Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation

Falko Kellner, Benjamin Bolte, Gerd Bruder, Ulrich Rautenberg, Frank Steinicke, Markus Lappe, and Reinhard Koch

Abstract—Head-mounted displays (HMDs) allow users to observe virtual environments (VEs) from an egocentric perspective. However, several experiments have provided evidence that egocentric distances are perceived as compressed in VEs relative to the real world. Recent experiments suggest that the virtual view frustum set for rendering the VE has an essential impact on the user's estimation of distances. In this article we analyze if distance estimation can be improved by calibrating the view frustum for a given HMD and user. Unfortunately, in an immersive virtual reality (VR) environment, a full per user calibration is not trivial and manual per user adjustment often leads to mini- or magnification of the scene. Therefore, we propose a novel per user calibration approach with optical see-through displays commonly used in augmented reality (AR). This calibration takes advantage of a geometric scheme based on 2D point – 3D line correspondences, which can be used intuitively by inexperienced users and requires less than a minute to complete. The required user interaction is based on taking aim at a distant target marker with a close marker, which ensures non-planar measurements covering a large area of the interaction space while also reducing the number of required measurements to five. We found the tendency that a calibrated view frustum reduced the average distance underestimation of users in an immersive VR environment, but even the correctly calibrated view frustum could not entirely compensate for the distance underestimation effects.

Index Terms—Optical see-through, HMD calibration, distance perception.

◆

1 INTRODUCTION

Virtual environments (VEs) are often characterized by head-mounted displays (HMDs) and a tracking system for measuring a user's position and orientation. Such head- or helmet-mounted displays are head-worn devices, which present different images for the left and right eye of the user by employing two miniature display units. Commercially available HMDs typically have fields of view (FOVs) which range from 20 to 80 degrees diagonally. Whereas, the effective visual field of humans is approximately 200 degrees horizontally and 150 degrees vertically [34]. In contrast to the *display's field of view* (DFOV), the *geometric field of view* (GFOV) defines the horizontal and vertical boundaries of the virtual viewing volume along with the aspect ratio. Usually, user movements measured by the tracking system are mapped one-to-one to the position and orientation of the virtual camera. This allows users to explore a VE similar to the real world. The projection of the virtual camera defines the *camera view frustum*, which is used for rendering, whereas the *display view frustum* refers to its physical equivalent in real world. For simplicity, usually the horizontal and vertical GFOVs of the camera view frustum are chosen in a way that the view frustum is arranged symmetrically around the view direction according to the DFOV as specified by the HMD manufacturer. However, usually such a setup does not reflect the real viewing situation, and in the optimal case the camera view frustum should match the display view frustum of the HMD. Only then, the viewport is mapped from virtual space onto the physical display in such a way that users are presented with imagery that will have "correct" perspective.

- *Falko Kellner, University of Kiel, E-mail: fkellner@mip.informatik.uni-kiel.de.*
- *Benjamin Bolte, University of Muenster, E-mail: b.bolte@uni-muenster.de.*
- *Gerd Bruder, University of Wuerzburg, E-mail: gerd.bruder@uni-wuerzburg.de.*
- *Ulrich Rautenberg, Group Research Virtual Technologies Volkswagen AG, E-mail: Ulrich.Rautenberg@volkswagen.de.*
- *Frank Steinicke, University of Wuerzburg, E-mail: frank.steinicke@uni-wuerzburg.de.*
- *Markus Lappe, University of Muenster, E-mail: mlappe@uni-muenster.de.*
- *Reinhard Koch, University of Kiel, E-mail: rk@mip.informatik.uni-kiel.de.*

Several experiments provide evidence that subjects judge egocentric distances as significantly compressed in VEs — in some cases up to 50% — relative to distance judgments in the real world [17, 20, 23, 25, 35]. Sources for parts of the observed distance compression effects have been identified in experiments, but an explanation for all the effects remains unknown. In the main body of literature on distance estimation, the camera view frustum was not completely calibrated to the display view frustum, which may be one reason for the significantly different spatial perception of a virtual world through a HMD in contrast to the real world. To achieve such a complete calibration, the virtual camera needs to be calibrated both intrinsically and extrinsically relatively to the tracked head position. In augmented reality (AR) systems, in which the real world environment is enriched with computer generated graphics, this calibration can be done and evaluated on a per user basis. Here, the individual user can calibrate the virtual camera by aligning tracked real world markers with virtual target positions. This calibration allows for a perfect overlap of real and virtual objects in an AR environment, which can be validated by the user directly. In such augmented reality systems, so called *optical see-through head mounted displays* (OSTHMDs) are used. Such displays do not present the images to the user directly, but typically the images are projected into the natural view of the user by, for instance, semi-transparent mirrors.

In this article we introduce a novel calibration scheme for OSTHMDs with short and intuitive user interaction, which can be used for both VR and AR environments. We evaluated the quality of the calibration and time required to perform a calibration using theoretical calculations and user studies. Afterwards, we conducted a blind walking experiment and used our method to calibrate the camera view frustum for each individual user in order to analyze if the often observed distance underestimation in VR environments is mainly caused by miscalibration of the camera view frustum. Section 2 recalls work related to our approach. Section 3 describes and evaluates the geometric calibration technique that we developed to determine the display view frustum of a given HMD. Section 4 explains the distance estimation experiments. Section 5 concludes the paper.

2 RELATED WORK

2.1 Scene Perception

Since most commercially available HMDs have a relatively narrow field of view in comparison to the effective visual field of humans, HMD users can only see a rather restricted portion of the virtual world

if the camera view frustum matches the view frustum of the HMD. In the real world, a narrow field of vision has been shown to degrade spatial awareness and human performance in navigation and manipulation tasks [15, 19] as well as visual search tasks [3]. Similarly, in the virtual world a restricted FOV may lead to perceptual, visual, and motor decrements in various kinds of performance tasks [1, 13]. Deviation between the camera and display view frustum occurs, for example, when the actual display view frustum varies from the nominal values specified by the HMD manufacturers. A deviation can also be induced intentionally. An increased view frustum allows for the inclusion of more information in the 3D view like a larger physical display would. A larger view frustum is essential in some applications and has the potential to improve and increase the user's sense of presence [2, 28]. On the other hand, a large GFOV may be unnecessary for tasks, which are localized within a small spatial region of interest [15], and it may aggravate simulator sickness effects, particularly those caused by visual-vestibular mismatch [28, 29].

Researchers have found evidence that distances in virtual worlds are underestimated in comparison to the real world [16, 17, 23], that visual speed during walking is underestimated in VEs [5] and that the distance one has traveled is also underestimated [7]. As experiments have shown, only a small portion of the observed compression can be caused by hardware issues, and an explanation for the larger portion of the effect remains unknown. It is often assumed that the human visual system would be most comfortable with HMD viewing conditions that match natural viewing of the real world [27]. However, recent insight into subjective judgment of natural FOVs used for rendering virtual scenes on HMDs showed desired discrepancies of up to 50% compared to such projections that are calibrated to match the correct FOV of the HMD [30]. Furthermore, it has been shown that different fields of view have a significant impact on distance estimation [22, 30].

Such results lead to the conclusion that calibrating not only the field of view, but the entire camera view frustum may be required for users to correctly judge distances in VEs. A great body of literature motivated that even if the distal size and depth of objects match, people's perceptions and actions in a VE may differ from those perceived and performed in an equivalent real environment.

2.2 Calibration Techniques

The FOV of an immersive HMD can be determined, for instance, with the psychophysical calibration method proposed by Steinicke et al. [30] or Kuhl et al. [22], where subjects have to compare a real-world object with a virtual object by raising and lowering the HMD. This technique incorporates a psychophysical two-alternative forced-choice task to accurately measure the relation between stimulus intensity and perception reported by a human observer. Another approach was proposed by Ellis and Nemire [6], who displayed vertical poles in the HMD, with the subjects' task to point at the perceived location of the poles in the real world, allowing to compute the angular difference between visual cues and proprioceptive responses for estimating the actual FOV of the HMD. However, full calibration of the view frustum both intrinsically and extrinsically as well as its validation per user is difficult without directly visible correspondences from real and virtual environment. For OSTHMDs, calibration methods for the view frustum can be roughly split into multi-point and single-point approaches. In multi-point approaches, the user of the display is asked to align a multi-dimensional calibration object of known size with a 2D target projection which he sees on one or both displays [18, 24]. Then, the 2D-3D correspondences are used to deduce all parameters of the system. A notable break-through in usability and generality has been achieved with the "Single Point Active Alignment Method" (SPAAM), presented by Tuceryan et al. [32]. Here, the alignment is reduced to one marker per shot and the user aligns it multiple times from different viewpoints. Different variants of calibration methods, focusing on SPAAM variants, were evaluated by Tang et al. [31]. Genc et al. [10] extend this approach to directly calibrate the displays for both eyes, but at the cost of a more demanding user interaction where the user aligns a stereoscopically presented disc with a real world 3D point. Genc et al. [11] later split the SPAAM approach into an offline

and online phase, where the user is presented with pre-calculated calibrations from many alignments and chooses the most suitable. The user then only needs to record a few more measurements to complete calibration. In 2004, Owen et al. [26] presented another two phase calibration for OSTHMDs, in which most characteristics of the display are calibrated offline in the first phase, further reducing measurements conducted by the individual user. Recently, Gilson et al. [12] presented a calibration method based on photogrammetry related to the approach by Owen et al. [26], which they say is independent of the user currently wearing the display as long as the display is not manipulated. The latter two methods require additional hardware like cameras, rigs and calibration patterns but can achieve high precision with respect to the reprojection error. The method we propose is most closely related to Fuhrmann et al. [8, 9], who also use three dimensional lines as geometric entities to deduce the camera center and an up-vector. All other parameters are fixed guesses which serve as a first approximation. This approximation is then refined by numerical methods. In contrast, we deduce all camera parameters directly from the geometrical model, showing that OSTHMD camera calibration can be done without iterative optimization of the parameters.

3 GEOMETRIC CALIBRATION

Our approach to OSTHMD calibration consists of an online data acquisition phase and an offline calibration phase. In the data acquisition phase, a user wears an OSTHMD, which is tracked via a 6 degrees of freedom (DOF) marker, attached at the back of the head, in a controlled environment. Through the HMD display, the user can see a stationary 3-DOF marker in a distance of approximately 2.5 meters. The user holds another 3-DOF marker in his hand. The 6D pose of the head-attached marker as well as the 3D positions of the 3-DOF markers are given in a world coordinate system defined by the external tracker. HMD, 6-DOF head marker and the user's eye are rigidly coupled as long as the user does not take off the display, so we aim to calibrate both the relative transform from 6-DOF marker to the user's eye as well as the imaging properties of display and eye. Display and eye can be jointly regarded as a pinhole camera with the typical parameters: The optical axis corresponds to the plane normal of the projection of the display's physical screen as observed by the user. The distance of the user's eye to this physical screen corresponds to the focal length. Lastly, the intersection point of optical axis and display screen is the principal point. The experimental setup is displayed in Figure 1. The OSTHMD is a nVisor ST60 from NVIS, Inc. The nVisor features two 1280x1024 pixel displays and has 60° DFOV according to the manufacturer specifications. Notably, the user can manually adjust the distance of the individual displays according to his interocular distance. The display itself exhibits almost no geometric distortion (less than 1% according to specifications), so no additional effort will be put into calibrating distortion effects. The marker tracking system in use is the ARTrack2 system by advanced real time tracking GmbH, which delivers tracked positions at 60 hertz with sub-millimeter accuracy for both 3-DOF and 6-DOF marker targets. Tracking systems that exhibit significant systematical or non-systematical errors would possibly require special treatment.

3.1 2D Point to 3D Line Correspondences

As in many OSTHMD calibration procedures, we use external tracker information to deduce 3D positions and orientations of several markers. All positions and orientations are given in an arbitrary tracker coordinate system. One 6-DOF marker is attached rigidly to the display, always reporting translation and rotation of the "head coordinate system" as a 3x3 rotation matrix R_h and a 3D translation vector C_h. When we transform all other markers in the scene into this head coordinate system, every marker position is relative to the head. The transformation is done by inverting the transformations R_h and C_h on the marker positions m_i as

$$m_{ih} = R_h^T (m_i - C_h). \tag{1}$$

In [8] a marker is held in different distances to the camera to create 2D-3D correspondences. When this is repeated for different distances

Fig. 1. A user aims at a stationary and a hand-held marker. Positions of 3-DOF markers and 6-DOF HMD-attached marker are tracked by infrared cameras.

for the same 2D point, the 3D positions form a 3D line through the 2D target. Our approach is similar to this, yielding a 3D line. The user is presented a 2D target on the HMD screen. By moving his body, for example rotating his head, the user aligns the 2D target with a distant stationary marker. Simultaneously, he aligns the hand-held marker with the target. The two marker positions now correspond to a line of sight through the target 2D position relatively to the pose of the head marker. Naturally, the user cannot hold perfectly still while targeting the markers. To compensate for this, we record the tracker data for a short amount of time (half a second, giving approximately 30 measurements) and take the robust mean of all measured markers in the head coordinate system. The targeting procedure is repeated with distinct 2D positions to generate a set of five lines in a pattern we describe in detail in the next section.

To determine the center of projection, in our case the position of the eye relatively to the head marker, it suffices to measure two distinct lines of sight. In theory, both lines intersect at the center of projection in the pinhole camera model. In practice however, the quality of measurements differs and two lines do not intersect in 3D space. In our method, we adapt [8] and use five specific distinct lines and find the point closest to all lines in a least squares sense. To simplify further calculations and notations, the calculated 3D position is memorized as C_{eye} and all lines are translated to intersect with the coordinate system's point of origin.

We denote the full transformation from world coordinate system into the virtual camera as $T_{full} = T_h \cdot T_{eye}$, where T_h is the head coordinate transform as a 4x4 matrix comprising of R_h and C_h reported by the tracker and T_{eye}, the 4x4 transformation matrix comprising of C_{eye} and a yet unknown rotation R_{eye}, which is inferred in the following.

3.2 Determining the Display Rotation

With all lines going through the point of origin, we can now identify the lines with their (normalized) vectors of direction. When determining the display plane normal, we make use of a special target point pattern. The target pattern consists of five targets, forming the typical "five" pattern which can be found on common dice. The pattern has one target on the center pixel and one target in each image corner. The lines through these targets are denoted L_0 to L_4 with L_0 being the center line and lines L_1 to L_4 being the corner targets from left to right, top to bottom. The corresponding vectors of direction are denoted $\vec{v}_0$ to $\vec{v}_4$. Finally, let P_0 to P_4 denote the intersection points of each line with the display plane E. Ideally, the display normal corresponds to the line measured through the center target. Using $\vec{v}_0$ as the display normal, we can calculate the intersection points of all lines with the assumed display plane

$$E_0 : p \cdot \vec{v}_0 = 1. \tag{2}$$

In this ideal case, the optical axis of the virtual camera corresponds to the viewing ray through the center pixel and there is no offset of the principal point. This also means we could directly use the direction of $\vec{v}_0$ to compute two degrees of freedom of the unknown rotation R_{eye}.

But, depending on how the user wears the HMD display, the five-pattern is distorted perspectively, as can be seen in Figure 2 (a). If a plane with the correct display normal is used for intersection with the lines, the distortion vanishes as is depicted in Figure 2 (b). Here, we can exploit that in the undistorted case all distances $\overline{P_iP_0}$ are equal. To find a display normal $\vec{n}$ with this characteristic, we reduce the dimensionality of the problem from 3D to two 2D problems, which can be solved easily. The 3D intersection points of each diagonal are transferred into their corresponding coordinate frame of a 2D plane. Each plane is formed by the camera center and three intersection points lying on a diagonal of the five-pattern. To simplify the calculations in the coordinate frame of the plane, we rotate the vector of the center line to the x-axis and then rotate the other intersection points into the xy-plane. The configurations in 3D and 2D are shown in Figure 3 for one of the two planes. In a plane, we can calculate the correct intersection points A and B such that the distances d_a and d_b are of equal length. Once we have these intersection points, we transfer them back into 3D space by inverting the prior transform. For the two 2D planes this computation yields different vectors $(A - B)$ which form linearly independent vectors in the desired 3D plane, so their cross product finally delivers the desired normal $\vec{n}$.

Computing the intersection points A and B in a plane is straightforward: We know that A is located somewhere along $\vec{v}_1$ and similar for B, in parametric form this means

$$A = s \cdot \begin{pmatrix} 1 \\ tan(\alpha) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad B = t \cdot \begin{pmatrix} 1 \\ tan(-\beta) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{3}$$

where α is the angle between v_0 and v_1 and β is the angle between v_0 and v_2 respectively. Additionally, A and B are located on some point of the unknown plane g (being a line in 2D). As in this coordinate frame the line g runs through $(1,0)^T$, we can write A and B as

$$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + d \cdot \vec{v}_g, \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - d \cdot \vec{v}_g \tag{4}$$

where $\vec{v}_g$ is the unknown direction of g. Here, we use that in the solution $d = d_a = d_b$ holds. When we insert equations (3) into equations (4), we can sum up both equations which results in

$$s \cdot \begin{pmatrix} 1 \\ tan(\alpha) \end{pmatrix} + t \cdot \begin{pmatrix} 1 \\ tan(-\beta) \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}. \tag{5}$$

This equation can be solved for s and t, finally yielding A and B with equations (3).

The normal $\vec{n}$ defines the rotation of the camera around x- and y-axis, making the common assumption that our virtual camera points in
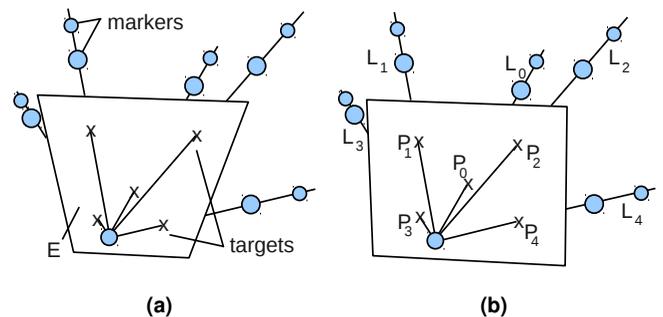

Fig. 2. (a) wrong display normal results in a distortion of the five dot pattern. (b) with the correct display normal, all lengths $\overline{P_iP_0}$ are equal.
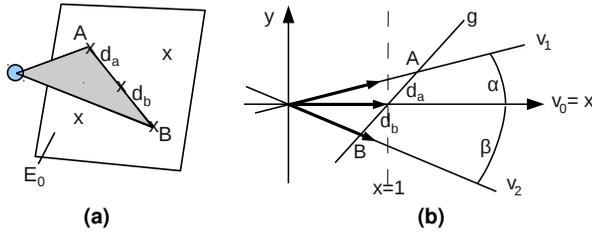
Fig. 3. (a) points on one diagonal and camera center form a plane. (b) In the plane, we look for the line g through $(1,0)^T$ for which $d_a = d_b$. For this, we calculate the intersection points A and B of g with $\vec{v_1}$ and $\vec{v_2}$.

z-direction. The wanted rotation therefore consists of rotating $\vec{n}$ onto $\vec{z} = (0,0,1)^T$. Since both vectors are normalized, the cosine of the rotation angle ϕ_{xy} is computed as the scalar product of the two vectors and the rotation axis $\vec{v}$ is the cross product of the vectors. We now transform all lines with this rotation such that their intersections with the plane E_0 have the correct distances $\overline{P_i P_0} = d$.

From these intersections, the missing rotation around the z-axis can be computed using each horizontal and vertical line of the target pattern individually (see Figure 4 (a)). The mean angle from the resulting angles is used as the missing rotation. Each of the angles is computed as in the following example: $\vec{x}$ being the vector $(1,0,0)^T$, we compute the angle between $\vec{x}$ and the segment $\overline{P_1 P_2}$ from

$$\vec{v_r} = \frac{P_2 - P_1}{\|P_2 - P_1\|}, \quad \phi_z = arccos(\vec{x}^T \cdot \vec{v_r}). \qquad (6)$$

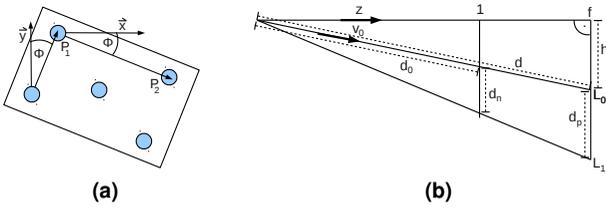
Fig. 4. (a) Computing rotation angle of z-axis from x-axis and y-axis. (b) Computation of focal length and principal point by intercept theorems.

Composition of the rotations on x-, y- and z-axis yields the final rotation matrix R_{eye}, completing the calculation of the external calibration in the head coordinate system, represented as the 4x4 transformation matrix T_{eye}.

3.3 Determining Focal Length and Principal Point

Now that translation and rotation of the display are known, we will estimate focal length and principal point of the virtual camera. The focal length is equivalent to the distance of eye to display, the principal point is the intersection of optical axis and display. Both measures can be estimated in pixel without knowledge of physical display or pixel size.

After applying the transformations in section 3.2, the optical axis corresponds to the z-axis and the intersection point of all lines with the plane at $z = 1$ are known. Looking at pairs of rays L_0, L_i, we can compute the focal length in pixel by intercept theorems. With the nomenclature of Figure 4 (b)

$$d = d_0 \cdot \frac{d_p}{d_n} \qquad (7)$$

where d_p is the distance of 2D points in pixel. d_n is the distance of the 3D points from intersecting the rays with the $z = 1$ plane. d_0 is the distance of the intersection point of line L_0 to the camera center. We define a plane in 3D with normal $\vec{z}$ that includes $(d \cdot \vec{v_0})$. This plane is

intersected with the line $f \cdot \vec{z}$, giving the focal length f as the distance of the intersection point to the camera center:

$$f = \frac{d}{\vec{z}^T \cdot \vec{v_0}} \qquad (8)$$

The distance of this intersection point to the intersection point of L_0 with this plane gives the displacement $\vec{h_d}$ of the principal point in pixel,

$$\vec{h_d} = d \cdot \vec{v_0} - f \cdot \vec{z}. \qquad (9)$$

For our specific device the pixel coordinates of the principle point are therefore $\vec{h} = (1280/2, 1024/2)^T - \vec{h_d}$. To represent the internal camera calibration, we use the camera calibration matrix K, which maps 3D coordinates onto 2D pixel positions on the image plane. K is a 3x3 matrix with the parameters focal length f, principal point (h_x, h_y) and skew s, which we set to $s = 0$. The K matrix together with the extrinsics can be used to form a projection matrix P according to [14]

$$P = K \cdot (T_{full\ 3x4}), \qquad (10)$$

which defines a mapping from 3D world coordinates into the virtual camera view. This representation is used in the following validation on synthetic data to measure reprojection errors. For the actual rendering of augmented or virtual environments, the K matrix and extrinsic matrix are transformed into projection matrix and model view matrix of an OpenGL renderer.

3.4 Validation on Synthetic and Real Data

To validate our approach and to evaluate the influence of noise on the system, we generated input data with synthetic camera parameters. The camera parameters were used to generate ground-truth and noisy 3D points, where normal distributed noise was added. The influence of distorted data was evaluated in one thousand trials per noise amplitude. Figure 5 shows the influence of noisy measurements on different parameters of the calibration. On the x-axis one can see the chosen σ that is used to distort the data. On the y-axis mean and standard deviation of a thousand calibrations is shown in the respective measure. The marker distances were fixed at 80cm for the near marker and 2.5m for the far marker. The synthetic camera's field of view was set to 40 degrees. Notably, noisy measurements have little influence on the estimation of the focal length and camera center estimation, but higher influence on the calculation of the principal point as can be seen in Figure 5 (c). This is justified by the fact that a small error in rotation estimation results in large pixel errors given the small display size.

Figure 5 (d) shows the influence of increasing stationary and hand-held marker distance on the camera center estimation. The hand-held marker thereby had a fixed distance to the head of 80cm. The distance to the stationary marker was increased in steps of 20cm in a range of one to four meters. The evaluation shows that after 2.5m the error is not reduced significantly. We therefore consider a distance of 2.5m as optimal.

Finally, Figure 6 shows scatter plots of the estimated camera centers of a thousand calibrations, projected onto the xy-, xz- and yz- planes. A principal component analysis shows mean and standard deviation of 0.32mm and 0.39mm in x- and y- axis directions. We clearly see that the z-component is the most error prone with mean and standard deviation of 0.79mm and 0.95mm. This is clear intuitively considering the influence of a small error in the xy-plane on the intersection point of two lines in 3D. Comparable simulations of user calibration tasks conducted by Axholt et al. [4] show much higher errors for calibrations with such a minimal number of correspondence points. This is due to our constraints on high marker distance and the unique method of line of sight intersection.

To evaluate our method in real data experiments, we conducted a user study with 12 individual persons. The group comprises of both expert users, who are familiar with wearing HMD gear and the calibration method, and novice users who use such a system for the first time. Each user conducts two or three acquisitions per eye, resulting in 70 data sets in total. At first, this data is used to compare our
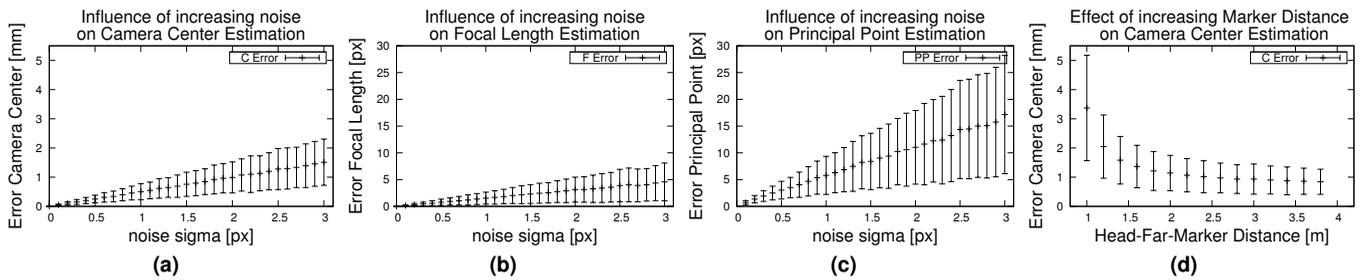
Fig. 5. (a) error of camera center (eye position), (b) focal length (distance eye to display) with increasing noise. (c) Error of principal point (Tilting of display to the eye) with increasing noise, and (d) influence of stationary marker distance while hand-held marker distance is fixed at 80cm. Each bar in the diagrams shows mean and standard deviation for one thousand calibrations.
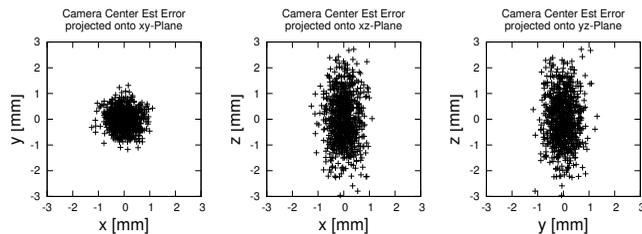

Fig. 6. Projections of estimated camera center for one thousand calibrations. Display FOV is 40 degrees, input data is distorted with $\sigma = 2.0$ pixel.

geometric calibration method to the result of a maximum likelihood estimator. As the latter calibration uses all observations individually in an optimization process, we refer to this calibration as "SPAAM", although we again estimated camera calibration matrix K and extrinsics separately which is in contrast to the recommended procedure in [32]. The mean errors for each target and the root mean squared error (RMSE) of all targets for all datasets is presented in Table 1. Here, we first show the reprojection error in pixel for the center target. As we used a near and a far marker this results in two distinct errors. In our approach, near center marker and far center marker are forced onto one line which results in a low error which is distributed uniformly. Then, we show the reprojection error in pixel for the other targets of the five-dot pattern. Lastly, we show the RMSE for all ten targets. In the SPAAM approach the RMSE is slightly reduced compared to our geometric approach. However, distributing the error uniformly for all targets, this also degrades the calibration accuracy in the center of the image, where the user of the head mounted display will naturally put most of his attention. As we used the same measurements for both our geometric calibration and SPAAM, the latter method also benefits from the proposed acquisition technique. The high marker distance guarantees non-degraded, non-planar measurements which are spread over a large distance. The reprojection errors from Table 1 give a quantitative result of the expected accuracy of the overlap of a real with a

virtual object for each eye individually. Additionally, we evaluated the effect of a misalignment in both eyes' displays on the triangulated object distance. For this evaluation we simulated inaccurate calibrations with specific RMSE and re-triangulated virtual objects at different distances. Triangulation was done by casting rays from the camera centers through the inaccurate pixel positions in the virtual displays and intersecting the rays in 3D space. Given our specific HMD display resolution and GFOV, a pixel reprojection RMSE of 2 pixel for each camera results in about 2 percent depth deviation for a focused object and 10 pixel for each camera results in about 10 percent depth deviation respectively. A subset of 6 users was additionally asked to try both single-point and two-point acquisition techniques and the time needed to complete data acquisition was measured. Table 2 shows that our acquisition technique is faster for both novice and expert users.

Table 2. Mean time required per eye for a calibration with ten acquisitions (single point) compared to five acquisitions (two points).

Users	Experiment	Mean Time (s)
Novice	SPAAM	64.24
	our Method	37.49
Expert	SPAAM	46.18
	our Method	31.84

4 EXPERIMENTAL EVALUATION ON DISTANCE JUDGMENTS

In this section we describe the experiment that we conducted to analyze the effect of the calibrated view frustum on distance estimation in VEs as derived by the calibration technique described in Section 3.

We performed all experiments in a 10m × 7m darkened laboratory room. In order to track the 6-DOF head position and orientation, an InertiaCube 3 (InterSense, 180Hz update rate) and an active infrared (IR) marker were attached on top of the OSTHMD. The Precision Position Tracker (PPT) X4 (WorldViz, 60Hz update rate) determined the position of the IR marker in the tracking space with sub-millimeter precision and sub-centimeter accuracy.

In the experiment subjects were seated in front of a table with a mounted chin rest and carton boxes on the tabletop (see Figure 7(a)) in order to perform the calibration process as described in Section 3. Both the chin rest and the boxes supported the subjects at stabilizing their head posture and the position of the hand-held marker to increase the data accuracy. The stationary marker was attached to a tripod standing in 2.5m distance to the chin rest.

During the experiment the room was darkened and a black curtain was fixed around the OSTHMD in order to reduce the subjects' perception of the real world, except in the baseline condition. The visual stimulus consisted of the virtual replica of the laboratory room and the target marker (see Figure 7(c)). The subjects received instructions on slides presented on the OSTHMD and responded via a Nintendo Wii remote controller. In order to focus subjects on the tasks no communication between experimenter and subject was performed during the experiment.

Table 1. Mean reprojection error in pixel for 70 calibration data sets performed by 12 users. We compare our geometric approach to a maximum-likelihood estimation on the same data set.

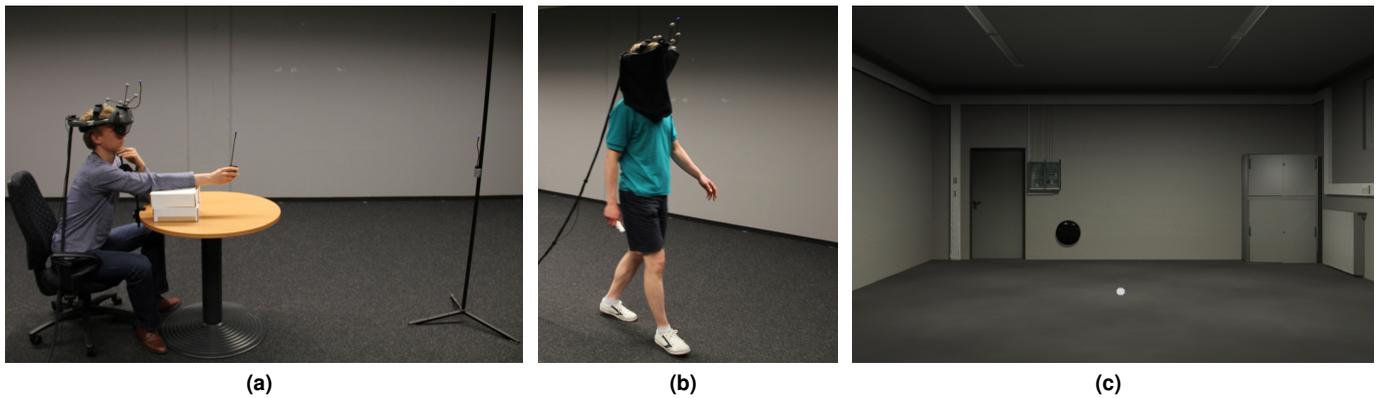|	geometric		SPAAM	
	near target	far target	near target	far target
Center	2.43	2.43	5.91	5.11
Upper left	7.77	7.10	5.67	4.96
Upper right	9.40	9.31	5.69	5.34
Lower left	7.06	7.53	5.07	5.61
Lower right	9.45	8.54	5.71	5.13
RMSE all	8.317		5.936	

Fig. 7. Illustration of the experiment setup: (a) subject during the calibration of the OSTHMD, (b) subject performing the blind walking task according to the judged distance of (c) the white target in the virtual laboratory replica presented on the OSTHMD.

4.1 Material and Methods

The angle of declination and the convergence angle of the eyes serve as cues to estimate the distance towards a shown target. Consequently, disregarding the user's individual position of the eyes probably affects the distance judgment in VEs. In order to analyze whether there is a difference in the user's ability to judge distances considering a fixed interpupillary distance (IPD), the user's individual position of the eyes or a calibrated camera view frustum, we tested the following experimental conditions:

- **Geometric calibrated rendering (GC)**: The intrinsic and extrinsic parameters of the OSTHMD were determined using the geometric calibration process (see Section 3) and applied to the rendering.

- **Eye position calibrated rendering (EC)**: Solely the eye position given by the extrinsic parameters was applied to the rendering without considering the intrinsic parameters. Instead, we used symmetric view frustums with three varying GFOVs, i. e., 50° (EC50), 60° (EC60) and 70° (EC70), which result in a magnified, native and minified representation of the VE.

- **Uncalibrated rendering (UC)**: The position of the tracked head marker was mapped one-to-one to the camera position assuming a fixed IPD of 65mm. We used the same GFOVs as in condition EC, i. e., 50° (UC50), 60° (UC60) and 70° (UC70).

- **Baseline (B)**: Subjects saw their real surrounding through the displays of the OSTHMD while wearing the same equipment as in the other conditions.

We used a within-subject design in this experiment and pseudo-randomized the order of the conditions. The target distances for each condition were 3m, 4m and 5m, and were each presented 4 times in randomized order resulting in 12 trials for each condition, i. e., $8 \times 12 = 96$ trials in total for each subject. In addition, subjects performed 2 training trials in each condition before the actual trials began.

At the beginning of each trial, subjects had to assume a specific initial position and orientation within the laboratory room with the support of two displayed markers on the HMD. One marker indicated the position and orientation of the destination and one marker indicated the current user's position and orientation. After the subjects confirmed a proper position and orientation by a button press on a Nintendo Wii remote controller, we displayed a target on the floor of the virtual laboratory replica at one of the aforementioned distances (see Figure 7(c)). The subjects were instructed to look at the target and afterwards press a button on the Wii controller. Following this, the screen went black and the subjects had to walk to the position of the target without vision (see Figure 7(b)). The next trial starts after

the subjects confirmed their final position by a button press on the Wii controller. We measured the Euclidean distance between the tracked IR marker at the initial and final position of the user. In the baseline condition, the target distances were indicated with a physical marker similar to the marker displayed in the other conditions.

In order to identify potential influences on the results, subjects filled out Kennedy's simulator sickness questionnaire (SSQ) [21] immediately before and after the experiments, as well as the Slater-Usoh-Steed (SUS) presence questionnaire [33], and a questionnaire collecting anthropologic data and informal responses.

4.2 Participants

6 female and 11 male (age 22-42, M $= 28.9$, SD $= 4.7$) subjects participated in the experiment. Subjects were students or members of the computer science, mathematics, psychology, and physics departments. All had normal or corrected to normal vision; 5 wore glasses and 2 wore contact lenses. 11 of the subjects had experience with walking in VR environments using a HMD setup. 9 had much video game experience, 2 some, and 6 none. All subjects were naïve to the experimental conditions. The total time per subject including pre-questionnaire, instructions, training, experiment, breaks, and debriefing took 75 minutes.

4.3 Results

Figure 8 shows the pooled results for the blind walking experiment with the target distances on the horizontal axis, and the mean walked distances on the vertical axis. The black plots show the results from the baseline condition and the colored plots show the walked distances for the conditions GC, EC and UC. The gray line corresponds to optimal distance judgments. The error bars show the standard errors.

In order to test for any significant main or interaction effects, we performed an 8 (viewing condition: GC, EC70, EC60, EC50, UC70, UC60, UC50, B) $\times$ 3 (target distance: 3m, 4m, 5m) repeated measure ANOVA on the ratio between the subject's estimated distance and the target distance. We analyzed the ratio between the subject's walked distance and the target distance in order to test if subjects were significantly better or worse in distance judgment for specific target distances. Since Mauchly's test of sphericity indicated that both the main effect of the viewing condition ($\chi^2(27) = 89.35, p < .001$) and the target distance ($\chi^2(2) = 14.45, p < .001$) violated the assumption of sphericity, we corrected the degrees of freedom using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .46$ for viewing condition, $\varepsilon = .62$ for target distance). All effects are reported as significant at $p < .05$. We found no significant main effect of the target distance ($F(1.24, 19.77) = 1.57, p = .229$) and no interaction effect ($F(14, 224) = .963, p = .492$), but a significant main effect of the viewing condition ($F(3.19, 51.0) = 29.13, \eta_p^2 = .65$) on the ability of subjects to judge distances.
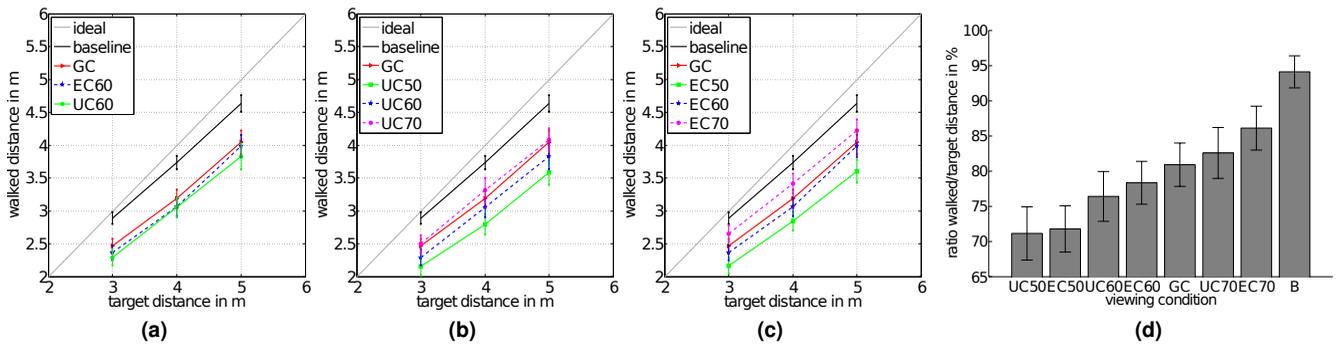
Fig. 8. Pooled results of the blind walking experiment. (a–c) show the mean walked distance on the y-axis and the target distance on the x-axis. The gray diagonal line represents an optimal distance judgment and the black plot corresponds to the baseline. (a) comparison of the geometric calibrated rendering condition (GC) with the eye position calibrated (EC) and uncalibrated (UC) rendering for the native DFOV of the OSTHMD. (b) shows condition GC in comparison with the uncalibrated rendering condition for the GFOVs 50° (UC50), 60° (UC60) and 70° (UC70). (c) compares condition GC with condition EC for the GFOVs 50° (EC50), 60° (EC60) and 70° (EC70). (d) shows the mean ratio between subject's walked distance and target distance. Since we found no significant main effect of the target distance, we pooled the data of the three target distances 3m, 4m and 5m. The error bars show the standard errors.

Figure 8(d) depicts the mean ratio between walked and target distance for all viewing conditions. We pooled the data of the three target distances, since we found no significant main effect of the target distance. The error bars show the standard errors. Subjects were quite accurate at judging distances in the baseline condition (M = 94.13%, SE = 2.28), even so they were equipped with the wired OSTHMD which provided a restricted FOV.

4.3.1 Impact of Geometric Calibration on Distance Judgment

Figure 8(a) shows the mean distance judgment of subjects for the applied levels of calibration, i. e., uncalibrated, eye position calibrated and geometric calibrated rendering without mini- or magnification of the VE. Considering subject's individual position of the eyes resulted in a decreased average distance underestimation of subjects in condition EC60 (M = 78.35%, SE = 3.05) compared to the uncalibrated rendering condition UC60 for a native DFOV (M = 76.41%, SE = 3.54). When the intrinsic parameters determined by the geometric calibration were additionally applied, the average distance underestimation of subjects decreased further in condition GC (M = 80.92%, SE = 3.08).

Although the geometric calibration improved the average distance judgment of subjects, we found a significant underestimation of distances for the condition GC, EC60 and UC60 compared to the baseline condition in a post hoc pairwise comparison with Bonferroni correction. Furthermore, the post hoc tests lack of any significant difference in distance judgment in the pairwise comparison of the conditions GC, EC60 and UC60.

4.3.2 Impact of GFOVs on Distance Judgment

Figure 8(b) shows the mean walked distance of subjects for the three GFOV conditions UC50, UC60 and UC70 compared to the geometric calibrated rendering condition, and, respectively, Figure 8(c) for the conditions EC50, EC60 and EC70 compared to condition GC. We found that the differences between the walked distances of subjects and the target distances decreased with increasing the GFOV and vice versa. A Bonferroni corrected pairwise post hoc analysis revealed that each GFOV increase of 10° results in significantly decreased differences between the walked distances of subjects and the target distances. That means, the walked distances of subjects were significantly closer to the target distances in condition UC70 (M = 82.59%, SE = 3.62) compared to UC60 (M = 76.41%, SE = 3.54) as well as in condition UC60 compared to UC50 (M = 71.15%, SE = 3.78), and respectively the same significant results were found for the conditions EC70 (M = 86.13%, SE = 3.12), EC60 (M = 78.35%, SE = 3.05) and EC50 (M = 71.79%, SE = 3.29).

4.3.3 Questionnaires

Subjects judged the task as easy (M = 0.88, SE = 0.23) on a 5-point Likert-scale (0 corresponds to very easy, 4 corresponds to very difficult). On a comparable Likert-scale subjects stated that they were not able to orientate themselves in the real laboratory room by seeing cues of the real surrounding (M = 0.35, SE = 0.21).

We found that subjects rated their sense of presence in the displayed virtual world on average with a score of 4.44 in the SUS questionnaire. The post-experiment SSQ score (M = 22.66, SE = 4.08) was significantly larger than the pre-experiment SSQ score (M = 10.34, SE = 3.03) of subjects ($t(16) = -2.646, p < .05, r = 0.55$). However, the mean difference between pre-SSQ and post-SSQ score of 12.32 is comparable to previously conducted studies in our laboratory and none of the subjects reported serious problems after the experiment.

4.4 Discussion

Substantially increasing the GFOV results in walked distances of users which are significantly closer to the shown target distances at the expense of a considerably distorted representation of the VE, which is an unsatisfactory situation in, for instance, training and simulation applications. Since a distorted representation of the VE might have an influence on the users' spatial impression of the laboratory replica, the smaller difference between the walked distances of users and the target distances does not necessarily indicate a reduced distance underestimation of users. Instead, users might perceive different distances of the shown targets when the GFOV is increased. The proposed geometric calibration method avoids large scene distortions, while improving the distance judgment of users by 4.5% on average for the measured target distances compared to an uncalibrated rendering.

Although the proposed geometric calibration improved the average distance judgment of users, subjects still significantly underestimate distances in the VE. Consequently, the calibration of the entire camera view frustum does not compensate for the distance underestimation effect, i. e., wrong or uncalibrated view frustums for stereoscopic rendering do not seem to be the main cause for distance underestimation in HMD setups.

5 CONCLUSION AND FUTURE WORK

In this paper we have proposed a calibration method for optical see-through head mounted displays and investigated the effect of calibrated view frustums on users' distance judgment in virtual environments. The calibration method requires no additional hardware such as cameras, rigs or calibration patterns and can be finished in approximately one minute for both eyes even by novice users. While classical optimizers minimize the reprojection error, our method by design

shows the smallest error in the center of the image and forces consistency along the center viewing ray. The error is therefore distributed in a way which fits better to the natural demands of viewing through a head mounted display.

In the conducted experiment to analyze the effect of the calibrated view frustums on users' distance judgment, we found the trend that the calibrated view frustum improves the distance estimation of users on average. However, the distance estimation of users still differs significantly from the real world baseline, which indicates that wrong or uncalibrated view frustums are probably not the main cause for distance underestimation in VR environments.

In the future we will investigate possible improvements of the calibration method to further reduce the needed user interaction and to make the calibration more robust against noise in the user acquired data. In addition, we will examine if the combination of calibrated view frustum with more realistically rendered virtual environments account for a greater extent of the distance underestimation effect.

ACKNOWLEDGMENTS

REFERENCES

[1] P. L. Alfano and G. F. Michel. Restricting the field of view: perceptual and performance effects. *Perceptual and Motor Skills*, 70:35–45, 1996.

[2] R. S. Allison, I. P. Howard, and J. E. Zacherx. Effect of field size, head motion and rotational velocity on roll vection and illusory self-tilt in a tumbling room. *Perception*, 28:299–306, 1999.

[3] K. Arthur. Effects of field of view on task performance with head-mounted displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 29–30, 1996.

[4] M. Axholt, M. Skoglund, S. D. Peterson, M. D. Cooper, T. B. Schoen, 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. *Human Factors and Ergonomics Society Annual Meeting Proceedings (HFES)*, 54(28):2427–2431, 2010.

[5] T. Banton, J. Stefanucci, F. Durgin, A. Fass, and D. Proffitt. The perception of walking speed in a virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(4):394–406, 2005.

[6] S. Ellis and K. Nemire. A subjective technique for calibration of lines of sight in closed virtual environment viewing systems. *Proceedings of the Society for Information Display*, 1993.

[7] H. Frenz, M. Lappe, M. Kolesnik, and T. Bührmann. Estimation of travel distance from visual motion in virtual environments. *ACM Transactions on Applied Perception (TAP)*, 4(1):1–18, 2007.

[8] A. Fuhrmann, D. Schmalstieg, and W. Purgathofer. Fast calibration for augmented reality. In *Proceedings of the Symposium on Virtual Reality Software and Technology (VRST)*, pages 166–167, 1999.

[9] A. Fuhrmann, D. Schmalstieg, and W. Purgathofer. Fast calibration for augmented reality. Technical Report TR-186-2-99-16, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/186, A-1040 Vienna, Austria, May 1999.

[10] 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 of IEEE/ACM International Symposium on Augmented Reality (ISAR)*, pages 165–174, 2000.

[11] Y. Genc, M. Tuceryan, and N. Navab. Practical solutions for calibration of optical see-through devices. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 169–175, 2002.

[12] S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster. Spatial calibration of an optical see-through head-mounted display. *Journal of Neuroscience Methods*, 173(1):140–146, 2008.

[13] M. A. Hagen, R. K. Jones, and E. Reed. On a neglected variable in theories of pictorial perception: truncation of the visual field. *Perception and Psychophysics*, 23:326–330, 1978.

[14] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[15] S. Hassan, J. Hicks, L. Hao, and K. Turano. What is the minimum field of view required for efficient navigation? *Vision Research*, 47(16):2115–2123, 2007.

[16] V. Interrante, L. Anderson, and B. Ries. Distance perception in immersive virtual environments, revisited. In *Proceedings of Virtual Reality (VR)*, pages 3–10. IEEE Press, 2006.

[17] V. Interrante, B. Ries, J. Lindquist, and L. Anderson. Elucidating the factors that can facilitate veridical spatial perception in immersive virtual environments. In *Proceedings of Virtual Reality (VR)*, pages 11–18. IEEE Press, 2007.

[18] A. Janin, D. Mizell, and T. Caudell. Calibration of head-mounted displays for augmented reality applications. In *Proceedings of the Virtual Reality Annual International Symposium (VRAIS)*, pages 246–255, 1993.

[19] S. E. M. Jansen, A. Toet, and N. J. Delleman. Effects of horizontal field-of-view restriction on manoeuvring performance through complex structured environments. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 189–189. ACM Press, 2008.

[20] J. A. Jones, J. E. Swan II, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of ACM SIGGRAPH Applied Perception in Graphics and Visualization (APGV)*, pages 29–35, 2011.

[21] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[22] S. Kuhl, W. Thompson, and S. Creem-Regehr. HMD calibration and its effects on distance judgments. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV)*. ACM Press, 2008.

[23] J. M. Loomis and J. M. Knapp. *Virtual and Adaptive Environments*, chapter Visual Perception of Egocentric Distance in Real and Virtual Environments, pages 21–46. Lawrence Erlbaum Associates, Mahwah, NJ., 2003.

[24] E. McGarrity and M. Tuceryan. A method for calibrating see-through head-mounted displays for AR. In *Proceedings of the IEEE/ACM International Workshop on Augmented Reality (IWAR)*, pages 75–84. IEEE Computer Society, 1999.

[25] R. Messing and F. H. Durgin. Distance perception and the visual horizon in head-mounted displays. *ACM Transactions on Applied Perception (TAP)*, 3(2005):234–250, 2.

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

[27] W. Robinett and J. P. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators and Virtual Environments*, 1(1):45–62, 1992.

[28] A. F. Seay, D. M. Krum, L. Hodges, and W. Ribarsky. Simulator sickness and presence in a high field-of-view virtual environment. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 784–785. ACM Press, 2002.

[29] K. M. Stanney and R. S. Kennedy. The psychometrics of cybersickness. *Communications of the ACM*, 40(8):66–68, 1997.

[30] F. Steinicke, G. Bruder, M. Lappe, P. Willemsen, and K. H. Hinrichs. Natural perspective projections for head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(7):888–899, 2011.

[31] A. Tang, J. Zhou, and C. B. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 161–168, 2003.

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

[33] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 1999.

[34] R. Warren and A. H. Wertheim. *Perception & Control of Self-Motion*. Lawrence Erlbaum Associates, 1990.

[35] P. Willemsen and A. A. Gooch. Perceived egocentric distances in real, image-based, and traditional virtual environments. In *Proceedings of Virtual Reality (VR)*, pages 275–276. IEEE Press, 2002.