R., Simone, M., Barile, G., De Luca, R., Cartanese, C., and Grassi, G. (2020). Experimental setup employed in the operating room based on virtual and mixed reality: Analysis of pros and cons in open abdomen surgery. *J. Healthc. Eng.* 2020, 1–11. doi:10.1155/2020/8851964

García-Vázquez, V., Von Haxthausen, F., Jäckle, S., Schumann, C., Kuhlemann, I., Bouchagiar, J., et al. (2018). Navigation and visualisation with hololens in endovascular aortic repair. *Innov. Surg. Sci.* 3, 167–177. doi:10.1515/iss-2018-2001

Garon, M., Boulet, P. O., Doironz, J. P., Beaulieu, L., and Lalonde, J. F. (2017). "Real-time high resolution 3D data on the HoloLens," in Adjunct Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2016, Merida, Mexico, September 19–23, 2016, 189–191. doi:10.1109/ISMAR-Adjunct.2016.0073

Gasques, D., Johnson, J. G., Sharkey, T., Feng, Y., Wang, R., Robin, Z., et al. (2021). "Artemis: A collaborative mixed-reality system for immersive surgical telementoring," in Conference on Human Factors in Computing Systems - Proceedings, Yokohama, Japan, May 8–13, 2021 (Association for Computing Machinery). doi:10.1145/3411764.3445576

Gsaxner, C., Pepe, A., Wallner, J., Schmalstieg, D., and Egger, J. (2019). "Markerless image-to-face registration for untethered augmented reality in head and neck surgery," in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) 11768 LNCS*, 236–244. doi:10.1007/978-3-030-32254-0_27

Gsaxner, C. (2021). *HoloLens2-Unity-ResearchModeStreamer*. [Dataset]. Available at: Https://github.com/cgsaxner/HoloLens2-Unity-ResearchModeStreamer (Accessed Feburary 24, 2022).

Gudivada, V., Apon, A., and Ding, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *In: Int. J. Adv. Softw*, 10.1, 1–20.

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., et al. (2015). The spatial presence experience scale (spes). *J. Media Psychol.* 1, 1–15. doi:10.1027/1864-1105/a000137

Haxthausen, F. V., Chen, Y., and Ernst, F. (2021). Superimposing holograms on real world objects using HoloLens 2 and its depth camera. *Curr. Dir. Biomed. Eng.* 7, 111–115. doi:10.1515/cdbme-2021-1024

Jiang, H., Xu, S., State, A., Feng, F., Fuchs, H., Hong, M., et al. (2019). Enhancing a laparoscopy training system with augmented reality visualization. *Simul. Ser.* 51. doi:10.23919/SpringSim.2019.8732876

Jung, J., Lee, J.-Y., Jeong, Y., and Kweon, I. S. (2015). Time-of-flight sensor calibration for a color and depth camera pair. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1501–1513. doi:10.1109/TPAMI.2014.2363827

Kowalski, M., Nasarzewski, Z., Galinski, G., and Garbat, P. (2018). HoloFace: Augmenting human-to-human interactions on HoloLens. Tech. Rep. arXiv:1802.00278, arXiv. ArXiv:1802.00278 [cs] type: article

Lawrence, J., Goldman, D. B., Achar, S., Blascovich, G. M., Desloge, J. G., Fortes, T., et al. (2021). *Project starline: A high-fidelity telepresence system*. New York, NY: Association for Computing Machinery

Lee, D. T., and Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Inf. Sci.* 9, 219–242. doi:10.1007/BF00977785

Leuze, C. M. J., and Yang, G. (2018). "Marker-less co-registration of brain MRI data to a subject's head via a mixed reality device," in Proceedings of the Society for Neuroscience, Paris, France, June 16–21, 2018.

Liebmann, F., Roner, S., von Atzigen, M., Scaramuzza, D., Sutter, R., Snedeker, J., et al. (2019). Pedicle screw navigation using surface digitization on the microsoft hololens. *Int. J. Comput. Assist. Radiol. Surg.* 14, 1157–1165. doi:10.1007/s11548-019-01973-7

Long, Y., Li, C., and Dou, Q. (2022). *Robotic surgery remote mentoring via ar with 3d scene streaming and hand interaction*. arXiv preprint arXiv:2204.04377.

Lorensen, W. E., and Cline, H. E. (1987). "Marching cubes: A high resolution 3D surface construction algorithm," in Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987, Anaheim, CA, July 27–31, 1987, 163–169. doi:10.1145/37401.37422

Mach, P., and Becvar, Z. (2017). Mobile edge computing: A survey on architecture and computation offloading. *IEEE Commun. Surv. Tutorials* 19, 1628–1656. doi:10.1109/comst.2017.2682318

Microsoft (2022). *MixedReality-WebRTC*. [Dataset]. Available at: Https://github.com/microsoft/MixedReality-WebRTC (Accessed Feburary 24, 2020).

Mitsuno, D., Ueda, K., Hirota, Y., and Ogino, M. (2019). Effective application of mixed reality device hololens: simple manual alignment of surgical field and holograms. *Plastic Reconstr. Surg.* 143, 647–651. doi:10.1097/prs.0000000000005215

Moezzi, R., Krcmarik, D., Cýrus, J., Bahri, H., and Koci, J. (2022). "Object detection using microsoft HoloLens by a single forward propagation CNN," in *Proceedings of the international conference on intelligent vision and computing (ICIVC 2021)*. Editors H. Sharma, V. K. Vyas, R. K. Pandey, and M. Prasad (Cham: Springer International Publishing), Proceedings in Adaptation, Learning and Optimization. 507–517. doi:10.1007/978-3-030-97196-0_42

Mohr, P., Tatzgern, M., Langlotz, T., Lang, A., Schmalstieg, D., and Kalkofen, D. (2019). "Trackcap: Enabling smartphones for 3d interaction on mobile head-mounted displays," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, May 4–9, 2021.

Navab, N., Blum, T., Wang, L., Okur, A., and Wendler, T. (2012). First deployments of augmented reality in operating rooms. *Computer* 45, 48–55. doi:10.1109/MC.2012.75

Nowak, J., Fraisse, P., Cherubini, A., Daurès, J.-P., and Daures, J.-P. (2021). *Point clouds with color: A simple open library for matching RGB and depth pixels from an uncalibrated stereo pair point clouds with color: A simple open library for matching RGB and depth pixels from an uncalibrated stereo pair point clouds with color: A.* Karlsruhe, Germany: IEEE, 1–7. doi:10.1109/MFI52462.2021.9591200ï

ORBBEC (2022). Astra — orbbec 3D. [Dataset]. Available at: Https://shop.orbbec3d.com/Astra (Accessed April 14, 2022]

Park, B. J., Hunt, S. J., Nadolski, G. J., and Gade, T. P. (2020). Augmented reality improves procedural efficiency and reduces radiation dose for ct-guided lesion targeting: a phantom study using hololens 2. *Sci. Rep.* 10, 18620–18628. doi:10.1038/s41598-020-75676-4

Park, S., Bokijonov, S., and Choi, Y. (2021). Review of microsoft hololens applications over the past five years. *Appl. Sci. Switz.* 11, 7259. doi:10.3390/app11167259

Pfister, H., Zwicker, M., Van Baar, J., and Gross, M. (2000). "Surfels: Surface elements as rendering primitives," in Proceedings of the ACM SIGGRAPH Conference on Computer Graphics, New Orleans, LA, July 23–28, 2000, 335–342.

Pratt, P., Ives, M., Lawton, G., Simmons, J., Radev, N., Spyropoulou, L., et al. (2018). Through the hololens™ looking glass: augmented reality for extremity reconstruction surgery using 3d vascular models with perforating vessels. *Eur. Radiol. Exp.* 2, 1–7.

Qian, L., Song, T., Unberath, M., and Kazanzides, P. (2020). AR-Loupe: Magnified augmented reality by combining an optical see-through head-mounted display and a loupe. *IEEE Trans. Vis. Comput. Graph.* 2626, 2550–2562. doi:10.1109/TVCG.2020.3037284

Roberts, D. W., Strohbehn, J. W., Hatch, J. F., Murray, W., and Kettenberger, H. (1986). A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *J. Neurosurg.* 65, 545–549. doi:10.3171/jns.1986.65.4.0545

Rojas-Muñoz, E., Cabrera, M. E., Lin, C., Andersen, D., Popescu, V., Anderson, K., et al. (2020a). The system for telementoring with augmented reality (STAR): A head-mounted display to improve surgical coaching and confidence in remote areas. *Surg. (United States)* 167, 724–731. doi:10.1016/j.surg.2019.11.008

Rojas-Muñoz, E., Lin, C., Sanchez-Tamayo, N., Cabrera, M. E., Andersen, D., Popescu, V., et al. (2020b). Evaluation of an augmented reality platform for austere surgical telementoring: a randomized controlled crossover study in cricothyroidotomies. *Npj Digit. Med.* 3, 75–79. doi:10.1038/s41746-020-0284-9

Rosser, J. C., Herman, B., and Giammaria, L. E. (2003). Telementoring. *Semin. Laparosc. Surg.* 10, 209–217. doi:10.1177/107155170301000409

Rosser, J. C., Young, S. M., and Klonsky, J. (2007). Telementoring: An application whose time has come. *Surg. Endosc.* 21, 1458–1463. doi:10.1007/s00464-007-9263-3

Schlosser, P. D., Grundgeiger, T., Sanderson, P. M., and Happel, O. (2019). An exploratory clinical evaluation of a head-worn display based multiple-patient monitoring application: impact on supervising anesthesiologists' situation awareness. *J. Clin. Monit. Comput.* 33, 1119–1127. doi:10.1007/s10877-019-00265-4

Sielhorst, T., Feuerstein, M., and Navab, N. (2008). Advanced medical displays: A literature review of augmented reality. *J. Disp. Technol.* 4, 451–467. doi:10.1109/JDT.2008.2001575

Stearns, L., DeSouza, V., Yin, J., Findlater, L., and Froehlich, J. E. (2017). "Augmented reality magnification for low vision users with the microsoft hololens and a finger-worn camera," in Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, Baltimore, MD, October 30–November 1, 2017 (New York, NY, USA: Association for Computing Machinery), 361–362. ASSETS '17. doi:10.1145/3132525.3134812

Stearns, L., Findlater, L., and Froehlich, J. E. (2018). "Design of an augmented reality magnification aid for low vision users," in Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, October 22–24, 2018 (New York, NY, USA: Association for Computing Machinery), 28–39. ASSETS '18. doi:10.1145/3234695.3236361

Stereolabs (2022). *ZED mini - mixed-reality camera — Stereolabs.* [Dataset]. Available at: Https://www.stereolabs.com/zed-mini/ (Accessed April 14, 2022).

Sylos Labini, M., Gsaxner, C., Pepe, A., Wallner, J., Egger, J., and Bevilacqua, V. (2019). "Depth-awareness in a system for mixed-reality aided surgical procedures," in *Intelligent computing methodologies*. Editors D.-S. Huang, Z.-K. Huang, and A. Hussain (Cham: Springer International Publishing), 716–726.

Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., et al. (2020). *HoloLens 2 research mode as a tool for computer vision research* arXiv.

Vassallo, R., Rankin, A., Chen, E. C. S., and Peters, T. M. (2017). "Hologram stability evaluation for microsoft HoloLens," in *Medical Imaging 2017: Image Perception Observer Performance, and Technology Assessment*, 1013614. doi:10.1117/12.2255831

Wenhao (2021). *HoloLens2-ResearchMode-Unity*. [Dataset]. Available at: Https://github.com/petergu684/HoloLens2-ResearchMode-Unity (Accessed Feburary 24, 2022).

Wilson, A. D. (2017). "Fast lossless depth image compression," in Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, Brighton, UK, October 17–20, 2017, 100–105.

XRGO (2021). Stylus XR – AR/MR/VR input device. [Dataset]. Available at: Https://xrgo.io/produkt/stylus-xr (Accessed Feburary 24, 2022)

Xue, H., Sharma, P., and Wild, F. (2019). User satisfaction in augmented reality-based training using microsoft HoloLens. *Computers* 8, 9. doi:10.3390/computers8010009

Yi, X., Adams, S., Babyn, P., and Elnajmi, A. (2020). Automatic catheter and tube detection in pediatric X-ray images using a scale-recurrent network and synthetic data. *J. Digit. Imaging* 33, 181–190. doi:10.1007/s10278-019-00201-7

Yoon, J. W., Chen, R. E., ReFaey, K., Diaz, R. J., Reimer, R., Komotar, R. J., et al. (2017). Technical feasibility and safety of image-guided parieto-occipital ventricular catheter placement with the assistance of a wearable head-up display. *Int. J. Med. Robot. Comput. Assist. Surg.* 13, e1836. doi:10.1002/rcs.1836

Yoon, H. (2021). Opportunities and challenges of smartglass-assisted interactive telementoring. *Appl. Syst. Innov.* 4, 56. doi:10.3390/asi4030056

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1330–1334. doi:10.1109/34.888718

Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R., et al. (2018). State of the art on 3D reconstruction with RGB-D cameras. *Comput. Graph. Forum* 37, 625–652. doi:10.1111/cgf.13386