

Article

Off-Line Camera-Based Calibration for Optical See-Through Head-Mounted Displays

Fabrizio Cutolo ^{1,*}, Umberto Fontana ¹, Nadia Cattari ² and Vincenzo Ferrari ¹

¹ Information Engineering Department, University of Pisa, 56126 Pisa, Italy; fabrizio.cutolo@endocas.unipi.it (F.C.); umbertofontana93@gmail.com (U.F.); vincenzo.ferrari@unipi.it (V.F.)

² EndoCAS Center, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56126 Pisa, Italy; nadia.cattari@endocas.unipi.it (N.C.)

* Correspondence: fabrizio.cutolo@endocas.unipi.it;

Received: 15 December 2019; Accepted: 22 December 2019; Published: 25 December 2019



Featured Application: We here present a completely automated and fast method to calibrate optical see-through head-mounted displays based on the use of a calibrated camera replacing the user's eye. The method can work with any type of optical see-through display and it is easy to replicate in both laboratory environments and real-world settings.

Abstract: In recent years, the entry into the market of self contained optical see-through headsets with integrated multi-sensor capabilities has led the way to innovative and technology driven augmented reality applications and has encouraged the adoption of these devices also across highly challenging medical and industrial settings. Despite this, the display calibration process of consumer level systems is still sub-optimal, particularly for those applications that require high accuracy in the spatial alignment between computer generated elements and a real-world scene. State-of-the-art manual and automated calibration procedures designed to estimate all the projection parameters are too complex for real application cases outside laboratory environments. This paper describes an off-line fast calibration procedure that only requires a camera to observe a planar pattern displayed on the see-through display. The camera that replaces the user's eye must be placed within the eye-motion-box of the see-through display. The method exploits standard camera calibration and computer vision techniques to estimate the projection parameters of the display model for a generic position of the camera. At execution time, the projection parameters can then be refined through a planar homography that encapsulates the shift and scaling effect associated with the estimated relative translation from the old camera position to the current user's eye position. Compared to classical SPAAM techniques that still rely on the human element and to other camera based calibration procedures, the proposed technique is flexible and easy to replicate in both laboratory environments and real-world settings.

Keywords: augmented reality; calibration; head mounted displays; optical see-through display; computer vision

1. Introduction

Visual augmented reality (AR) technology aims to enhance the user's view of the real world by overlaying computer generated elements on it. Currently, optical see-through (OST) head mounted displays (HMDs) are at the leading edge of the AR technology, and they have the potential to become ubiquitous eventually in different fields of applications [1–3].

Nevertheless, their profitable usage across medical and industrial settings is still hampered by the complexity of the display calibration procedures required to ensure accurate spatial alignment

between a real-world scene and computer generated elements [4,5]. The display calibration procedures of consumer level OST devices are rather simplified to improve usability, and this is achieved at the expense of sub-optimal results that are not tolerable for those applications for which the accurate alignment between virtual content and perceived reality is of the utmost importance [6]. This aspect is pushing research towards the realization of standardized OST calibration procedures that are not only flexible and easy to replicate, but that can also provide reliable and accurate results.

Overall, in visual AR applications, the problem of defining the appropriate spatial location of the digital 3D content with respect to the real scene is the principal factor that provides the user with a sense of perceptual congruity [7]. This problem is particularly challenging in OST HMDs, for the solution of which, knowing of the position of the user's viewpoint(s) cannot be overlooked. On a first approximation, a single OST display can be modeled as an off-axis pinhole camera whose imaging plane corresponds to the semi-transparent virtual screen of the display and whose projection center corresponds to the nodal point of the user's eye [8].

The goal of the OST display calibration is therefore to estimate the projection parameters of the combined eye–display system that encapsulates the optical features of the display and whose values vary according to the position of the user's eye with respect to the imaging plane of the display.

Manual calibration procedures rely on user interaction to collect sets of 3D-2D correspondences by aligning, from different viewpoints, world reference points to image points shown on the HMD virtual screen [9]. These methods, particularly when aimed at estimating all the projection parameters simultaneously, are tedious, highly dependent on operator skill, time consuming, and should be repeated every time the HMD moves on the user's head.

To lessen the burden on users in terms of time and workload, the calibration process can then be broken down into two phases: a first phase in which all the eye–display projection parameters are determined by performing a sort of factory calibration, ideally in a controlled setup; a second prior-to-use phase in which the calibration is refined by adjusting just a small subset of projection parameters. This is the underlying rationale behind several simplified manual calibration procedures, also referred to as semi-automatic calibration methods [10].

Finally, interaction-free calibration methods that exploit eye tracking techniques are ideally the preferred option for those AR applications that demand for accurate virtual-to-real spatial alignment over an extended period of time [8,11]. Such is the case of OST HMDs used as an aid to high precision tasks (e.g., for surgical or industrial applications). However, and as we illustrate in more detail in the next section, none of these methods is thoroughly automatic, since they also rely upon an off-line calibration phase performed manually or through a calibrated camera that replaces the user's eye in a controlled setup.

Therefore, irrespective of the approach chosen, the off-line calibration step is paramount to minimize the human element, and for this reason, it ought to be highly accurate and reliable.

This paper presents an approach for performing accurate off-line camera based calibrations of OST HMDs. In our procedure, the projection parameters of the eye–display model are estimated by means of standard camera calibration and computer vision techniques. To this aim, the OST HMD must be mounted over a rigid and adjustable holder equipped with a camera that replaces the user's eye.

The main contributions of our work are as follows:

1. An off-line calibration procedure that is easy to replicate in both laboratory environments and real-world settings.
2. A calibration procedure that can work with any type of OST HMD: with finite and infinite focal distances, with complex or simple optical combiners, featuring inside-out or outside-in tracking mechanisms.
3. A calibration procedure that is completely automated.
4. A calibration procedure that entails a simple on-line refinement step to account for the user's eye position.
5. A detailed formulation of all the mathematical steps involved in the procedure.

The remainder of this paper is structured as follows: Section 2 surveys different two phase approaches for OST HMD calibration. Section 3 provides a general overview of our camera based OST HMD calibration procedure together with its mathematical formulation starting from the standard projection matrix of the off-axis pinhole camera model of the eye–display. Section 4 illustrates the technical implementation of the calibration procedure. Section 5 describes the experiments and discusses the results. Finally, Section 6 concludes with a summary and future work directions.

2. Related Works

Research on how to achieve correct calibration of OST displays has been conducted over many years. In 2017, Grubert et al. [7] presented a comprehensive survey of all the calibration procedures proposed up to that time. In their work, the authors provided also useful insights into the fundamentals of calibration techniques, grouping them into three main categories: manual methods, semi-automatic methods, and automatic methods. We here provide a more focused overview of the different techniques in which the calibration workflow explicitly relies on an off-line phase.

In 2002, Genc et al. [12] proposed a simplified version of the SPAAM (a two phase SPAAM) in which the calibration process was split into two phases: an off-line stage that involves the estimation of the fixed projection parameters and a second on-line stage that aims to update the existing projection matrix partially by applying a 2D image warping of the screen image that includes scaling and shift.

A vision based robust calibration (ViRC) method was proposed in 2013 [13]. The method clearly distinguishes two types of parameters: the ones estimated off-line that are associated with the device dependent parameters and that are based on an approximate position of the user’s eye and those measured on-line that are instead related to the actual position of the user’s eye (4 DoF user dependent parameters). These latter parameters are refined through a perspective-n-point algorithm whose inputs are: the image coordinates of a cross-hair cursor displayed on the see-through screen at different positions and the 3D coordinates of a fiducial marker that the user must visually align to such a cursor. The 3D coordinates of the marker are determined by querying a tracking camera attached to the HMD (i.e., adopting an inside-out tracking approach).

In 2004, Owen et al. [14] presented display relative calibration (DRC), a two phase camera based calibration method. In the first phase of the method, the authors replace the user’s eye with a camera and use a mechanical jig to determine the projection parameters of the display system. In the second phase, the eye–display parameters are optimized on the position of the user’s eye(s) by means of a SPAAM-like procedure.

In 2005, Figl et al. [15] presented a camera based method for calibrating a medical binocular OST HMD, the Varioscope™ M5. The calibration system uses a precision spindle moving a mobile calibration grid via a stepping motor and a tracker probe to localize the moving grid in space with respect to the HMD (i.e., adopting an outside-in tracking approach). However, the proposed method does not have any calibration refinement step that accounts for the user’s eye position.

A similar camera based procedure was presented by Gilson et al. [16] that also employs an inside-out tracking mechanism. After the calibration, the authors did not include any user centered refinement step to account for the eye position, since in their method: “the camera remains stationary by design, so we have constrained our method to minimize for only one camera pose” (page 6). Overall, the authors declared to have achieved an average virtual-to-real alignment accuracy of about two pixels for three camera positions different from the calibration one.

In 2014, Itoh and Klinker [8,17] proposed an automatic method with inside-out tracking that employed an eye tracker to measure the position of the user’s eye. The relative pose between the external camera and the eye tracking camera was predetermined during an off-line calibration session. The dynamic tracking of the user’s eye was used for continuously refining the eye–display projection parameters, whose initial values once again must be computed off-line through a standard SPAAM-like method or a camera based procedure.

A similar approach was proposed in 2015 by Plopski et al. [11,18]. The method refines on-line the pre-determined calibration parameters through a corneal imaging calibration (CIC) procedure that relies on corneal imaging to obtain correspondence pairs from a pre-calibrated HMD screen with its content reflection on the cornea of the user.

In 2016, Zhang et al. [19] presented an interesting two phase method in which a depth camera refined on-line the eye–display projection parameters performing the optimization of the hand registration through a genetic algorithm. Again, a first off-line phase was needed to provide a baseline for the subsequent optimization.

Finally, in 2017, Klemm et al. [20] presented an off-line automated camera based method that resulted in an optimized rendering frustum together with an arbitrary number of distortion maps determined from different viewpoints. Unfortunately, their triangulation approach only worked for OST HMDs with simple optics for which, as stated by the authors on page 57, the “pixel locations are independent from the eye position”.

As anticipated above, all these methods, require an off-line calibration step performed manually or via a camera replacing the user’s eye. This off-line calibration, to be effective and accurate, should be performed in a controlled setup for the following two main reasons: to reduce the errors due to human interaction and to ease and speed-up the subsequent prior-to-use calibration refinement done by the user.

In line with this, this paper presents a camera based OST calibration method capable of accurately estimating the intrinsic matrix of the eye–display pinhole model, and we provide a detailed and substantiated formulation of each step involved in the calibration procedure.

3. Methods

3.1. Rationale

In a previous work [21], we described a closed-loop calibration method specifically suited for OST displays with the focal plane at infinity. This particular optical feature reduces the complexity of the problem of estimating the projection matrix of the off-axis eye–display pinhole model, whose parameters can be modeled irrespective of the specific eye position with a procedure based on a simple homography based 2D warping of the ideal on-axis imaging plane. In this paper, we extend the scope of such a method, so that it can be applied to any kind of OST display (i.e., with finite focal distances), and we model the contribution of the eye position to the projection parameters.

3.2. Notation

The following notation is used throughout the paper. Uppercase letters denote spatial coordinate systems, such as the eye–display coordinate system E. Lowercase letters denote scalar values, such as the focal length f_u . Both 3D and 2D points are represented by homogeneous vectors. Vectors are denoted by lowercase bold letters with a superscript denoting the reference coordinate system (e.g., a 3D point in eye–display coordinates \mathbf{v}^E or a 2D image point in virtual screen coordinates \mathbf{i}^E). Vectors can also be expressed in component form, with a bold subscript indicating the correspondence (e.g., $\mathbf{v}^E = (x_v, y_v, z_v, 1)^E$). Matrices are denoted by uppercase typewriter letters (e.g., the intrinsic camera matrix of the eye–display ${}^E K$).

The 6 DoF transformations from one coordinate system to another are so defined. Given two coordinate systems A and B, the transformation from A to B is defined as ${}^B T = ({}^B R, {}^B t)$ where ${}^B R$ is the rotation matrix and ${}^B t$ is the translation vector. Therefore, we have:

$$\mathbf{v}^B = {}^B R \mathbf{v}^A + {}^B t \quad (1)$$

3.3. The Off-Axis Pinhole Camera Model of the Eye–Display

As mentioned in Section 1, the eye–display system is commonly modeled as an off-axis pinhole camera (i.e., the associated virtual rendering camera) where the nodal point of the user’s eye

corresponds to the center of projection and the imaging plane is the semi-transparent virtual screen. This model describes the projection transformation between the coordinates of a point in the 3D space and the associated 2D point displayed on the imaging plane of the see-through display (Figure 1).

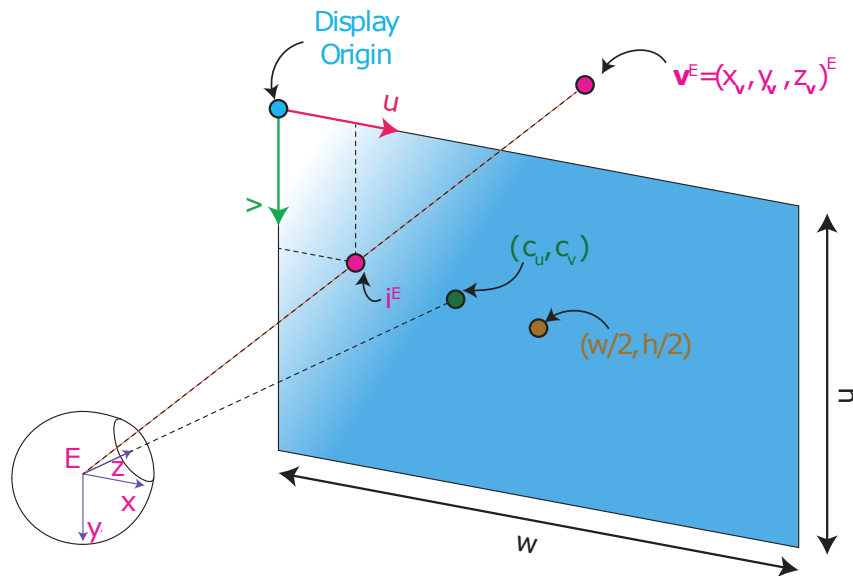


Figure 1. 3D representation of the off-axis eye–display pinhole model comprising the eye as the projection center and the see-through virtual screen as the imaging plane.

The intrinsic matrix of the off-axis eye–display model is:

$${}^{\text{off-E}}\mathbf{K} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (2)$$

where f_u and f_v are the focal lengths in pixels, and they denote the distances between the imaging plane of the display (i.e., the virtual screen) and the pinhole camera projection center (i.e., the nodal point of the user’s eye); c_u and c_v are the coordinates in pixels of the principal point that corresponds to the intersection between the principal axis and the virtual screen.

Using Equation (2), the perspective projection between a 3D point in the eye–display coordinate system \mathbf{v}^E and its associated 2D projection \mathbf{i}^E , both expressed in homogeneous coordinates, is:

$$\lambda \mathbf{i}^E = {}^{\text{off-E}}\mathbf{K} [\mathbf{I}_{3 \times 3} \quad \mathbf{0}_{3 \times 1}] \mathbf{v}^E = {}^E\mathbf{P} \mathbf{v}^E \quad (3)$$

The above formulation assumes a special choice of the world coordinate system W , with $W \hat{=} E$. In more general terms, by plugging the 6 DoF transformation from W to E into Equation (3), we obtain the projection transformation that maps world points onto the imaging plane of the eye–display ${}^E\mathbf{P} \in \mathbb{R}^{3 \times 4}$:

$$\lambda \mathbf{i}^E = {}^{\text{off-E}}\mathbf{K} \begin{bmatrix} {}^E\mathbf{R} & {}^E\mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{v}^W = {}^E\mathbf{P} \mathbf{v}^W \quad (4)$$

The goal of all calibration methods is to provide an accurate estimation of the matrix ${}^E\mathbf{P}$, either by solving it as a whole or by determining each of the four entries of ${}^{\text{off-E}}\mathbf{K}$ and each of the six DoF of ${}^E\mathbf{T}$ individually.

3.4. Formulation of the Calibration Procedure

As mentioned in the Introduction, calibration procedures of OST displays aim to achieve a perfect spatial alignment between a 3D real-world scene and the computer generated scene perceived through the display. To this end, we need to estimate the perspective projection matrix of the off-axis rendering

camera that yields a correct mapping of each 3D vertex \mathbf{v}^W of the virtual object onto the imaging plane space \mathbf{i}^E . This 3D-2D mapping (i.e., process of rasterization) must be done in a way that all 3D vertexes along a ray that originates from the rendering camera center (i.e., the eye center E) is rendered at the same location on the virtual screen and perfectly aligned to their real counterpart. Our calibration procedure measures such a projection matrix by means of a camera that acts as a replacement of the user’s eye. Hereafter, we use the term viewpoint camera to refer to such a camera.

In our formulation, as well as in the implementation of the procedure, we exploit a vision based inside-out tracking technique as done in [14,16,22]. Nonetheless, the same methodological guidelines that we here describe would also apply with an outside-in tracking technique, provided that it measures the pose of the sensor reference system, attached to the HMD, S with respect to the world scene W.

Figure 2 shows the spatial relationships between the coordinate systems associated with the tracking camera S, the internal viewpoint camera C (the user’s eye replacement), and the rendering camera E.

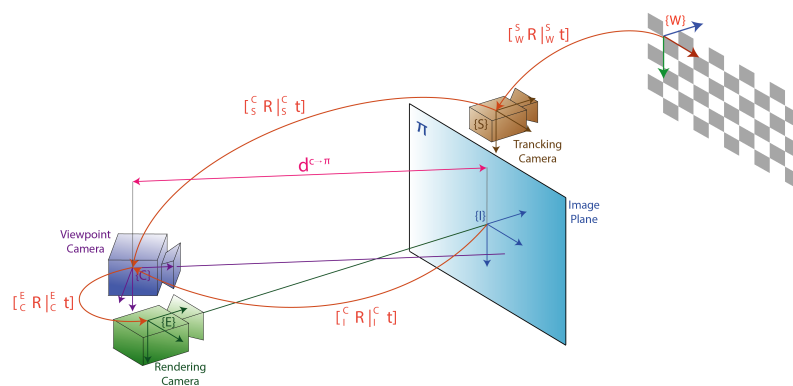


Figure 2. Illustration of the setup for the off-line calibration of the see-through display.

The first step in our calibration procedure is the estimation of the viewpoint camera projection matrix that maps world points \mathbf{v}^W onto camera image points \mathbf{i}^C :

$$\begin{aligned} \lambda \mathbf{i}^C &= {}^C K \begin{bmatrix} {}^C R & {}^C t \\ {}^S R & {}^S t \end{bmatrix} \begin{bmatrix} {}^S R & {}^S t \\ {}^W R & {}^W t \end{bmatrix} \mathbf{v}^W = \\ &= {}^C K \begin{bmatrix} {}^C R & {}^C t \\ {}^W R & {}^W t \end{bmatrix} \mathbf{v}^W = {}^C P \mathbf{v}^W \end{aligned} \tag{5}$$

We determine the intrinsic matrix of the viewpoint camera ${}^C K$ and the relative pose between the viewpoint camera and the tracking camera $\begin{bmatrix} {}^C R & {}^C t \\ {}^S R & {}^S t \end{bmatrix}$ with conventional calibration routines [23] described in more detail in Section 4.2, whereas the pose of the world scene W with respect to S is measured on-line by querying the tracking camera $\begin{bmatrix} {}^S R & {}^S t \\ {}^W R & {}^W t \end{bmatrix}$. We used an OpenCV checkerboard as the target object for the inside-out tracking (see Section 4.1).

Next, the key step in the procedure is the computation of the planar homography ${}^E H$ between the two pinhole cameras C and E that enables the image points of the viewpoint camera \mathbf{i}^C to be mapped onto the imaging plane of the see-through display (i.e., the virtual screen) \mathbf{i}^E .

$$\lambda \mathbf{i}^E = {}^E H \mathbf{i}^C \tag{6}$$

where the planar homography ${}^E H$ can be broken down as follows [24]:

$${}^E H = {}^{on-E} K \left({}^E R + \frac{{}^E \mathbf{t} (n^C)^T}{d^{C \rightarrow \pi}} \right) {}^C K^{-1} \tag{7}$$

In Equation (7), ${}^{\text{on-E}}\mathbf{K}$ is the intrinsic matrix of the ideal on-axis camera model of the eye–display, whose entries are established considering the manufacturer’s specifications of the OST display. Notably, we assumed that the focal lengths on both the x-axis and y-axis were equal ($f_x = f_y$), meaning the display pixels were considered as being perfectly square:

$${}^{\text{on-E}}\mathbf{K} = \begin{bmatrix} \frac{w}{2 \tan(\frac{hfov}{2})} & 0 & w/2 \\ 0 & \frac{h}{2 \tan(\frac{vfov}{2})} & h/2 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

where $\frac{hfov}{2}$ and $\frac{vfov}{2}$ are the horizontal and vertical field-of-view (FOV) of the OST display and w and h are its width and height in pixels.

The other entries of Equation (8) are:

- the rotation matrix between the imaging planes of the viewpoint camera and of the OST display ${}^{\text{E}}\mathbf{R}_C$;
- the virtual translation vector between the viewpoint camera and the rendering camera ${}^{\text{E}}\mathbf{t}$. We also label such a vector as parallax contribution, given that it models the transformation between ideal ${}^{\text{on-E}}\mathbf{K}$ and real ${}^{\text{off-E}}\mathbf{K}_{p_0}$, based on a specific eye position (subscript p_0). This contribution is described in more detail in Section 3.4.3;
- the normal unit vector to the OST display screen in the viewpoint camera reference system n^C ;
- the distance between the viewpoint camera center C and the imaging plane of the OST display $d^{C \rightarrow \pi}$.

In the next two subsections, we explain how we estimate each of the above listed unknown variables.

3.4.1. Estimation of the Imaging Plane Pose with Respect to the Viewpoint Camera

We determine the position ${}^{\text{I}}\mathbf{t}$ and orientation ${}^{\text{I}}\mathbf{R} \equiv {}^{\text{E}}\mathbf{R}$ (i.e., the virtual pose) of the imaging plane of the OST display with respect to the viewpoint camera by computing the location of a calibration checkerboard projected onto the imaging plane of the see-through display (see Section 4).

The physical size of the checkerboard square is established by arbitrarily dictating the distance from the projection center to the imaging plane $d^{E \rightarrow \pi}$ and adopting the intrinsic parameters of the ideal on-axis eye–display model (Equation (8)), as also suggested in [25]. By observing such a checkerboard with the calibrated viewpoint camera, we are able to compute both ${}^{\text{I}}\mathbf{t}$ and ${}^{\text{I}}\mathbf{R}$.

Notably, the translation vector ${}^{\text{I}}\mathbf{t}$ is linearly proportional to $d^{E \rightarrow \pi}$ (${}^{\text{I}}\mathbf{t} \propto d^{E \rightarrow \pi}$).

3.4.2. Estimation of the Rendering Camera Pose with Respect to the Viewpoint Camera

Given the pose of the imaging plane of the OST display with respect to a particular position of the viewpoint camera (i.e., C), we can retrieve the pose of the rendering camera relative to the viewpoint camera by considering the transformation that brings the imaging plane to the focal plane. This pose encapsulates the rotational contribution caused by the different orientations of the rendering and of the viewpoint camera ${}^{\text{E}}\mathbf{R} \equiv {}^{\text{I}}\mathbf{R}$ and the parallax contribution, which is also proportional to the distance from the projection center to the imaging plane (${}^{\text{E}}\mathbf{t} \propto d^{E \rightarrow \pi}$).

Geometrically, it is easy to demonstrate the following relation, valid $\forall C_i$ position (Figure 3):

$$d^{E \rightarrow \pi} = d^{C \rightarrow \pi} + z_C^E = d^{C_i \rightarrow \pi} + z_{C_i}^E \quad (9)$$

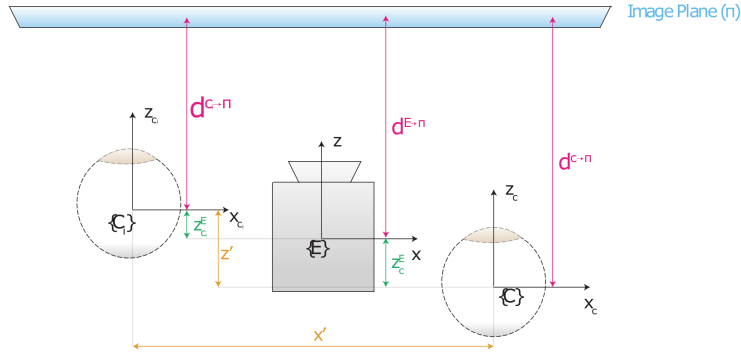


Figure 3. The x-z plane of off-axis pinhole model with the rendering camera and the eye in two different positions.

3.4.3. Meaning of the Parallax Contribution for the Off-Axis Eye–Display Pinhole Model

By plugging Equation (5) into Equation (6), we obtain:

$$\begin{aligned} \lambda \mathbf{i}^E &= {}^E \mathbf{H} \mathbf{i}^C = {}^{\text{on-E}} \mathbf{K} \left({}^E \mathbf{R} + \frac{{}^E \mathbf{t}(n^C)^T}{d^{C \rightarrow \pi}} \right) [{}^C \mathbf{R} \quad {}^C \mathbf{t}] \mathbf{v}^W = \\ &= {}^{\text{on-E}} \mathbf{K} \left(\mathbf{I}_{3 \times 3} + \frac{{}^E \mathbf{t}(n^C)^T}{d^{C \rightarrow \pi}} {}^E \mathbf{R}^{-1} \right) {}^E \mathbf{R} [{}^C \mathbf{R} \quad {}^C \mathbf{t}] \mathbf{v}^W = \\ &= {}^{\text{on-E}} \mathbf{K} \left(\mathbf{I}_{3 \times 3} + \frac{{}^E \mathbf{t}(n^C)^T}{d^{C \rightarrow \pi}} {}^E \mathbf{R}^{-1} \right) [{}^E \mathbf{R} \quad {}^E \mathbf{t}] \mathbf{v}^W \end{aligned} \quad (10)$$

Since $(n^C)^T {}^E \mathbf{R}^{-1} = ({}^E \mathbf{R} n^C)^T = (n^E)^T = (0, 0, 1)$, which is the unit vector of the principal axis of the display, Equation (10) becomes:

$$\lambda \mathbf{i}^E = \overbrace{{}^{\text{on-E}} \mathbf{K} \left(\mathbf{I}_{3 \times 3} + \frac{{}^E \mathbf{t}(n^E)^T}{d^{C \rightarrow \pi}} \right)}^{\mathbf{E} \mathbf{P}} [{}^E \mathbf{R} \quad {}^E \mathbf{t}] \mathbf{v}^W \quad (11)$$

This can be rewritten as:

$$\lambda \mathbf{i}^E = {}^{\text{on-E}} \mathbf{K} \underbrace{\begin{bmatrix} 1 & 0 & x_C^E/d^{C \rightarrow \pi} \\ 0 & 1 & y_C^E/d^{C \rightarrow \pi} \\ 0 & 0 & 1 + z_C^E/d^{C \rightarrow \pi} \end{bmatrix}}_{\text{off-E} \mathbf{K}_{p_0}} [{}^E \mathbf{R} \quad {}^E \mathbf{t}] \mathbf{v}^W \quad (12)$$

$\text{off-E} \mathbf{K}_{p_0}$ can be further processed using Equation (9) and imposing $\text{off-E} \mathbf{K}_{p_0}(3,3) = 1$, given that $\mathbf{E} \mathbf{P}$ is defined up to a scale factor λ :

$$\text{off-E} \mathbf{K}_{p_0} = {}^{\text{on-E}} \mathbf{K} \begin{bmatrix} \frac{d^{C \rightarrow \pi}}{d^{E \rightarrow \pi}} & 0 & \frac{x_C^E}{d^{E \rightarrow \pi}} \\ 0 & \frac{d^{C \rightarrow \pi}}{d^{E \rightarrow \pi}} & \frac{y_C^E}{d^{E \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \quad (13)$$

Equation (13) tells us that, to achieve an accurate spatial alignment between 3D real-world points and computer generated elements rendered by an OST display and observed by a specific viewpoint (p_0), each virtual 3D vertex v^W must be observed by an off-axis rendering camera whose projection matrix ${}^E P$ is modeled as follows:

- The extrinsic parameters of the camera are: $[\begin{smallmatrix} {}^E R \\ {}^E t \end{smallmatrix}]$. Unsurprisingly, the center of projection of the virtual rendering camera is the user’s eye (in our procedure, the viewpoint camera), whereas its orientation matches the display virtual screen’s orientation.
- The intrinsic parameters of the off-axis camera are obtained by applying a homographic transformation to the ideal on-axis intrinsic matrix:

$${}_{\text{on-}E}^{\text{off-}E} H_{p_0} = H_{p_0} = \begin{bmatrix} \frac{d^{C \rightarrow \pi}}{d^{E \rightarrow \pi}} & 0 & \frac{x_C^E}{d^{E \rightarrow \pi}} \\ 0 & \frac{d^{C \rightarrow \pi}}{d^{E \rightarrow \pi}} & \frac{y_C^E}{d^{E \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

This homography allows computing the intrinsic matrix of the real off-axis eye–display system, and it encapsulates the shift and scaling effect (i.e., the parallax contribution) due to a particular position of the user’s eye.

The intrinsic matrix for a different viewpoint position p_i is:

$${}_{\text{on-}E}^{\text{off-}E} K_{p_i} = {}_{\text{on-}E} K = \begin{bmatrix} \frac{d^{C_i \rightarrow \pi}}{d^{E \rightarrow \pi}} & 0 & \frac{x_{C_i}^E}{d^{E \rightarrow \pi}} \\ 0 & \frac{d^{C_i \rightarrow \pi}}{d^{E \rightarrow \pi}} & \frac{y_{C_i}^E}{d^{E \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \tag{15}$$

Therefore, from simple algebraic manipulations, we can obtain the transformation matrix that enables us to pass from ${}_{\text{on-}E}^{\text{off-}E} K_{p_0}$ to ${}_{\text{on-}E}^{\text{off-}E} K_{p_i}$:

$$\begin{aligned} {}_{\text{on-}E}^{\text{off-}E} K_{p_i} &= {}_{\text{on-}E}^{\text{off-}E} K_{p_0} H_{p_0}^{-1} H_{p_i} = \\ &= {}_{\text{on-}E}^{\text{off-}E} K_{p_0} \begin{bmatrix} \frac{d^{C_i \rightarrow \pi}}{d^{C \rightarrow \pi}} & 0 & \frac{x_{C_i}^E - x_C^E}{d^{C \rightarrow \pi}} \\ 0 & \frac{d^{C_i \rightarrow \pi}}{d^{C \rightarrow \pi}} & \frac{y_{C_i}^E - y_C^E}{d^{C \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{16}$$

From this, we derive the same relation presented in [8] and in [7]:

$${}_{\text{on-}E}^{\text{off-}E} K_{p_i} = {}_{\text{on-}E}^{\text{off-}E} K_{p_0} \begin{bmatrix} 1 + \frac{z'}{d^{C \rightarrow \pi}} & 0 & \frac{-x'}{d^{C \rightarrow \pi}} \\ 0 & 1 + \frac{z'}{d^{C \rightarrow \pi}} & \frac{-y'}{d^{C \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

where ${}^C_i \mathbf{t} = [x', y', z']$ is the translation from the old viewpoint position to the new viewpoint position and where $d^{C_i \rightarrow \pi} = d^{C \rightarrow \pi} + z'$.

3.4.4. Relation between the Optical Properties of the OST Display and Intrinsic Matrix of the Eye–Display Model

In any HMD, the role of the collimation optics is to display the microdisplay image so that it appears magnified at a comfortable viewing distance [26]. Specifically, the imaging plane of many consumer level OST HMDs is at infinity (i.e., at a very far distance), and this happens if the microdisplay is located in close proximity of the focal point of the collimation optics. In this case, the light rays coming from each pixel of the microdisplay are arranged in a parallel pattern, and the virtual image perceived by the user, to a first approximation, is not subjected to any shift or scaling effect for different positions of the user’s eye within the display eye-motion-box area.

Thereby, for those OST displays, the position of the user’s eye does not influence the intrinsic matrix of the off-axis eye–display pinhole camera model ${}^{off-E}K_p$, and this implies that the intrinsic linear parameters of the eye–display can be considered as parallax-free:

$${}^{off-E}K_p = {}^{off-E}K_{p_0} \begin{bmatrix} 1 + \frac{\overset{\approx 0}{z'}}{d^{C \rightarrow \pi}} & 0 & \frac{\overset{\approx 0}{-x'}}{d^{C \rightarrow \pi}} \\ 0 & 1 + \frac{\overset{\approx 0}{z'}}{d^{C \rightarrow \pi}} & \frac{\overset{\approx 0}{-y'}}{d^{C \rightarrow \pi}} \\ 0 & 0 & 1 \end{bmatrix} \approx {}^{off-E}K_{p_0} \tag{18}$$

since $x' \setminus y' \setminus z' \ll d^{C \rightarrow \pi}$.

In this case, the on-line refinement step accounting for the user’s eye position is not needed. Differently, for those OST displays with shorter focal distances (e.g., Microsoft HoloLens, Magic Leap, Epson Moverio BT-200), the light rays within the eye-motion-box of the display are not parallel. For this reason, the contribution of the position of the user’s eye to the intrinsic matrix cannot be neglected, and the update contribution expressed by Equation (17) must be taken into consideration (i.e., the intrinsic matrix is not parallax-free).

4. Technical Implementation of the Procedure

4.1. Hardware Setup

Our calibration procedure is applicable to any OST HMD, with finite or infinite focal distance, based on waveguides or on large spherical or semispherical optical combiners.

We tested our method on a commercial binocular OST HMD (ARS.30 by Trivisio [27]) appropriately reworked to embody a vision based inside-out tracking mechanism. The visor was provided with dual SXGA OLED panels with $1280 \times 101,024$ resolution and with a pair of standard flat optical combiners tilted at about 45° . The two panels were controlled via HDMI. The diagonal FOV of the HMD was of 30° , which corresponded to ≈ 1.11 arcmin/pixel angular resolution. The eye-relief of the display was 30 mm, and the size of the eye-motion-box was 8×10 mm. The imaging plane of the OST display was projected at 500 mm ($d^{E \rightarrow \pi} = 500$ mm).

Following a similar approach to our previous research works [28,29], we embedded the HMD in a 3D printed plastic shell; the plastic shell was designed to house a pair of liquid crystal (LC) shutters and a pair of stereo RGB cameras (Leopard Imaging LI-OV4689) for the inside-out tracking

mechanism (Figure 4). The LC panels were placed in front of the optical combiners of the HMD, and they could be electronically controlled so as to modify the transparency of the see-through display and switch from see-through state to occluded state. We used the occluded modality for the second step of the calibration, in which we projected a calibration checkerboard onto the imaging plane of the see-through display.

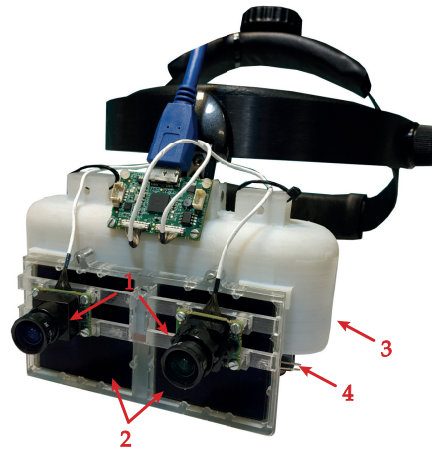


Figure 4. The custom-made hybrid video-optical see-through head mounted display used in the experimental session. 1→Pair of stereo cameras for the inside-out tracking. 2→Pair of liquid-crystal optical shutters to drive the switching mechanism between occluded to optical see-through state. 3→Plastic shell that incorporates all the components of the optical see-through visor. 4→Optical combiner of the optical see-through visor.

Figure 5 shows the experimental setup. In our procedure, we used a single camera inside-out tracking mechanism. The tracking cameras had 109° diagonal FOV, which, was associated with a 1280×720 image resolution, yielding ≈ 4.8 arcmin/pixel angular resolution. As we will discuss in Section 5, this particular feature enabled us to have a reference limit in the evaluation of the overall calibration accuracy.

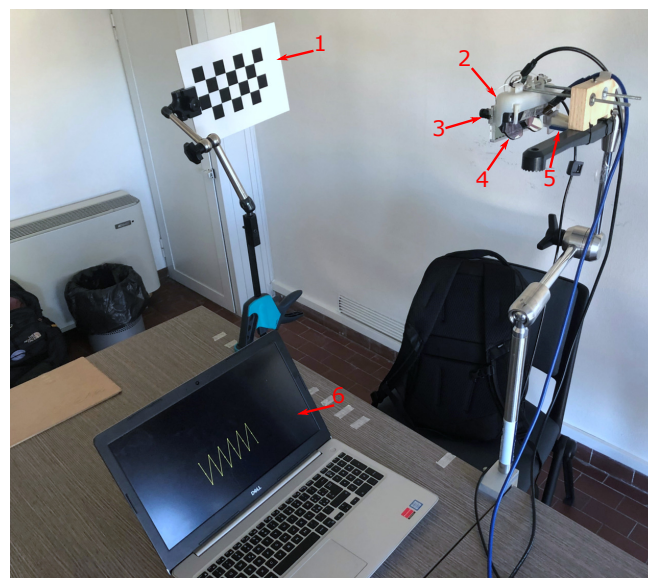


Figure 5. Experimental environment. 1→Validation checkerboard. 2→Optical see-through head mounted display. 3→Tracking camera. 4→Optical combiner. 5→3D printed mounting template with the embedded viewpoint camera. 6→Laptop where the calibration routines and the AR application run.

In our tests, we performed the calibration on the right display. Therefore, we placed the viewpoint camera behind the right display at approximately the eye-relief distance from the flat combiner (i.e., ≈ 30 mm).

The viewpoint camera was a SONY FCB-MA130, which had a $1/2.45''$ CMOS sensor, a 1280×720 resolution, and a 59° diagonal FOV, which corresponded to ≈ 2.67 arcmin/pixel.

Both the HMD and the validation checkerboard were locked to two rigid and adjustable holders. The validation checkerboard was a standard 7×4 OpenCV calibration checkerboard with a square size of 30 mm. The checkerboard was placed at approximately 650 mm of distance from the viewpoint camera. We attached the viewpoint camera to a 3D printed mounting template. The mounting template was equipped with fixing holes for placing the camera in eight pre-set positions radially arranged within the eye-motion-box of the see-through display (Figure 6). The template and the camera were both anchored to the translation bar of the HMD holder.

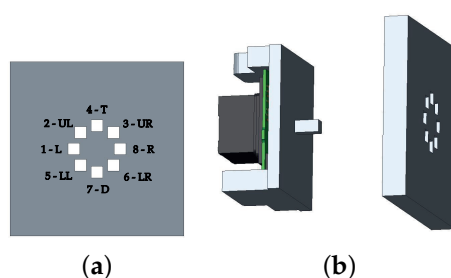


Figure 6. 3D CAD of the mounting template used for placing the viewpoint camera in eight different positions within the eye-motion-box of the see-through display. (a) The eight fixing holes for camera positioning. (b) The mounting template.

We calibrated the display for the left position (reference position), and then, we computed the intrinsic matrix of the eye–display for the remaining seven positions using Equation (17), where $d^{C \rightarrow \pi}$ is given by Equation (9) and where the translation vector ${}^C_i \mathbf{t}$ is computed using two stereo calibrations between the tracking camera (fixed) and the viewpoint camera in the reference and in the new position:

$${}^C_i \mathbf{t} = [x', y', z'] = {}^S_i \mathbf{t} - {}^S_1 \mathbf{t} \quad (19)$$

4.2. Calibration Software

As outlined in Section 3.4, the calibration workflow was broken down into four main steps, each of which was associated with a different software routine:

- Estimation of the intrinsic camera parameters of both the viewpoint camera and the tracking camera (intrinsic camera calibrations).
- Estimation of the pose of the tracking camera with respect to the viewpoint camera (heterogeneous stereo calibration).
- Estimation of the pose of the rendering camera with respect to the viewpoint camera (parallax contribution estimation).
- Estimation of the final eye–display projection matrix (eye–display matrix calculation).

The calibration workflow is depicted in Figure 7. The details of the calibration steps are presented in the next subsections.

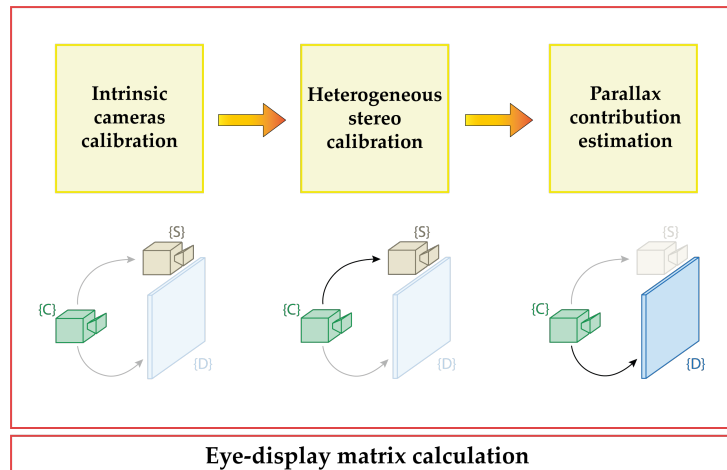


Figure 7. Conceptual schematic of the optical see-through display calibration procedure.

4.2.1. Intrinsic Camera Calibrations

The viewpoint camera and the tracking camera were calibrated with a conventional calibration technique [23] that required storing multiple camera views of a planar pattern (i.e., OpenCV checkerboard). Linear parameters (i.e., intrinsic camera matrix) and non-linearities due to camera lens distortion (i.e., distortion parameters) were computed using non-linear least-squares minimization (i.e., the Levenberg–Marquardt algorithm). This procedure was performed using the MATLAB camera calibration toolbox (R2018b MathWorks, Inc., Natick, MA, USA).

4.2.2. Heterogeneous Stereo Calibration

The relative pose between the viewpoint camera and the tracking camera was estimated through a stereo calibration routine specific for two different camera types. This algorithm was developed in C++ under the Linux Operating System (Ubuntu 16.04) and using the OpenCV API 3.3.1 [30].

4.2.3. Parallax Contribution and Eye–Display Matrix Estimation

The output of the heterogeneous stereo calibration algorithm, together with the intrinsic parameters of the two physical cameras and the manufacturer’s ideal projection parameters of the display were all fed into the final calibration routine. This routine was developed in MATLAB.

As described in Section 3.4, the pose of the OST display imaging plane with respect to the viewpoint camera was determined as follows: a calibration checkerboard was projected onto the imaging plane of the occluded see-through display in full-screen modality and acquired by the viewpoint camera; the virtual pose of the imaging plane was determined through a standard perspective-n-point algorithm [31]. These final data enabled us to compute the projection parameters of the off-axis eye–display model.

5. Experiments and Results

A dedicated software application was developed in MATLAB to validate the accuracy of the calibration technique. In the application, we generated a virtual scene whose virtual viewpoint (i.e., rendering camera) was controlled according to the extrinsic and the intrinsic parameters of the eye–display model determined from the calibration. In this way, we created an AR OST visualization, whose overlay accuracy could be considered as an objective evaluation metric for the calibration procedure. We therefore elaborated the image frames of the viewpoint camera and measured the overlay error, or reprojection error, between real and virtual features.

As ground truth real features, we considered the corners of the validation checkerboard, whereas as virtual features, we used virtual circles. The overlay error was computed as the Euclidean distance (o_{2D}) between the image coordinates of the real and virtual features (Figure 8).

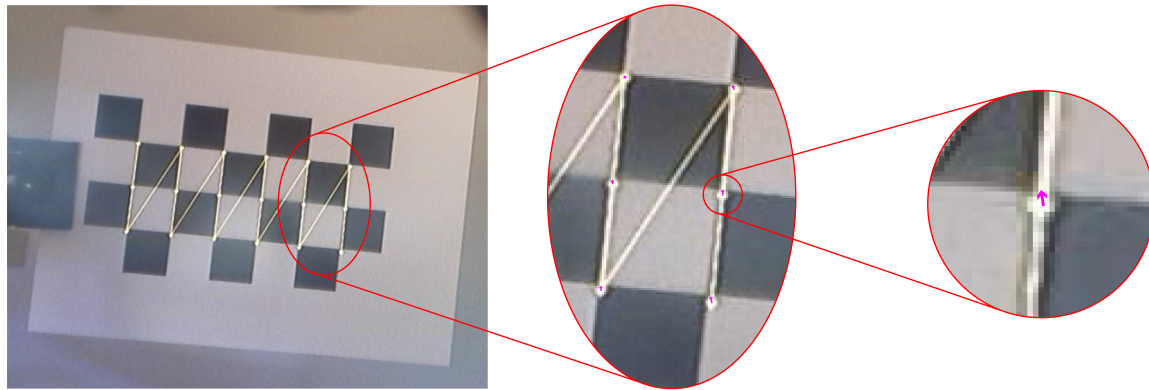


Figure 8. Image grabbed from the viewpoint camera after the calibration. The accuracy of the calibration is measured as the overlay error between the ground truth real corners of the validation checkerboard and the centers of the virtual yellow circles. The zoomed details of the image shows the on-image displacements between the real and the virtual landmarks. Red arrows indicate the direction and magnitude of overlay error in pixels.

We exploited the hybrid nature of the HMD, which can work both under OST and the video see-through modality [28], to measure the overlay error as follows:

- The image coordinates of the checkerboard corners were extracted by processing a frame of the viewpoint camera acquired with both the display and the optical shutter turned off. In order to do this, we used the MATLAB corner detection algorithm (Figure 9b).
- The image coordinates of the virtual landmarks were extracted by processing the viewpoint camera image without the real-world background, hence with both the display and the optical shutter turned on. The coordinates of the centroids of the virtual circles were retrieved using a semi-automatic centroid detection algorithm (Figure 9c).

5.1. Quantitative Evaluation of Virtual-to-Real Registration Accuracy

Quantitative results are presented in terms of the average value, standard deviation, and max value of the overlay error (o_{2D}^{err}) onto the imaging plane of the viewpoint camera over the eight positions on which the calibration was performed. Similarly, we also measured the values of the subtended angular error (α^{err}). Then, knowing the overlay error in pixels and given the distance z_C^g and the focal length of the viewpoint camera f_C , we also computed the associated absolute error in mm at the validation checkerboard plane by using the following relation: [32,33]:

$$a_{3D}^{err} \approx \frac{o_{2D}^{err} z_C^g}{f_C} \tag{20}$$

5.2. Results and Discussion

Table 1 shows the results of the eight calibrations in terms of overlay error (o_{2D}^{err}), angular error (α^{err}), and absolute error (a_{3D}^{err}). The overall mean, standard deviation, and max values of o_{2D}^{err} were 2.23, 0.96, and 5.03 px; the same values for α^{err} were 5.98, 2.58, and 13.47 arcmin and for a_{3D}^{err} 1.13, 0.49, and 2.55 mm.

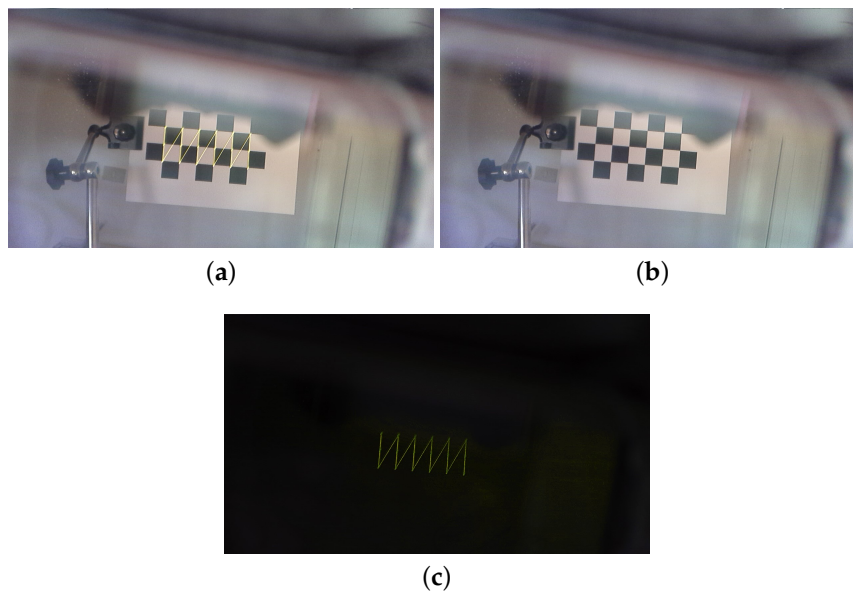


Figure 9. Viewpoint camera frames. (a) Camera frame of the augmented scene. (b) Camera frame of the real-world scene. (c) Camera frame of the virtual scene.

Table 1. Results of the accuracy analysis in pixels (overlay error), arcmin (angular error), and mm (absolute error).

| Camera Position | Overlay Error (pixel) | | | Angular Error (arcmin) | | | Absolute Error (mm) | | |
|-----------------|-----------------------|---------|------|------------------------|---------|-------|---------------------|---------|------|
| | Mean | Std Dev | Max | Mean | Std Dev | Max | Mean | Std Dev | Max |
| 1-L | 1.08 | 0.52 | 1.98 | 2.88 | 1.39 | 5.3 | 0.55 | 0.26 | 1 |
| 2-UL | 1.43 | 0.78 | 3.23 | 3.82 | 2.08 | 8.66 | 0.72 | 0.39 | 1.64 |
| 3-UR | 2.03 | 0.87 | 3.48 | 5.43 | 2.32 | 9.33 | 1.03 | 0.44 | 1.76 |
| 4-T | 2.06 | 0.58 | 3.08 | 5.52 | 1.56 | 8.25 | 1.04 | 0.3 | 1.56 |
| 5-LL | 2.74 | 0.82 | 4.34 | 7.35 | 2.2 | 11.61 | 1.39 | 0.42 | 2.19 |
| 6-LR | 3.29 | 0.6 | 5.03 | 8.82 | 1.6 | 13.47 | 1.67 | 0.3 | 2.55 |
| 7-D | 2.61 | 0.66 | 4 | 6.99 | 1.76 | 10.72 | 1.32 | 0.33 | 2.03 |
| 8-R | 2.61 | 0.67 | 4.05 | 6.99 | 1.8 | 10.85 | 1.32 | 0.34 | 2.05 |

The average overlay error for the reference position was comparable to that obtained by Owen et al. [14] and by Gilson et al. [16], whereas for the remaining positions, the average overlay error was comparable to that obtained in [9] and in [8]. It should be also noted that the results of the angular error are reported considering the rather low angular resolution of the viewpoint camera (≈ 2.67 arcmin/pixel). This explained the values of the angular errors obtained. If we considered the absolute error at the distance where the validation checkerboard was placed (≈ 650 mm), our results were comparable to those obtained with the camera based triangulation approach proposed by Klemm et al. [20]. The experiment results suggested that the calibration refinement step accounting for the user's eye position was paramount to achieve accurate results also for eye positions different from the reference one for all those OST HMDs with finite focal distances (i.e., not parallax-free OST HMDs). This was in accordance with what was suggested by Owen et al. [14] and contrary to what was hypothesized by Gilson et al. and by [34] et al. It should be also stressed that this assumption was clearly different from what was suggested in [20] (page 58), where the classification of the OST HMDs was made based on the level of complexity of the OST optical combiner: "these results show that the triangulation approach only works for OSTG with simple optics and that it is not a universal approach".

Overall, there were at least two sources of calibration inaccuracies in our method, both of which could be easily addressed in the future. The first one was not due to the calibration procedure per se, but it was due to the intrinsic tracking accuracy of the inside-out mechanism adopted in our study, which, as anticipated, was affected by the angular resolution of the tracking camera (i.e., ≈ 4.8 arcmin/pixel). This choice was dictated by the need to offer sufficient stereo overlap between the two tracking cameras, albeit at the expense of a reduced pixel density. It should be noted that this value of angular resolution had a detrimental effect on the calibration results as it increased the uncertainty in estimating the position of the validation checkerboard. This measure was used both for calibrating the eye–display model for the reference position of the viewpoint camera and also for computing the translation vector \mathbf{c}_t . We believe that in the future, the accuracy of the off-line phase of the calibration would benefit from the selection of tracking cameras with a higher value of angular resolution, achievable by selecting cameras with a higher image resolution or with longer focus lenses.

A second source of inaccuracy arose from not considering, in our solution, the contribution of the display optical distortion. Indeed, the collimation optics of any OST HMDs, despite being essential to provide a reasonable FOV and a comfortable viewing distance for the virtual image, is a source of optical distortions such as radial and tangential distortions [35]. The same can be said for the optical combiner, particularly for waveguide based OST displays [20]. These optical distortions not only affect the shapes of the virtual objects perceived by the user, but they also distort the pattern of binocular disparities between the left and right virtual images, and therefore, they alter the stereoscopic perception of the virtual content [36]. To counter this problem, we are currently working on a solution that integrates a non-linear optimization step in the calibration algorithm. This further step can enable us to estimate the optical distortion parameters (radial and tangential). By doing so, we can correctly compensate for the optical distortions through a predistortion technique based on a 2D non-linear mapping of the virtual image on the imaging plane, as done in [25].

In order for the distortion correction, as well as the transformation (17) to be efficient and real time, a GPU based texture mapping approach (i.e., late warping) should be also adopted [37]. This mechanism would require the use of an eye tracking camera. On account of this, our calibration procedure potentially can be integrated in the low level rendering mechanism of commercial OST HMDs to encompass the linear and non-linear 2D warping of the ideal imaging plane yielded by the intrinsic optical properties of the display and by the user's eye position [38].

6. Conclusions and Future Work

In this paper, we presented a method to easily calibrate OST HMDs based on the use of a calibrated camera replacing the user's eye. The method, which could work with any type of OST HMD, only required the camera to observe a planar calibration pattern displayed on the imaging plane of the see-through display.

The method exploited standard camera calibration and computer vision techniques for estimating the projection parameters of the off-axis display model for a generic position of the viewpoint camera. In a prior-to-use second phase, the projection parameters were then refined through a planar homography that encapsulated the shift and scaling effect (i.e., the parallax contribution) associated with the relative translation from the old camera position to the current user's eye position.

To evaluate the efficacy of the calibration technique objectively, we measured the overlay error between real and virtual features observed by the viewpoint camera from different positions within the eye-motion-box of the display. Experimental results indicated that the average overlay error of our calibration was 2.23 pixel.

Our future work involves the integration of an additional step in the off-line calibration procedure devoted to the estimation of the radial and tangential distortion coefficients due to the lens system of the OST display. This calibration step can be performed through a non-linear optimization method such as the Levenberg–Marquardt algorithm.

We are also working on the integration of an eye tracking camera with our custom made OST HMD to perform user studies to evaluate the efficacy of the run-time refinement of the linear and non-linear calibration parameters and to design possible algorithm improvements that take into consideration the eye model.

Finally, the flexibility of our camera based calibration procedure can pave the way toward the profitable use of consumer level OST HMDs also in those applications that demand high accuracy in the spatial alignment between computer generated elements and the real-world scene (e.g., in surgical or high precision industrial applications).

Author Contributions: Conceptualization, F.C.; methodology, F.C., U.F., and N.C.; software, F.C., U.F., and N.C.; validation, F.C., U.F., N.C., and V.F.; formal analysis, F.C.; investigation, F.C. and U.F.; resources, F.C. and V.F.; data curation, U.F. and N.C.; writing, original draft preparation, F.C.; writing, review and editing, F.C., U.F., and N.C.; visualization, F.C.; supervision, F.C. and V.F.; project administration, V.F.; funding acquisition, V.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the HORIZON2020 Project VOSTARS (Video-Optical See Through AR surgical System), Project ID: 731974. Call: ICT-29-2016 Photonics KET2016.

Acknowledgments: The authors would like to thank Renzo D’Amato for his support in designing and assembling the AR visor.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; nor in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|--------------------------------------|
| AR | Augmented reality |
| HMD | Head mounted display |
| OST | Optical see-through |
| SPAAM | Single point active alignment method |
| FOV | Field-of-view |

References

1. Liu, H.; Auvinet, E.; Giles, J.; Rodriguez y Baena, F. Augmented Reality Based Navigation for Computer Assisted Hip Resurfacing: A Proof of Concept Study. *Ann. Biomed. Eng.* **2018**, *46*, 1595–1605, doi:10.1007/s10439-018-2055-1.
2. Chen, L.; Day, T.W.; Tang, W.; John, N.W. Recent Developments and Future Challenges in Medical Mixed Reality. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2017), Nantes, France, 9–13 October 2017; pp. 123–135, doi:10.1109/ISMAR.2017.29.
3. Zhou, F.; Duh, H.B.; Billingham, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In Proceedings of the 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008), Cambridge, UK, 14–18 September 2008; pp. 193–202, doi:10.1109/ISMAR.2008.4637362.
4. Qian, L.; Azimi, E.; Kazanzides, P.; Navab, N. Comprehensive tracker based display calibration for holographic optical see-through head-mounted display. *arXiv* **2017**, arXiv:1703.05834.
5. Cutolo, F. Letter to the Editor on “Augmented Reality Based Navigation for Computer Assisted Hip Resurfacing: A Proof of Concept Study”. *Ann. Biomed. Eng.* **2019**, *47*, 2151–2153, doi:10.1007/s10439-019-02299-w.
6. Jones, J.A.; Edewaard, D.; Tyrrell, R.A.; Hodges, L.F. A schematic eye for virtual environments. In Proceedings of the 2016 IEEE Symposium on 3D User Interfaces (3DUI 2016), Greenville, SC, USA, 19–20 March 2016; pp. 221–230, doi:10.1109/3DUI.2016.7460055.
7. Grubert, J.; Itoh, Y.; Moser, K.; Swan, J.E. A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 2649–2662, doi:10.1109/TVCG.2017.2754257.

8. Itoh, Y.; Klinker, G. Interaction-free calibration for optical see-through head-mounted displays based on 3D Eye localization. In Proceedings of the 2014 IEEE Symposium on 3D User Interfaces (3DUI 2014), Minneapolis, MN, USA, 19–20 March 2014; pp. 75–82, doi:10.1109/3DUI.2014.6798846.
9. Tuceryan, M.; Navab, N. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. In Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), Munich, Germany, 5–6 October 2000; pp. 149–158, doi:10.1109/ISAR.2000.880938.
10. Navab, N.; Zokai, S.; Genc, Y.; Coelho, E.M. An on-line evaluation system for optical see-through augmented reality. In Proceedings of the IEEE Virtual Reality 2004 (IEEE VR 2004), Chicago, IL, USA, 27–31 March 2004; pp. 245–246, doi:10.1109/VR.2004.1310091.
11. Plopski, A.; Itoh, Y.; Nitschke, C.; Kiyokawa, K.; Klinker, G.; Takemura, H. Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays. *IEEE Trans. Vis. Comput. Graph.* **2015**, *21*, 481–490, doi:10.1109/TVCG.2015.2391857.
12. Genc, Y.; Tuceryan, M.; Navab, N. Practical solutions for calibration of optical see-through devices. In Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002), Darmstadt, Germany, 30 September–1 October 2002; pp. 169–175, doi:10.1109/ISMAR.2002.1115086.
13. Makibuchi, N.; Kato, H.; Yoneyama, H. Vision based robust calibration for optical see-through head-mounted displays. In Proceedings of the IEEE International Conference on Image Processing (ICIP 2013), Melbourne, Australia, 15–18 September 2013; pp. 2177–2181, doi:10.1109/ICIP.2013.6738449.
14. Owen, C.B.; Zhou, J.; Tang, A.; Xiao, F. Display-relative calibration for optical see-through head-mounted displays. In Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), Arlington, VA, USA, 2–5 November 2004; pp. 70–78, doi:10.1109/ISMAR.2004.28.
15. Figl, M.; Ede, C.; Hummel, J.; Wanschitz, F.; Ewers, R.; Bergmann, H.; Birkfellner, W. A fully automated calibration method for an optical see-through head-mounted operating microscope with variable zoom and focus. *IEEE Trans. Med Imaging* **2005**, *24*, 1492–1499, doi:10.1109/TMI.2005.856746.
16. Gilson, S.J.; Fitzgibbon, A.W.; Glennerster, A. Spatial calibration of an optical see-through head-mounted display. *J. Neurosci. Methods* **2008**, *173*, 140–146, doi:10.1016/j.jneumeth.2008.05.015.
17. Itoh, Y.; Klinker, G. Performance and sensitivity analysis of INDICA: INteraction-Free Display CALibration for Optical See-Through Head-Mounted Displays. In Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2014), Munich, Germany, 10–12 September 2014.
18. Plopski, A.; Orlosky, J.; Itoh, Y.; Nitschke, C.; Kiyokawa, K.; Klinker, G. Automated Spatial Calibration of HMD Systems with Unconstrained Eye-cameras. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2016), Merida, Mexico, 19–23 September 2016; pp. 94–99, doi:10.1109/ISMAR.2016.16.
19. Zhang, Z.; Weng, D.; Liu, Y.; Yongtian, W. A Modular Calibration Framework for 3D Interaction System Based on Optical See-Through Head-Mounted Displays in Augmented Reality. In Proceedings of the 2016 International Conference on Virtual Reality and Visualization (ICVRV), Tokyo, Japan, 11–13 May 2016; pp. 393–400, doi:10.1109/ICVRV.2016.72.
20. Klemm, M.; Seebacher, F.; Hoppe, H. High accuracy pixel-wise spatial calibration of optical see-through glasses. *Comput. Graph.* **2017**, *64*, 51–61, doi:10.1016/j.cag.2017.02.001.
21. Fontana, U.; Cutolo, F.; Cattari, N.; Ferrari, V. Closed - Loop Calibration for Optical See-Through Near Eye Display with Infinity Focus. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct 2018), Munich, Germany, 16–20 October 2018; pp. 51–56, doi:10.1109/ISMAR-Adjunct.2018.00033.
22. Genc, Y.; Tuceryan, M.; Khamene, A.; Navab, N. Optical see-through calibration with vision based trackers: propagation of projection matrices. In Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), New York, NY, USA, 29–30 October 2001; pp. 147–156, doi:10.1109/ISAR.2001.970524.
23. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334, doi:10.1109/34.888718.
24. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge books online; Cambridge University Press: Cambridge, UK, 2003.
25. Lee, S.; Hua, H. A Robust Camera-Based Method for Optical Distortion Calibration of Head-Mounted Displays. *J. Disp. Technol.* **2015**, *11*, 845–853, doi:10.1109/JDT.2014.2386216.

26. Holliman, N.S.; Dodgson, N.A.; Favalora, G.E.; Pockett, L. Three-Dimensional Displays: A Review and Applications Analysis. *IEEE Trans. Broadcast.* **2011**, *57*, 362–371, doi:10.1109/TBC.2011.2130930.
27. Trivisio, Lux Prototyping sarl. Available online: <https://www.trivisio.com/> (accessed on 24 December 2019).
28. Cutolo, F.; Fontana, U.; Carbone, M.; D'Amato, R.; Ferrari, V. [POSTER] Hybrid Video/Optical See-Through HMD. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct 2017), Nantes, France, 9–13 October 2017; pp. 52–57, doi:10.1109/ISMAR-Adjunct.2017.31.
29. Cutolo, F.; Fontana, U.; Ferrari, V. Perspective Preserving Solution for Quasi-Orthoscopic Video See-Through HMDs. *Technologies* **2018**, *6*, 9, doi:10.3390/technologies6010009.
30. OpenCV, Open Source Computer Vision library. Available online: <https://opencv.org/> (accessed on 24 December 2019).
31. Gao, X.-S.; Hou, X.-R.; Tang, J.; Cheng, H.-F. Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 930–943, doi:10.1109/TPAMI.2003.1217599.
32. Müller, M.; Rassweiler, M.C.; Klein, J.; Seitel, A.; Gondan, M.; Baumhauer, M.; Teber, D.; Rassweiler, J.J.; Meinzer, H.P.; Maier-Hein, L. Mobile augmented reality for computer-assisted percutaneous nephrolithotomy. *Int. J. Comput. Assist. Radiol. Surg.* **2013**, *8*, 663–675, doi:10.1007/s11548-013-0828-4.
33. Ferrari, V.; Vigliani, R.M.; Nicoli, P.; Cutolo, F.; Condino, S.; Carbone, M.; Siesto, M.; Ferrari, M. Augmented reality visualization of deformable tubular structures for surgical simulation. *Int. J. Med Robot. Comput. Assist. Surg.* **2016**, *12*, 231–240, doi:10.1002/rcs.1681.
34. Oishi, T.; Tachi, S. Methods to Calibrate Projection Transformation Parameters for See-through Head-mounted Displays. *Presence Teleoper. Virtual Environ.* **1996**, *5*, 122–135, doi:10.1162/pres.1996.5.1.122.
35. Robinett, W.; Rolland, J.P. Computational model for the stereoscopic optics of a head-mounted display. In *Stereoscopic Displays and Applications II*; Merritt, J.O., Fisher, S.S., Eds.; International Society for Optics and Photonics: Bellingham, WA, USA, 1991; Volume 1457; pp. 140–160, doi:10.1117/12.46303.
36. Cattari, N.; Cutolo, F.; D'amato, R.; Fontana, U.; Ferrari, V. Toed-in vs Parallel Displays in Video See-Through Head-Mounted Displays for Close-Up View. *IEEE Access* **2019**, *7*, 159698–159711, doi:10.1109/ACCESS.2019.2950877.
37. Watson, B.A.; Hodges, L.F. Using texture maps to correct for optical distortion in head-mounted displays. In Proceedings of the Virtual Reality Annual International Symposium '95 (VRAIS '95), Research Triangle Park, NC, USA, 11–15 March 1995; pp. 172–178, doi:10.1109/VRAIS.1995.512493.
38. Bax, M.R. Real-time lens distortion correction: 3D video graphics cards are good for more than games. *Stanford ECJ* **2004**, *1*, 9–13, doi:10.1109/ITSC.2003.1252685.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).