IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 2, 2019, accepted October 21, 2019, date of publication October 31, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950877

Toed-in vs Parallel Displays in Video See-Through Head-Mounted Displays for Close-Up View

NADIA CATTARI[1], FABRIZIO CUTOLO[2], (Member, IEEE), RENZO D'AMATO[2], UMBERTO FONTANA[1], AND VINCENZO FERRARI[2], (Member, IEEE)

[1]EndoCAS Centre, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56124 Pisa, Italy
[2]Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

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

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

ABSTRACT In non-orthostereoscopic video see-through (VST) head-mounted displays (HMDs), the perception of the three-dimensional space is negatively altered by geometrical aberrations, which may lead to perceptual errors, problems of hand-eye coordination, and discomfort for the user. Parallax-free VST HMDs have been proposed, yet their embodiments are generally difficult to create. The present study investigates the guidelines for the development of non-orthostereoscopic VST HMDs capable of providing perceptually coherent augmentations for close-up views, hence specifically devoted to guide high-precision manual tasks. Our underlying rationale is that, under VST view, a perspective-preserving conversion of the camera frames is sufficient to restore the natural perception of the relative depths around a pre-defined working distance in non-orthostereoscopic VST HMDs. This perspective conversion needs to account for the geometry of the visor and the working distance. A simulation platform was designed to compare the on-image displacements between the direct view of the world and the perspective-corrected VST view, considering three different geometrical arrangements of cameras and displays. A user study with a custom-made VST HMD was then conducted to evaluate quantitatively and qualitatively which of the three configurations was the most effective in mitigating the impact of the geometrical aberrations around the reference distance. The results of the simulations and of the user study both proved that, in non-orthostereoscopic VST HMDs, display convergence can be prevented, as the perspective conversion of the camera frames is sufficient to restore the correct stereoscopic perception by the user in the peripersonal space.

INDEX TERMS Head-mounted display, stereoscopic displays, augmented reality, video see-through displays, orthoscopic view, optical aberrations.

I. INTRODUCTION

In ideal visual augmented reality (AR) systems there should not be any perceivable difference between the user's natural view of the world and his/her augmented view through the display. This is especially true when they are designed to aid complex manual tasks that demand great dexterity (e.g., surgical procedures). Avoiding this difference entails an accurate AR registration and ergonomic interaction with the augmented scene. According to Azuma *et al.* [1], "*The basic goal of an AR system is to enhance the user's perception of and interaction with the real world through supplementing the real world with 3D virtual objects that appear to coexist in the same space as the real world*".

In line with this, wearable AR systems based on head-mounted displays (HMDs) provide the user with an egocentric and natural viewpoint which is why they are deemed as the most ergonomic and effective solutions to guide those tasks manually performed under the user's direct vision [2], [3]. To ensure a consistent perception of the augmented world and to improve the interaction with it, the real world and the virtual enrichment must be perfectly integrated with each other photometrically, temporally, and spatially.

In stereoscopic video see-through (VST) HMDs, the user's visual perception of the 3D world is mediated by two different optical systems: the acquiring camera and the visualization display. The stereoscopic images of the real scene

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

are recorded by a pair of cameras rigidly anchored to the visor with an anthropometric interaxial distance, and then reproduced onto the left and right displays of the visor after being accurately merged with the virtual content [4]. Both the acquisition and visualization stages adversely alter the visual perception of the real world due to different monoscopic and stereoscopic optical aberrations. Such aberrations are influenced by:

- The intrinsic linear and non-linear optical properties of the two optical systems (i.e. the estimated pinhole camera model of the cameras and of the displays);
- How these two optical systems relate to each other;
- The geometrical relations between these two and the user's visual system.

In this work, we identify and mitigate the perceptual problems caused by the optical aberrations inevitably present in non-orthostereoscopic VST HMDs that are designed to guide high-precision manual tasks. The main aim is to obtain a system that provides the user with visual stimuli that are as consistent as possible with the natural view of the world, and thus to provide an effective and comfortable stereoscopic vision of the world.

II. RELATED WORKS

There are two different types of optical aberrations: chromatic ones, which affect the colours of the image, and geometrical ones, resulting in the spatial distortion of the reproduced image. With a pair of 2-Dimensional (2D) displays, in which the displayed images are projected in front of the user's eyes at an optically pre-set distance, we are intrinsically prone to the well-known vergence-accommodation mismatch [5]–[7]. With a stereo VST HMD, as with a stereo virtual reality (VR) HMD, the human visual system matches the vergence and accommodation distance only for fixation points at the fixed display plane, whereas for depths within accommodative-effective distances, this mismatch cannot be neglected.

Various solutions have been proposed to address this problem [8], such as displays consisting of an array of focal planes on which the images captured by the cameras are projected [9], [10] or the use of near-eye light field displays [11], [12]. However, these solutions based on integral imaging technology or on stacked LCD panels are still characterized by a non-sufficiently wide depth-of-field, a low spatial resolution, and a reduced light throughput of the display. In this paper, we analyze the perceptual problems caused solely by geometrical aberrations, modelling the user's visual system as a pair of pinhole cameras. Hereafter we use the term optical aberrations to refer only to geometrical aberrations, considering each optical system (camera, display, and eye) with a focal length that enables them to be modelled as ideal pinhole cameras.

In broad terms, in standard 2D displays, optical aberrations are due to the inherent incongruity between the user's sensation of the real-world light field and that generated by the 2D display image whose optical path to the eye is being disturbed and diverted due to passing through the lenses of the two optical systems (acquiring camera and visualization display) [13]. The differences between the intrinsic projection properties of the two optical systems (e.g., focal length and principal point) can be resolved via software after calibration. Similarly, the non-linear distortions introduced by the lenses of both the optical systems (e.g., radial distortion, tangential distortion,) can be compensated for by first undistorting the cameras frames. Then, a pre-distortion function can be used on those frames, before projecting them onto the two displays of the visor [14], [15], in order to compensate for the distortion introduced by the lenses of the visor itself [16]. By doing so, the remaining geometric aberrations are mostly due to the displacement between the ideal centers of three optical systems (eye, camera, and display), which may cause differences between the natural view and the VST-mediated view in terms of a magnification/demagnification factor and a parallax effect. This parallax distorts the patterns of horizontal and vertical binocular disparities, which can result in a distorted sensation of absolute and relative depths of the objects in the scene and introduce discomfort for the user [17]–[19].

Theoretically, to prevent the effect of parallax, orthostereoscopic VST HMDs should be adopted. An optical system is defined as orthoscopic if it can provide images devoid of geometric aberrations. In such a system, the user perceives the object in the scene with correct proportions, dimensions and spatial localization. This means that when using a rigorously orthostereoscopic VST HMD, the view mediated by cameras and displays does not present any difference with the NE view. For a binocular HMD to be considered orthostereoscopic, it has to comply with the following specifications [20]:

- The center of projection of the cameras and that of the displays must coincide. These must in turn coincide with the centers of projection of the wearer's eye.
- The left and the right optical axes of the displays must be coincident with the left and the right optical axes of the cameras, respectively.
- The distance between the left and the right cameras and between the left and the right displays must be equal to the distance between the observer's eyes (IPD).
- The field of view (FOV) of the displays must coincide with the FOV of the cameras.

There are various solutions for implementing claimed parallax-free VST HMDs. In 1998 Fuchs *et al.* [21] were the first to present the idea of an orthostereoscopic VST HMD for use in the medical field. In their system, the projection center of the camera is optically moved close to the nodal point of the observer's eye by means of a pair of mirrors. In a work by Takagi et al. published in 2000 [20], the authors presented an analysis of all the possible distortions in depth perception due to non-rigorous orthostereoscopic configurations. Here too, the authors tried to create a parallax-free VST HMD by means of a set of mirrors and optical prisms whose task was to optically fold the optical axes of the displays to those of the two cameras. However, this solution was still characterized by

an offset of approximately 30 mm between the camera center and the exit pupil of the display. Therefore, the conditions of rigorous orthostereoscopy were not fully met. The system was thus labelled as quasi-orthoscopic, as we will explore further later in this section.

In 2005, State *et al.* [22] presented an innovative VST HMD specifically designed to generate zero eye-camera offset. Their prototype was developed as a result of a design and optimization work through a software simulator. Yet also there, due to the constructive complexity, the actual embodiment did not meet all the requirements for implementing an orthoscopic VST visor that they had described in their simulated scenario. Their final system could provide a parallax-free perception of the reality only for user-specific and constant settings in terms of eye position, interpupillary distance (IPD), and eye convergence.

Finally, in 2009 Bottecchia *et al.* [23] proposed a prototype of orthoscopic monocular VST HMD, in which the correction of the parallax was made using software. Unfortunately, the authors did not provide further details in their paper on how this computer-based correction was done.

From the above-mentioned works, it is clear that the creation of rigorously orthoscopic VST systems is challenging. Mirrors are needed to bring the center of projection of the camera onto the center of projection of the human eye. However, this complicate the design, which is bulkier not only because of the mirrors, but also because of the cables needed to convey the electrical signals for the rotation of both displays and cameras. Furthermore, since the projection center of the camera needs to exactly overlap the projection center of the eye (specification 1), each time the user adjusts the visor's position on his/her head, the positions of the eyes need determining and the mirrors re-calibrated accordingly, making this type of solution unpractical. Sub-optimal prototype solutions have thus been proposed, which can be categorized as quasi-orthoscopic. These VST HMDs were typically manufactured by assembling commercial binocular displays, designed for VR applications, with stereo cameras rigidly anchored on the top [24]–[26] or front of the HMD [27]. The prototypical embodiment with the cameras mounted centrally and on-axis with the displays yield a parallax only along the anterior-posterior direction [2], [28]. In this case, there is above all a magnification effect that can be partially compensated for by applying a scaling factor referring to a specific distance from the observer. This planar transformation (i.e., plane-induced homography) readjusts the size of the camera's image with that of a reference plane perceived by the naked eye [29].

Nevertheless, most VST HMDs are designed for applications in which the cameras are mounted parallel to each other. By contrast, for those AR applications in which the user is asked to interact with the augmented scene within the peripersonal space (i.e., tasks performed within arm's reach), the stereo cameras with parallel optical axes are not able to provide sufficient stereo overlaps for stereo fusion. To cope with this, both hardware and software solutions have been proposed [30]. As an example of a hardware solution, the excessive retinal disparity can be overcome by physically converging the cameras [31], [32] so as to increase the overlapping zones between the left and right images and thus yielding a correct stereoscopic vision even of the closest objects without diplopia. However, as previously mentioned, the optical axes of cameras and displays need to be coincident to prevent geometric and stereoscopic distortions, such as keystone distortion and depth plane curvature. Therefore, if the cameras are converging, the displays should also physically converge by the same angle. A valid alternative is represented by a purely software approach [33], [34], in which the stereo overlap is maximized via software by dynamically handling the convergence or the shearing of the display frustum based on a heuristic estimation of the working distance. The benefits and shortcomings of both approaches are discussed more extensively in our previous work [30].

In this paper, we aim to provide a conclusive answer as to whether, in order to mitigate the perspective distortions due to the optical aberrations in VST HMDs, the convergence of the cameras should be associated with an equivalent physical rotation of the displays, or a perspective correction using software is sufficient. The latter hypothesis would simplify the implementation of perceptually coherent quasi-orthoscopic VST HMDs designed for close-range tasks. We investigated both ideal orthoscopic and realistic quasi-orthoscopic VST visors, in three different configurations: parallel cameras and display (PA) configurations, toed-in cameras and display (TI) configurations, and toed-in cameras and parallel display (STI) configurations. The TI configuration is used to simulate the purely hardware solution, with the physical rotation of both the cameras and the displays, whereas the STI configuration simulates a matched hardware-software solution, where the images caught by the convergent cameras are warped before being rendered onto the corresponding parallel displays. To test the quasi-orthoscopic configurations, we designed a dedicated simulation platform in MATLAB (R2018b MathWorks, Inc., Natick, Massachusetts, US), which enabled us to test different arrangements of cameras and displays. The outcomes of the simulations were finally validated through a user study conducted with a realistic quasi-orthoscopic VST device assembled with commercial components.

III. ORTHOSCOPIC CASE ANALYSIS

Under close-up viewing conditions, such as performing tasks within arm's reach, the PA configuration is not able to provide sufficient stereo overlaps. In this case, as shown in Fig. 1, the pattern of horizontal disparities between the left and the right images is excessive, to such an extent that the visual cortex is not able to integrate the two images, regardless of the eye convergence, causing diplopic vision [35].

In Fig. 1, the eyes have been arbitrarily oriented in parallel. Nonetheless, if the eyes converged, the reduced area of stereo overlap surrounding the fixation point, would make it rather unlikely for the user to correctly merge the two images into a single stereo-view.

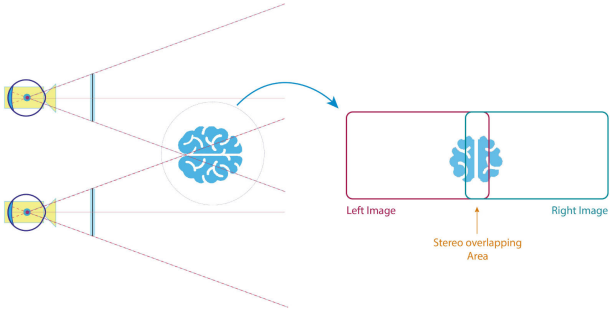


FIGURE 1. Parallel configuration in the orthoscopic case: if the target is close to the user, this configuration cannot provide a sufficient stereo overlap area. The horizontal disparities between the left and the right images are liable to cause diplopia.

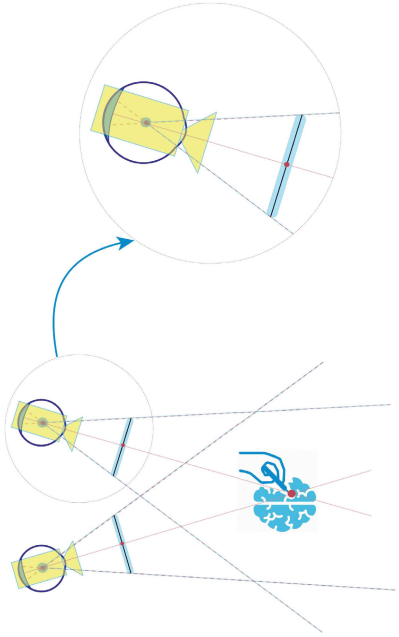


FIGURE 2. Purely hardware solution for orthoscopic video see-through head-mounted displays: both cameras and displays are rotated around the projection center of the respective eye. The condition of naked eye view is efficiently restored, as the optical axes of camera, display and eye are all coincident.

Hardware and software solutions are illustrated in Fig. 2 and Fig. 3. In both figures we intentionally omitted the optical elements needed for virtually moving the camera centers of projection centered on the observer's eye.

In the purely hardware solution, both cameras and displays must be rotated so that the rotations go around the projection center of the associated eye. The camera frame is projected onto the respective display with one-to-one mapping (i.e., the camera frame is projected on the display as it is). In geometric terms, when implementing this TI configuration, the stimuli necessary for the correct sensation of all the points in space can be correctly restored when the eyes are converged. The naked-eye (NE) viewing condition is thus restored, as the optical axes of eye, display and camera are all coincident and converge at the fixation point. On the other hand, in the mixed software/hardware approach (Fig. 3), only

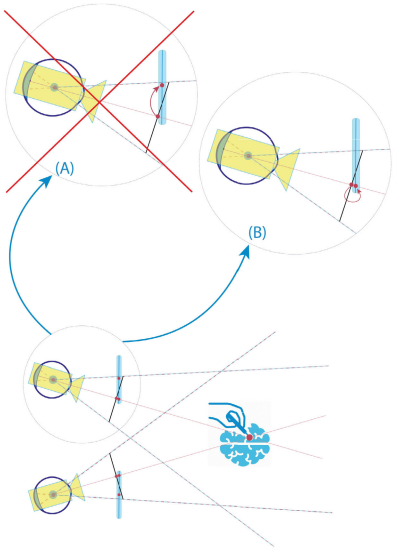


FIGURE 3. Mixed software/hardware solution for orthoscopic video see-through head-mounted displays. The cameras are rotated around the projection center of the respective eye, whereas the displays are kept parallel to each other. In this case, to restore the natural pattern of retinal disparities, a simple copy of the acquired image is not sufficient, unlike the previous case (A). Thus, a rotation-induced homography needs to be applied to the camera frames before projecting them onto the displays (B).

the cameras are made to converge, so as to mimic the natural eye convergence in the case of close working distances. Conversely, displays are fixed in the PA configuration, as in the case of commercial HMD devices where the displays are normally set to be parallel. Here, in order to restore the NE viewing condition, projecting the frame acquired by the camera on the display with a one-to-one mapping is insufficient, and a rotation-induced homography is needed first, as shown in Fig. 3. A perspective-preserving view of the world is thus restored at any distance from the observer by retrieving the correct pattern of image disparities between rotated camera frames and parallel display frames. The retinal disparities obtained are geometrically and therefore sensationally equivalent to those experienced by the user whilst looking directly at the world (without the mediation of the HMD). Therefore, in the orthoscopic case, the software rotation of the images is intrinsically equivalent to the hardware rotation because the display, cameras and eye projection centers are coincident [29].

In the case of quasi-orthoscopic configurations, the displays, cameras and the eye projection centers are not coincident. In the next section, we describe the software simulator developed for comparing different arrangements of cameras and displays in a quasi-orthoscopic VST HMD.

IV. SIMULATOR

MATLAB enables an ad hoc virtual scenario to be created, in which surfaces and objects can be arranged and then observed through virtual cameras that can be freely placed and oriented inside the virtual space as required. We developed an application for the simulation of the entire

acquisition and visualization process underpinning the whole VST mechanism of quasi-orthoscopic HMDs. The virtual scenario comprises the following elements: a planar grid as the reference object to be observed, a pair of virtual cameras representing the real cameras, a pair of virtual cameras representing the user's eyes, and a pair of planar surfaces representing the images screened on the displays 4 at a certain distance with respect to their centers of projection (i.e., we modelled the two displays also as two virtual on-axis cameras).

The virtual cameras simulating the real ones are modelled on commercial Leopard Imaging cameras (LI_OV4689) in terms of intrinsic parameters (resolution set at 1280 × 720 pixels; diagonal FOV of 109°. The planar surfaces representing the images projected on the displays are modelled using the specifications of a commercial OST optical engine: LUMUS OE-33 [36] with resolution 1280 × 720 pixels, diagonal FOV of 40°, angular resolution of about 1.7 arcmin/pixel [34], [37] eye-relief of 22 mm, and eye motion box of 10 × 8 mm. The placement of the displayed images at the eye relief distance does not affect the validity of the simulations, since we restored the correct perspective view, which is independent of the real viewing distance, which for the LUMUS OE-33 is infinite.

In our simulations, we centered the cameras in front of the displays considering the footprint of the real components in order to reduce the eye-to-camera parallax. Using this virtual scenario, we can simulate what the user's eyes would see with and without the mediation of the VST display.

We defined the coordinate systems for virtual eyes, displays, virtual cameras and the world as shown in Fig. 4. For the sake of simplicity, we initially considered the projection centers of the eyes as being coincident with those of the displays, thus simulating an ideal condition. We considered the origin and orientation of the world's coordinate system as the reference system of the virtual scene being created, thus the subsequent placement of all the elements in question within this virtual scenario (reference plane, cameras, displays and eyes) refers to this system.

For both the left and the right sides of the stereoscopic HMD, the reference systems of the two optical elements (camera and display) were determined from the CAD file of a modified version of a VST HMD visor previously presented in [34]. Together with the coordinate systems, we also defined the transformations between the coordinate systems associated with all the elements in the scene.

The geometrical relations between the elements in the scene can be modified during an initialization stage. The user can choose from the following parameters:

- The position of the reference plane for which the homographic transformation is computed.
- The position and orientation of a testing grid in the scene reference system.
- Whether the cameras and/or the displays are to be set in PA or TI configuration (i.e., to focus at the center of the testing grid).

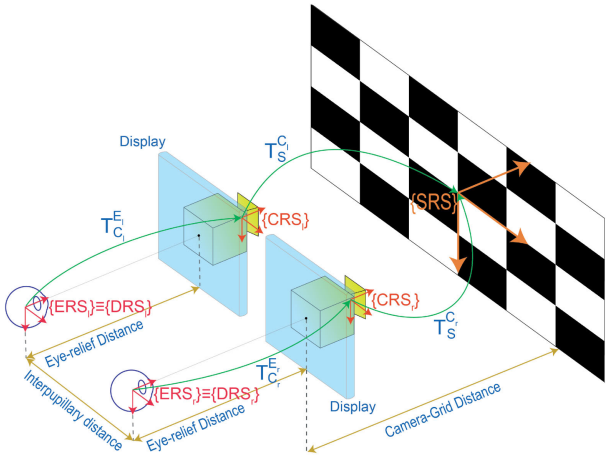


FIGURE 4. The virtual scenario within the simulator. The reference systems of the elements in the scenario and the geometrical relation between them are shown. SRS is the reference system of the virtual scenario; CRS (left and right) are the reference systems of the two virtual cameras; DRS (left and right) are the reference systems of the two displays; ERS (left and right) are the two reference systems of the eyes; T_C^E (left and right) is the transformation matrix that relates the camera to the eye; T_S^C (left and right) is the transformation matrix that relates the virtual scenario to the camera.

- The fixation point defining the point in space where the observer's eyes are focused.
- The position of both the eyes in the scene reference system.
- The IPD

The IPD constrains both the distance between the optical axes of the acquiring cameras and of the displays. For each side, the simulator acquires and outputs three images: the first frame represents the NE view of the testing grid (without VST mediation); the second the acquiring camera view of the grid; the third what the eye would perceive when gazing at the display, on which the camera frame, appropriately warped, is being projected. The NE acquisition is used as a benchmark: the aim of the plane-induced homography applied to the image captured by the camera is to eliminate the parallax on a reference plane between the cameras and the eyes/displays [34]. In the next section, we illustrate this homographic transformation in more detail.

A. HOMOGRAPHY

This section describes the procedure to compute the homographic transformation that relates two perspective views of a planar scene placed at a pre-defined distance. In broad terms, the homography is defined as a bijective function between the elements of two R^2 spaces, whereby each point of the first space corresponds to one and only one point of the second space. The homography we work with is a transformation that models the correspondences between two perspective views (one from the acquiring camera and one from the display) of the same reference plane. The following variables and symbols are used hereafter:

- The homographic transformations H_W^D and H_W^C, which relate respectively the points of the reference plane in the world W to their projections onto the image planes of both the display D and the camera C.
- The distance $d^{C\to\pi}$ between the origin of the camera reference system (CRS) and the reference plane π.
- All the reference systems are right-handed, with the Z axis forward oriented.
- R_C^D and $\vec{t}_C^D$, which are respectively, the rotation matrix and the translation vector of the transformation matrix T_D^C between the camera reference system (CRS) and the display reference system (DRS).
- K_C and K_D, which are respectively the intrinsic matrices of the camera and display projective models. To determine the intrinsic parameters of the camera, encapsulated by K_C, we used the MATLAB Computer Vision Toolbox tool for calibration, which implements Zhang's method [38]. K_D encapsulates the parameters of the frustum of the near-eye display. We derived the focal length of the display (f) using the manufacturer's specifications regarding the horizontal and vertical FOV of the display considered in the simulator (*hfov* and *vfov*). We assumed that the focal length on both the x-axis and y-axis were equal ($f_x = f_y$), meaning the display pixels were considered as being perfectly square. We also assumed the coordinates of the principal point were exactly half of the display resolution ($C_u = W/2 = 640, C_v = H/2 = 360$. In summary we assumed:

$$K_D = \begin{bmatrix} \frac{W}{2\tan(\frac{hfov}{2})} & 0 & W/2 \\ 0 & \frac{H}{2\tan(\frac{vfov}{2})} & H/2 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

- $\vec{n}$, which is the normal unit vector to the grid plane, expressed in the CRS.
- H_C^D, which is the perspective preserving homography, induced by the plane of the grid placed at distance $d^{C\to\pi}$ from the camera.

The homography H_D^C describes the bijective relation between the camera viewpoint and the user's viewpoint, such that:

$$\lambda p_D = H_C^D(R_C^D, \vec{t}_C^D, K_C, K_D, \pi)p_C \quad (2)$$

where $p_C = (x_C.y_C)$ are the coordinates of a pixel in the camera image, whereas $p_D = (x_D, y_D)$ are the coordinates of the corresponding pixel in the display image. The points are expressed in homogeneous coordinates, thus λ is a generic scale factor due to the equivalence of the homogeneous coordinates rule. The parenthesis means that the homography H_C^D is a function of the relative pose between CRS and DRS $(R_C^D, \vec{t}_C^D)$, the intrinsic parameters of the camera and display (K_C, K_D), and the position and orientation of the reference plane in the scene (π) [29], respectively.

$$H_C^D = K_D\left(R_C^D + \frac{\vec{t}_C^D\vec{n}}{d^{C\to\pi}}\right)K_C^{-1} \quad (3)$$

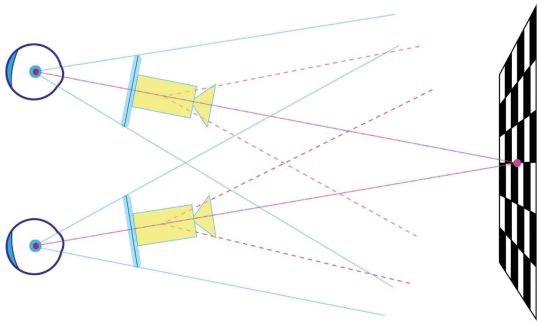


FIGURE 5. Toed-in configuration in non-orthoscopic case: both cameras and displays are convergent.

The homographic transformation is only valid at a fixed plane with normal unit vector $\vec{n}$ and distance $d^{C\to\pi}$ from the acquiring camera. Hence, if the scene observed is not planar or if the plane under observation is different from the homography plane, the direct view of the scene will not perfectly match with the rendered image on the display (i.e., direct view and VST-mediated view are not orthoscopically registered). In the next subsection, we analyze the simulations carried out in detail.

B. SIMULATION RESULTS IN THE CASE OF A QUASI-ORTHOSCOPIC HMD

The simulations were conducted considering the distance between the reference plane and the observer equal to 50 cm (an average working distance during manual procedures) and an IPD of 6.5 cm. To answer the question as to whether or not the rotation of the displays is needed in the case of quasi-orthoscopic systems, we tested three configurations with our simulator: TI configuration (Fig. 5), STI configuration (Fig. 6) and PA configuration (Fig. 7). This simulation aims to confirm whether or not the vergence of the cameras itself is necessary.

The outcomes of these three configurations are analyzed separately. In all three cases, for each eye, we show the overlapping views between the NE view of the grid (used as a benchmark) and what the eye can see when staring at the display, where the camera frame, appropriately warped, is being projected. The composite images corresponding to the overlapping views were formed using the imfuse function by MATLAB. In each configuration, we analyzed the results by placing the grid not only exactly at the reference plane distance (50 cm), but also at 40 and 60 cm. In this way, we intended to analyze how the pattern of image disparities is distorted around the reference plane.

1) TI CONFIGURATION

As illustrated in Fig. 8a, there is a perfect overlap between the composite image between image corresponding to the NE view of the grid (in pink in both figures) and the image observed on the display (shown in green in both figures) for both eyes with the grid placed on the reference

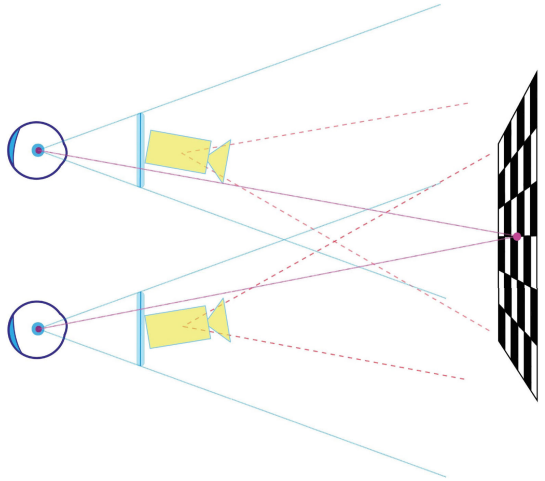


FIGURE 6. Semi toed-in configuration in non-orthoscopic case: only the cameras are made to converge on the reference grid, whereas the displays are kept parallel to each other.

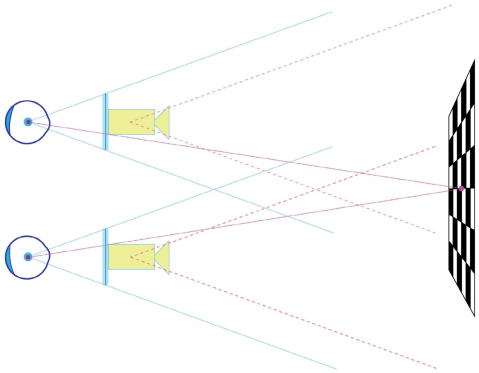


FIGURE 7. Parallel configuration in non-orthoscopic case: both cameras and displays are parallel. This configuration was added to the simulations to determine whether or not the rotation of the cameras is actually needed.

plane (50 cm). This shows that the homographic transformation was correctly evaluated and applied. This also confirms that, as expected, no optical aberrations are introduced on the view of the reference plane after applying the perspective conversion of the camera frames. This thus proves that we have restored the ideal orthoscopic vision of that plane.

2) STI CONFIGURATION

As in the TI configuration, we applied the perspective homography to achieve a perfect alignment between the VST-mediated view and the NE view of the reference plane at 50 cm (Fig. 8b). However, there is one difference compared to the previous case: in the TI configuration, unlike the STI configuration, most of the stereo FOV of the displays is preserved. Other than this, the two simulated cases are identical. In fact in both configurations the homography enables us to view the reference plane as if the camera center of projection coincided with that of the eye. Thus, on the homography plane, the view of the grid is not affected by any optical aberration.

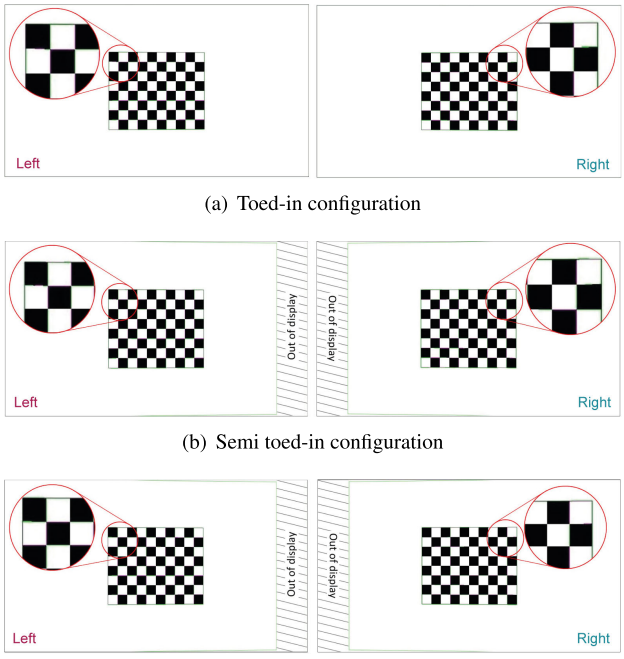


FIGURE 8. Overlapping between direct view and video see-through view of the reference plane (50 cm). (a) Toed-in configuration. (b) Semi toed-in configuration. (c) Parallel configuration.

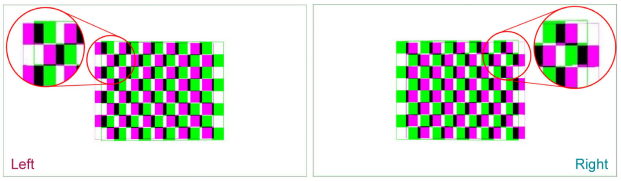
3) PA CONFIGURATION

The physical rotation of the cameras can be avoided if replaced by an equivalent virtual rotation of the camera frustums in order to achieve the maximum stereo overlap for close fixation points as suggested by [33]. Obviously, this approach works for those VST systems that have wide FOV cameras. The outcomes of the simulations in Fig. 8c confirm that, again, there is a perfect overlap between the NE view of the grid and the display-mediated view. In addition, as in the STI configuration, a small portion of the stereo FOV is lost due to the parallel arrangement of the displays.

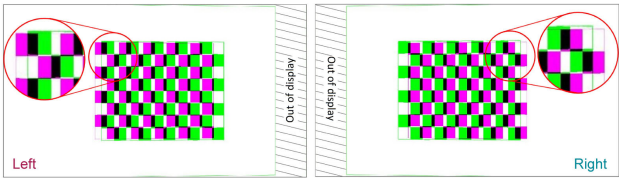
Below we report the results of the simulation only for the TI and STI configurations, as the aim of this work was to assess whether in quasi-orthoscopic VST HMDs, display rotation is needed to facilitate a close-up view. However, the results of the simulation for the PA configuration are the same as those reported below for the other two configurations.

4) TI VS STI CONFIGURATION OUTSIDE THE REFERENCE PLANE

As for the area outside the homography plane, as expected, we lose the perfect match between the NE and the display-mediated view, as shown in Fig. 9 and Fig. 10, moving the grid at 40 and 60 cm. In general, the presence of the optical aberrations outside the reference plane may compromise the user's perception of the objects in the scene, both in terms of depth and size. However, systems affected by such optical aberrations have already been proposed to assist in manual tasks [26], [39], [40]. In all these studies, the users were



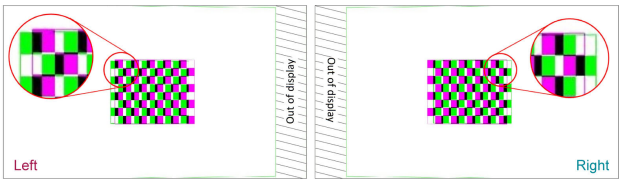
(a) Toed-in configuration



(b) Semi toed-in configuration

FIGURE 9. On-image displacements between direct view and video see-through view of the target grid. The grid is placed at 40 cm from the user, whereas the homography is evaluated for a plane at 50 cm. (a) Toed-in configuration. (b) Semi toed-in configuration.

(a) Toed-in configuration



(b) Semi toed-in configuration

FIGURE 10. On-image displacements between direct view and video see-through view of the target grid. The grid is placed at 60 cm from the user, whereas the homography is evaluated for a plane at 50 cm. (a) Toed-in configuration. (b) Semi toed-in configuration.

able to perform the procedures without any difficulty, despite highlighting the lack of naturalness in the view of the real scenario through the HMD device.

We evaluated the pattern of image disparities (d_{dx},d_{dy}) by comparing, for each side, two 1280 × 720 pixel images: the NE view of the reference grid and its display-mediated view. In both images, we used the corners of the chessboard as reference features and we compared their coordinates in pixels. Table 1 shows the results obtained in arcmin. The value of the retinal disparity (d_{rx},d_{ry}) in arcmin was computed by plugging $W = 1280$, $H = 720$, $\alpha = 35.2°$ (display *hfov*), and $\beta = 20.2°$ (diplay *vfov*) into the following relations.

$$d_{rx} = 2\arctan\left(\frac{d_{dx}\tan(\alpha/2)}{W}\right) \tag{4}$$

$$d_{ry} = 2\arctan\left(\frac{d_{dy}\tan(\beta/2)}{H}\right) \tag{5}$$

TABLE 1. Disparity between the naked eye and VST mediated views in arcmin.

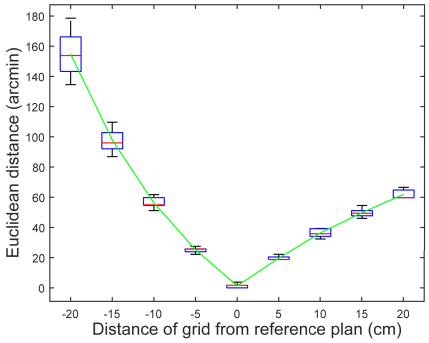
Config.			Mean		Std. Dev.		Max value		Eucl. Dist.	
			d_{rx}	d_{ry}	d_{rx}	d_{ry}	d_{rx}	d_{ry}	*Mean*	σ
Close Grid (40 cm)	Toed-in	*Right*	61.2	3.4	3.9	2.2	66.5	8.5	61.2	3.7
		Left	61.5	3.6	3.9	2	68.2	8.3	61.7	3.9
	Semi Toed-in	*Right*	61.5	3.7	4.4	2.4	68.2	8.5	61.7	4.4
		Left	61.7	3.6	4.1	2	68.2	8.3	61.9	4.1
Far Grid (60 cm)	Toed-in	*Right*	38.5	0.7	1.9	0.9	40.9	1.7	38.5	1.9
		Left	38.8	1.7	2.2	1.5	42.6	5.1	38.8	2.2
	Semi Toed-in	*Right*	38.5	0.7	1.9	1	40.9	3.4	38.5	1.9
		Left	39	1.5	2.6	1.4	42.6	3.4	39	2.6

The values in the table are the errors in terms of mean value, standard deviation (σ), maximum value and Euclidean distance between real point and virtual point (i.e., $e_r = \sqrt{d_{rx}^2 + d_{ry}^2}$) for the right and the left sides in all the cases analyzed.

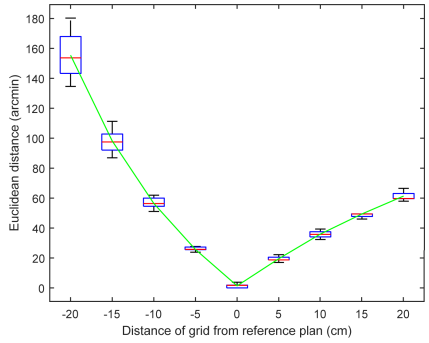
The simulation data shown in Table 1 clearly show that there are no substantial differences in terms of disparity between the two configurations, even outside the reference plane. Therefore, we can reasonably assert that, in order to preserve a coherent 3D perception around a pre-defined reference plane, rotating the displays can be prevented if we apply an appropriate perspective conversion to the camera frames. As further confirmation, we also measured the pattern of disparities for eight planes between 30 and 70 cm. The results of these simulations are reported in Fig. 11.

5) TI VS STI WITH THE EYES OUTSIDE THE CENTERS OF THE EYEBOXES OF THE DISPLAYS

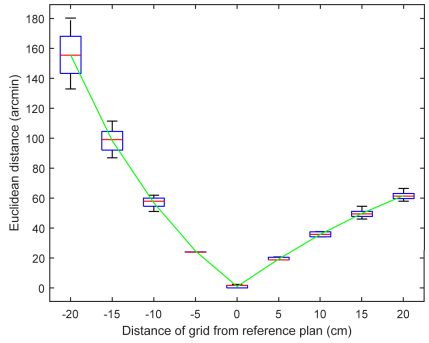
If the projection centers of the eyes are maintained within the eye motion box of the displays, although not exactly coincident, the user perceives the same 2D images. In any case, since the homography is calculated to map the cameras images for the center of the eye-box and at the eye-relief distance, any displacement of the eyes introduces a parallax between the NE view and the VST-mediated view. We tested such condition for both TI and STI configurations by simulating an extreme anti-symmetric displacement comprising a 5 mm anti-symmetric displacement of the eyes along the interocular axis (x-axis) and a 4 mm displacement along the vertical axis (y-axis). In this case, the calculated homography, cannot restore the condition of ideal orthoscopy on the reference plane. The pattern of horizontal and vertical disparities between the left and right images is



(a) Toed-in configuration



(b) Semi toed-in configuration



(c) Parallel configuration

FIGURE 11. Boxplot of the simulated misalignment in terms of Euclidean distance for eight planes between 30 and 70 cm . (a) Toed-in configuration. (b) Semi toed-in configuration. (c) Parallel configuration.

not consistent with that of the NE view which may maintain the sense of visual discomfort for the user for close-up views [18], [19] (Fig. 12). Again, there are no significant differences between the disparity patterns obtained on each side in both configurations.

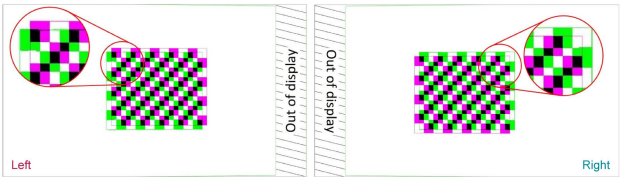
V. USER STUDY

To confirm the results obtained with the simulator, we carried out a user study with a custom-made quasi-orthoscopic VST HMD (Fig. 13), which enabled us to deploy the three aforementioned viewing configurations.

A. EXPERIMENTAL SETUP

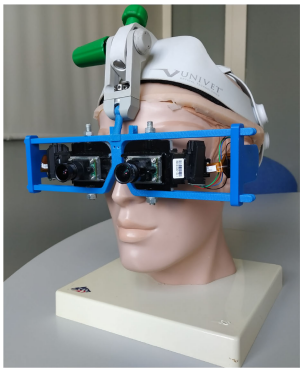
The device used for the experimental tests was assembled by reworking a commercial binocular OST visor with a similar approach to our previous work [34]. The specifics of the OST HMD are set out in Section IV. Each waveguide of the HMD was anchored to a 3D-printed rigid shell, which also comprised a support for the on-axis camera. The translation and the rotation of each waveguide/camera block are ensured by a horizontal guide and obtained by means of two coaxial

(a) Toed-in configuration

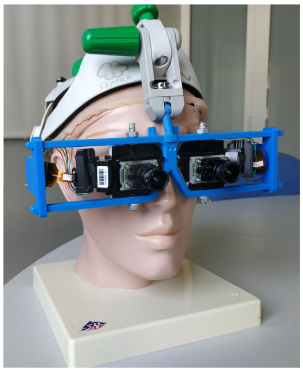


(b) Semi toed-in configuration

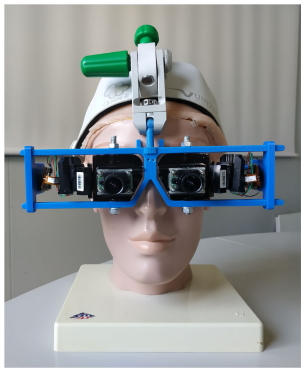
FIGURE 12. On-image displacements between direct view and video see-through view of the target grid when the eyes are outside the ideal position. The user's interpupillary distance is 7.5 cm, and the offset on the y-axis between the left and the right eye is 8 mm.. (a) Toed-in configuration. (b) Semi toed-in configuration.



(a) Right side view

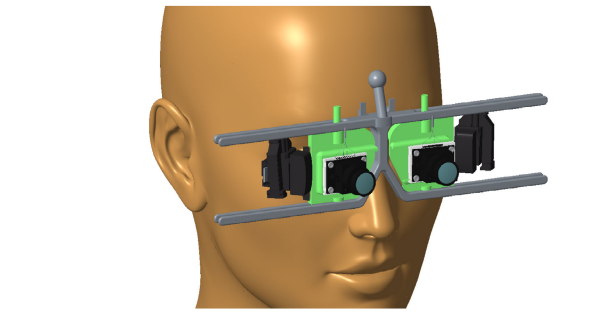


(b) Left side view

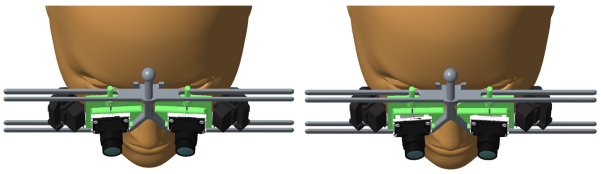


(c) Front view

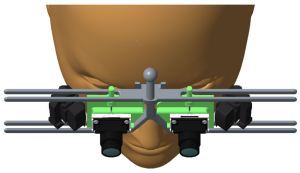
FIGURE 13. The custom-made quasi-orthoscopic video see-through visor. (a) Right side view. (b) Left side view. (c) Front view.



(a) Three-quarter view



(b) Toed-in configuration (c) Semi toed-in configuration



(d) Parallel configuration

FIGURE 14. CAD design of the three video see-through viewing configurations. (a) Three-quarter view. (b) Toed-in configuration. (c) Semi toed-in configuration. (d) Parallel configuration.

pivot axes (one above and one below the waveguide). The pivot axes are constrained to the rigid shell, which allows the user to set the inter-axial distance between the displays and their convergence, which is established depending on the working distance (i.e., the stereo cameras must converge at the pre-fixed distance). Once the final configuration has been set up, both the blocks can be rigidly anchored to the guide by means of four nuts. All the rigid components were CAD designed using PTC Creo Parametric 3D Modelling software (vers. 3.0) (Fig. 14).

For the effective operation of the VST mechanism, we developed a dedicated software application based on OpenCV machine vision libraries (vers. 3.3.1). The software was built in C++ on Linux (Ubuntu 16.04) and runs on a standard workstation class PC with the following specifications: Intel Core i7-4770 CPU @ 3.40 GHz with 4 cores and 12 GB RAM, Nvidia GeForce GTX 1050 (2GB). The core function of the software implementing the VST mechanism processes images recorded by the pair of external RGB cameras before rendering them on the two microdisplays of the visor. These frames are initially undistorted to eliminate the non-linearities due to optical distortions, and then warped according to the perspective-preserving homography (as explained in Section IV.A). The Fig. 15 shows the experimental setup.

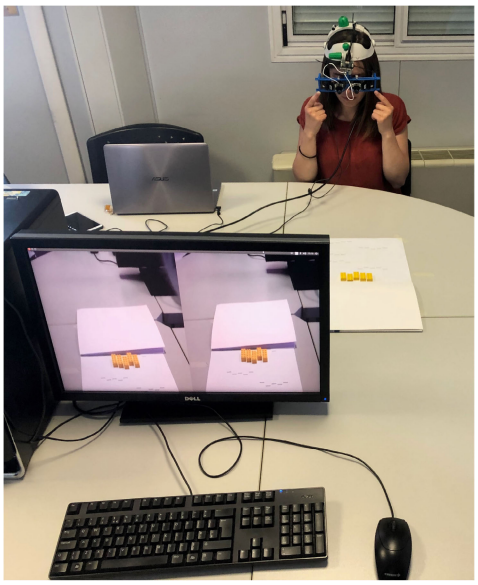


FIGURE 15. Experimental setup.

TABLE 2. Demographics of the ten participants in the user study.

General info	Value
Gender (male; female; non-binary)	(3; 7; 0)
Age (min; max; mean; STD)	(26; 40; 30.2; 4.9)
IPD (mm) (min; max; mean; STD)	(52.4; 67.3; 57.1; 4.5)
Uncorrected poor vision (no, yes)	(8, 2)

B. EXPERIMENTAL DESIGN

Ten participants were recruited from university students, staff, and faculty members. The demographic information on the users is listed in Table 2. All participants had normal vision acuity or corrected visual acuity with the aid of prescription glasses or contact lenses. In the tests, each participant wearing the VST HMD was asked to estimate the relative depth relations between five objects of the same size and colour (five yellow Lego® 9.6 × 32 × 16 mm bricks) adopting a similar approach to that presented in [34]. The five bricks were laid on a A3 paper sheet (size 297 × 420 mm), which had ten sequences of demarcation lines distributed as follows: each sequence of demarcation lines was associated with an absolute depth from the observer covering a range from about 34 to 70 cm. The relative distances between the bricks within each sequence were decided randomly, with a defined relative resolution of 3 mm.

The perceptual tests were repeated twice (using two different A3 sheets) for the three HMD configurations and without the HMD (NE condition). The ten subjects therefore each performed the four groups of stereoscopic fusion tests, resulting in $10 \times 4 \times (10 \times 2) = 800$ sequences of demarcation lines overall. The tests with the HMD were all performed maintaining the same homography transformation, which referred to a reference plane placed at a distance of 500 mm from the ideal

position of the wearer's eyes. The final goal of the tests was to assess how this aspect would have a detrimental effect on perceiving relative depths around the reference plane. Before each session, each participant was instructed about the test and asked to slide the waveguide/camera blocks to suit his her own IPD if he/she could not see all the images on the displays.

1) QUALITATIVE ASSESSMENT OF THE VST VIEWING CONFIGURATIONS

After completing the four test sessions, all participants were asked to fill in the demographic survey and a Likert questionnaire to investigate differences in the perceived workload and visual comfort among the different VST viewing configurations. The Likert questionnaire, shown in Table 3, comprises six items, each evaluated using a five-point monotone Likert scale (from 1 = strongly disagree, to 5 = strongly agree).

2) QUANTITATIVE EVALUATION OF THE PERCEPTION OF RELATIVE DEPTHS IN THE PERIPERSONAL SPACE

The goal of the quantitative evaluation was to measure the degree of accuracy in perceiving depth relations in the peripersonal space with the HMD under the three different VST configurations. For each sequence of five demarcation lines, we considered three possible scores: correct sequence (score = 1), wrong sequence (score = 0), half-correct sequence (score = 0.5). The latter score was assigned when the subject guessed a sequence of three consecutive bricks. For each test session, the overall score was reported in terms of success rate (i.e., S_r) and expressed as a percentage (e.g., a score of $\frac{15}{20} = 75\%$). Each trial was analyzed by a blinded observer. The time for completing the tasks under the different viewing conditions was also measured. The results and the statistical analysis were both processed in MATLAB.

3) STATISTICAL ANALYSIS

Responses to the Likert questionnaire were summarized using median with dispersion measured by the interquartile range. The Kruskal-Wallis test was used to understand whether answers differed based on the VST viewing configuration experienced. A p-value<0.05 was considered statistically significant.

Quantitative results were presented in terms of the average value, standard deviation, and max value of the success rate and completion time over the ten subjects with the HMD under the three VST configurations and without the HMD (NE condition). With the first statistical analysis, we conducted two Kruskal-Wallis tests to evaluate whether the success rate and the completion time for viewing configurations originated from the same distribution. One Kruskal-Wallis test was performed to analyze the impact of the VST viewing configuration over the success rate and time to completion, whereas the second test also included the NE condition in the statistical evaluation. For both tests, a p-value <0.05 was considered statistically significant.

TABLE 3. Qualitative comparison between VST configurations according to the Likert questionnaires. (1: Strongly Disagree; 5: strongly agree).

Questions	TI Conf. median (iqr)	STI Conf. median (iqr)	PA Conf. median (iqr)	p-value
You were able to merge the stereo images	4.5 (4∼5)	4 (4∼5)	4 (4∼5)	0.98
You were able to perceive relative depths	4 (4∼4)	4 (3∼4)	4 (4∼4)	0.48
You were not affected by a significant distortion of the scene through the visor (e.g.,double vision, blurred vision)	4 (2∼5)	3 (2∼4)	4 (3∼4)	0.61
You felt no discomfort while performing the task (e.g., visual fatigue, dizziness, nausea)	3 (2∼4)	3 (2∼4)	3.5 (3∼4)	0.46
Overall, the task was not difficult to complete	4 (3∼5)	4 (3∼5)	4 (3∼4)	0.83
You experienced no stress, irritation, or frustration while performing the task	4 (3∼5)	3.5 (3∼4)	4 (3∼4)	0.70

A second statistical analysis was conducted to verify whether the success rates of the three VST configurations were differently influenced by the depth. We therefore clustered the data associated with the ten subjects over the same depth planes (ten) where the sequences of demarcation lines were located. Data were once again presented in terms of average value, standard deviation, and max value. A Kruskal-Wallis test was conducted between the three VST viewing conditions in terms of the success rate at different depths. A p-value <0.05 was considered statistically significant.

VI. RESULTS AND DISCUSSION

Table 3 shows the results of the Likert questionnaire. Subjects expressed a positive opinion regarding their ability to stereo fuse (item 1) and perceive relative depths (item 2) with all the three VST configurations. In addition, there was no statistically significant difference in the answers regarding the distortion of the scene with the three VST configurations (item 3). The participants expressed a neutral opinion regarding the discomfort while performing the depth judgement tasks (item 4), however they expressed an overall positive opinion regarding the enjoyability and friendliness of the user-study (items 5 and 6).

The results in terms of mean ($S_{r_{MEAN}}$, t_{MEAN}), standard deviation (S_{r_σ}, t_σ) and max values ($S_{r_{MAX}}$, t_{MAX}) of the success rates and completion times among all the participants are reported in Table 4. The p-values of the Kruskal-Wallis tests are also reported. The first test (p-value 3 conf.) revealed there was no statistically significant difference between the results obtained with the three VST configurations. The second test (p-value 4 conf.) revealed that the VST HMD did not appear to have any statistically relevant impact on the perception of

TABLE 4. Quantitative evaluation results: success rates (in %) and completion times (s.) over the ten participants.

	TI	STI	PA	NE	p-value 3 conf.	p-value 4 conf.
$S_{r_{MEAN}}$	80.5	87.5	80	92.5		
S_{r_σ}	16.7	10.9	11.7	9.2	0.3786	0.0673
$S_{r_{MAX}}$	100	100	95	100		
t_{MEAN}	12.6	12.8	13.5	8.83		
t_σ	4	4.2	6.5	4.2	0.9875	0.0694
t_{MAX}	18.4	22.9	28.9	17.8		

TABLE 5. Quantitative evaluation results: success rates (in %) over the ten depth planes between 34 and 70 cm.

	TI conf.	STI conf.	PA conf.	p-value
$S_{r_{MEAN}}$	80.5	87.5	80	
S_{r_σ}	8.9	8.6	9.6	0.1316
$S_{r_{MAX}}$	92.5	95	90	

the relative depths compared with the NE viewing condition. Table 5 shows the results of the second statistical analysis. The p-value of the Kruskal-Wallis test revealed that the VST configuration had no statistically significant influence over the success rates, clustering the success ratios over the ten different working distances around the reference plane.

Overall, the results of the qualitative and of the quantitative analyses of the user study suggest that the three VST configurations are statistically equivalent in terms of their ability to restore the natural perception of the relative depths around the pre-defined working distance. The quantitative analysis also revealed that the depth perception distortions under the VST view are not as severe. This is in line with findings experienced in [33] and suggested in [41], where it was speculated that the distortion of the visual space derived from the mathematical models underpinning artificial stereo vision is significantly higher than what the user actually perceives while wearing such devices. We believe that the joint effect of other depth cues partially compensate for such distortions. Also with a misalignment between the real and the ideal position of the user's eyes, for which the homography is calculated, there were no significant differences reported by the simulations between the images obtained with parallel displays and those obtained with convergent displays. Finally, and once again in accordance with [33], we can reasonably hypothesize that the physical rotation of the cameras can also be prevented if replaced by an equivalent virtual rotation of the camera frustums, providing wide-angle head-mounted cameras are used that ensure a sufficient stereo overlap at close distances. However, such wide-angle cameras should also include sensors with a resolution capable of maintaining the same pixel density. This latter feature is unfortunately not compatible with a reasonably high sampling rate (at least 60 fps), which is an essential pre-requisite for AR VST applications that must ensure low latency.

VII. CONCLUSION

We have described the development of AR stereoscopic VST HMDs for close-up views, designed as an aid to the execution of manual tasks. Our aim was to identify and mitigate the geometrical aberrations that in such systems affect depth perception, and instead provide the users with stimuli as consistent as possible with a natural view. Given the fact that the creation of orthostereoscopic VST HMDs for close-up view is difficult to achieve, we tried to provide a final answer as to whether in quasi-orthoscopic devices, the convergence of the cameras (required for ensuring sufficient stereo overlaps at close distances) is associated with an equivalent physical rotation of the displays or a purely software perspective correction is sufficient. This indication is intended to simplify the development of perceptually coherent quasi-orthoscopic VST HMDs. We thus developed a dedicated simulation platform that enabled us to compare the simulated NE view with the simulated VST view of a real scene, considering three different geometrical arrangements of cameras and displays. The simulations suggest that no additional geometrical aberrations are introduced if we apply a plane-induced perspective conversion to the camera frames, regardless of whether or not the displays are rotated. This perspective transformation is able to restore the right geometric correspondence between the NE view and the VST-mediated view of the reference plane, whereas the on-image displacements outside the reference plane are mostly the same for the three simulated VST configurations.

The results of the simulations were confirmed by a user study performed with a custom-made quasi-orthoscopic VST HMD assembled with commercial components. The results of the real tests proved that the distortion in the patterns of binocular disparities at different distances from the reference plane, does not appear to have a detrimental effect on the perception of the relative depths around the reference plane.

In conclusion, our work offers a conclusive answer: in developing quasi-orthoscopic VST HMDs, it is possible to opt for parallel displays, since the perspective distortions due to the parallax can be mitigated by appropriately warping the camera frames by a perspective transformation. This is computed by considering the geometry of the visor and the intrinsic parameters of the cameras and displays.

This indication will simplify the implementation of perceptually coherent VST HMDs and will pave the way to their profitable use as aid to close-range tasks that demands for great dexterity such as for surgical or industrial applications.

ACKNOWLEDGMENT

(Nadia Cattari and Fabrizio Cutolo are co-first authors.) The authors would like to thank Carmen Sabato for providing support during the testing session.

REFERENCES

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov. 2001.
[2] F. Cutolo, C. Freschi, S. Mascioli, P. D. Parchi, M. Ferrari, and V. Ferrari, "Robust and accurate algorithm for wearable stereoscopic augmented reality with three indistinguishable markers," *Electronics*, vol. 5, no. 3, p. 59, 2016.
[3] H. Liu, E. Auvinet, J. Giles, and F. R. Y. Baena, "Augmented reality based navigation for computer assisted hip resurfacing: A proof of concept study," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1595–1605, 2019.
[4] J. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence*, vol. 9, no. 3, pp. 287–309, Jun. 2000.
[5] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," *Proc. SPIE*, vol. 7690, May 2010, Art. no. 76900B.
[6] H. Hua, "Enabling focus cues in head-mounted displays," *Proc. IEEE*, vol. 105, no. 5, pp. 805–824, May 2017.
[7] Koheiüshima, K. R. Moser, D. C. Rompapas, J. E. Swan, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato, "SharpView: Improved clarity of defocused content on optical see-through head-mounted displays," in *Proc. IEEE VR*, Greenville, SC, USA, Mar. 2016, pp. 173–181.
[8] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 362–371, Jun. 2011.
[9] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *J. Vis.*, vol. 10, no. 8, p. 22, 2010.
[10] J. P. Rolland, M. W. Krueger, and A. A. Goon, "Dynamic focusing in head-mounted displays," *Proc. SPIE*, vol. 3639, pp. 463–470, May 1999.
[11] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013, Art. no. 220.
[12] F.-C. Hung, K. Chen, and G. Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graph.*, vol. 34, p. 24, Aug. 2015.
[13] E. M. Calabrò, F. Cutolo, M. Carbone, and V. Ferrari, "Wearable augmented reality optical see through displays based on integral imaging," in *Proc. Int. Conf. Wireless Mobile Commun. Healthcare*, 2017, pp. 345–356.
[14] S.-H. Lee, S.-K. Lee, and J.-S. Choi, "Correction of radial distortion using a planar checkerboard pattern and its image," *IEEE Trans. Consum. Electron.*, vol. 55, no. 1, pp. 27–33, Feb. 2009.
[15] B. Prescott and G. F. Mclean, "Line-based correction of radial lens distortion," *Graph. Models Image Process.*, vol. 59, pp. 39–47, Jan. 1997.
[16] S. Lee and H. Hua, "A robust camera-based method for optical distortion calibration of head-mounted displays," *J. Display Technol.*, vol. 11, no. 10, pp. 845–853, Oct. 2015.
[17] R. Allison, "Analysis of the influence of vertical disparities arising in toed-in stereoscopic cameras," *J. Imag. Sci. Technol.*, vol. 51, pp. 317–327, Jul./Aug. 2007.
[18] M. S. Banks, J. C. A. Read, R. S. Allison, and S. J. Watt, "SMPTE periodical—Stereoscopy and the human visual system," *Smpte Motion Imag. J.*, vol. 121, no. 4, pp. 24–43, May 2012.
[19] C. Vienne, J. Plantier, P. Neveu, and A. E. Priot, "The role of vertical disparity in distance and depth perception as revealed by different stereo-camera configurations," *I-Perception*, vol. 7, no. 6, Nov./Dec. 2016, Art. no. 2041669516681308.
[20] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi, "Development of a stereo video see-through HMD for AR systems," in *Proc. IEEE ACM Int. Symp. Augmented Reality*, Munich, Germany, Oct. 2000, pp. 68–77.
[21] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer, "Augmented reality visualization for laparoscopic surgery," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 1496, Cambridge, MA, USA, 1998, pp. 934–943.
[22] A. State, K. P. Keller, and H. Fuchs, "Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display," in *Proc. 4th IEEE ACM Int. Symp. Mixed Augmented Reality*, Vienna, Austria, Oct. 2005, pp. 28–31.
[23] S. Bottecchia, J.-M. Cieutat, C. Merlo, and J.-P. Jessel, "A new AR interaction paradigm for collaborative teleassistance system: The POA," *Int. J. Interact. Des. Manuf.*, vol. 3, no. 1, pp. 35–40, Feb. 2009.
[24] M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya, "A stereoscopic video see-through augmented reality system based on real-time vision-based registration," in *Proc. IEEE Virtual Reality*, New Brunswick, NJ, USA, Mar. 2000, pp. 255–262.
[25] G. Caruso and U. Cugini, "Augmented reality video see-through HMD oriented to product design assessment," in *Proc. Int. Conf. Virtual Mixed Reality*, San Diego, CA, USA, 2009, pp. 532–541.
[26] F. Cutolo, M. Carbone, P. D. Parchi, V. Ferrari, M. Lisanti, and M. Ferrari, "Application of a new wearable augmented reality video see-through display to aid percutaneous procedures in spine surgery," in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graph.*, Lecce, Italy, 2016, pp. 43–54.
[27] J. P. Rolland, R. L. Holloway, and H. Fuchs, "Comparison of optical and video see-through, head-mounted displays," *Proc. SPIE*, vol. 2351, pp. 293–307, Dec. 1995.
[28] F. Cutolo, P. D. Parchi, and V. Ferrari, "Video see through AR head-mounted display for medical procedures," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Munich, Germany, Sep. 2014, pp. 393–396.
[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[30] F. Cutolo and V. Ferrari, "The role of camera convergence in stereoscopic video see-through augmented reality displays," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 12–17, 2018.
[31] V. Ferrari, F. Cutolo, E. M. Calabrò, and M. Ferrari, "HMD video see though AR with unfixed cameras vergence," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Munich, Germany, Sep. 2014, pp. 265–266.
[32] K. Matsunaga, T. Yamamoto, K. Shidoji, and Y. Matsuki, "Effect of the ratio difference of overlapped areas of stereoscopic images on each eye in a teleoperation," *Proc. SPIE*, vol. 3957, pp. 236–243, May 2000.
[33] A. State, J. Ackerman, G. Hirota, J. Lee, and H. Fuchs, "Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space," in *Proc. IEEE ACM Int. Symp. Augmented Reality*, New York, NY, USA, Oct. 2001, pp. 137–146.
[34] F. Cutolo, U. Fontana, and V. Ferrari, "Perspective preserving solution for quasi-orthoscopic video see-through HMDs," *Technologies*, vol. 6, no. 1, p. 9, Mar. 2018.
[35] R. A. Akka, "Automatic software control of display parameters for stereoscopic graphics images," *Proc. SPIE*, vol. 1669, pp. 31–38, Jun. 1992.
[36] *Lumus Optical*. Accessed: Sep. 19, 2019. [Online]. Available: http://lumusvision.com/products/oe33/
[37] D. Gadia, G. Garipoli, C. Bonanomi, L. Albani, and A. Rizzi, "Assessing stereo blindness and stereo acuity on digital displays," *Displays*, vol. 35, pp. 206–212, Oct. 2014.
[38] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[39] 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," *J. Cranio-Maxillofacial Surg.*, vol. 42, pp. 1970–1976, Dec. 2014.
[40] F. Cutolo, A. Meola, M. Carbone, S. Sinceri, F. Cagnazzo, E. Denaro, N. Esposito, M. Ferrari, and V. Ferrari, "A new head-mounted display-based augmented reality system in neurosurgical oncology: A study on phantom," *Comput. Assist. Surg.*, vol. 22, pp. 39–53, Dec. 2017.
[41] P. Milgram and M. Krueger, "Adaptation effects in stereo due to on-line changes in camera configuration," *Proc. SPIE*, vol. 1669, pp. 122–134, Jun. 1992.

NADIA CATTARI received the M.Sc. degree in biomedical engineering from the University of Pisa, where she is currently pursuing the Ph.D. degree in clinical and translational science with the EndoCAS Centre. Her research interests include 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.

FABRIZIO CUTOLO (M'17) received the B.Sc. and M.Sc. degrees in electrical and computer engineering and the Ph.D. degree in translational medicine from the University of Pisa, Pisa, Italy, in 2006 and 2015, respectively. He is currently a Postgraduate Research Associate with the Department of Information Engineering, University of Pisa. His research interests include 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).

RENZO D'AMATO received the Ph.D. degree in aerospace engineering from the University of Pisa. He is currently a Postgraduate Research Associate with the Department of Information Engineering, University of Pisa. His research interests include unmanned non-military aerial vehicles, flight simulations, rapid prototyping, and 3D model visualization.

UMBERTO FONTANA received the B.Sc. degree in biomedical engineering from the University of Pisa, where he is currently pursuing the master's degree in robotic and automation engineering. His research interests include vision augmentation, calibration in augmented reality, and computer graphics.

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 the coordinator of the EndoCAS Centre of the University of Pisa. His research interests involve image-guided surgery and simulation, computer vision and augmented reality devices and applications in medicine. He is involved in several national and international research projects. He is the coordinator of the HORIZON2020 project VOSTARS (Call ICT-29-2016).