Spatial calibration of an optical see-through head-mounted display

Article

Accepted Version

Spatial calibration of an optical see-through head mounted display

Stuart J. Gilson[1], Andrew W. Fitzgibbon[2], Andrew Glennerster[3]
May 21, 2008

[1]: Department of Physiology, Anatomy and Genetics
Parks Road,
Oxford
OX1 3PT
United Kingdom
[2]: Microsoft Research Ltd
7 JJ Thomson Avenue
Cambridge CB3 0FB
United Kingdom
[3]: School of Psychology and Clinical Language Scienes,
University of Reading,
Earley Gate,
Reading
RG6 6AL
United Kingdom

Pages (including figures and tables): 25
Figures: 9
Tables: 0
Corresponding author: Stuart Gilson
Department of Physiology, Anatomy and Genetics
Parks Road,
Oxford,
OX1 3PT
United Kingdom
Phone: +44 (0)1865 272460
Fax: +44 (0)1865 272469
Email: stuart.gilson@physiol.ox.ac.uk

Abstract

We present here a method for calibrating an optical see-through Head Mounted Display (HMD) using techniques usually applied to camera calibration (photogrammetry). Using a camera placed inside the HMD to take pictures simultaneously of a tracked object and features in the HMD display, we could exploit established camera calibration techniques to recover both the intrinsic and extrinsic properties of the HMD (width, height, focal length, optic centre and principal ray of the display). Our method gives low re-projection errors and, unlike existing methods, involves no time-consuming and error-prone human measurements, nor any prior estimates about the HMD geometry.

Keywords: HMD; optical see through; calibration; photogrammetry.

1 Introduction

Virtual Reality is a tool used increasingly frequently in vision research. Technology has traditionally limited research to graphically simplistic stimuli viewed under head-fixed conditions, but the affordability and increased availability of head mounted displays (HMDs), accurate tracking systems, and computers with sufficient power for real-time 3D graphics has allowed vision researchers to tackle many exciting new questions involving stereo and free head movements that they were unable to address before. However, these new technologies require new calibration techniques. The issue we address here is calibration of an optical see-through HMD. Unlike calibrating a monitor, it is not easy to access the small display screens within the HMD and make the angular measurements needed to calibrate the displays.

For some applications of virtual reality (e.g. navigation, gaming, architectural walk-throughs) a precise calibration may be unnecessary, although even here the minimum requirements of inter-ocular separation (Howarth, 1999; Mon-Williams et al., 1993) and appropriate gaze direction (Mon-Williams et al., 1998) must be met to avoid nausea, fatigue and eye discomfort. The degree of accuracy of calibration required to avoid discomfort and to provide a fusible stereo image (Wann et al., 1995; Mon-Williams et al., 1996) is less than that required for some other applications. For example, any time that a user has to interact (using tracked devices) with virtual objects, correspondence between visual and proprioceptive feedback should be as close as possible, permitting the usual hand-eye coordinated movements needed for such interactions. A calibration that fails to deliver this correspondence will impair the user's ability to complete the task, or may even make it impossible. Similarly, the correct portrayal of space is critical for visual psychophysics (e.g. Bingham et al., 2001; Sahm et al., 2005; Messing and Durgin, 2005; Tcheang et al., 2005; Glennerster et al., 2006) where it is often vital for the

validity of the experiment that the projection of stimuli at the eye is as close as possible to that from a real scene.

HMDs may be a long way from providing a perfect recreation of the visual input experienced in a natural environment (one of which is the mismatch between vergence, which is free to vary, and accommodation, which is fixed in the display (see Watt et al., 2005; Winterbottom et al., 2007). Nevertheless, it is worthwhile to attempt as faithful a rendition of the 3D geometry of the scene as possible, which requires an accurate calibration. The type of HMD determines which calibration method to use.

See-through HMDs fall into two categories: "video see-through" and "optical see-through". Video see-through HMDs employ a head-mounted camera (two cameras for stereo displays) which capture the scene as a digital image, upon which the computer generated images are overlaid and then fed to the display screens inside the HMD. The key strength of video see-through is that the combined real and virtual image is available as a single digital composite and, hence, various calibration techniques can be used to ensure a good overlay (Tuceryan et al., 1995; Azuma et al., 2001). This is clearly advantageous for many augmented reality applications which rely on the interaction of virtual and real objects.

On the other hand, optical see-through HMDs allow the observer to view the real world directly, with the computer-generated image super-imposed via a half-silvered mirror. These HMDs lead to the most natural viewing conditions for the observer, since there are no spatial offsets to consider when viewing the real world, but because there is no digital record of the scene, conventional computer vision calibration methods cannot be used to align the real and virtual scenes.

Until 2004, most optical see-through calibration methods required judgements from a human operator wearing the HMD. Tuceryan et al. (2002) presented a method (Single Point

Active Alignment Method, SPAAM) in which the user aligned pre-selected 2D markers in the display with a single, fixed location in the real world (of known position). Given a sufficient number of samples, they could estimate the transformation between the tracked centre of the HMD and the optic centre and orientation of the display. A similar stereo calibration could be carried out using a 3D marker. The authors discussed at length the ergonomics of expecting a human to do these kinds of alignment tasks, finding a solution that involved the minimum of time and skill. In order to quantify their calibration procedure they placed a digital camera inside the HMD and measured registration errors between real- and virtual-points. However, they did not use the camera itself for the calibration.

Genc et al. (2000, 2002) extended the SPAAM method to stereo, emphasizing just how difficult this extension was. Kato and Billinghurst (1999) and McGarrity and Tuceryan (1999) showed functionally similar techniques with similar results and problems. Fuhrmann et al. (2001) showed a calibration for augmented reality displays and devices that was essentially another refinement of the same technique, exploiting a human operator to align real- and virtual objects.

Like our present work, Owen et al. (2004) described a calibration method based on camera calibration (photogrammetry), placing a camera inside the HMD and recording an image of a calibration grid drawn in the HMD display. Since fiducial points in the calibration grid are known in both HMD coordinates and camera coordinates they were able to synthesize a 3D calibration rig and use that data for conventional camera calibration. Once calibrated, they then described an optional “phase II” user-centred refinement to the calibration, requiring users to align features in the HMD with real-world features.

Our method is similar to Owen’s but differs in a number of key ways. First, all measured positions are obtained directly from the tracker or computed within the tracker’s coordinate

frame and there are no error-prone measurements of object positions to be performed by a human operator. This greatly reduces errors in calibration. Second, our method uses images taken through the HMD optics, allowing us to include any optical distortions in our solution. Third, for the calibrations we tested in our HMD, we found no need to perform a "phase II" calibration, thus obviating the need for further human-centred, error-prone visual alignments. Finally, unlike Owen and the majority of other HMD calibration papers, we provide a quantitative evaluation of our calibration, by re-projecting known 3D positions back into the calibration images and measuring the pixel error between these and the original projections.

A summary of our method is as follows. First we capture the 3D coordinate locations of tracked markers, and their 2D projections in the camera image (section 2.1). We convert those 2D projections from camera coordinates to HMD coordinates (section 2.2) and then perform photogrammetry to obtain a calibration of the HMD (section 2.3). To measure the success of our calibration method, we re-project the 3D marker locations into the images via the new calibration and examine re-projection errors (section 3). Finally, we discuss our results and future work in section 4.

2 Methods

In order to display appropriate images that were correct for a virtual scene, images must be rendered using appropriate frustums. This goal was divided into two components: finding the intrinsic parameters of the display and finding the extrinsic parameters with respect to the HMD tracked centre. That is, we sought an estimate of the optic centre and direction of the principal ray (extrinsic parameters) and the width, height, focal length and principal point (the point through which the principal ray passes) of both displays (intrinsic parameters, see figure 1).

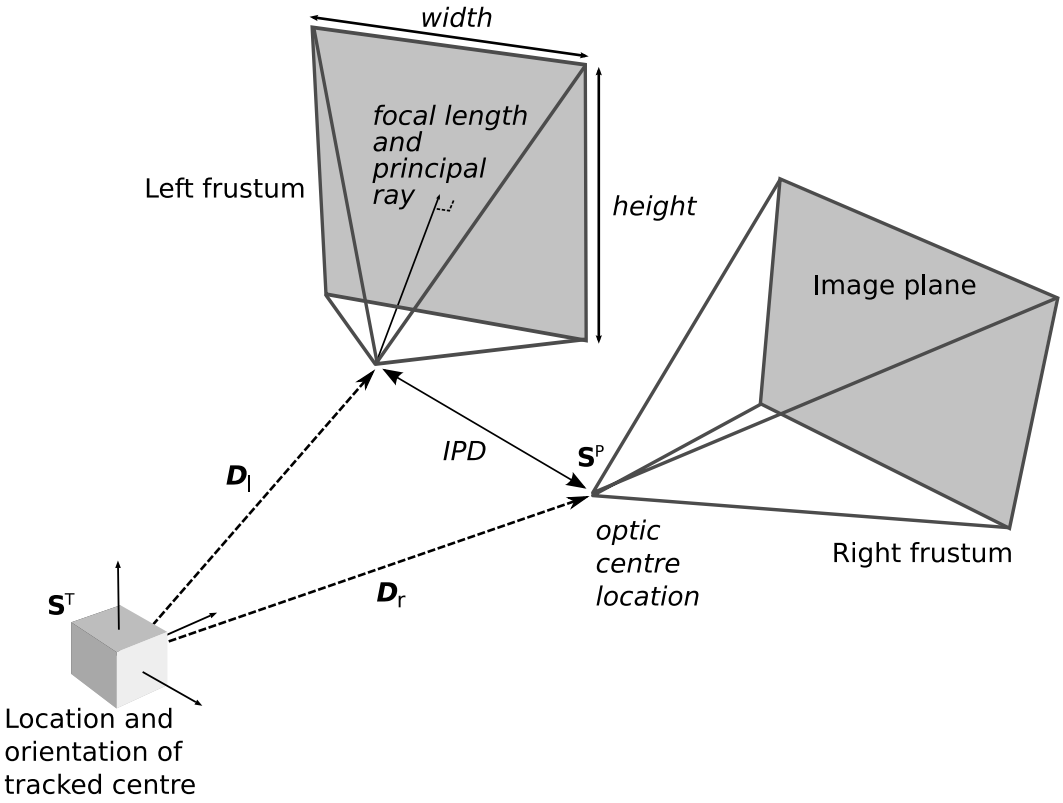


Figure 1: The goals of calibration are: to find the relationship between the tracked centre and the pose of each frustum (extrinsic calibration) with respect to the tracked centre; and to ensure that each frustum has the correct width, height and focal length for the left- and right-displays (intrinsic calibration).

The principal ray is normal to the image plane of the HMD and passes through the optic centre, denoting the orientation of the frustum. This is the information that is required, given an estimate of the pose and orientation of the head, to correctly render images to the left and right eyes. The resulting intrinsic and extrinsic matrices correspond directly to the *projection* and *model-view* matrices used in typical 3D graphics rendering libraries (such as OpenGL), and makes incorporating the calibration into a virtual reality system easy.

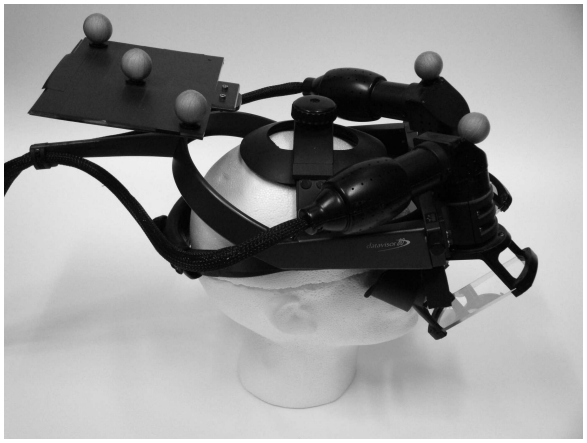
Our calibration procedure was essentially an enhanced form of a conventional camera calibration (Tsai, 1986). We placed a camera inside the HMD and captured multiple images of

a planar calibration object on which were mounted tracked markers. The calibration object was moved in successive images relative to the fixed HMD (section 2.1). The 3D locations of the tracked markers and their 2D projections in HMD coordinates are inputs to the calibration method. Calibration yielded the intrinsic parameters of the HMD (section 2.3.1), and the position of the optic centre in absolute tracker coordinates. Since we knew the position and orientation of the HMD tracked centre in the same coordinate frame it was straight forward to compute the difference between the two (section 2.3.2). This difference can then be combined with any future tracker coordinate of the headset to produce the dynamic, real-time optic centre location.

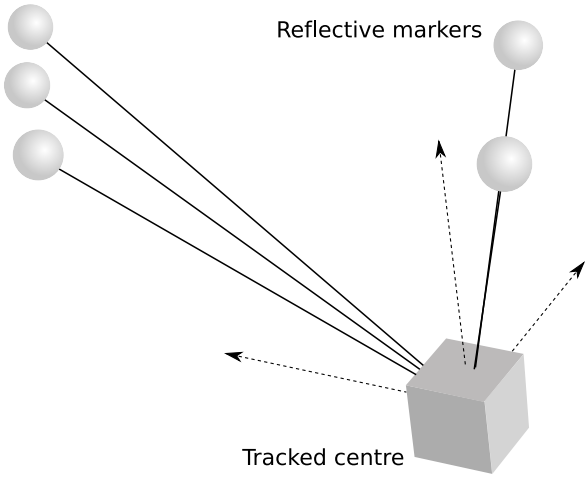
The tracked centre will be at some arbitrary point on the HMD. Our method did not require this point to be aligned with any structures on the HMD, nor did it need to be measured with respect to the HMD displays. The only requirement was that it remained in the same position with respect to the two optic centre locations of the HMD display — if the tracked centre is moved (e.g. because of upgrading equipment) then the HMD would require recalibration.

We calibrated an nVision Industries DataVisor80 HMD (figure 2), which has a nominal 80° field of view for each eye, and a 1280×1024 resolution. Pixel size was 3.4 arcmin. We placed the HMD over a mannequin head that had an Allied Vision Technologies (AVT) Pike camera (1388×1038 pixels resolution) mounted inside (figure 3). We were then able to capture multiple digital images of both the room and the images produced by the HMD (figures 4 and 5). The HMD image was a simple 21×21 regular array of dots, acting as vertices of a grid, and covered the entire 1280×1024 pixel display.

A Vicon Motion Systems MX3 real-time optical tracker was used to track the position of both the HMD (figure 2(b)) and the calibration object. The Vicon system has a nominal spatial accuracy of 0.1mm and orientation accuracy of 0.15°.



(a)



(b)

Figure 2: nVision DataVisor80 head mounted display with reflective tracker markers attached (a) and the HMD as modeled by the Vicon tracker software (b). All tracked models consist of 3 or more spherical markers rigidly attached to the object. The software reports the position and orientation of a 'tracked centre' which is rigidly related to the markers via an arbitrary transform. Our calibration method does not require any special positioning of this tracked centre, other than it remain rigid with respect to the HMD.

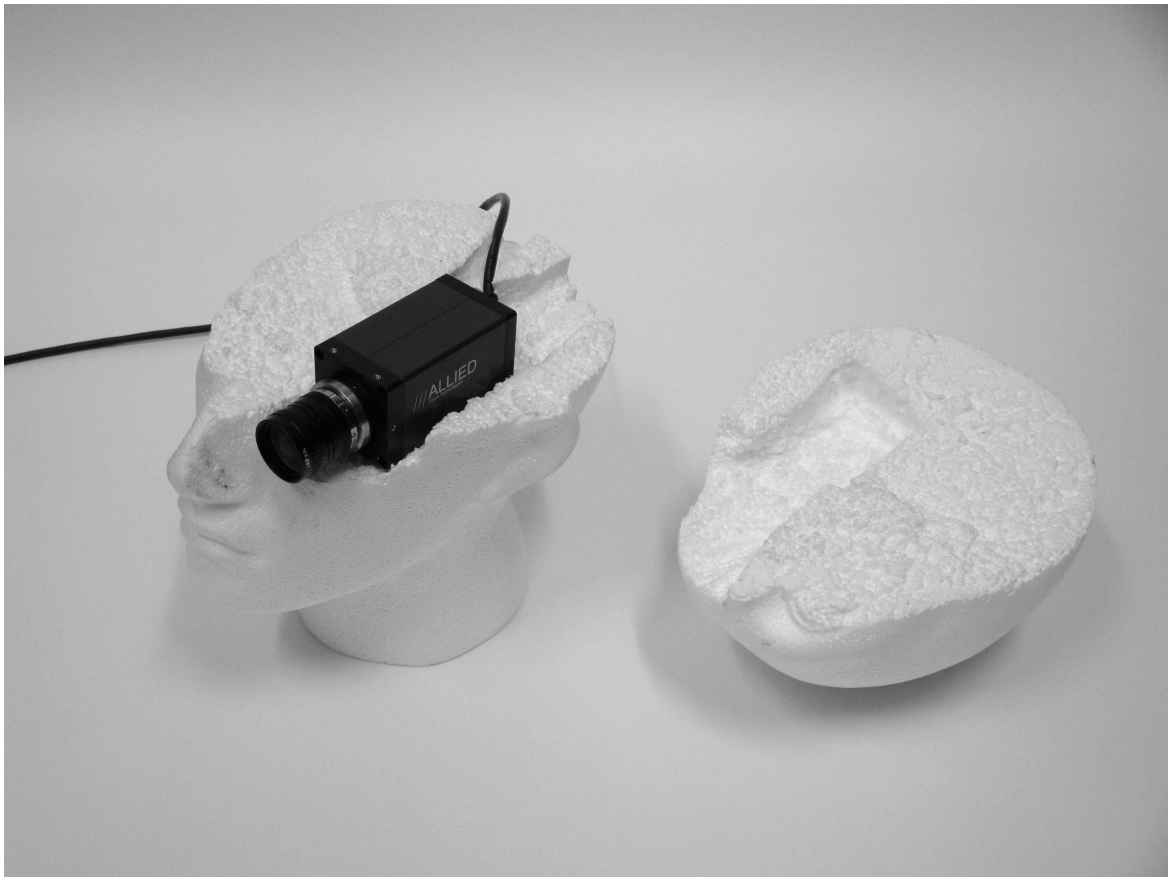


Figure 3: Mannequin head with camera in position.

In the remainder of this section, we discuss the calibration of only one HMD display. The steps required to calibrate the remaining display are identical and can be performed in a separate independent step. We do not believe there is a need to perform an additional stereo calibration.

2.1 Data capture

The camera was placed rigidly inside the mannequin head and the HMD placed on the head. The camera was positioned such that it could capture as much of the HMD image as possible, but especially the nasal region of each display where the virtual scene will be viewed binocu-

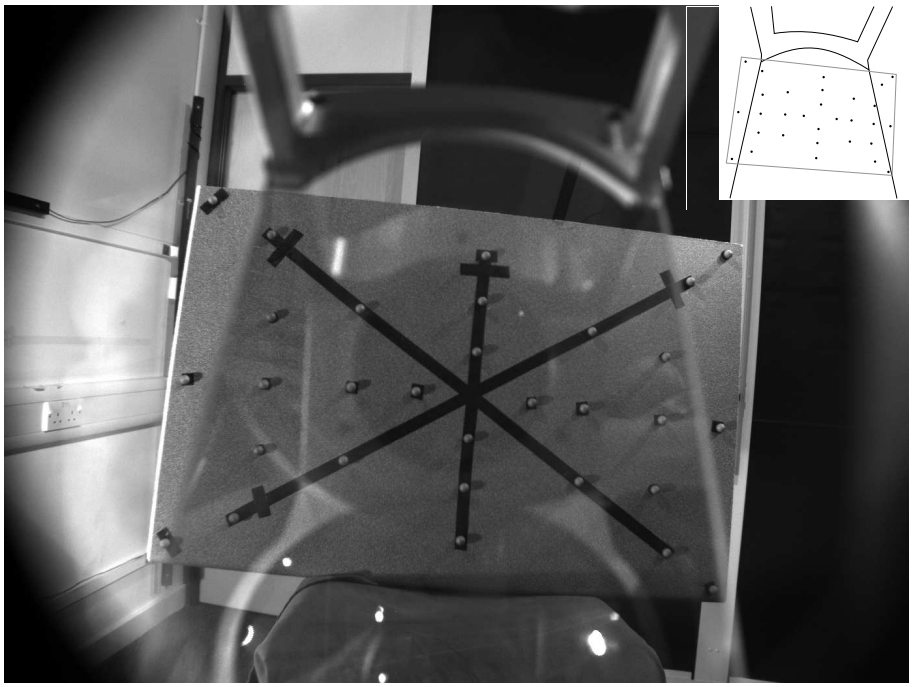


Figure 4: Typical image scene taken through the HMD image plane by the mannequin-mounted camera. The planar calibration object and tracked markers can be seen in the scene as small light-grey/white spots. Also visible is the half-silvered mirror that projects the HMD image onto the image plane, and the associated mounting bracket at the top of the image. Bright white patches are reflections from the optical tracker illumination and these did not affect calibration. Inset shows basic image composition, including the outline of the half-silvered mirror and support bracket.

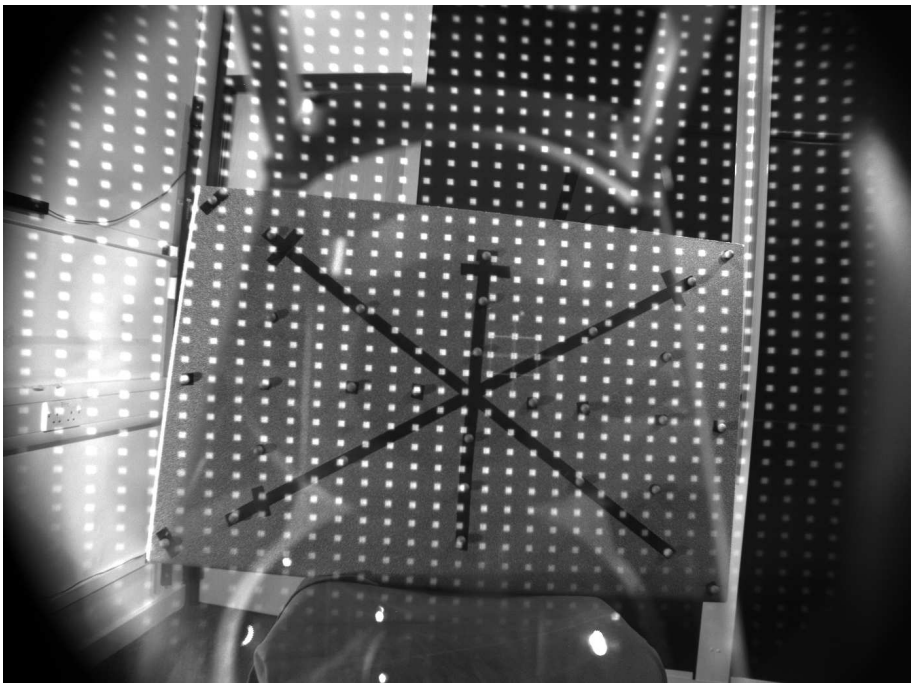


Figure 5: The same scene as figure 4 but with HMD grid image visible. Notice how the full HMD image is not visible through the circular aperture. In practice, we captured the image of the grid while opaque felt covered the HMD – occluding the world while retaining the grid image permitted easy automatic extraction of the grid vertices.

larly. The whole assembly then remained untouched throughout the data capture process. We recorded this HMD position and orientation, $\mathbf{S}^T$, as reported by the Vicon tracking system.

The calibration object was a rigid board upon which tracked markers were placed. The number of markers did not appear to be critical and between 10 and 35 have been used in the course of writing this paper. The calibration object was held steady during capture of each image and associated locations of the tracked markers. It was mounted rigidly on a wheeled trolley, which was moved across the floor between image captures.

For each image captured, we used the tracker to record the position of the markers which we denote $\mathbf{X} = \{(x_j, y_j, z_j) \,|\, j = 1 \ldots m\}$ for the m markers that were visible in all images and

where x, y and z are in the coordinate frame of the tracking system. We also extracted the corresponding projections of the markers into the camera images, $\mathbf{x}^{\mathrm{CAM}} = \{(x_j, y_j) \,|\, j = 1 \ldots m\}$.

We captured images of the calibration object at different places in the room giving n images of the markers and the corresponding n sets of 3D marker positions. The calibration object was always positioned to be entirely visible within each image and effort was made to obtain images of the object in as many portions of the camera image as possible, and at a variety of distances from the camera.

For users of tracking systems comprising only a few 6 degree-of-freedom (DOF) markers, an alternative calibration could be as follows. A Tsai grid (Tsai, 1986) could be mounted to a rigid board and a 6 DOF tracked marker rigidly attached to the same board. The spatial translation between the marker and centre of the Tsai grid would have to be measured. $\mathbf{X}$ would be comprised of synthetically generated grid vertices transformed by the position and orientation information obtained from the tracker and the measured translation. Although this is a possible route, this introduces a human measurement which our method avoids.

We stress here the importance of using 3D inputs to the calibration which are in the same coordinate frame as the tracker. This forces the photogrammetry to return a solution in the coordinate frame of the tracker. This becomes important in section 2.3.2 where we compare the extrinsic parameters from calibration with the recorded parameters of the HMD tracked centre.

In this paper we captured 4 data sets for each display, with each data set consisting of between 8 and 10 images and the corresponding tracked marker positions. Between captures, both the HMD and camera were removed from the mannequin head and then replaced. Doing so ensured that all data sets were independent, and better reflects the inter- and intra-personal positional variations the HMD would be subjected to. Removing and replacing the HMD on

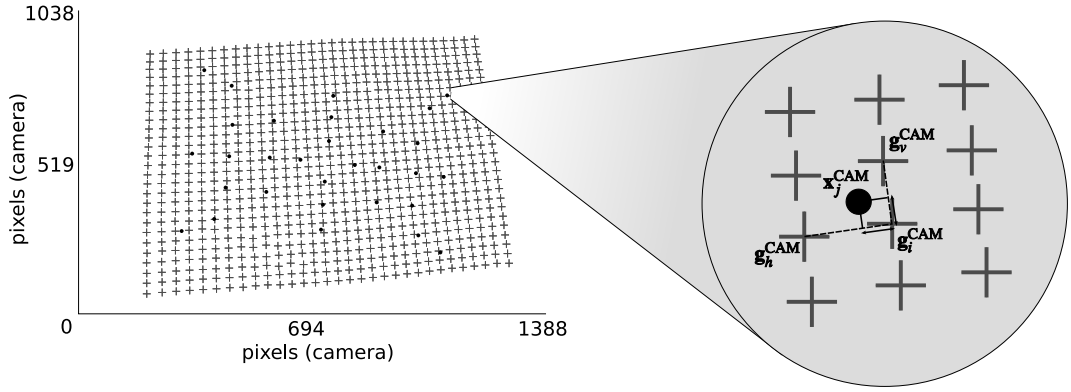
the mannequin in this manner altered the spatial relationship between the HMD displays and the camera by a few millimetres.

2.2 Converting from camera to HMD space

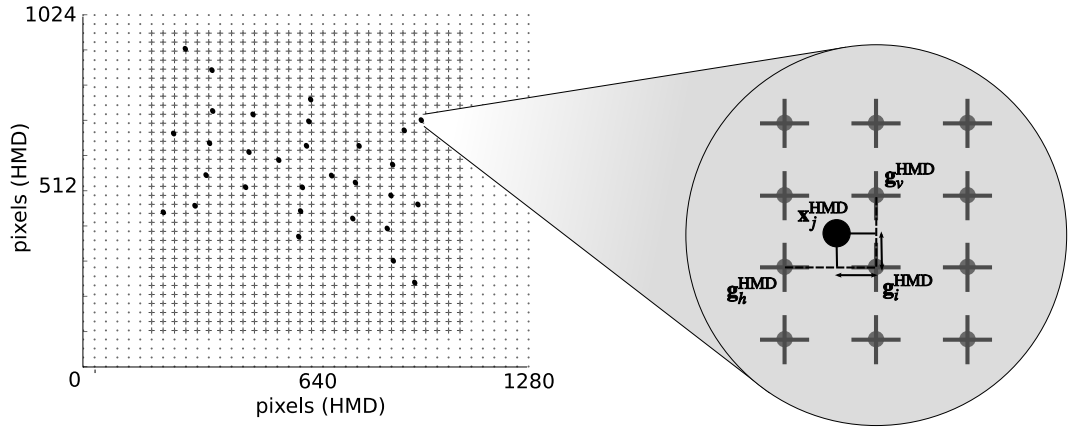
If camera calibration techniques are to be used to deliver a calibration of the HMD frustum rather than the camera, then the image positions of markers must be converted from camera coordinates to image positions expressed in the coordinate frame of the HMD display. The principle is as follows. Photogrammetry will deliver an estimate of the camera optic centre, which we will take as a good approximation to the HMD frustum optic centre. Thus, all the rays involved are the same as those in standard camera calibration. However, since the intersection of those rays with the HMD image plane are different, the intrinsic properties of the HMD frustum can be quite different, as can the principal ray defining the image plane orientation. Mapping camera image locations to HMD image locations for each marker is therefore the first step.

The HMD image was a regular array of dots whose vertices we denote by: $\mathbf{g}^{\mathrm{HMD}} = \{(x_i^{\mathrm{HMD}}, y_i^{\mathrm{HMD}}) \,|\, i = 1 \ldots 441\}$, where x^{HMD} and y^{HMD} are in HMD display coordinates. These grid vertices were extracted from the camera image (in figure 5) using a semi-automated centroid-detection algorithm, which delivered sub-pixel accuracy, to give: $\mathbf{g}^{\mathrm{CAM}} = \{x_i^{\mathrm{CAM}}, y_i^{\mathrm{CAM}} | i = 1 \ldots 441\}$ where x^{CAM} and y^{CAM} were camera image coordinates (figure 6).

Importantly, each grid vertex i in camera space had to be associated with the corresponding vertex i in HMD space and, thus, every vertex in $\mathbf{g}^{\mathrm{CAM}}$ was paired with the appropriate vertex in $\mathbf{g}^{\mathrm{HMD}}$. We did this by manually identifying the central grid vertex in the HMD image and computing correspondences relative to that vertex. Using this mapping between camera and HMD grid vertices, we could re-express the list of marker coordinates (currently in camera



(a)



(b)

Figure 6: (a) Extracted grid vertices (grey crosses) and marker projections (black discs) in camera coordinates. The figure exemplifies the kinds of non-linear distortions present in the HMD image. Inset shows a single marker's projection and its neighbourhood of grid vertices. The marker projection is stored as an interpolated position with respect to this neighbourhood. (b) The same marker projections now expressed in HMD coordinates The small grey circles represent all vertices of the HMD grid, while those with crosses indicate which vertices were visible in the camera image.

space) in HMD coordinates. For each marker point, $\mathbf{x}_j^{\mathrm{CAM}}$, we found the smallest triangle of grid vertices which encompassed it. We could then use linear interpolation to re-express the coordinate $\mathbf{x}_j^{\mathrm{CAM}}$ in HMD coordinates, giving $\mathbf{x}_j^{\mathrm{HMD}}$. Specifically, as illustrated in figure 6, the basis vectors $(\mathbf{g}_h^{\mathrm{CAM}} - \mathbf{g}_i^{\mathrm{CAM}})$ and $(\mathbf{g}_v^{\mathrm{CAM}} - \mathbf{g}_i^{\mathrm{CAM}})$ were mapped onto their equivalents in the regular HMD array, $(\mathbf{g}_h^{\mathrm{HMD}} - \mathbf{g}_i^{\mathrm{HMD}})$ and $(\mathbf{g}_v^{\mathrm{HMD}} - \mathbf{g}_i^{\mathrm{HMD}})$. Expressed in terms of these basis vectors, $\mathbf{x}_j^{\mathrm{CAM}}$ and $\mathbf{x}_j^{\mathrm{HMD}}$ are equivalent points.

The resulting arrays, $\mathbf{x}^{\mathrm{HMD}}$, one for each image, represented the marker projections in HMD coordinates (figure 6b) — they are now largely isolated from rotations of the camera with respect to the HMD, and from some types of projective distortion caused by the camera not being mounted perpendicular to the normal to the HMD image plane. We used a dense grid (21×21 vertices) in order to minimize any distortions introduced by the camera.

2.3 Camera calibration (photogrammetry)

Photogrammetry recovers the intrinsic and extrinsic parameters of a camera or frustum in three stages: initial estimate of intrinsic parameters; computing the corresponding extrinsic parameters for the given images; and minimization of all parameters in order to find a solution with the smallest re-projection error (e.g. Strobl et al., 2007).

The initial intrinsic estimates are obtained by computing the homography that maps each pair of 3D coordinate data ($\mathbf{X}$) to the corresponding 2D image projections ($\mathbf{x}^{HMD}$). The principal point location is initially set to the middle of the image and the image is assumed to be square (i.e. aspect ratio was 1.0). With these estimates defining a frustum, the optic centre position and orientation can be computed for a sample image using the corresponding homography.

These initial estimates are then refined by minimizing the root-mean-square errors be-

tween the reprojected positions and the original projections extracted in section 2.1. The resulting 5 intrinsic and 6 extrinsic parameters are more tolerant to measurement errors and produce lower re-projection errors than the initial estimates. Figure 7 shows the intrinsic and extrinsic parameters in graphical form for a set of sample data.

Our method differs from conventional photogrammetry in two key ways. First, conventional photogrammetry assumes the input data to be multiple 2D projections of a Tsai grid (Tsai, 1986) or similar, for which the software would synthesize appropriate 3D planar data. Since the goal of conventional photogrammetry is just to obtain intrinsic parameters of the camera, such synthesis is appropriate. However, we had real 3D data and this information can be used to generate appropriate extrinsic parameters. So our method bypassed this synthesis and instead used the known marker coordinates (**X**, gathered in section 2.1). The procedure for obtaining initial estimates of the frustum remained the same, but now the estimates of the extrinsic parameters reflected the location of the frustum in the tracker's coordinate frame. Second, standard photogrammetry will generate extrinsic parameters for all *n* images as there is no requirement for the camera to remain stationary. In our method, the camera remains stationary by design, so we have constrained our method to minimize for only one camera pose.

2.3.1 Obtaining the intrinsic calibration

The 5 intrinsic parameters were reported directly by the calibration software. We show here, the steps required to assemble the *projection* matrix needed by most graphics rendering software (such as OpenGL). The focal length (horizontal and vertical, f) and principal point (horizontal and vertical, c) should be placed in the 4x4 projection matrix thus:

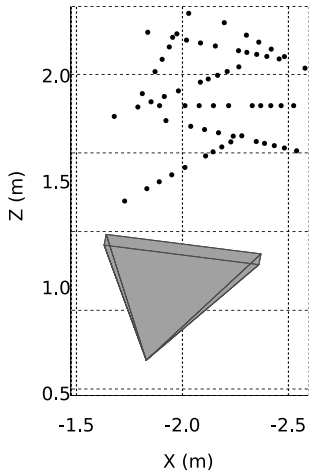


Figure 7: Plan view of frustum geometry produced by our photogrammetry method. Black filled circles represent the positions of the tracked markers.

$$\mathbf{P} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & \text{near} & \text{far} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The 5th intrinsic parameter (skew) is not used, and near and far clipping planes denoted by `near` and `far` will be application dependent and are not considered here.

2.3.2 Obtaining the extrinsic calibration

We denote the single extrinsic parameters returned from the photogrammetry as:

$$\mathbf{S}^P = \begin{bmatrix} \mathbf{R} & \mathbf{T}' \\ 0 & 1 \end{bmatrix}$$

where $\mathbf{S}^P$ was a 4×4 transformation matrix comprised of a 3×3 rotation matrix, $\mathbf{R}$, and a 1×3 translation vector, $\mathbf{T}$, which defines the frustum's position and orientation in the tracker

coordinate frame (see figure 7). However, a goal of calibration is to define the frustum relative to the HMD's tracked location and orientation. With this transform known, it is possible to take any real-time HMD tracked position and immediately find the optic centre from which the virtual scene can be rendered.

Using the HMD position and orientation recorded during data capture, $\mathbf{S}^T$, the transform between this and the photogrammetry optic centre, $\mathbf{S}^P$ is just:

$$\mathbf{D} = \mathbf{S}^P \times \text{inv}\left(\mathbf{S}^T\right)$$

This simplicity arose from $\mathbf{S}^P$ and $\mathbf{S}^T$ being in the same coordinate frame. $\mathbf{D}$ can then be applied to any future tracker transform ($\mathbf{S}_k^T$) to get a real-time, calibrated image in the HMD:

$$(x_j, y_j) = \mathbf{PDS}_k^T \left[\mathbf{X}_j \; 1\right]'$$

where (x_j, y_j) is the calibrated HMD image coordinate of marker $\mathbf{X}_j$ for the HMD at position $\mathbf{S}_k^T$ after it has passed through the standard rendering pipeline.

We can also break down $\mathbf{D}$ into its components, including the translation component along the optical axis of the HMD (the axis along which the inter-pupillary distance (IPD) adjustment is made). With this quantity known, we can modify our calibration to account for different users with differing IPDs, without any need to change the existing intrinsic parameters.

3 Results

We collected four independent data sets for each eye. In the following discussion, the data set used to obtain a calibrated frustum is called the *training* set, and the remaining three *test* sets. We obtained a calibrated frustum from each data set, and recorded the root-mean-square (RMS) pixel errors from re-projecting the tracked marker data through that frustum. The RMS

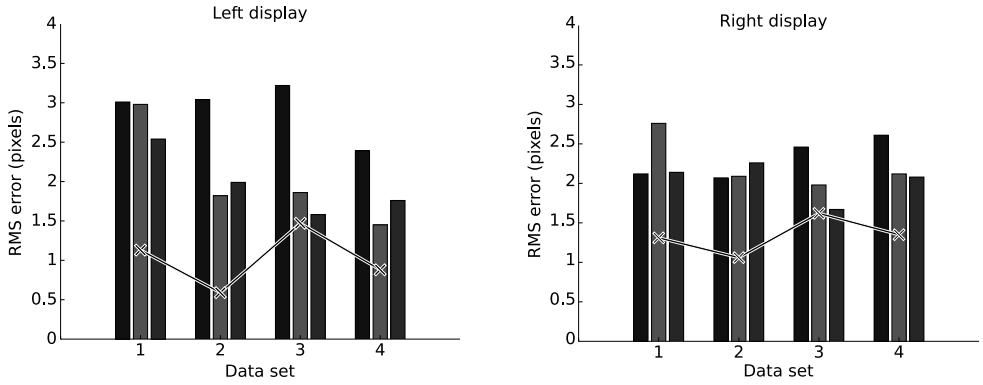


Figure 8: Root-mean-squared (RMS) errors between marker projections and re-projected points for left- and right-displays. Black crosses indicate RMS errors within each data set. Bars show RMS errors from each calibration tested on the images of the remaining three data sets.

errors from these training images are shown as black crosses in figure 8. The mean RMS error on each training data set was 1.02 pixels for the left eye, and 1.34 pixels for the right eye.

A more challenging performance measure was too see how well each frustum performed in the remaining three test data sets - low RMS errors here would indicate that the calibrated frustum was a good approximation to the optical properties of the HMD. Conversely, high RMS errors would indicate that the "training" data was not representative of the problem or that the minimization had specialized too much on the training data set and could not generalize to other data sets.

The bars in figure 8 show each calibrated frustum's RMS errors on the remaining three data sets. Here, the mean RMS errors were 2.3 pixels and 2.2 pixels for the left and right eyes respectively. Such low errors indicate that the calibrated frustums were good representations of the HMD displays. Figure 9 shows the original 3D marker positions ($\mathbf{X}$) reprojected through the new calibration and onto the HMD image plane for the left eye.

In addition to the quantitative measures summarized in figure 8, we have found a number

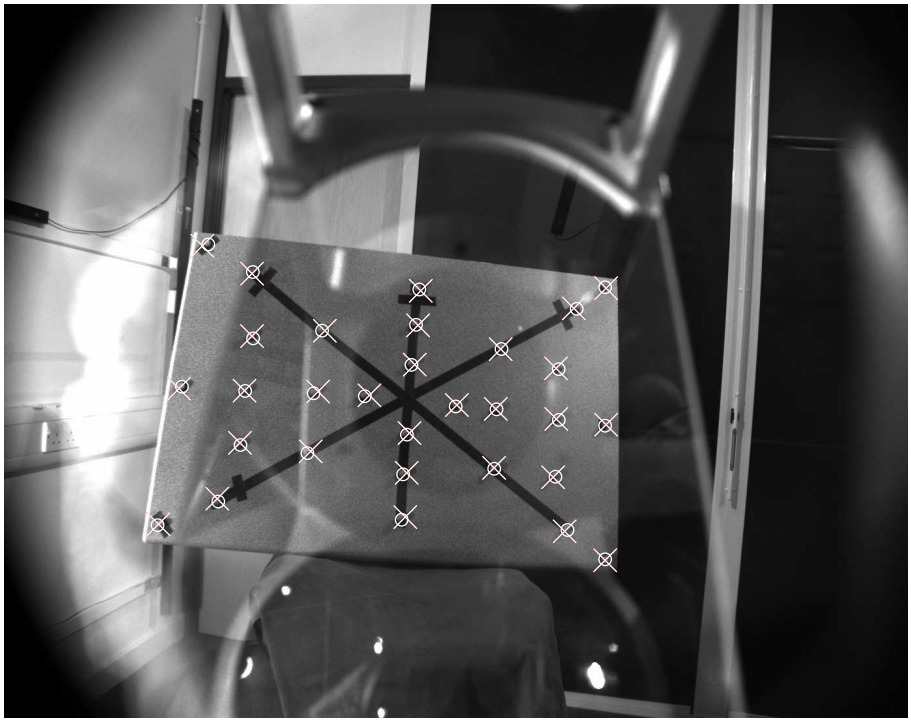


Figure 9: 3D marker positions (crosses) projected into original image after calibration and using the transform **D**. There was a RMS error of 1.0 pixels between these and the actual marker projections (circles). The underlying image is another example from the same training data set as the image in figure 4.

of informal, qualitative checks to be useful in assessing the quality of a calibration. First, torting the head around the line of sight produces a characteristic deformation of objects (elongation along one axis and compression along the orthogonal axis) if the aspect ratio is wrong. Second, trying to fuse an object rendered in the binocular region of the HMD should be comfortable and easy. Third, with the HMD in see-through mode, the virtual representation of real-world, tracked objects should overlie those real-world objects over the majority of the image plane. Successive HMD users confirmed these observations applied for our calibration.

Critically, we were able to obtain a good stereo image despite the fact that both frustums were calibrated completely independently of each other. Again, this suggests that the individual

calibrations are accurate. The only outstanding issues are that registration with real-world objects tended to deteriorate toward the extremities of the images, suggesting that the simple linear model of the image planes had not completely captured the actual distortions in each display. We intend to explore more complex non-linear distortion models in future work.

4 Conclusion

We have shown a robust and accurate calibration method applicable to optical see-through HMDs with accurate tracking systems. By rigidly placing a digital camera inside the HMD, we are able to capture images as seen through the HMD optics. With our modification to established photogrammetry techniques we obtained a HMD calibration that yielded accurate registration between real- and virtual-worlds. Additionally, we use the tracker to supply all 3D coordinates needed for calibration, and thus removed the need for any human measurements of continuous variables. We were able to use the camera images to measure the registration errors and report the numerical accuracy of our calibration.

Our method does not rely on human-supplied initial estimates, nor does it require any knowledge about the physical parameters of the HMD. Our method does not explicitly calibrate the left- and right-displays as a stereo pair. If the individual displays are calibrated accurately then an explicit stereo calibration should not be necessary, and our results support this notion.

In future work, we will extend this method to deal with non-linear distortions in the image display and the applicability of the method to non-see-through HMDs.

Acknowledgements

This work was funded by the Wellcome Trust.

References

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Comput. Graph.*, 21:34–47.

Bingham, G., Bradley, A., Bailey, M., and Vinner, R. (2001). Accommodation, occlusion and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *J. Exp. Psychol. Human*, 26:1314–34.

Fuhrmann, A. L., Splechtna, R., and Prikryl, J. (2001). Comprehensive calibration and registration procedures for augmented reality. In *Proc. Eurographics Workshop on Virtual Environ.*, pages 219–28.

Genc, Y., Sauer, F., Wenzel, F., Tuceryan, M., and Navab, N. (2000). Optical see-through HMD calibration: a stereo method validated with a video see-through system. In *Proc. Int. Symp. Augment. Real.*, pages 165–74.

Genc, Y., Tuceryan, M., and Navab, N. (2002). Practical solutions for calibration of optical see-through devices. In *Proc. Int. Symp. Mixed and Augment. Real.*

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., and Parker, A. J. (2006). Humans ignore motion and stereo cues in favour of a fictional stable world. *Curr. Bio.*, 16:428–43.

Howarth, P. A. (1999). Oculomotor changes within virtual environments. *Appl. Ergon.*, 30:59–67.

Kato, H. and Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. Int. Workshop Augment. Real.*, pages 85–94.

McGarrity, E. and Tuceryan, M. (1999). A method for calibrating see-through head-mounted displays for AR. In *Proc. Int. Symp. Mixed and Augment. Real.*, pages 75–84.

Messing, R. M. and Durgin, F. H. (2005). Distance perception and the visual horizon in head-mounted displays. *Trans. on Appl. Perc.*, 2:234–50.

Mon-Williams, M., Plooy, A., Burgess-Limerick, R., and Wann, J. (1998). Gaze angle: a possible mechanism of visual stress in virtual reality headsets. *Ergonomics*, 41:280–5.

Mon-Williams, M., Wann, J. P., and Rushton, S. (1993). Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthal. Physl. Opt.*, 13:387–91.

Mon-Williams, M., Wann, J. P., and Rushton, S. (1996). Design factors in stereoscopic virtual-reality displays. *J. Soc. Inf. Display*, 3:207–10.

Owen, C. B., Zhou, J., Tang, A., and Xiao, F. (2004). Display-relative calibration for optical see-through head-mounted displays. In *Proc. Int. Symp. Mixed and Augment. Real.*

Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B., and Willemsen, P. (2005). Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM T. App. Perc.*, 2(1):35–45.

Strobl, K., Sepp, W., Fuchs, S., Paredes, C., and Arbter, K. (2007). Camera calibration toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.

Tcheang, L., Gilson, S. J., and Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Res.*, 45:2177–89.

Tsai, R. (1986). An efficient and accurate camera calibration technique for 3D machine vision. *Proc. Comp. Vis. Pattern Recogni.*, pages 364–74.

Tuceryan, M., Genc, Y., and Navab, N. (2002). Single point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. *Presence-Teleop. Virt.*, 11:259–76.

Tuceryan, M., Greer, D., Whitaker, R., Breen, D., Crampton, C., Rose, E., and Ahlers, K. (1995). Calibration requirements and procedures for a monitor-based augmented reality system. *IEEE Trans. Vis. Comput. Gr.*, 1:255–73.

Wann, J. P., Rushton, S., and Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Res.*, 35(19):2665–799.

Watt, S. J., Akeley, K., Ernst, M. O., and Banks, M. S. (2005). Focus cues affect perceived depth. *J. Vision*, 5:834–62.

Winterbottom, M. D., Patterson, R., Pierce, B. J., Covas, C. M., and Winner, J. (2007). Depth of focus and visual recognition of imagery presented on simultaneously viewed displays: implications for head-mounted displays. *Hum. Factors*, 49:907–19.