Parallax Free Registration for Augmented Reality Optical See-through Displays in the Peripersonal Space

Vincenzo Ferrari *Member, IEEE*, Nadia Cattari, Umberto Fontana, and Fabrizio Cutolo *Member, IEEE*

Abstract—Egocentric augmented reality (AR) interfaces are quickly becoming a key asset for assisting high precision activities in the peripersonal space in several application fields. In these applications, accurate and robust registration of computer-generated information to the real scene is hard to achieve with traditional Optical See-Through (OST) displays given that it relies on the accurate calibration of the combined eye-display projection model. The calibration is required to efficiently estimate the projection parameters of the pinhole model that encapsulate the optical features of the display and whose values vary according to the position of the user's eye. In this work, we describe an approach that prevents any parallax-related AR misregistration at a pre-defined working distance in OST displays with infinity focus; our strategy relies on the use of a magnifier placed in front of the OST display, and features a proper parameterization of the virtual rendering camera achieved through a dedicated calibration procedure that accounts for the contribution of the magnifier. We model the registration error due to the viewpoint parallax outside the ideal working distance. Finally, we validate our strategy on a OST display, and we show that sub-millimetric registration accuracy can be achieved for working distances of ± 100 mm around the focal length of the magnifier.

Index Terms—Augmented reality, optical see-through, registration

1 Introduction

Wearable augmented reality (AR) systems based on head mounted displays (HMDs) have stimulated the technological growth of a broad range of consumer applications. AR HMDs preserve the user's egocentric viewpoint and, by smoothly augmenting the visual perception of the world, they represent the most ergonomic and reliable solutions to aid complex manual tasks performed under the user's direct vision, such as in surgical procedures [2, 10, 12, 36, 37].

In AR HMDs, the see-through capability can be accomplished either through a video see-through (VST) mechanism or through an optical see-through (OST) mechanism. In VST HMDs, the view of the real world is mediated by video cameras rigidly anchored to the visor. The camera frames are digitally blended with the virtual content and the resulting video streams are displayed on the microdisplays of the visor. In OST HMDs, the user's direct view of the world is optically merged, through an optical combiner, with the computer-generated content [34]. This aspect offers a clear advantage over VST solutions, particularly when used to interact with objects in the peripersonal space, since it allows users to maintain their own natural visual experience with almost no alterations [5, 15].

The OST paradigm implemented in commercial HMDs is fundamentally the same as that described by Benton back in 2001 [3] and successfully proposed in a wide range of commercial AR visors, both monocular and binocular: Google Glass, Microsoft HoloLens, Meta 2, Epson Moverio, Vuzix devices, Lumus Optical devices, Optinvent ORA2, ODG R-9, etc. The user's own view is augmented by rendering the virtual content on a two-dimensional (2D) microdisplay placed outside the user's field of view and by guiding the display light to the user's eye(s) by means of an optical combiner. In this way, the user perceives both the external scene light and the light emitted from the microdisplay offset from the viewing region. Collimation lenses are placed between the microdisplay and the optical combiner to focus the virtual 2D image so that it appears at a pre-defined and comfortable viewing distance on a virtual focal plane [19, 33].

OST HMDs are at the cutting edge of AR research and are the most promising output medium for AR applications. However, mass adoption is still far from becoming reality due to various technological and human-factor limitations. These include the sub-optimal form factor and weight, which prevents them from being worn for over long periods of time, the small augmentable fields of view they can offer, which cannot be expanded without losing pixel density, and the low brightness images they generate [26].

Human-factor issues result from the perceptual limitations typical of all traditional near-eye-displays (NEDs) with fixed focal planes, such as: focal rivalry (FR) and vergence-accommodation conflict (VAC) [7, 9, 30]. FR is related to the inability to render suitable focus cues to stimulate natural eye accommodation responses: when viewing through a traditional OST display users are forced to accommodate their eyes either to the real scene or to the virtual surface of projection [21, 22]. Unfortunately, the focal plane of many consumer-level OST NEDs is at infinity (e.g., Optinvent ORA2 and Lumus Optical) or at a distance that is incompatible with its use as an aid to manual activities (as it is far from the peripersonal space) [7, 9]. This not only leads to visual fatigue but also reduces user performance in completing manual tasks that require keeping both the real and the virtual information in focus simultaneously, such as integrating virtual and real information when reading [16], or connecting points with a line [7].

The well-known VAC [25] is inherent in all conventional AR and VR binocular displays with fixed focal planes. In these displays, stereopsis is stimulated from a pair of 2D perspective images. In fact, when viewing through a binocular display, the generated disparities force users to fix on the object with the same vergence as in natural viewing, but, at the same time, their eyes are forced to focus on the imaging plane of the display and not where the objects are placed in the space. The correlation between vergence and accommodation is thus disrupted with consequent discomfort for the user [11].

- *Vincenzo Ferrari is with the Information Engineering Department, University of Pisa, Via G. Caruso 16, 56122 Pisa, and with the EndoCAS Centre, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56124 Pisa, Italy E-mail: vincenzo.ferrari@unipi.it.*
- *Nadia Cattari is with the Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56124 Pisa, Italy. E-mail: nadia.cattari@endocas.unipi.it.*
- *Umberto Fontana is with the Information Engineering Department, University of Pisa, Via G. Caruso 16, 56122 Pisa. E-mail: umbertofontana93@gmail.com.*
- *Fabrizio Cutolo is with the Information Engineering Department, University of Pisa, Via G. Caruso 16, 56122 Pisa, and with the EndoCAS Centre, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56124 Pisa, Italy E-mail: fabrizio.cutolo@endocas.unipi.it.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

In any traditional OST HMD, it is difficult to achieve accurate real-to-virtual image registration. This is a significant limitation, given that a key prerequisite for considering the OST HMD as a reliable device for guiding manual tasks is its ability to provide an accurate (in terms of spatial alignment) augmented view of the world. All the above shortcomings of the standard OST HMDs are due to the incompatibility between the nature of the 4D light field, related to the real scene, and the nature of the virtual content, rendered as a 2D image through the NED [4].

To counter these issues, light field (LF) AR displays have been proposed [20–22, 28]. These displays can potentially solve the limitations of standard OST displays by providing a complete volume overlay between computer-generated and real LFs. However, prototype solutions based on integral imaging technology or on stacked LCD panels [4, 24, 29] have: an insufficiently wide depth-of-field, a low spatial resolution, and a reduced light throughput of the see-through display. Moreover, the full specifications of the LF mechanisms of two soon-to-be-marketed and/or claimed LF OST displays, Magic Leap 1 and AVEGANT, have not yet been revealed.

Overall, in a traditional OST NED, the real-to-virtual image registration relies on a precise and dedicated calibration procedure to estimate the intrinsic and extrinsic parameters of the virtual rendering camera. These parameters account for the eye position with respect to the OST display and encapsulate the projection properties of the eye-NED pinhole model.

This calibration step often entails tedious and error-prone virtual-to-real alignments performed manually by the user [17]. Moreover, this process should be in theory repeated whenever the HMD moves and causes a change in the relative position between the OST display and the user's eye.

Alternative interaction-free methods have been proposed [23, 32], which incorporate a user-specific and real-time estimation of the position of the user's eye with respect to the NED by means of an inward-looking eye-tracking camera attached to the HMD. Unfortunately, eye-tracking cameras are only able to estimate the position of the center of rotation of the user's eye, while in reality an OST-to-eye calibration entails estimating the eye's first nodal (focal) point [18].

For these reasons, none of the approaches completely remove the effects of parallax contribution from the whole registration error. In fact, during his keynote speech at ISMAR 2017 [1] Georg Klein stated that the Microsoft HoloLens has a maximum static registration error of < 10 mrad which results in an error of about 5 mm at a distance of 500 mm from the user, as experimentally verified in [8].

This paper presents a hardware/software strategy that removes the parallax error at a pre-defined working distance. We also describe a geometrical model that computes the registration error due to viewpoint parallax in real use cases when the OST configuration falls outside the ideal configuration. Finally, we describe the implementation of an OST HMD based on this idea and we report the results of an experimental study that measures the AR registration accuracy by means of an inner camera looking through the display.

2 The OST Display Optics

The key elements of an OST NED are the microdisplay and the optical engine. The optical engine in turn can be broken down into two main components: the collimation optics and the optical combiner (Figure 1). The optical combiners that characterize most commercial embodiments of OST HMDs are: either standard half-mirror-based or waveguide-based, both in their different implementations [35].

The collimation optics acts as a magnifier for the microdisplay image, and projects the image enlarged at a fixed distance from the eye. In most commercially available displays, this focal distance is predetermined by the manufacturer.

We now summarize some basic concepts of geometrical optics applied to a simple magnifier consisting of a positive lens placed in front of the eye. In optics, the term "virtual image" refers to an image that can only be seen by looking through an optics (the magnifier in our case): placing a magnifier between the eye and a real object, we visually perceive a virtual image associated with the real object. In a NED,

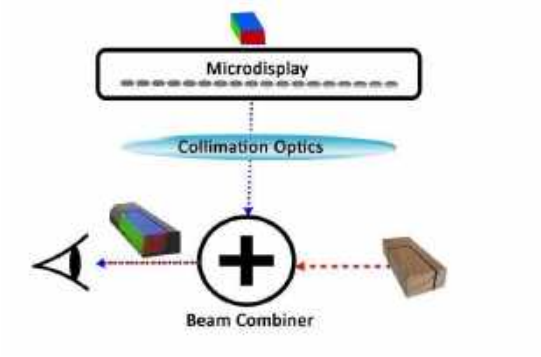

Fig. 1. Key elements of an optical see-through near-eye display.

the real object corresponds to the computer-generated image rendered on the microdisplay: the role of the magnifier is therefore to display the microdisplay image so that it appears magnified at a comfortable viewing distance [19].

In the AR community, "virtual image" usually refers to the computer-generated content merged with the real world scene. Hereafter, the virtual content rendered onto the microdisplay will thus be referred to as a "computer-generated image", whereas, in order to avoid confusion, the image seen through the magnifier, which in the optics community is called a virtual image as well, will be referred to as "relayed image", whether it is related to a real object or to a computer-generated image.

Given a magnifier consisting of a positive lens, with its two focal points F and F' each one at distance f from the nodal point N, an object whose real height is h placed at the distance d is projected and magnified at the distance D with a relayed height H, as expressed as follows:

$$D = -\frac{fd}{d-f} \tag{1}$$

$$H = h\frac{f}{f-d} \tag{2}$$

Figure 2 geometrically represents all the variables reported in the previous equations.

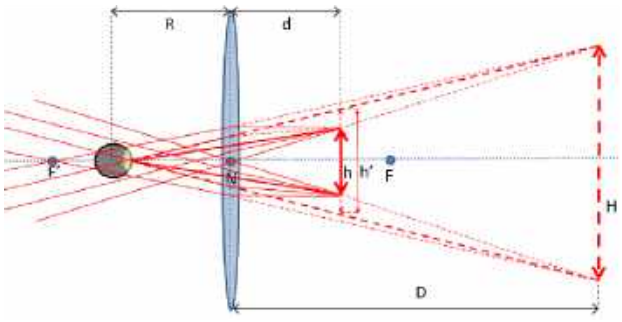

Fig. 2. Magnifier optical rules. Solid lines represent the real object (arrow) and its light rays; Dashed lines represent the corresponding relayed image with its relayed light rays. Each point of the relayed image is associated to a bundle of outgoing relayed light rays, all arranged in a radial pattern.

The magnification equation relates the apparent size h' of a relayed image to the real size h of an object at the same distance:

$$h' = (d+R)\frac{H}{D+R} \tag{3}$$

The scale factor I between h' and h can be calculated as:

$$I = \frac{h'}{h} = \frac{-f(d+R)}{dR-fR-fd} \tag{4}$$

Given the basic laws of geometrical optics, each point in the relayed image is associated with a bundle of outgoing relayed light rays, arranged in a radial pattern, emitted (or reflected) by the point of the relayed image itself. The observer's pupil intercepts some of these light rays and the eye optical elements drive them to the retina; the photoreceptors of the retina convert the photons into electrical impulses that are transmitted through the optic nerve to the brain.

If a real object is placed exactly at the distance $d = f$, its relayed image is projected to infinity with a scale factor $I = \frac{(f+R)}{f}$. In this case, the light rays associated with each point in the relayed image become parallel (Figure 3).

Thus, if the microdisplay of a NED is placed close to the focal point of the collimation optics, the projected light rays become parallel and the relayed image is projected to infinity (i.e., to a very far distance). Therefore, with this optical arrangement, the user perceives the same computer-generated relayed image regardless of the eye's position within a certain region - the NED "eye-box", as represented in Figure 4.

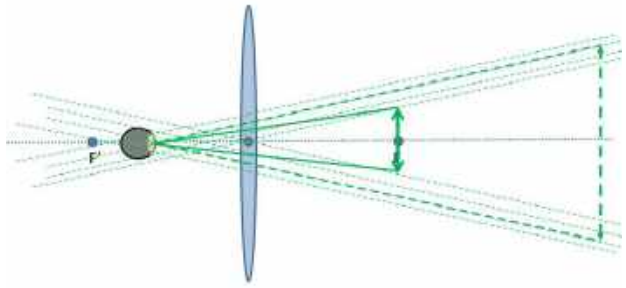

Fig. 3. If an object is placed close to the focal point F, the light rays associated to each point of its relayed image are arranged in a parallel pattern.

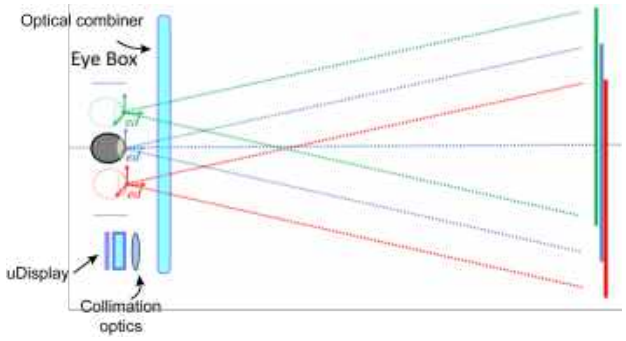

Fig. 4. Near-eye display with the focus distance at infinity: moving the eye within the display eye-box, the user perceives the same relayed image as if a "virtual screen was anchored to the eye". The shifted green, blue, and purple virtual screens (all projected to infinity) are depicted as not perfectly overlapped to better clarify their relative shift.

3 How to Obtain a Coherent Virtual-to-Real Content Registration in OST-NED with Infinity Focus

In a traditional OST HMD, the correct registration between the real scene view and the computer-generated content, projected on the imaging plane of the see-through display, requires a calibration procedure that estimates the projection parameters of the virtual rendering camera. For a survey of many the calibration methods of OST displays, see Grubert et al. [17]. The intrinsic matrix K_{ed} that yields a correct mapping of each 3D point $\hat{P}^{ed}$ of the virtual object, represented in the coordinate system of the eye's first nodal point (ed), onto the imaging plane of the see-through display $\hat{p}^d$ is (points are expressed in homogeneous coordinates):

$$\lambda \hat{p}^d = K_{ed}\hat{P}^{ed} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \hat{P}^{ed} \tag{5}$$

where f_x and f_y are the focal lengths of the eye-NED pinhole model, and c_x and c_y are the coordinates of the principal point all expressed in pixels, whereas λ is a generic scale factor due to the equivalence between points in homogeneous coordinates.

Given that the imaging plane of many consumer-level OST NEDs is at infinity, the position of the eye does not have any impact on the projection model of the combined eye-NED pinhole model Figure 4. In fact, K_{ed} remains the same irrespectively of the eye position and therefore the relayed image perceived by the user is not subjected to any shift or scaling effect for different positions of the user's eye within the display eye-box area.

On the other hand, a different position of the eye only affects the pose of the virtual object relative to the rendering camera (i.e., the extrinsic parameters) ${}^{ed}\mathtt{T}_W$ between world (W) and eye-NED (ed) reference system:

$$\lambda \hat{p}^d = K_{ed}\,\hat{P}^{ed} = K_{ed}\;{}^{ed}\mathtt{T}_W\;\hat{P}^W \tag{6}$$

Where the pose represented by the rigid transformation matrix can be broken down as a rotation matrix followed by a translation vector: ${}^{ed}\mathtt{T}_W = [{}^{ed}\mathtt{R}_W | {}^{ed}\mathtt{t}_W]$.

The registration of the virtual content to the real world view is therefore achieved by tracking the world scene to be enhanced with respect to ed (i.e., estimate ${}^{ed}\mathtt{T}_W$) by means of a tracking sensor t. The rigid transformation ${}^{ed}\mathtt{T}_t$ between t and ed is estimated through a calibration procedure. We performed this calibration by means of a camera c acting as a replacement of the user's eye pointing to the display and placed in a generic, and unknown, position within the eye-box of the see-through display. More details on the calibration procedure are given in [14].

The calibration is split into two phases:

1. First, we determined the rigid transformation ${}^{c}\mathtt{T}_t$ between t and c by localizing with both devices, a physical structured marker placed in front of the display (Figure 5).

2. Secondly, the rigid transformation ${}^{ed}\mathtt{T}_c$, representing the pose between the reference system c and ed, was computed by rendering on the display a virtual structured marker of known size (a checkerboard) and by localizing it with c.

Finally, we have:

$${}^{ed}\mathtt{T}_t = {}^{ed}\mathtt{T}_C\;{}^{c}\mathtt{T}_t \tag{7}$$

Considering that the viewpoint camera position corresponds to the actual viewpoint position ($c = ed \Rightarrow {}^{ed}\mathtt{t}_c = 0$), equation 7 becomes:

$${}^{ed}\mathtt{T}_t = {}^{ed}[\mathrm{R}|0]_c\;{}^{c}\mathtt{T}_t \tag{8}$$

Therefore, knowing K_{ed}, the line-of-sight alignment between a pixel of the display and its associated 3D real point both observed from the calibration point of view, is ensured if:

$$\lambda \hat{p}^d = K_{ed}\;{}^{ed}\mathtt{T}_t\;{}^{t}\mathtt{T}_W\;\hat{P}^W \tag{9}$$

In a real use case, with the user's eye looking through the display, an offset O between the actual viewpoint and the calibration point generates a registration error due to the viewpoint parallax, as depicted in Figure 6. If we knew exactly the extent of the offset O, we would be

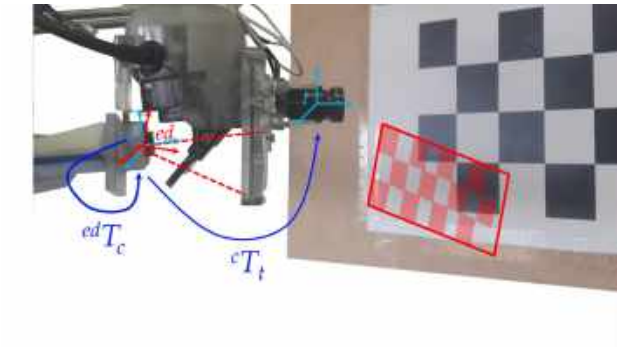

Fig. 5. Calibration between the tracking camera t and the camera c acting as a replacement of the user's eye and pointing to the display. First ${}^{c}\mathbf{T}_{t}$ is determined through a standard stereo calibration, then ${}^{ed}\mathbf{T}_{c}$ is estimated by localizing a virtual checkerboard (red checkerboard) rendered on the display.

able to compensate for the parallax and maintain the correct virtual-to-real AR registration simply by applying an opposite shift to the virtual content.

This would entail estimating the exact position of the eye's first nodal point (i.e., the center of projection of the user's eye) with respect to the NED display. Unfortunately, eye-tracking devices and/or manual calibration procedures are error-prone and only capable of directly or indirectly estimating the center of rotation of the user's eye(s) and not the actual center(s) of projection, which is usually shifted by 7-8 mm [18]. These are thus the main reasons why common eye-tracking mechanisms are unable to fully compensate for the registration error due to viewpoint parallax, particularly when the eye is rotated to look outside the center of the OST display. An error in the estimation of the position of the nodal point of the eye contributes to the registration error due to the viewpoint parallax (Figure 6).

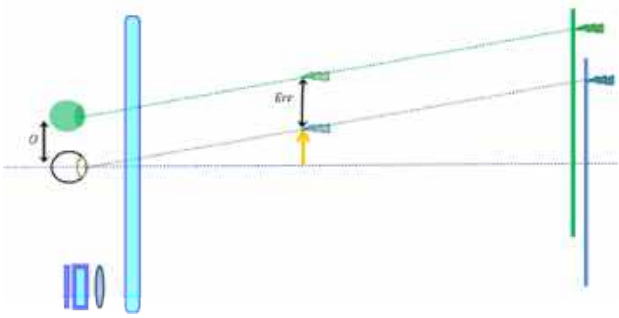

Fig. 6. Virtual-to-real registration error in case of a traditional optical see-through display with infinity focus. Given a real object (the orange arrow), it is possible to render a virtual content (the blue flash) in order to obtain a coherent virtual-to-real augmented reality registration from a well-defined calibration point (grey eye). If the actual viewpoint is shifted (green eye) by an offset O, the virtual content is perceived as shifted by O (green flash), while the absolute position of the real content does not change. The effect of the viewpoint parallax is therefore a registration error Err=O.

4 How to Mitigate the Registration Error due to the Viewpoint Parallax in the Peripersonal Space

To mitigate the AR registration error due to the viewpoint parallax, we exploit an additional magnifier placed in front of the OST display. This optical element modifies the visual perception of the real scene and projects a pre-defined plane of the scene to infinity, just like the relayed one. Our rationale is that the focal distance f of the magnifier should be selected so as to be close to the user's average fixation point distance d. By doing so, the light rays emitted from each point of the real reference plane and perceived "through the lens" are arranged in a parallel pattern just like the computer-generated ones perceived through the OST display. A parallax-free view of the real plane is thus obtained (Figure 7).

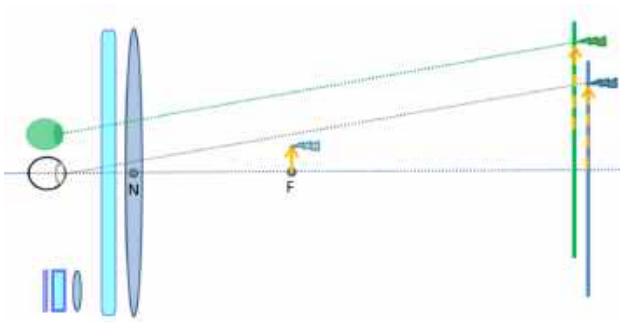

Fig. 7. Placing a magnifier in front of the OST display, the registration error due to the viewpoint parallax is removed for those points that are at the magnifier focal distance f.

Adding the magnifier to the OST display entails adjusting the projection model of the virtual rendering camera, since, according to equation 4 in the case of $d = f$, the magnifier makes the user perceive the real scene as being magnified by a scale factor of $I_f = \frac{(f+R)}{f}$. For this reason, the virtual content must be magnified by the same scale factor. This is done by tuning the focal length of the intrinsic matrix. Furthermore, in a non ideal scenario, the non perfect alignment of the magnifier optical axis with that of the OST display causes an additional registration error, which can be compensated for by adjusting the principal point components of K_{ed}. These parameters are not user-dependent, and are only related to a steady relation between the OST display and the magnifier.

We now exploit various equations to prove that in our solution, the registration error due to the viewpoint parallax is significantly mitigated also for working distances within the peripersonal space, namely for objects placed outside the focal distance f, such as those that occur in realistic scenarios. Specifically, considering the maximum length of the human arm, we do not consider fixation points at distances greater than 700 mm. Similarly, considering the minimum accommodation distance of the human eye, we do not consider fixation points at distances of less than 150 mm. We thus consider a 150 - 700 mm range of possible fixation points.

The relayed images associated with objects placed outside the focal distance of the magnifier ($d \neq f$) are magnified with a scale factor of I_d that changes in relation to the distance (equation 4). Given a magnifier with a certain focal distance f, we can calculate for each distance d the relative percentage difference (R_d) between the actual scale factor and the scale factor for the optimal case with $d = f$:

$$R_d = 100 * (\frac{I_d - I}{I}) \tag{10}$$

Figure 8 shows the trend of this relative percentage difference considering a reasonable eye-to-magnifier distance $R = 50$ mm. Figure 9 shows the trend of the same difference for a given lens with $f = 500$ mm in relation to the eye-to-magnifier distance R and in relation to the object distance d.

The different scale factors can be estimated and adjusted in the virtual rendering, by scaling the 3D content as a function of the distance. However, as shown in Figures 8 and 9, there are wide depth ranges in which the relative percentage difference between the scale factors is lower than 1%. For working distances of ± 100 mm around the reference distance (i.e., between 400 and 600 mm), $R_d < 0.7$ %.

Those points of the real objects that are not placed at the focal distance, emit and/or reflect light rays that are not arranged in a parallel

pattern whereas the light rays associated with the computer-generated image remain in a parallel pattern. In this case, if the eye is along the magnifier principal axis, there is only the scale factor but no parallax contribution: the user can perceive a perfectly registered AR scene (Figure 10).

Moreover, as depicted in Figure 11, if the eye is not aligned with the magnifier principal axis and it is shifted by an unknown offset O, the relayed image of the real object (dotted arrow) is perceived at the finite distance D, with a viewpoint shifted by O (because the associated light rays are not arranged in a parallel pattern). The projection of the virtual content (blue flash) is therefore perceived at infinity and shifted by O (green flash) (because in this case the associated light rays are arranged in a parallel pattern).

Owing to the different arrangement of the patterns of light rays associated with the relayed images of the virtual and real content, the user perceives a registration error due to the parallax (the green flash does not appear on the top of the red arrow). This parallax can be geometrically calculated at the real object distance d by applying simple geometrical rules:

$$\alpha = atan\left(\frac{H}{D+R}\right); \alpha' = atan\left(\frac{H-O}{D+R}\right) \tag{11}$$

$$Err = (d+R)\left[tan(\alpha) - tan(\alpha')\right] \tag{12}$$

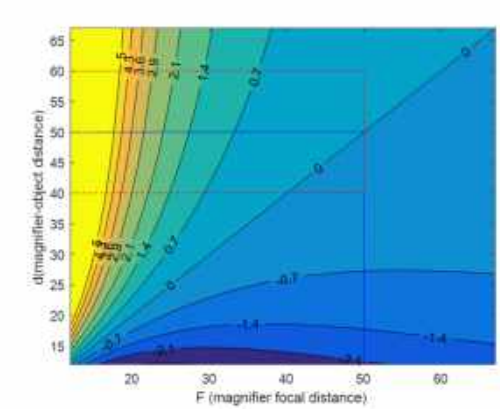

Fig. 8. Relative percentage difference between scale factors of an object at distance $d \neq f$ and the ideal scale factor for $d = f$. Data are calculated for $R = 50$ mm. Notably, for the confidence interval ± 100 mm around the $d = 500$ mm, the relative percentage difference is well below 0.7%.

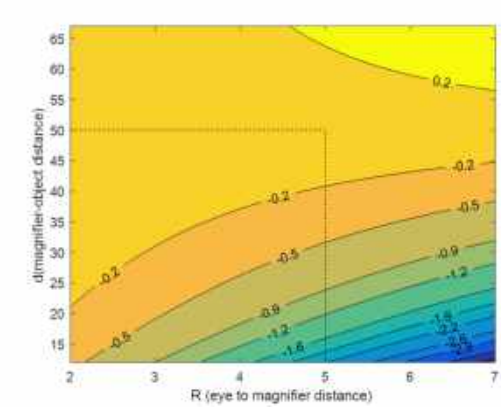

Fig. 9. Relative percentage difference between scale factors of an object at distance $d \neq f$ and the ideal scale factor for $d = f$. Data are calculated for $f = 500$ mm.

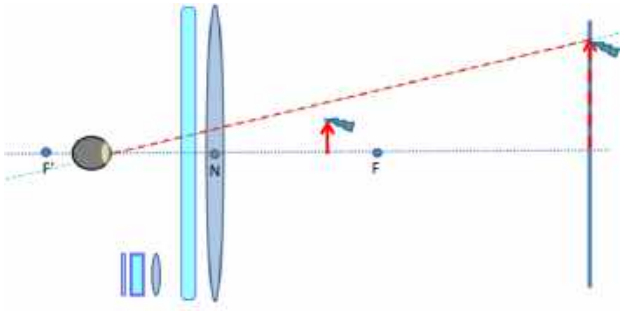

Fig. 10. If an object (red arrow) is outside the magnifier focal distance, given the optics rules we can estimate its associated relayed image (dashed red arrow) in order to render the virtual content on the display (blue flash) as registered with the real one.

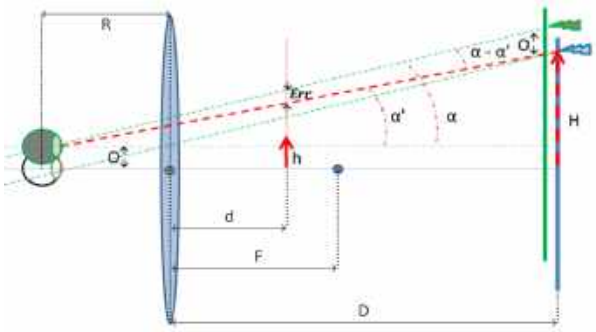

Fig. 11. Parallax error if $d \neq f$ and with an eye offset O with respect to the magnifier optical axis. For legibility reasons, we omitted the representation of the optical see-through display.

$$Err = O\frac{d+R}{D+R} = O\frac{(d+R)(d-f)}{dR+f(d+R)} \tag{13}$$

where α is the angle between the axes passing through the eye and the virtual object projection, α' is the angle between the axes passing through the eye and the real object projection, Err is the linear error at the real distance.

Notably, the registration error does not depend on the position of the target object with respect to the center of the imaging plane (i.e., the "off-axis" contribution, whose 1D component is denoted by H).

Figure 12 shows this registration error due to the viewpoint parallax Err at different distances d and with different values of the magnifier focal length f, considering: 1) a reasonable eye to magnifier distance $R = 50$ mm; and 2) a worst case eye displacement with respect to the calibration point O of 5 mm (considering that NED eye-box dimensions are seldom > 10 mm and that the calibration point can be empirically and approximately set in its center).

5 Experimental Setting and Calibration Procedure

The Figure 13 shows the experimental setup. In order to demonstrate the feasibility of our approach, we developed a prototype by reworking a commercial OST HMD (the ARS30 by Trivisio, Luxemburg). The ARS30 consists of a pair of 1280x1024 OLED microdisplays, a pair of half-silvered mirrors, and two collimation optics capable of collimating the relayed images to infinity (the manufacturer states that there is a focal distance of about 150 m). Each NED has a 30° diagonal angle of view, which corresponds to ≈ 1.11 arcmin/pixel angular resolution, and an eye-box dimension of about 8x10 mm. We only needed to use one display of the HMD. The ARS30 was integrated with a tracking USB camera placed above the display (Leopard Imaging LI-OV580); the tracking camera has a 109° diagonal angle of view, and a 1280x720 resolution.

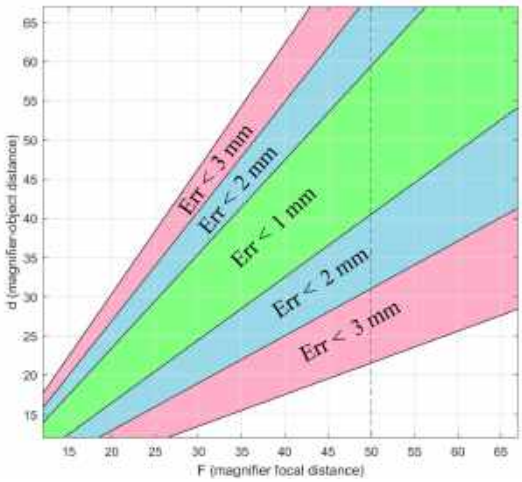

Fig. 12. Parallax error Err for $O = 5$ mm and $R = 50$ mm.

Following a similar approach to our previous works [12, 13], we embedded the HMD into a 3D printed plastic shell; the plastic shell was designed to hold a liquid crystal optical shutter that enables us to hide the see-through view upon request.

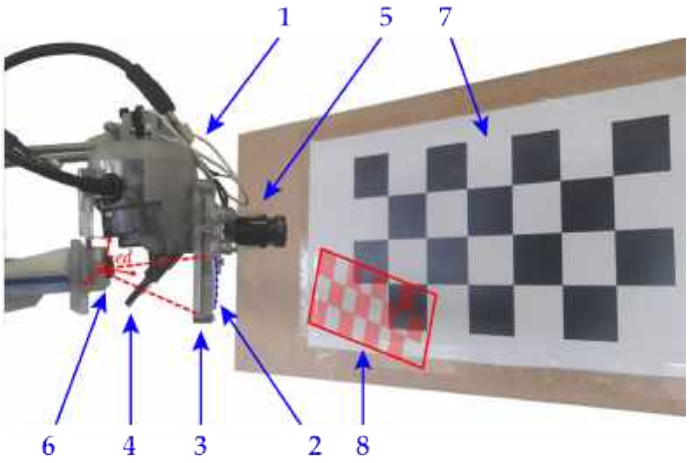

Fig. 13. Experimental setting for the calibration procedure and the experimental session. 1→ The optical see-through head-mounted display (HMD). 2→ The magnifier used to mitigate the registration error due to the viewpoint parallax. 3→ The optical shutter. 4→ The optical combiner. 5→ The tracking camera. 6→ The viewpoint camera. 7→ The real checkerboard used as target object for calibration and validation. 8→ The projected checkerboard used for the calibration procedure.

In [14], we presented all the algebraic steps involved in the camera-based calibration procedure - see Section 3. The calibration workflow was as follows. We placed the viewpoint camera c at the calibration point, at a position that was empirically and approximately set in the center of the NED eye-box and at the eye relief distance.The viewpoint camera was a SONY FCB-MA130, which has a $1/2.45''$ CMOS sensor, a 1280x720 resolution, and a 59° diagonal FOV, which corresponds to ≈ 2.67 arcmin/pixel angular resolution. The rigid transformation ${}^{c}\mathbf{T}_{t}$ was determined using both devices to trace a real checkerboard placed in front of the display, with the optical shutter in open state (i.e., phase 1 of the calibration).

Secondly, the transformation ${}^{ed}\mathbf{T}_{c}$ was computed by rendering a virtual checkerboard onto the display whose pose in the virtual scenegraph is known, and by estimating its pose in the c reference system. This second calibration step was performed with the optical shutter closed to prevent the influence of the light rays coming from the real environment (the same result can be obtained with a mechanical optical shutter). All these routines were developed in Matlab with the Computer Vision Toolbox.

Finally, we placed a magnifier in front of the OST HMD. The magnifier is an ophthalmic lens with f = 500 mm. The distance R between the camera and magnifier was about 50 mm. We implemented a Matlab routine to calibrate the intrinsic parameters K_{ed}. Compared to the method we described in [14], we added a third phase of the calibration procedure with an AR routine that enabled us to refine the eye-display projection parameters manually in order to account for the contribution of the magnifier. In this routine, the intrinsic parameters of the eye-display model (focal distance and principal point) were recursively refined by looking through the display and considering both the up-scale factor due to the magnifier, as illustrated with equation 4, and the off-center placement of the magnifier.

Once the parameters had been refined, the same AR routine was used to measure the virtual-to-real registration accuracy. According to the estimated projection parameters, the AR routine renders 18 virtual spots registered at the inner corners of the real checkerboard placed at distances of around 500 mm (Figure 14).

Fig. 14. Viewpoint camera view through the OST display that highlights that the virtual spots (small cyan coloured circles) are perfectly aligned with the inner corners of the real checkerboard. This view also enables the one-time calibration refinement of the parameters of the virtual rendering camera that account for the contribution of the external magnifier.

6 Real Setup Results

Given that the scale factor does not significantly change around the magnifier focal distance (see Figure 9 for $f = 500$ mm), in our AR routine the virtual objects were not re-scaled in function of the distance.

To evaluate the accuracy and the robustness of our approach, we first performed a demo by moving the inner camera outside the calibration point and moving the checkerboard before and behind the optimal distance $d = 500$ mm. As shown in the supplementary material video, the registration is also maintained at distances $d \neq 500$ mm.

Afterwards, we quantitatively evaluated the virtual-to-real registration accuracy by displacing the viewpoint camera in pre-set positions using the dedicated camera holder shown in Figure 15. There are nine preset positions, eight of which are radially arranged at 4 mm around the calibration position (i.e., the central position). For each camera position, we placed the checkerboard at nine different distances from 300 to 700 mm with a 50 mm step.

For each displaced position of the viewpoint camera with respect to the calibration position, and for each distance of the checkerboard, we captured two camera frames: the first was a real view of the physical checkerboard, the second showed the virtual content projected onto the display at the corners of the checkerboard. This virtual content was captured with see-through display occluded to remove the real world background. We thus estimated the virtual-to-real image registration error in pixels for all the 18 checkerboard inner corners. Table 1 reports the mean and standard deviation of the on-image registration error expressed in pixels for all the nine viewpoint camera positions and

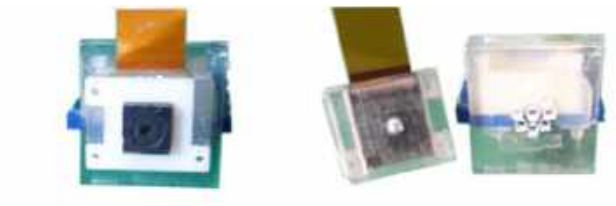

Fig. 15. Viewpoint camera (left) and its holder (right) used adjust the camera in nine pre-set positions within the eye-box of the optical see-through display.

the nine distances of the checkerboard. The errors are also reported in visual angles (in arcmin) (Table 2) and in Euclidean distances (i.e., the absolute registration error in mm). The absolute registration error was measured by backprojecting the on-image error at the current chessboard distance (Table 3).

The overall mean virtual-to-real registration error was 4.06 pixels with a 2.49 pixels standard deviation, which corresponds in terms of angular error to 4.55 arcmin and 2.8 arcmin, and in terms of absolute error to 1.51 mm and 1.08 mm respectively.

Table 1. REGISTRATION ERROR IN PIXEL

Mean (σ)		Checkerboard distances (mm) 300	350	400	450	500	550	600	650	700	All
Camera Positions	**Cent**	3.75 (2.12)	2.19 (1.33)	1.65 (0.79)	2.77 (0.95)	2.35 (0.92)	2.34 (0.68)	1.56 (0.82)	2.38 (0.83)	2.68 (1.2)	2.39 (1.2)
	1	7.94 (1.09)	6.02 (0.76)	3.79 (0.59)	2.04 (0.95)	1.88 (0.93)	1.86 (0.88)	2.89 (1.32)	3.24 (1.63)	4.47 (1.69)	3.37 (2.01)
	2	6.72 (1.57)	3.94 (1.07)	2.16 (1.02)	2.43 (1.24)	3.27 (1.4)	3.84 (1.61)	5.2 (1.93)	6.5 (2.07)	7.9 (2.03)	4.64 (2.48)
	3	5.95 (1.49)	3.57 (1.5)	1.88 (0.95)	2.7 (1.34)	3.42 (2.01)	4.73 (1.99)	5.69 (2.37)	7.01 (2.44)	7.83 (2.49)	4.79 (2.7)
	4	5.26 (2.44)	2.4 (1.5)	1.68 (0.85)	2.22 (1.29)	3.48 (1.49)	4.73 (1.46)	6.32 (1.51)	8.24 (1.46)	7.71 (1.45)	4.76 (2.73)
	5	8.62 (3.57)	3.36 (2.03)	2.39 (1.57)	2.83 (0.98)	3.13 (0.91)	4.9 (0.86)	5.79 (0.97)	6.63 (0.85)	7.14 (0.4)	4.94 (2.54)
	6	6.14 (3.88)	3.82 (2.19)	2.55 (1.62)	2.23 (0.8)	2.74 (0.74)	3.47 (0.67)	3.87 (0.73)	4.54 (0.65)	5.12 (0.78)	3.79 (1.96)
	7	6.56 (2.74)	4.91 (2.43)	3.07 (1.35)	2.01 (0.76)	1.74 (0.81)	1.48 (0.6)	1.7 (0.69)	1.86 (0.59)	2.12 (0.78)	2.67 (1.99)
	8	8.45 (1.87)	6.55 (1.25)	4.14 (0.9)	2.93 (1.04)	2.14 (0.89)	2.34 (1.23)	1.77 (0.77)	2.04 (0.86)	1.94 (0.93)	3.48 (2.41)
	All	6.61 (2.88)	4.41 (2.15)	2.64 (1.4)	2.46 (1.08)	2.68 (1.32)	3.29 (1.72)	3.87 (2.26)	4.72 (2.68)	5.21 (2.78)	4.06 (2.49)

Table 2. ANGULAR REGISTRATION ERROR IN ARCMIN

Mean (σ)		Checkerboard distances (mm) 300	350	400	450	500	550	600	650	700	All
Camera Positions	**Cent**	4.21 (2.37)	2.46 (1.49)	1.85 (0.89)	3.11 (1.07)	2.64 (1.03)	2.62 (0.76)	1.75 (0.92)	2.67 (0.94)	3.01 (1.35)	2.68 (1.35)
	1	8.91 (1.22)	6.75 (0.86)	4.25 (0.66)	2.28 (1.07)	2.1 (1.04)	2.08 (0.99)	3.24 (1.48)	3.64 (1.83)	5.02 (1.9)	3.78 (2.25)
	2	7.54 (1.76)	4.42 (1.2)	2.42 (1.14)	2.72 (1.39)	3.67 (1.57)	4.31 (1.8)	5.83 (2.17)	7.29 (2.32)	8.87 (2.28)	5.2 (2.78)
	3	6.68 (1.67)	4.01 (1.68)	2.11 (1.06)	3.03 (1.5)	3.84 (2.25)	5.31 (2.23)	6.39 (2.65)	7.87 (2.73)	8.78 (2.79)	5.37 (3.03)
	4	5.91 (2.74)	2.69 (1.68)	1.88 (0.96)	2.49 (1.44)	3.9 (1.67)	5.31 (1.64)	7.1 (1.7)	9.25 (1.64)	8.65 (1.63)	5.34 (3.06)
	5	9.68 (4.01)	3.77 (2.28)	2.68 (1.77)	3.18 (1.1)	3.51 (1.02)	5.49 (0.97)	6.5 (1.08)	7.44 (0.95)	8.02 (0.45)	5.54 (2.85)
	6	6.89 (4.35)	4.29 (2.46)	2.86 (1.82)	2.51 (0.89)	3.07 (0.83)	3.9 (0.76)	4.34 (0.82)	5.1 (0.73)	5.75 (0.87)	4.25 (2.2)
	7	7.37 (3.07)	5.51 (2.72)	3.44 (1.52)	2.25 (0.85)	1.96 (0.9)	1.66 (0.67)	1.92 (0.78)	2.09 (0.66)	2.38 (0.88)	2.99 (2.24)
	8	9.48 (2.09)	7.35 (1.4)	4.65 (1.01)	3.29 (1.17)	2.4 (1.00)	2.59 (1.37)	1.99 (0.86)	2.29 (0.97)	2.18 (1.05)	3.9 (2.71)
	All	7.78 (3.12)	4.82 (2.38)	3.06 (1.59)	2.72 (1.23)	3.06 (1.53)	3.83 (1.99)	4.66 (2.49)	5.62 (3.02)	6.02 (3.1)	4.55 (2.8)

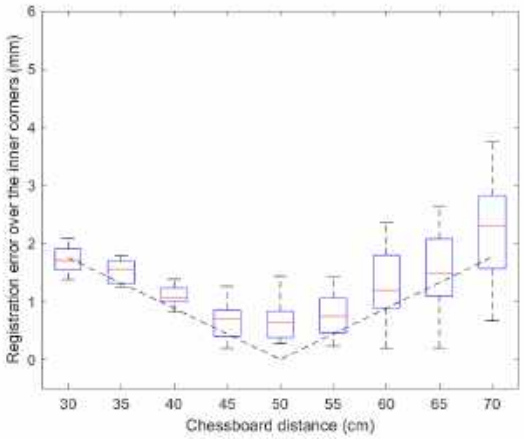

Fig. 16. Boxplot of the registration error for all the 18 corners of the checkerboard for a single viewpoint camera position and for nine different checkerboard distances. The dotted line represents the theoretical registration error due to the viewpoint parallax as calculated through the equation 13

.

Figure 16 reports the box plot of the absolute registration error (in mm) for a specific viewpoint camera position at the various checkerboard distances. The dotted line represents the theoretical registration error due to the viewpoint parallax as calculated through equation 13.

7 Discussion

As shown in Figure 7, in a traditional OST display with infinity focus, without the additional magnifier, and calibrated for a generic eye position within the display eye-box, an eye displacement of value O generates a viewpoint parallax that contributes at the same value to the virtual-to-real registration error. Consequently, an accurate calibration of a traditional OST display, obtained for instance using a camera to replace the user's eye for a specific viewpoint position, is not sufficient to ensure an accurate virtual-to-real registration if the user's eye does not perfectly match the calibration position.

As reported in Table 3, our solution substantially reduces the absolute registration error to < 1 mm for target objects placed at $d = f = 500$ mm even if the eye shifts by 4 mm with respect to the calibration point.

Even so, based on our geometrical simulation, the expected theoretical parallax error should be 0 if the checkerboard were placed at exactly the magnifier focal distance. The estimated absolute registration error is due to inaccuracies in tracking and in calibration. For instance, in our calibration procedure, we did not consider the non-linear distortions due to the optics of the NED and of the magnifier, whereas there is certainly some image distortion (such as the radial distortion). As suggested in [27], better results in terms of registration accuracy would likely be provided by a camera-based calibration method that tackles this problem and also estimates the non-linearities in the projection model of the see-through display due to the optical distortions.

For checkerboard distances $d > f$, the registration error increases (the last row in Tables 1,2, and 3 except for the rightmost cell). This increment is due to a decreased tracking accuracy, which decreases with the distance, and it is also caused by the increased contribution of the parallax to the registration error as modelled in Section 4.

By contrast, for checkerboard distances $d < f$, taking into account

Table 3. ABSOLUTE REGISTRATION ERROR IN MM

Camera Positions	Mean (σ)	Checkerboard distances (mm) 300	350	400	450	500	550	600	650	700	All
	Cent	0.81 (0.46)	0.55 (0.34)	0.48 (0.23)	0.9 (0.31)	0.85 (0.33)	0.93 (0.27)	0.67 (0.36)	1.12 (0.39)	1.36 (0.31)	0.88 (0.45)
	1	1.72 (0.24)	1.52 (0.19)	1.1 (0.17)	0.66 (0.31)	0.68 (0.33)	0.74 (0.35)	1.25 (0.57)	1.52 (0.77)	2.26 (0.86)	1.23 (0.73)
	2	1.46 (0.34)	0.99 (0.27)	0.62 (0.29)	0.79 (0.4)	1.18 (0.51)	1.53 (0.64)	2.25 (0.84)	3.05 (0.97)	3.99 (1.03)	1.79 (1.26)
	3	1.29 (0.32)	0.9 (0.38)	0.54 (0.27)	0.88 (0.44)	1.24 (0.72)	1.88 (0.79)	2.5 (1.03)	3.3 (1.15)	3.96 (1.26)	1.89 (1.38)
	4	1.14 (0.53)	0.61 (0.38)	0.48 (0.25)	0.72 (0.42)	1.26 (0.54)	1.88 (0.58)	2.74 (0.66)	3.87 (0.69)	3.9 (0.74)	1.91 (1.4)
	5	1.87 (0.77)	0.85 (0.51)	0.69 (0.46)	0.92 (0.32)	1.13 (0.33)	1.95 (0.34)	2.51 (0.42)	3.12 (0.4)	3.62 (0.2)	1.87 (1.09)
	6	1.33 (0.84)	0.97 (0.55)	0.74 (0.47)	0.73 (0.26)	0.99 (0.27)	1.38 (0.27)	1.68 (0.32)	2.14 (0.31)	2.59 (0.39)	1.4 (0.75)
	7	1.42 (0.59)	1.24 (0.61)	0.88 (0.39)	0.65 (0.25)	0.63 (0.29)	0.59 (0.24)	0.74 (0.3)	0.88 (0.28)	1.07 (0.39)	0.88 (0.45)
	8	1.83 (0.4)	1.66 (0.32)	1.2 (0.26)	0.95 (0.34)	0.77 (0.32)	0.92 (0.49)	1.77 (0.33)	0.96 (0.41)	0.98 (0.47)	1.1 (0.51)
	All	1.43 (0.63)	1.04 (0.54)	0.76 (0.41)	0.8 (0.35)	0.97 (0.48)	1.31 (0.68)	1.68 (0.98)	2.22 (1.26)	2.64 (1.41)	1.51 (1.08)

all the nine viewpoint camera positions, the mean registration error at first decreases (e.g., for 400 mm $\leq d \leq$ 450 mm) and then increases (e.g., for 300 mm $\leq d \leq$ 350 mm). We hypothesize that the initial decrease may be due to an increased tracking accuracy, which exceeds the initially negligible parallax contribution to the registration error. The subsequent increase in the error (for distances between 350 and 400 mm) is likely due to the substantial increment in the parallax contribution. The impact of the parallax contribution to the global registration error is highlighted in Figure 16, which shows both the parallax contribution obtained by equation 13 and the boxplot of the absolute registration error for a specific viewpoint camera position.

Contrary to what modelled by equation 13, our tests showed a slight tendency of the error to increase as the target object was moved off-axis with respect to the center of the display. We can hypothesize that, since image distortions are larger at the image periphery, the tendency of the AR overlay accuracy to degrade as the target object moves off-axis is mostly due to the fact that we modeled the optical elements as perfect thin lenses; in reality, there is a bit of radial distortion introduced by both the magnifying lens and display optics. This contribution, which only depends on the direction of the light rays and not on the real content shown in front of the HMD, if accurately parametrized during the calibration phase, could be easily compensated for by distorting the virtual image by the same amount.

As highlighted in Tables 1, 2, and 3, the global registration error trend is slightly different over the eight viewpoint camera positions around the central one, although the theoretical parallax error contribution should be the same giving that the camera offset is the same (4 mm for each radially arranged position). This discrepancy could be due to the non-perfect parallelism between the optical axis of the viewpoint camera and the magnifier principal axis. In order to better validate our equation for the registration error due to the parallax, the magnifier principal axis and the optical axis of the viewpoint camera should be arranged parallel to each other. In this way, the results of the calibration would apply for all the radially arranged positions of the viewpoint camera.

In our setup, we obtained an average absolute registration error of 0.97 mm and an average angular error of 3.37 arcmin for working distances of $\pm$100 mm around the reference distance (i.e., between 400 and 600 mm). This registration error is reasonably low for the modified OST display to be considered as sufficiently reliable to guide high-precision manual procedures. Alternatively, and as outlined in Section 3 and depicted in Figure 6, in a traditional OST display with an infinity focus and without the magnifier, the viewpoint parallax alone would contribute by at least 4 mm to the global registration error if not compensated for through tedious and/or user-dependent calibration procedures [17]

The results of our real tests confirm the results of the geometrical model presented in Section 4. According to that model and as illustrated in Figure 14, although outside the magnifier focal distance, the registration error due to the parallax is not completely removed. In fact, its contribution is significantly mitigated, and there is a large interval of distances where it is $<$ 1 mm (i.e. for $|d-f| \leq$ 100 mm for $f =$ 500 mm). Specifically, an error lower than 1 mm is generally enough to guide several manual procedures. Furthermore, in a real application, being able to estimate the registration error due to the parallax means that the user can be warned when the AR view cannot ensure a virtual-to-real registration with the required accuracy. In such cases, the application could disable the rendering of the virtual content for those working areas that fall outside the confidence interval.

The existence of a confidence interval means that the user must maintain the working distance in this range. In fact some users, such as surgeons or industrial operators, are already accustomed to operating with a similar working distance constraint whilst wearing loupe magnifiers. In such cases, the user is normally required to maintain the same working distance in order to preserve an in focus view, given the limited depth of field offered by the hand free wearable magnifiers, which is however commonly $<\pm$100 mm [31].

As described in Section 5, the calibration procedure comprises a manual refinement of the virtual rendering camera parameters in order to be able to correctly set the augmented view through the display and the magnifier. The projection parameters are refined by looking through the display and considering both the upscale factor due to the magnifier and the non-ideally centered placement of the magnifier.

In a real implementation, the magnifier generates a scaled view of the real scene, which could be an advantage for high precision manual procedures (in fact some users, such as surgeons or industrial operators, often wear loupe magnifiers).

In our simulations and in our test sessions, we also considered distances $d > f$. In this configuration, the relayed image of a real object is not formed in front of the eye but behind it, and, for this reason, it cannot be perfectly viewed in focus. The greater the distance of the object beyond the focal length of the magnifier, the more blurred the image perceived by the user will be.

In addition, in terms of image focusing, the magnifier placed in front of the OST display projects the relayed image to infinity, where the see-through display also projects the computer-generated image. This not only mitigates the registration error due to viewpoint parallax, but also contributes to reduce the focus rivalry problem typical of OST HMDs mentioned in the introduction.

8 Conclusion

In this work, we have described a strategy to prevent the virtual-to-real AR registration error due to the viewpoint parallax at a given working distance in the peripersonal space. Our solution is based on an OST display with infinity focus coupled with a magnifier with a focal distance equal to the working distance, and it features an accurate parameterization of the virtual rendering camera based on a dedicated calibration routine aimed at estimating the intrinsic parameters of the OST display coupled with the magnifier.

We have geometrically demonstrated that, although outside the magnifier focal distance the registration error due to the viewpoint parallax is not completely removed, its contribution is thereby substantially mitigated ensuring results that are significantly better than those obtained with no magnifier. We tested our strategy on a reworked commercial OST HMD and we verified its accuracy. There is a useful range of working areas, both in terms of spatial displacements between the current viewpoint and the calibration point as well as in terms of the relative distance between the fixation point and the magnifier focal distance, where the contribution of the viewpoint parallax to the global registration error is sub-millimetric.

Until now, the accurate and robust virtual-to-real registration on an AR display has only been ensured with the VST mechanism. Here, we have demonstrated that high registration accuracy can also be obtained with OST displays. We believe that our strategy could thus pave the way for the adoption of OST HMDs to guide high precision manual procedures.

Our technique features the use of an additional optical component (i.e., a magnifying lens) to be integrated into the HMD and placed at a given distance away from the user's eye. Overall, the additional lens of roughly 2 mm thickness only slightly increases the overall footprint and weight of the HMD: its impact in terms of bulkiness would be evident only for those visors with a glasses-like form factor (e.g., the Lumus Optics).

Our approach is designed for OST displays with infinity focus in which the computer-generated content is projected to infinity and for many of the waveguide-based OST displays that are currently implemented with a glass-like form factor, since this is a key feature to support manual tasks (e.g., the Lumus Optics DK32 and the Optinvent ORA2).

Future work will involve evaluating the efficacy of adaptive positive lenses (e.g., liquid lenses) that can tune the focal length in real time by means of an external stimulus [6]. The external stimulus could be associated with the estimated distance of the fixation point obtained by querying a tracking camera or a dedicated depth-camera sensor.

Moreive, a rigorous validation of the proposed method would also entail separating the contribution to the registration error due to the viewpoint parallax from the contribution due to the tracking inaccuracies, from the non-linearities due to the optics, and from the non-ideal placement of the magnifier relative to the calibration camera.

To counter some of the limitations of our solution, we also plan to integrate a fully-automatic off-line calibration procedure, so that we can also estimate the radial and tangential distortions caused by the lens and/or by the collimation optics of the OST display respectively. This interaction-free procedure would also prevent the human-induced errors of the manual process through which, in our solution, we refine the eye-display projection parameters in order to account for the contribution of the magnifier. The method could exploit, for instance, a standard non-linear optimization technique, such as the Levenberg–Marquardt algorithm. We believe that this calibration refinement step could further improve the overall virtual-to-real registration accuracy.

Finally, with regard to the methods used to evaluate the accuracy of the virtual-to-real registration, future work will also include experimental tests involving actual users, and based on the manipulation of three-dimensional objects. In this way, we will be able to also account for the deformed visual perception of the real 3D world "through the magnifier lens". These geometrical aberrations are also owing to the different scale factors yielded by the transverse magnification (linear) and the longitudinal magnification (quadratic). With this user study, we will be able to also evaluate the efficacy of the proposed strategy in mitigating the focus rivalry problem typical of OST HMDs. To this aim, we will exploit the recently developed software framework designed to support the development of AR applications with HMDs as aid to high-precision manual tasks [12].

Acknowledgments

This research was partly supported by the HORIZON2020 Project VOSTARS (Video-Optical See Through AR surgical System), Project ID: 731974. Call: ICT-29-2016 Photonics KET 2016. This work was also partly supported by the Italian Ministry of Education and Research (MIUR) within the framework of the CrossLab project (Departments of Excellence) of the University of Pisa, laboratory of Augmented Reality

References

[1] Keynotes. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. xxv–xxvi, Oct 2017. doi: 10.1109/ISMAR-Adjunct.2017.14

[2] G. Badiali, V. Ferrari, F. Cutolo, C. Freschi, D. Caramella, A. Bianchi, and C. Marchetti. Augmented reality as an aid in maxillofacial surgery: Validation of a wearable system allowing maxillary repositioning. *Journal of Cranio-Maxillofacial Surgery*, 42(8):1970 – 1976, 2014. doi: 10.1016/j.jcms.2014.09.001

[3] S. Benton. *Selected Papers on Three-dimensional Displays*. SPIE milestone series. SPIE Optical Engineering Press, 2001.

[4] E. M. Calabrò, F. Cutolo, M. Carbone, and V. Ferrari. Wearable augmented reality optical see through displays based on integral imaging. In P. Perego, G. Andreoni, and G. Rizzo, eds., *Wireless Mobile Communication and Healthcare*, pp. 345–356. Springer International Publishing, Cham, 2017.

[5] N. Cattari, F. Cutolo, R. D'amato, U. Fontana, and V. Ferrari. Toed-in vs parallel displays in video see-through head-mounted displays for close-up view. *IEEE Access*, 7:159698–159711, 2019. doi: 10.1109/ACCESS.2019.2950877

[6] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs. Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2906–2916, Nov 2018. doi: 10.1109/TVCG.2018.2868532

[7] 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 Transactions on Biomedical Engineering*, 67(2):411–419, Feb 2020. doi: 10.1109/TBME.2019.2914517

[8] S. Condino, G. Turini, P. D. Parchi, R. M. Viglialoro, N. Piolanti, M. Gesi, M. Ferrari, and V. Ferrari. How to build a patient-specific hybrid simulator for orthopaedic open surgery: Benefits and limits of mixed-reality using the microsoft hololens. In *Journal of healthcare engineering*, 2018.

[9] F. Cutolo. Letter to the editor on "augmented reality based navigation for computer assisted hip resurfacing: A proof of concept study". *Annals of Biomedical Engineering*, 47(11):2151 – 2153, November 2019. doi: 10.1007/s10439-019-02299-w

[10] 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 L. T. De Paolis and A. Mongelli, eds., *Augmented Reality, Virtual Reality, and Computer Graphics*, pp. 43–54. Springer International Publishing, Cham, 2016.

[11] F. Cutolo and V. Ferrari. The role of camera convergence in stereoscopic video see-through augmented reality displays. *International Journal of Advanced Computer Science and Applications*, 9(8), 2018. doi: 10.14569/ijacsa.2018.090803

[12] F. Cutolo, B. Fida, N. Cattari, and V. Ferrari. Software framework for customized augmented reality headsets in medicine. *IEEE Access*, 8:706–720, 2020. doi: 10.1109/ACCESS.2019.2962122

[13] F. Cutolo, U. Fontana, M. Carbone, R. D'Amato, and V. Ferrari. [poster] hybrid video/optical see-through hmd. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 52–57, Oct 2017. doi: 10.1109/ISMAR-Adjunct.2017.31

[14] F. Cutolo, U. Fontana, N. Cattari, and V. Ferrari. Off-line camera-based calibration for optical see-through head-mounted displays. *Applied Sciences*, 10(1), 2019. doi: 10.3390/app10010193

[15] F. Cutolo, U. Fontana, and V. Ferrari. Perspective preserving solution for quasi-orthoscopic video see-through hmds. *Technologies*, 6(1), 2018. doi: 10.3390/technologies6010009

[16] J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2228–2241, June 2019. doi: 10.1109/TVCG.2018.2832633

[17] J. Grubert, Y. Itoh, K. Moser, and J. E. Swan. A survey of calibration methods for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2649–2662, Sept. 2018. doi: 10.1109/tvcg.2017.2754257

[18] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6):1124–1133, June 2006. doi: 10.1109/TBME.2005.863952

[19] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett. Three-dimensional displays: A review and applications analysis. *IEEE Transactions on Broadcasting*, 57(2):362–371, June 2011. doi: 10.1109/TBC.2011.2130930

[20] Hongen Liao, N. Hata, S. Nakajima, M. Iwahara, I. Sakuma, and T. Dohi. Surgical navigation by autostereoscopic image overlay of integral videography. *IEEE Transactions on Information Technology in Biomedicine*, 8(2):114–121, June 2004. doi: 10.1109/TITB.2004.826734

[21] H. Hua. Enabling focus cues in head-mounted displays. In *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, p. JTu1F.2. Optical Society of America, 2017.

[22] H. Hua. Optical methods for enabling focus cues in head-mounted displays for virtual and augmented reality. In B. Javidi, J.-Y. Son, and O. Matoba, eds., *Three-Dimensional Imaging, Visualization, and Display 2017*, vol. 10219, pp. 92 – 101. International Society for Optics and Photonics, SPIE, 2017. doi: 10.1117/12.2264157

[23] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 75–82. IEEE, Mar. 2014. doi: 10.1109/3dui.2014.6798846

[24] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods. An occlusion capable optical see-through head mount display for supporting co-located collaboration. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 133–141, Oct 2003. doi: 10.1109/ISMAR.2003.1240696

[25] G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931, July 2016. doi: 10.1109/TVCG.2015.2473855

[26] D. Lanman and D. Luebke. Near-eye light field displays. *ACM Trans. Graph.*, 32(6), Nov. 2013. doi: 10.1145/2508363.2508366

[27] S. Lee and H. Hua. A robust camera-based method for optical distortion calibration of head-mounted displays. *Journal of Display Technology*, 11(10):845–853, Oct 2015. doi: 10.1109/JDT.2014.2386216

[28] H. Liao, T. Dohi, and K. Nomura. Autostereoscopic 3d display with long visualization depth using referential viewing area-based integral photography. *IEEE Transactions on Visualization and Computer Graphics*, 17(11):1690–1701, Nov 2011. doi: 10.1109/TVCG.2010.267

[29] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 29–38, Oct 2013. doi: 10.1109/ISMAR.2013.6671761

[30] 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 *2016 IEEE Virtual Reality (VR)*, pp. 253–254, March 2016. doi: 10.1109/VR.2016.7504749

[31] D. Pieptu and S. Luchian. Loupes-only microsurgery. *Microsurgery*, 23(3):181–188, 2003. doi: 10.1002/micr.10126

[32] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura. Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):481–490, April 2015. doi: 10.1109/TVCG.2015.2391857

[33] J. P. Rolland and O. Cakmakci. The past, present, and future of head-mounted display designs. In Y. Wang, Z. Weng, S. Ye, and J. M. Sasian, eds., *Optical Design and Testing II*, vol. 5638, pp. 368 – 377. International Society for Optics and Photonics, SPIE, 2005. doi: 10.1117/12.575697

[34] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, 2000. doi: 10.1162/105474600566808

[35] K. Sarayeddine and K. Mirza. Key challenges to affordable see-through wearable displays: the missing link for mobile ar mass deployment. In *Defense, Security, and Sensing*, 2013.

[36] T. Sielhorst, C. Bichlmeier, S. M. Heining, and N. Navab. Depth perception – a major issue in medical ar: Evaluation study by twenty surgeons. In R. Larsen, M. Nielsen, and J. Sporring, eds., *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pp. 364–372. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[37] P. Vávra, J. Roman, P. Zonča, P.Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi. Recent development of augmented reality in surgery: A review. *Journal of healthcare engineering*, 2017:4574172, 2017. doi: 10.1155/2017/4574172

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 imageguided 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.

Nadia Cattari has an MSc in Biomedical Engineering from the University of Pisa. Currently she is a PhD student in Clinical and Translational Science at the EndoCAS Centre of the University of Pisa. 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.

Umberto Fontana has a BSc degree in Biomedical Engineering from the University of Pisa. He is currently a M.Sc student in Robotics and Automation Engineering at the University of Pisa. His research interests are machine vision, mixed reality, calibration in mixed reality, robotics and automation, and computer graphics.

Fabrizio Cutolo (Member IEEE from 2017) has BSc and M.Sc. degrees in Electrical and Computer Engineering from the University of Pisa, Pisa, Italy, in 2006 and a Ph.D. degree in Translational Medicine from the University of Pisa, Pisa, Italy, in 2015. He is a post-graduate research associate at the Department of Information Engineering of the University of Pisa, Italy. 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, and he is currently WP leader of the HORIZON2020 project VOSTARS, Call ICT-29-2016.