Assessment of Optical See-Through Head Mounted Display Calibration for Interactive Augmented Reality

Giorgio Ballestin
giorgio.ballestin@dibris.unige.it

Manuela Chessa
manuela.chessa@unige.it

Fabio Solari
fabio.solari@unige.it

Department of Informatics, Bioengineering, Robotics and System Engineering
University of Genoa - Italy

Abstract

Interaction in Augmented Reality environments requires the precise alignment of virtual elements added to the real scene. This can be achieved if the egocentric perception of the augmented scene is coherent in both the virtual and the real reference frames. To this aim, a proper calibration of the complete system, composed of the Augmented Reality device, the user and the environment, should be performed. Over the years, several calibration techniques have been proposed, and objectively evaluating their performance has shown to be troublesome. Since only the user can assess the hologram alignment fidelity, most researchers quantify the calibration error with subjective data from user studies. This paper describes the calibration process of an optical see-through device, based on a visual alignment method, and proposes a technique to objectively quantify the residual misalignment error.

1. Introduction

Starting from the definition given by Azuma [2], Augmented Realty (AR) systems should have the following three characteristics: (i) they should combine real and virtual; (ii) be interactive in real time, and (iii) registered in 3D. In [13], the authors define the concept of *locational realism*. Specifically, the location of virtual objects has to be perceived as equally real, solid and believable as actual physical objects. This is particularly important, when interaction among the users and the virtual and real objects is required [3], e.g. in industrial environments, when AR system are used to teach procedures or assembling tasks, or in rehabilitation contexts, where the users should perform precise actions with their limbs. To achieve a satisfactory *locational realism*, the head-mounted display (HMD) used must be properly calibrated.

The current state of the art calibration techniques for Optical See-Through (OST) devices have been collected in a recent survey [13]. The main problem behind OST HMD calibration can be defined as obtaining the transformation between the HMD display and the user's eyes, together with its intrinsic parameters, which will define the virtual camera projection used to render the scene. This task can be fully automated assuming an eye tracker is integrated in the HMD, obtaining in real time all the parameters needed to achieve a fully defined projective geometrical model. If no eye tracker is available, however, Manual [24][9][10] or Semi-Automatic [11][18][21] calibration methods must be used instead. These methods always require some manual user interaction (e.g. alignments of points), which must be performed before the user starts to use the HMD for its intended purpose. The difference between the various techniques is usually linked to the number of alignments needed during the calibration phase.

An important problem related to OST HMD calibration is the difficulty of quantitatively measure the residual reprojection error. Most studies, as stated in the survey [13], quantify the calibration reprojection error by analyzing data collected from user studies. Several data collection schemes have been proposed, but rarely any objectively quantifiable data is provided. Sometimes, a picture of the alignment is provided, but single images are not foolproof evidence of a genuine calibration since the alignment can be the result of multiple roto-translation errors which compensate one another in that specific geometrical scenario. In [24], a mannequin head with cameras has been used to display a picture of the obtained alignment, but no further analysis is performed on feed obtainable from the cameras: the single picture of the alignment is used as metric for the calibration quality. However, a single alignment cannot be considered a good indicator of a genuine calibration, as it can be the

result of several errors which compensate themselves.

The scope of this paper is to introduce a technique to evaluate the calibration performance, by using the objective data obtained from a stereo camera mounted on a mannequin head, in order to overcome the limits of an evaluation performed by users only. The proposed solution is to measure the reprojection error by comparing the distance between the real and virtual 3D positions of several points, which are observed through the HMD lenses by the stereo camera. The residual positional error between the virtual camera and the real camera is also obtained, together with the intrinsic parameter drift. With current user-based methods, none of these parameters can be estimated, making different calibration techniques hard to compare.

This paper thus proposes a system that addresses the calibration of a commercial OST HMD in a general way, without relying on the specific calibration procedure of the producer, and by allowing the computation of all the necessary transformation to achieve the desired *locational realism*. In particular, we exploited a standard calibration procedure, i.e. the Single-Point Active Alignment Method (SPAAM) technique [13] and a commercial tracking system, i.e. the HTC Vive Lighthouse system, as in [20]. In this way, we are able to fully calibrate the OST device, i.e. the Meta2 by Metavision, by computing all the transformations that link the users' eyes, the camera device, the real world and the virtual environment reference frames. In case we need to track real objects in the scene, in order to augment them with virtual elements, it is possible to add more HTC trackers. Since a single system is used to track all the required poses, it is easy to define a common reference frame, unlike other setups where further transformations are required [5].

The main advantage of the devised solution is that it is independent from the specific hardware choice we made, indeed the same approach can be extended to any OST HMD and to any tracking system.

The paper is organized as follows. Section 2 briefly summarizes the related work behind the calibration technique. In Section 3, we describe the specific experimental setups we consider, though the generality of the approach must be considered. In Sections 4 and 5 we describe our calibration implementation and procedure, while in Section 6 we describe our validation method. The quantitative results of our technique are shown in Section 7. Finally, in Section 8, we discuss the obtained results and the limitation of the proposed approach. Further improvements will be necessary to obtain a system where egocentric perception of real and virtual elements is coherent and interaction is possible.

2. Related Work

In OST devices, for a proper rendering we need to know the 6-DoF pose of the virtual cameras in the physical world, which coincides with the 6-DoF pose of the user's eyes. For this reason, unlike Video See-Through (VST) HMDs, a calibration is always needed.

As summarized in a recent review [13], OST calibration techniques can be split between Manual, Semi-Automatic and Automatic methods.

Manual calibration methods require the test subject to perform a manual alignment task, which is needed to compute all the required parameters that define the projective geometry of the user-HMD system.

Semi-Automatic methods seek to simplify the calibration process by reducing the number of alignments required. This is normally achieved by computing only the parameters that change between different users (e.g. [21]) or between different sessions for the same user (e.g. [10][11]). Semi-Automatic methods seek to reduce the reprojection error by being less reliant on the user precision during the alignment task of the calibration. It must be noted that implementing these calibration techniques is usually more elaborate with respect to traditional Manual methods, and sometimes requires additional hardware. As an example, in [21] the parameters were split between the eye model and the display model, which was calibrated separately using a mechanical apparatus.

Automatic calibration methods do not require any user input, as they are able to obtain the 6-DoF pose of the eyes automatically, mostly by using eye tracker sensors. Of course, Automatic calibration represents the best option when available, as it accommodates for changes of the geometry during run time (although with a processing overhead). The presence of an eye tracker integrated in the HMD also enables enhanced gaze interactions (as opposed to the fixed crosshair-centered based ones) and more realistic rendering techniques (e.g. foveated rendering). Currently, however, most commercial HMDs do not provide integrated eye tracking functions, thus implementing Automatic calibration methods can be troublesome due to the difficulties related with detecting eye movements using external sensors. For this reason, these methods will not be covered in this study.

In [13], several studies [1][14][22] investigating the difference between calibration procedures are discussed. Reducing the impact of human error by simplifying the calibration process has shown to also reduce the amount of reprojection error; however, there is evidence that the procedures are still unable to obtain accurate results with naive users [22][13]. To obtain uniform data, the test subjects must be trained accordingly to be precise during the calibration process.

Independently of the type of calibration, the purpose is always to find the transformation matrix which can map HMD pixel coordinates to the user's eye's coordinates. In the common pinhole camera model, this transformation $\mathbf{G}$ (Eq. 1) maps a 3D world point (x_w, y_w, z_w) into 2D pixel

coordinates (u, v). In this study we will refer to the intrinsic matrix as $\mathbf{K}$ (Eq.3) and to the extrinsic matrix as $(\mathbf{R}, \mathbf{t})$. The camera model formulation has been widely discussed, for further details refer to [16][25][7].

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{G} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1}$$

$$\mathbf{G} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \tag{2}$$

$$\mathbf{K} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Our calibration is based on the SPAAM technique [24], a popular calibration method, adapted to current mainstream systems. In the original study, a tracker receiver (defined "*mark*") was rigidly attached to the HMD. A tracker transmitter obtained the pose of the *mark* in the transmitter coordinate system. The camera Matrix defining the projective transformation was defined as in Eq. 4, where $\mathbf{F}$ is the 4x4 homogeneous transformation matrix which defines the transformation between the transmitter and the mark, $\mathbf{C}$ is the pose of the transmitter with respect to world coordinates and $\mathbf{G}$ is the 3x4 projection matrix of the camera-mark transformation.

$$\mathbf{A} = \mathbf{GFC} \tag{4}$$

3. Experimental setup

In our experimental setup, the used OST HMD was a Meta2 by Metavision. The Meta2 HMD has 90° field of view and a resolution of 2560x1440 pixels (1280x1440 per eye). The Meta2 provides its internal SLAM, which has been disabled and replaced with a Vive-tracker based localization. This choice was led by the inconsistency of the SLAM localization, which is unable to compensate the drift caused by the error accumulation over time. The tracking provided by the Vive Lighthouses and trackers, on the other hand, has a good degree of stability and precision [19][4] (sub millimetric jitter on static objects).

To validate the calibration, a ZED mini camera has been mounted on a polystyrene mannequin head (Figure 1). The ZED mini is a stereo camera which can be set to run at a resolution of 1920x1080 pixel per eye (allowing 30 frames per second) or 1280x720 pixel (running at 60 frames per second). It has a baseline of 63 mm, which is in the middle of the mean Inter-pupillary Distance (IPD) of men (64±3.4 mm) and women (61.7±3.6 mm) according to [12]. The FoV of the ZED mini is of 90 degrees horizontally, 60 degrees vertically and 110 degrees diagonally. The focal length is 2.8 mm, 1400 pixels at full HD resolution and 700 pixels at HD720.

Figure 1. The mannequin head used to perform the calibration.

We used Unity as graphic engine.

Since we did not use any advanced functionality provided by the chosen hardware, we can assume without loss of generality that the following techniques can be applied reliably with any combination of similar sensors. Our choice of hardware leans towards easily available sensors in the current commercial context.

4. Calibration

In our setup (Figure 2), we used the tracking system of the HTC Vive, composed by the two Lighthouses paired with two Vive trackers. We made this choice to exploit the precision of the Vive tracking system [19][4] and its ability to retain the world space frame pose fixed over time, as opposed to SLAM systems which usually set the origin of the world frame in the initial pose obtained when starting the mapping process (and thus it varies on different sessions). The Vive trackers are already tracked in world coordinates, with the center of the world space defined during the room scale calibration. For this reason, $\mathbf{F}$ and $\mathbf{C}$ (Eq. 4) can be combined in a single matrix which represents the 6-DoF pose of the Vive Tracker rigidly attached to the HMD.

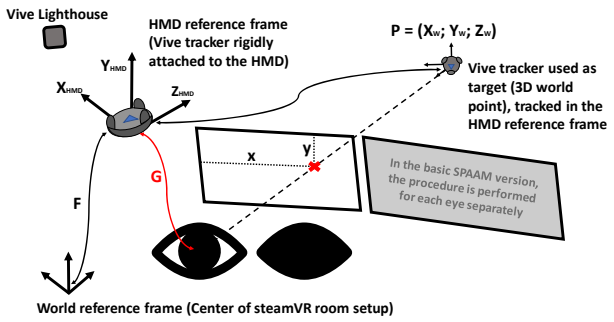


Figure 2. The SPAAM calibration setup in our implementation.

We used one of the trackers to track the HMD pose (Figure 1, right), and the other as 3D point for the alignment

task. The HMD tracker (which will be called *mark* for coherence with [24]) is rigidly attached (with a screw) to the HMD by means of a 3D printed support (Fig 3) specifically designed around the Meta2 HMD. We will refer to the same notation used in [24], with a few modifications.

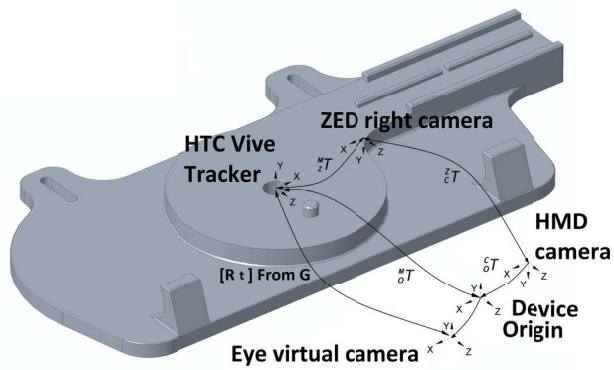


Figure 3. Schema of the transformations needed to obtain the projection alignment. The G matrix is obtained with the SPAAM calibration, the other transformations can be computed off-line.

Let $\mathbf{P_w} = [x_\mathbf{w}, y_\mathbf{w}, z_\mathbf{w}, 1]^T$ be the homogeneous coordinates of the known 3D point, which is represented by the Vive tracker which is positioned stationary attached to a tripod in front of the user, and $\mathbf{P_I} = [x, y]^T$ its projected image point. We define *mark* the coordinate system defined by the Vive tracker attached to the HMD. We assume $\mathbf{P_w}$ coordinates are already expressed in the mark coordinate system (as this transformation can easily be obtained by exploiting the hierarchical structure of the scene in Unity). The projection model (Eq. 1) can be solved with SVD to obtain all the elements of the 3x4 Camera matrix $\mathbf{G}$.

The calibration procedure thus involves displaying a crosshair (or other types of pointers) that needs to be aligned to the fixed 3D point in the world. The position of the 3D point in mark coordinates $\mathbf{P_w}$ is then saved together with the pixel coordinates $\mathbf{P_I}$ where the crosshair was displayed, creating two equations of the system (Eq. 1) for each alignment. In the original study, the 3x4 projection matrix $\mathbf{G}$ (in Hartley and Zisserman [16] notation) was converted in the 4x4 projection notation used by OpenGL by pushing the parameters into a 4x4 orthographic projection. For more insight on the process, refer to [24][23]. In our study we dissected the camera matrix into its intrinsic and extrinsic parameters by RQ decomposition, and applied them separately to a standard camera object in Unity, by using its *transform* (for the extrinsic parameters) and *physical camera* (for the intrinsic) properties.

This calibration procedure holds for the monocular case. It can be adapted to the stereo case [9] by displaying a 3D object (e.g. a disk) for the alignment task, in different positions in the two views (the shifted disparity should be tuned based on the user's measured IPD). Another approach is to simply perform the calibration separately for each eye.

In the original SPAAM study [24], it was not specified if any sort of point normalization (as shown in [6]) had been performed to avoid badly conditioned matrix during the singular value decomposition, nor any outlier rejection procedures. Since from preliminary results we observed that head jitter during the alignment task introduced a non-trivial amount of error, to reduce the variance between calibrations we implemented a RANSAC procedure [8]. Instead of collecting the minimum 6 alignments, we have been collecting $n = 15$ alignments per eye. Each alignment was considered an inlier when the reprojection error was under 0.1 mm. The considered pixel size (when projected on the lenses) was considered to be 0.059 mm. This value has been obtained by considering the surface difference between the projecting LCD surface and the lenses. Our stopping criteria is based on the number of iterations i, which is updated every time a new model with more inliers m is found, based on the probability β of finding a better model (Eq. 5). The value of β was set to 0.001.

$$i = \frac{\log \beta}{\log(1 - (m/n)^k)} \tag{5}$$

Increasing the number of alignments required in the procedure increases the calibration precision, at the cost of increased user strain. As pointed out in [13], it is advisable to track the workload increase by using subjective measurements such as NASA TLX [15]. Since the data used in this study during the validation is not provided by users, this analysis was not applicable, thus the increased user strain has not been considered as a limiting factor during calibration.

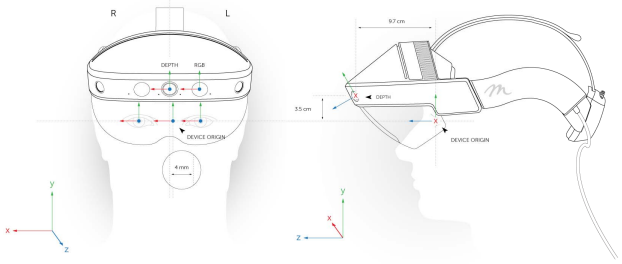


Figure 4. The inner reference frames transformations of the Meta2 HMD as reported by the manufacturer.

The HMD rendering is based upon the pose of the device origin tracked by its provided SLAM system. The SLAM and rotation tracking provided by the HMD has been disabled, since the pose tracking is already performed by the *mark* tracker. The Meta2 HMD is thus used merely as a visualization device, so that the whole procedure can be easily adapted to work with any HMD. Once the $\mathbf{G}$ matrix has been computed (Eq. 1), which represents the transformation between the *mark* tracker and the user's eyes, it is needed to

also obtain the transformation between the original tracking system and the *mark* tracker. The reference frames provided by the Meta2 SDK are oriented as in Figure 4: the virtual cameras' poses are expressed with respect to the device origin.

To obtain the relative position between the *mark* tracker and the Meta2 camera reference system, we fixed the tracker with a 1/4" screw on a 3D-printed part which was designed to precisely house the ZED mini stereo camera. The printed part can be easily designed and modified to adapt to any HMD. Moreover, the transformation matrix ${}^{M}_{Z}\mathbf{T}$ between the (ZED) camera and the *mark* tracker will be directly obtainable from the CAD drawing of the part.

Figure 5. Calibration of the ZED camera with the Meta2 camera, needed to obtain the transformation matrix ${}^{Z}_{C}\mathbf{T}$ that defines their relative pose.

The relative pose between the ZED mini and the Meta2 camera can be obtained by using traditional stereo camera calibration procedures (see [26][17]), by taking pictures of a checkerboard pattern from one of the two cameras of the ZED mini and the Meta2 Camera simultaneously. Before collecting the images, the ZED mini stereo system was calibrated with the same procedure, thus any image collected by means of the ZED mini was already undistorted. The calibration was performed with the Matlab Camera Calibration App, which is based on Zhang implementation [26]. Once the transformation ${}^{Z}_{C}\mathbf{T}$ between the Meta Camera and the ZED mini is obtained (see Figure 5), since the transformation ${}^{M}_{Z}\mathbf{T}$ between the ZED camera and the *mark* tracker is known from the CAD, and so is the transformation ${}^{C}_{O}\mathbf{T}$ between the Meta device origin and the Meta Camera (from the manufacturer documentation), we can then compute the transformation ${}^{M}_{O}\mathbf{T}$ between the Meta device origin and the *mark* tracker by combining the transformations in cascade (Eq. 6).

$$ {}^{M}_{O}\mathbf{T} = {}^{M}_{Z}\mathbf{T}\,{}^{Z}_{C}\mathbf{T}\,{}^{C}_{O}\mathbf{T} \tag{6} $$

The ${}^{T}_{V}\mathbf{T}$ matrix obtained from the extrinsic parameters of projection matrix G, is composed by the transformation ${}^{M}_{O}\mathbf{T}$ from the *mark* tracker to the device origin and the transformation ${}^{O}_{V}\mathbf{T}$ from the device origin and the Virtual (left or right) camera (Eq. 7).

$$ {}^{M}_{V}\mathbf{T} = {}^{O}_{V}\mathbf{T}\,{}^{M}_{O}\mathbf{T} \tag{7} $$

Since the *mark* tracker is rigidly attached to the HMD, ${}^{M}_{O}\mathbf{T}$ is fixed. Once ${}^{O}_{V}\mathbf{T}$ is computed for each eye, it would also be possible to switch back to the HMD SLAM tracking system, retaining the calibration effects (assuming the absence of headset slippage during use).

5. Calibration Procedure

To evaluate the calibration residual error, we performed a calibration session by using the mannequin head equipped with the ZED mini to simulate a human vision system.

In a normal scenario, the user would move his/her head to match the displayed crosshair with the fixed target. Since in our case the target is also a tracker, it would also be possible to perform the alignment with a combined movement of the head and the target tracker (e.g. kept in the user's hand). To perform the calibration with the fixed mannequin, the alignment is performed by moving the target tracker until the alignment is observed in the ZED video feed (see Figure 6). We placed the target tracker on a moving table to achieve precise alignments.

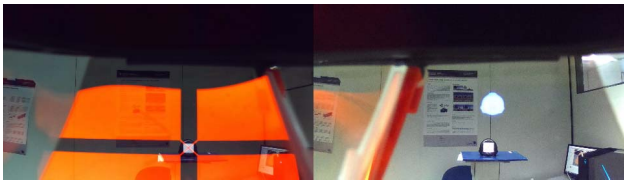


Figure 6. The view from inside the HMD during the alignment task while calibrating the left eye. On the right lens, the bright hologram floating several decimeters above the tracker shows the current uncalibrated reprojection.

Once both eyes (cameras) have been calibrated, to test the residual reprojection error, we measured the misalignment by using a tracked checkerboard. The checkerboard has been placed on a metal table, which was tracked by a third Vive tracker, rigidly attached by means of a 3D printed part (Figure 7).

The same tracker used as a target during the calibration could have been used, we have chosen to use a third one for the sake of convenience. The checkerboard position with respect to the attached tracker is known (from the custom nature of the support), thus if a perfect calibration is achieved, it is possible to display a virtual checkerboard perfectly aligned with the real one. We can measure the quality of the calibration from the offset between the virtual checkerboard and the real one.

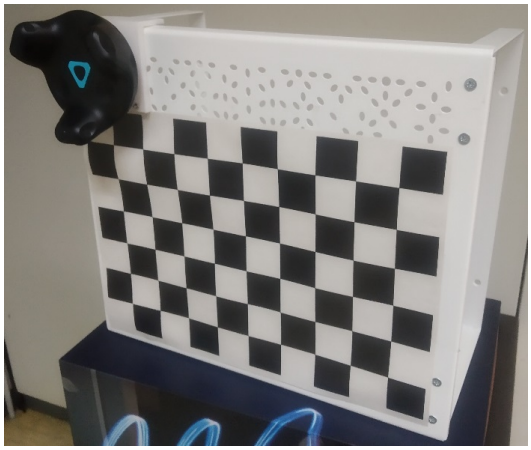


Figure 7. The tracked checkerboard.

We collected 19 images of the checkerboard from different angles, trying to cover as much as the image frame as possible. The checkerboard was placed at distances between 50-90 centimeters, which is the area of interest for human interaction in peripersonal space. The checkerboard was placed on a still support. For each checkerboard position, two stereo pairs were collected: one (pair) with the augmented checkerboard displayed, one without displaying the augmentation (Figure 8). The two stereo pairs (augmented and unaugmented) were thus not collected simultaneously, but since both the mannequin and the checkerboard were firmly fixed, we can assume there was no significant change between the two image pairs even if collected at slightly different time intervals (a few seconds apart from one another). The overall amount of collected images thus was 19 for each eye, both augmented and unaugmented versions, for a total of 76 images. From the original set of 19 stereo pairs, 8 were discarded as ill-defined (e.g. the checkerboard was not fully visible in both views, either in the real or the augmented case).

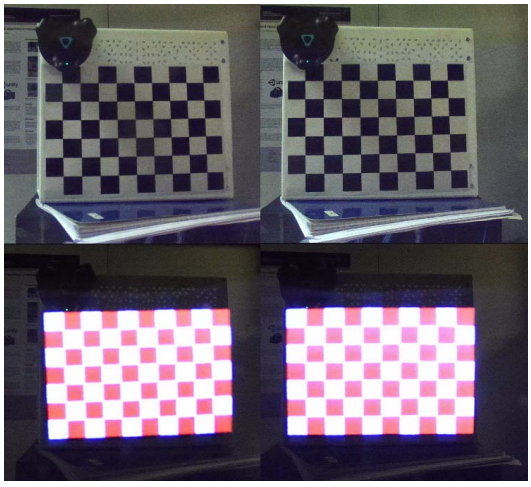


Figure 8. Left and right view of the augmentation after the calibration.

To have a better contrast in the augmented images, the lights were turned off when collecting the augmented pairs. In scarce lighting conditions, the augmentation completely covers the real checkerboard, simplifying the segmentation of the image.

The augmented checkerboard has been displayed in red and white, to be easily segmented by binarizing the images after performing a color based thresholding. Figure 9 shows one of the segmented views (b), together with its corresponding unaugmented image (a), and the detected checkerboard points.

6. The Validation Technique

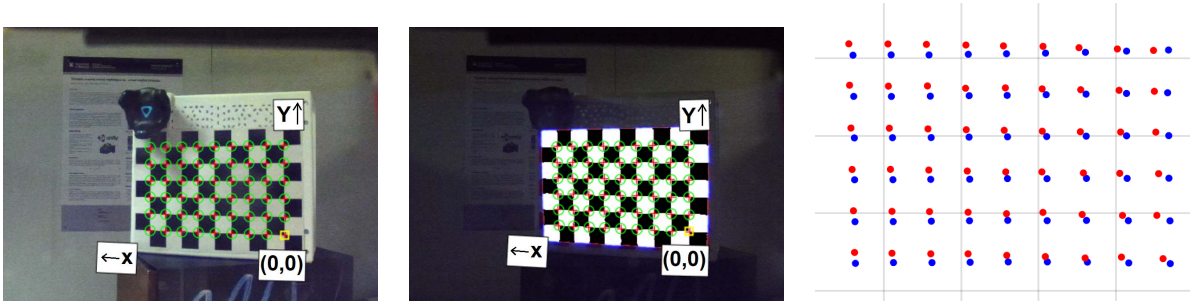
We measured the calibration error by performing a stereo calibration between the real cameras and their virtual counterparts. The set of unaugmented images from the left camera and the corresponding augmented set (still from the left camera) were thus considered as part of a stereo system, which was calibrated to find the relative displacement (residual positional drift). The same procedure was then repeated for the right camera.

To test the repeatibility of the system, we performed the calibration procedure another 8 times for each eye, and we computed the standard deviation of the parameters over all the iterations.

To quantify the misalignment error perceived by the user, we measured the distance between the projected 3D positions of the checkerboard points with the 3D positions displayed by the augmentation, as in (Figure 9). In (Figure 9 c), the blue dots represent the real position of the projected 3D points, while the red dots represent the positions of the corresponding virtual points. The distribution and magnitude of these distances over the image plane can be used as metric of the perceived error, which can be useful to define which areas of the work space are more suited for interaction. To achieve this representation, we obtained the real 3D positions by triangulating the points detected in the real stereo rig (e.g. Figure 8 top two views), and the perceived 3D positions by triangulating the points detected from the augmented views (Figure 8, bottom two views). Since the detected points are in the reference frame of the first camera used during the stereo calibration, we brought the virtual 3D points in the real left camera reference frame one by applying the transformation from the left virtual to the real camera obtained previously.

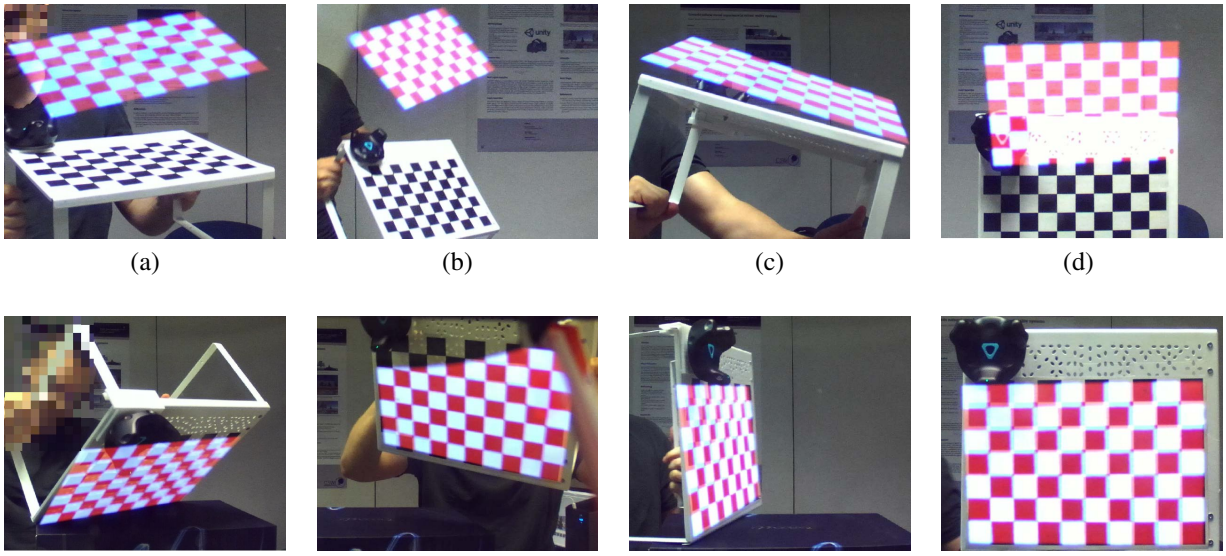
7. Results

The residual calibration error, and the standard deviation computed over 8 calibrations, are shown in Table 1. By comparing the deviation of the obtained parameters over the



(a) (b) (c)

Figure 9. The points triangulated in the real (a) and virtual (b) stereo rigs (only the left view is displayed). In (c) the 2D misalignment between the triangulated 3D points can be observed (grid size is 5cm). The red points are generated from the virtual stereo system, the blue ones are obtained from the unaugmented views.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 10. The top images (a,b,c,d) show the alignment error before the calibration. The bottom images (e,f,g,h) display the obtained alignment post-calibration. All the images are collected from inside the HMD, with the mannequin stereo system.

calibration attempts with the residual error of the first calibration, it is possible to see that the other calibrations are likely to produce similar residual errors, keeping the system stable over different sessions. We can observe how there is a residual rototranslational error: roughly in the volume where the calibration alignments were collected, the rotational drift in the vertical y axis is compensated by a horizontal translation.

The average euclidean distance error between perceived and real positions is 23 ± 11.5 mm, computed over 594 pairs of points (11 image pairs, 54 points per image). When considering the plane orthogonal to the optical axis, the average 2D euclidean distance error perceived is 8.5 ± 4.5 mm.

The heat map of the alignment error distribution along the x, y orthogonal plane (Figure 11) shows that the majority of the distortion is localized towards the edges of the

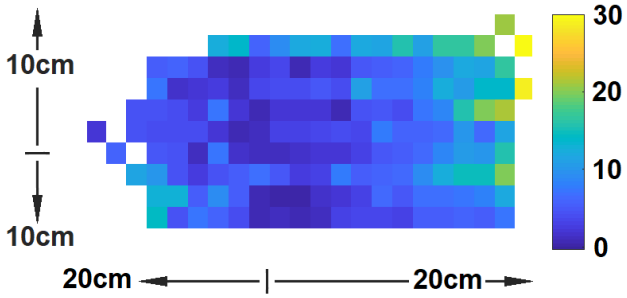


Figure 11. Heat map of the euclidean distance of alignment error along the plane orthogonal to the optical axis (colormap measures in mm).

field of view. This bias is rather common during stereo systems calibration as the edges of the cameras views are usu-

ally not overlapped, thus no alignments are performed in that area.

	Left Camera	R.E.L.	Right Camera	R.E.R.
f_x	± 7.2 mm	34 mm	± 10 mm	21 mm
f_y	± 11.5 mm	62 mm	± 11 mm	48 mm
c_x	± 5.4 mm	0.9 mm	± 2 mm	6.5 mm
c_y	± 5.4 mm	0.7 mm	± 5 mm	3.3 mm
s	± 2.8 mm	2.7 mm	± 2.2 mm	1.3 mm
t_x	± 32 mm	66 mm	± 24 mm	60 mm
t_y	± 14.1 mm	38 mm	± 22 mm	50 mm
t_z	± 31.9 mm	26 mm	± 37.4 mm	66 mm
R_x	$\pm 4.21°$	$0.48°$	$\pm 8.11°$	$0.54°$
R_y	$\pm 4.32°$	$3.28°$	$\pm 2.6°$	$3.98°$
R_z	$\pm 1.21°$	$7.57°$	$\pm 0.85°$	$6.9°$

Table 1. The Right/Left camera columns report the parameters standard deviation over the different calibrations. The other two columns report the absolute Residual Error for the Left (R.E.L.) and Right (R.E.R.) cameras.

The depth misalignment does not change significantly over distance (Figure 12), with a mean error of 20.5 ± 12.6 mm. Since such misalignment was hardly noticeable (e.g. Figure 10 g), this error is probably compensated by the scale factor introduced by the focal length drift. Considering the computed perceived misalignment, we consider that the obtained reprojection can be used for an effective AR user interaction.

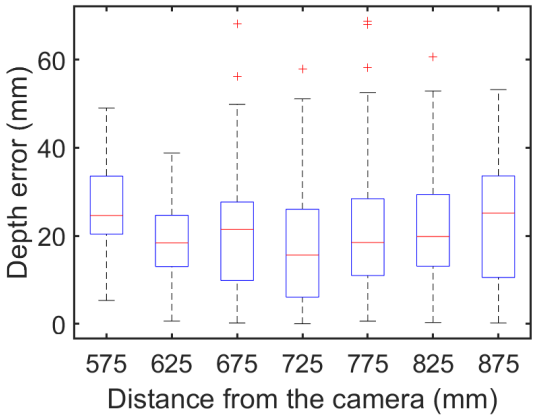


Figure 12. Progression of the alignment error over depth. Each box bins data from $\pm 25mm$.

8. Conclusion

In this paper, we discuss the calibration of an OST AR HMD, and propose a technique to analyze the quality of the obtained alignment. The calibration procedure uses an external and off-the-shelf tracking system, which provides precise measures, and it can be generalized for different tracking systems. In many studies the quality of the calibration is often validated by means of user studies, which can lead to: (i) perceptual bias introduced by the users; (ii) the dependency to subjective metrics, and (iii) calibration systems which are not easily comparable with each another without several comparative/repetition studies. In our proposed method, the calibration is validated by means of objectively quantifiable data obtained by a stereo camera. As metrics to quantify the degree of locational realism achieved, and thus the quality of the calibration, we propose to measure (i) the residual parameter error between the computed parameters of the virtual camera with respect to the real camera; (ii) the amount of perceived misalignment error in the work space area of interest, and (iii) the repeatability of the calibration expressed as the error variance over multiple calibration sessions.

As it was reported in [22], we also experienced that the projection matrices obtained with the SPAAM procedure are subject to small misalignment errors. The average alignment error however can be considered suitable for interaction bounded in peripersonal space, thus providing a coherent egocentric perception of the augmented scene in both the virtual and the real reference frames.

References

[1] M. Axholt, M. Skoglund, S. D. Peterson, M. D. Cooper, T. B. Schön, F. Gustafsson, A. Ynnerman, and S. R. Ellis. Optical see-through head mounted display direct linear transformation calibration robustness in the presence of user alignment noise. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 54, pages 2427–2431. SAGE Publications Sage CA: Los Angeles, CA, 2010.

[2] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.

[3] G. Ballestin, F. Solari, and M. Chessa. Perception and action in peripersonal space: A comparison between video and optical see-through augmented reality devices. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 184–189. IEEE, 2018.

[4] A. Borrego, J. Latorre, M. Alcañiz, and R. Llorens. Comparison of Oculus Rift and HTC Vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games for health journal*, 7(3):151–156, 2018.

[5] A. Canessa, M. Chessa, A. Gibaldi, S. P. Sabatini, and F. Solari. Calibrated depth and color cameras for accurate 3D interaction in a stereoscopic augmented reality environment. *Journal of Visual Communication and Image Representation*, 25(1):227–237, 2014.

[6] W. Chojnacki and M. J. Brooks. Revisiting Hartley's normalized eight-point algorithm. *IEEE transactions on pattern analysis and machine intelligence*, 25(9):1172–1177, 2003.

[7] O. Faugeras and O. A. FAUGERAS. *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image

analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[9] Y. Genc, F. Sauer, F. Wenzel, M. Tuceryan, and N. Navab. Optical see-through HMD calibration: A stereo method validated with a video see-through system. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 165–174. IEEE, 2000.

[10] Y. Genc, M. Tuceryan, A. Khamene, and N. Navab. Optical see-through calibration with vision-based trackers: Propagation of projection matrices. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 147–156. IEEE, 2001.

[11] Y. Genc, M. Tuceryan, and N. Navab. Practical solutions for calibration of optical see-through devices. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 169. IEEE Computer Society, 2002.

[12] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville. Anthropometric survey of US army personnel: methods and summary statistics 1988. Technical report, Anthropology Research Project Inc Yellow Springs OH, 1989.

[13] J. Grubert, Y. Itoh, K. Moser, and J. E. Swan. A survey of calibration methods for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 24(9):2649–2662, 2018.

[14] J. Grubert, J. Tuemle, R. Mecke, and M. Schenk. Comparative user study of two see-through calibration methods. *VR*, 10:269–270, 2010.

[15] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[16] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[17] J. Heikkila, O. Silven, et al. A four-step camera calibration procedure with implicit image correction. In *cvpr*, volume 97, page 1106, 1997.

[18] N. Navab, S. Zokai, Y. Genc, and E. M. Coelho. An on-line evaluation system for optical see-through augmented reality. In *IEEE Virtual Reality 2004*, pages 245–246. IEEE, 2004.

[19] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017.

[20] S.-T. Noh, H.-S. Yeo, and W. Woo. An hmd-based mixed reality system for avatar-mediated remote collaboration with bare-hand interaction. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*, pages 61–68. Eurographics Association, 2015.

[21] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative calibration for optical see-through head-mounted displays. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 70–78. IEEE Computer Society, 2004.

[22] A. Tang, J. Zhou, and C. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 161. IEEE Computer Society, 2003.

[23] E. Trucco and A. Verri. *Introductory techniques for 3-D computer vision*, volume 201. Prentice Hall Englewood Cliffs, 1998.

[24] M. Tuceryan, Y. Genc, and N. Navab. Single-point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. *Presence: Teleoperators & Virtual Environments*, 11(3):259–276, 2002.

[25] G. Xu and Z. Zhang. *Epipolar geometry in stereo, motion and object recognition: a unified approach*, volume 6. Springer Science & Business Media, 2013.

[26] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.