An automated calibration method for non-see-through head mounted displays

Stuart J. Gilson[a,*], Andrew W. Fitzgibbon[b], Andrew Glennerster[c]

[a]*Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK*
[b]*Microsoft Research Ltd., Cambridge, UK*
[c]*School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK*

Abstract

Accurate calibration of a head mounted display (HMD) is essential both for research on the visual system and for realistic interaction with virtual objects. Yet, existing calibration methods are time consuming and depend on human judgements, making them error prone. These methods are also limited to optical see-through HMDs. Building on our existing approach to HMD calibration Gilson et al. (2008), we show here how it is possible to calibrate a non-see-through HMD. A camera is placed inside an HMD displaying an image of a regular grid, which is captured by the camera. The HMD is then removed and the camera, which remains fixed in position, is used to capture images of a tracked calibration object in multiple positions. The centroids of the markers on the calibration object are recovered and their locations re-expressed in relation to the HMD grid. This allows established camera calibration techniques to be used to recover estimates of the HMD display's intrinsic parameters (width, height, focal length) and extrinsic parameters (optic centre and orientation of the principal ray). We calibrated a HMD in this manner and report the magnitude of the errors between real image features and reprojected features. Our calibration method produces low reprojection errors and involves no error-prone human judgements.

Key words: head mounted display, non-see-through, calibration, photogrammetry; immersive virtual reality

1. Introduction

A head mounted display (HMD) can be modelled in a similar way to a conventional camera. Like a camera, it has both *intrinsic* parameters – focal length, aspect ratio, centre pixel – and *extrinsic* parameters – position of the optic centre and orientation of the principal ray. It also has an *image plane*, upon which pixels are drawn which represent the rays of light from the scene striking the virtual film. Collectively, these parameters define the *viewing frustum* which determines how the vertices of virtual objects are projected onto the image plane. All 3D rendering systems include the notion of a software frustum at their core, and this must be configured appropriately for the physical display in order to project the 3D scene correctly onto the image plane.

The issue, then, is to find the projection parameters for each eye's display in a HMD. HMD manufacturer specifications tend to be inadequate for this task, so the only other solution is to attempt to measure these display properties. Unlike calibrating a monitor display, it is usually difficult to get sufficient physical access to a HMD display in order

---

*Corresponding author

to make accurate measurements. Instead, we describe here a method based on photogrammetry (camera calibration) techniques.

HMDs fall into two categories: see-through and non-see-through. Of the see-through variety, there are two sub-categories: optical-see-through and video-see-through. Video-see-through displays are very popular in augmented reality applications, where a video camera mounted within the HMD sends digitized images of the real world to the graphics computer, which can then overlay computer graphics onto the images before sending them to the HMD to be displayed to the observer. Such displays are generally straight-forward to calibrate (Tuceryan et al., 1995; Azuma et al., 2001), since the issue of calibrating a conventional camera is well understood (Hartley and Zisserman, 2001). However, the optic centre of the camera is not at the observer's eye, and the resulting calibrated display will differ from that which the observer would see if they removed the HMD. For some applications, this discrepancy is acceptable (e.g. navigation, gaming, architectural walk-throughs), while for other applications involving interaction with real and virtual objects the offset between hand and eye may be detrimental to the task.

Optical-see-through displays generally use a half-silvered mirror placed in front of the observer's eyes, with a display device (cathode-ray tube or liquid crystal) mounted on the HMD. The half-silvered mirror permits rays of light from the real world to reach the observer, while also reflecting images from the display device. The observer sees a composite of the two sources, but with several limitations. Notably, the computer graphics (CG) image is effectively blended with the real world image and, as such, can never completely obliterate the real world. Hence, making virtual objects occlude real ones is impossible. Also, dark details in the CG image will become washed out by bright areas of the real world. Of more relevance here, is that there is no digital record of the rays entering the HMD optics and, so, existing camera calibration methods cannot be used. Critically, without an accurate calibration, virtual objects will register poorly with real objects, making optical-see-through a poor choice for augmented reality.

Non-see-through HMDs usually place the display device directly in front of the observer's eye, and are thus optically much simpler than either of the other two types of HMD. This does not make them any easier to calibrate, though. While the real world is not visible to the observer, and so registering virtual objects with real world ones is not an issue, a correct calibration is still important. Failure to calibrate correctly can lead to observers misinterpreting the virtual world (for example, they often underestimate distances to objects which can be a symptom of an incorrect calibration (Kuhl et al., 2009). Inadequate calibration can also lead to users experiencing premature fatigue and possible onset of nausea (Mon-Williams et al., 1993, 1998; Regan, 1995).

Thus, there is a demand for a reliable calibration procedure for both optical-see-through and non-see-through HMDs. Several methods of calibration that have been described previously have relied on human judgements. For non-see-through HMDs this involves either the observer removing the headset and comparing the widths or locations of objects presented in the real world with those shown in the headset (Kuhl et al., 2009) or judging the separation of features in the HMD image with an after-image produced by a bright flash (Rinalducci et al., 1996). Both types of method suffer from the inevitable imprecision and inaccuracy of human judgements. These methods were also designed to calibrate a restricted range of parameters (horizontal and vertical scale, pitch and pin-cushion distortion).

Optical see-through HMDs have the advantage that the real world and computer-generated image can be viewed si-

multaneously. In the literature on optical see-through HMDs, the most extensively covered calibration method, SPAAM (single point active alignment method), uses a human observer to calibrate the display by wearing the HMD and positioning their head in order to align HMD image points with real world objects whose locations are known (Tuceryan et al., 2002). When this alignment is achieved, the HMD position and pose is recorded from the tracker, and the procedure is repeated with more image/world coordinate pairs until sufficient data has been gathered to estimate the projection parameters. This is a time-consuming process, requiring a skilled observer to make numerous, potentially erroneous judgements. Also, there can be high variability in the results due to the difficulty of performing such an alignment task with a free-moving head. Finally, images in the HMD are at a fixed accommodative distance but it is desirable to match image objects with real world objects at a range of distances in order to estimate the projection parameters correctly. Human observers can find it difficult to match the visual direction of real and virtual objects at substantially different distances (and, hence, with different accommodative demands).

Owen et al. (2004) used a calibration method that at first sight appears similar to ours, with at least one fundamental difference. In our method, all real-world coordinates (of both HMD location and visible markers) are reported by a single tracking system, which obviates the need for error-prone human measurements such as are used by Owen et al. (2004). Unlike Owen et al. (2004) and the majority of other HMD calibration papers, we provide a quantitative evaluation (i.e. root-mean-square error) of the extent to which our calibration has been successful.

We have described previously a calibration method applicable to see-through HMDs (Gilson et al., 2008). Here, we describe a method suitable to *non-see-through* HMDs, thus making our method appropriate for the majority of HMDs currently used for virtual reality applications. In the current method, we used a dynamically tracked object which the user could move freely within the volume visible through the HMD display. This made it easier for the user to cover a wide range of the HMD field of view and hence to obtain a more complete and accurate calibration than would be obtainable with a small number of statically positioned objects.

2. Methods

Our aim was to find estimates of the intrinsic and extrinsic matrices which define the HMD display (shown pictorially in figure 1). The extrinsic matrix describes the location of the optic centre of the HMD display and the orientation of the principal ray (in world coordinates):

$$\mathbf{S} = \begin{bmatrix} \mathbf{R} & \mathbf{T}^{\mathrm{T}} \\ 0 & 1 \end{bmatrix} \tag{1}$$

where $\mathbf{R}$ is a 3×3 rotation matrix and $\mathbf{T}$ is a 1×3 translation matrix, i.e. 6 extrinsic parameters.

The intrinsic matrix (c.f. Hartley and Zisserman, 2001, page 150) comprises the focal length (f, in both horizontal and vertical directions, thus denoting aspect ratio), centre pixel location (c), and image skew (s):

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$
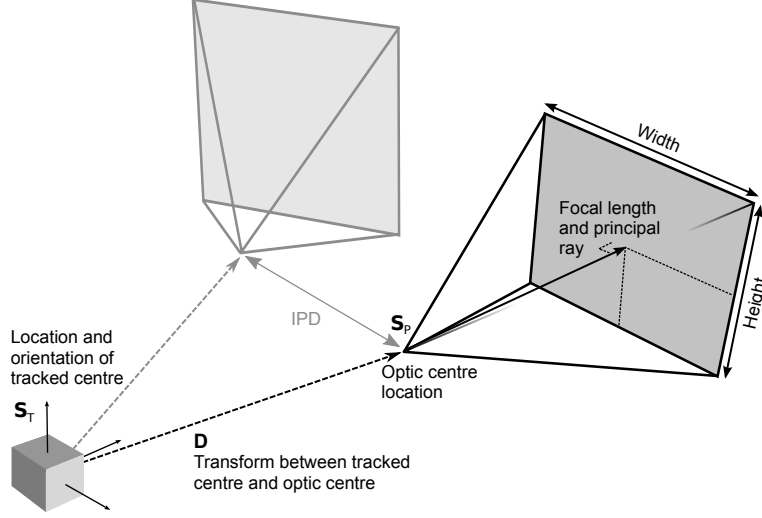
3

Figure 1: The purpose of calibration is to find values for the displays' internal projection parameters, and also the position and orientation of the displays with respect to the HMD's tracked centre.

These 11 parameters define a linear projection model, transforming 3D coordinates of virtual objects into image space. In order to obtain projected pixel coordinates (necessary for rendering), we reformulate $\mathbf{K}$ into a format used in computer graphics languages (like OpenGL):

$$\mathbf{P} = \begin{bmatrix} \frac{2 \times \mathrm{ncp}}{\mathrm{right} - \mathrm{left}} & s & \frac{\mathrm{right} + \mathrm{left}}{\mathrm{right} - \mathrm{left}} & 0 \\ 0 & \frac{2 \times \mathrm{ncp}}{\mathrm{top} - \mathrm{bottom}} & \frac{\mathrm{top} + \mathrm{bottom}}{\mathrm{top} - \mathrm{bottom}} & 0 \\ 0 & 0 & -\frac{\mathrm{fcp} + \mathrm{ncp}}{\mathrm{fcp} - \mathrm{ncp}} & -\frac{2 \times \mathrm{fcp} \times \mathrm{ncp}}{\mathrm{fcp} - \mathrm{ncp}} \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{3}$$

where:

$$\mathrm{left} = -\mathrm{ncp} \times \frac{c_x}{f_x}, \qquad \mathrm{right} = \mathrm{ncp} \times \frac{w - c_x}{f_x},$$

$$\mathrm{bottom} = -\mathrm{ncp} \times \frac{c_y}{f_y}, \qquad \mathrm{top} = \mathrm{ncp} \times \frac{h - c_y}{f_y}$$

define the borders of the near clipping plane (w and h are the pixel width and the height of the graphics viewport, and ncp and fcp are the near- and far-clipping planes – these are application specific and not covered further here).

Briefly, the calibration procedure was as below:

1. The HMD was rigidly mounted on a stable table, and it's position and orientation were recorded by the tracking system.

2. The camera was placed inside the HMD such that the camera was approximately fronto-parallel to the HMD display, and could capture as much of one of the displays as possible.

3. A chequerboard grid image was displayed in the HMD display, and captured by the camera. The image was post-processed to locate and record the grid vertices in the camera image.
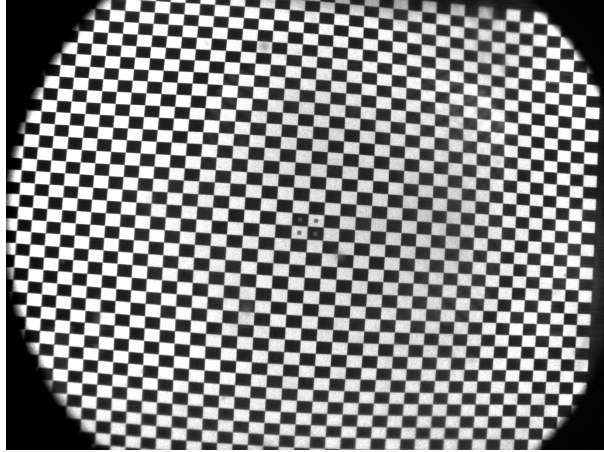
4

Figure 2: Chequerboard pattern displayed in the HMD and captured by the camera. Note spots indicating the logical centre of HMD display. The vignetting of the grid seen here was due to the camera lens.

4. The HMD was carefully moved away from the camera, without moving the camera or table.

5. The calibration object was then waved around within the field of view of the camera. The centroids of the markers on the calibration object were extracted from the camera images in real time. The 3D location of the same markers were also recorded by the tracking system.

6. The 3D marker positions and the 2D image centroids formed the input to the off-line calibration calculations.

7. Typically, this procedure would then be repeated for the other HMD display, although we only describe calibration of one display here.

To capture images for calibration, we used an AVT Pike (1280×1024 pixels resolution, ≈65° field of view). The camera was configured to have the smallest lens aperture possible to maximize its depth of field, while still enabling sufficient illumination for the necessary image processing. The shutter time was also minimized as much as illumination allowed, to increase the frame rate of the images and to minimize temporal blurring between frames of any moving objects.

We calibrated an nVis SX111 HMD, which has a nominal 76° horizontal field of view in each display, giving a binocular field of view of ≈102° with 66% stereo overlap. Each display comprises 1280×1024 pixels and refreshes at 60Hz. The camera was positioned inside the independently supported HMD in such a way that it could capture as much of the HMD image as possible. The location and pose of the HMD, $\mathbf{S}_T$, was recorded using a 6 degree-of-freedom real time optical tracking system (Vicon Motion Systems MX3). The HMD displayed a simple chequerboard pattern (41×41 vertices) and an image of this was captured using the camera (figure 2). The vertices of the chequerboard in this image were extracted using image processing. Using a salient feature in the middle of the HMD image, we were able to relate the known vertices of the HMD chequerboard to corresponding vertices in the camera image.

This allowed us to generate a mapping between camera image and HMD coordinates. If the HMD vertices are denoted by:

$$\mathbf{g}^{\mathrm{HMD}} = \left\{ \left(x_i^{\mathrm{HMD}}, y_i^{\mathrm{HMD}} \right) | i = 1...1681 \right\} \tag{4}$$

where x_i^{HMD} and y_i^{HMD} are HMD coordinates, then, for each vertex i there exists:

$$\mathbf{g}^{\mathrm{CAM}} = \left\{ \left(x_i^{\mathrm{CAM}}, y_i^{\mathrm{CAM}} \right) | i = 1...1681 \right\} \tag{5}$$

where x_i^{CAM} and y_i^{CAM} are the coordinates of the corresponding vertex in camera coordinates. This allowed any camera coordinate to be converted to a HMD coordinate using interpolation. If $\mathbf{x}_t^{\mathrm{CAM}}$ denotes a real world point captured by the same, unmoved camera at time t, then we found the smallest triangle of the chequerboard that encompassed it, whose vertices we denote $\mathbf{g}_i$, $\mathbf{g}_h$ and $\mathbf{g}_v$. We then used linear interpolation to re-express the camera coordinate in HMD coordinates using the basis vectors $(\mathbf{g}_h^{\mathrm{CAM}} - \mathbf{g}_i^{\mathrm{CAM}})$ and $(\mathbf{g}_v^{\mathrm{CAM}} - \mathbf{g}_i^{\mathrm{CAM}})$ and their equivalents $(\mathbf{g}_h^{\mathrm{HMD}} - \mathbf{g}_i^{\mathrm{HMD}})$ and $(\mathbf{g}_v^{\mathrm{HMD}} - \mathbf{g}_i^{\mathrm{HMD}})$ respectively. Expressed in terms of these basis vectors, $\mathbf{x_t}^{\mathrm{CAM}}$ and $\mathbf{x_t}^{\mathrm{HMD}}$ are equivalent points (Gilson et al., 2008).

This essentially rectifies the camera data and means calibration of the camera itself is not required. The HMD calibration procedure relies only on the assumption that the camera coordinates can be mapped onto HMD coordinates by a linear (affine) mapping within the region of a single square of the the chequerboard.

We removed the HMD from the camera and – crucially – ensured that the camera did not move. The camera then captured frames of the real world in which we moved a calibration object along a random trajectory within the field of view of the camera. The operator's objective when generating the trajectory was to 'paint' as much as possible of the camera image with projections of the markers. In addition, it is beneficial for the accuracy of the calibration results to include as wide a depth range as possible in the trajectory (figure 3). The calibration object consisted of several rigidly positioned markers forming an asymmetric planar pattern. The asymmetry allowed the Vicon tracking system to report the object's position unambiguously. By using a number of markers on a rigid object, the Vicon tracking system could report each marker's location with greater accuracy than if just one marker was used. Our calibration software extracted the 2D centroids of the markers from the camera images in real time, while also recording the markers' 3D locations reported by the Vicon tracker.

Aside from being attached to the same physical object, the data from each marker is entirely independent of the others in the subsequent stages of calibration, so we concatenated their coordinates together to form one large trajectory. For simplicity, the following explanation refers to this compound trajectory as if it were generated by just one marker.

We thus obtained, for each trajectory, approximately 8000 quintuples $\left[\mathbf{x}_t^{\mathrm{CAM}} \mathbf{X}_t \right]$ representing the instantaneous location of the marker at any moment ($\mathbf{X}_t$) and its projection on the camera image ($\mathbf{x}_t^{\mathrm{CAM}}$). Before these coordinates could form the input to the camera calibration routine, all 2D image locations were transformed into HMD coordinates using the basis vectors described above, to give $\mathbf{x}_t^{\mathrm{HMD}}$. This was a critical step, since without this, the subsequent photogrammetry would produce an intrinsic model of the camera, not the HMD.

We computed initial values for the intrinsic and extrinsic matrices by finding a single homography that mapped $\mathbf{x}_t^{\mathrm{HMD}}$ onto the corresponding $\mathbf{X}_t$ (Hartley and Zisserman, 2001, page 92). The resulting estimates for focal length,
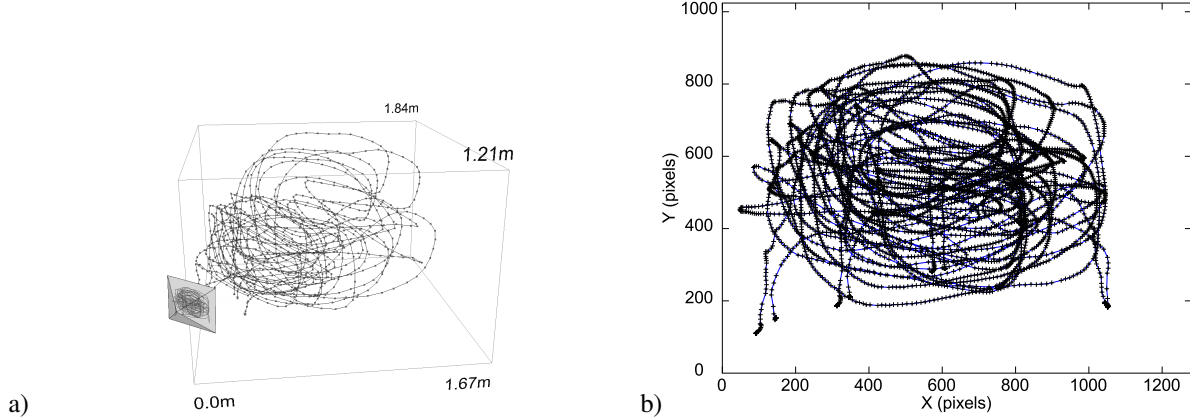
a)   b)

Figure 3: Plots of the 3D marker location and corresponding 2D centroid locations extracted from the camera images. a) Plot of 3D marker locations (circles, every fifth location shown for clarity) as the calibration object was moved along a trajectory (grey line). Camera only shown schematically, since at this stage the viewing parameters are not known. b) 2D marker centroids extracted from camera images as the object, consisting of 4 markers, was moved. The data points are shown linked to illustrate the trajectory the markers were moved through, but the ordering information is not needed for calibration since each data point is entirely independent.

aspect ratio, centre pixel, optic centre location and principal ray direction were then used as a starting point for a simplex minimization (Lagarias et al., 1998). The cost function was the reprojection error – that is, the root-mean-square (RMS) difference (in pixels) between the original projections $\mathbf{x}^{\mathrm{HMD}}$ and the new projections computed by:

$$(x_t, y_t, z_t, w_t) = \mathbf{P}\mathbf{S}_{\mathrm{P}}[\mathbf{X}_t 1]^{\mathrm{T}} \tag{6}$$

(the depth component, z_t, of the homogeneous coordinate can be discarded leaving a simple difference vector between $\mathbf{x}_t^{\mathrm{HMD}}$ and (x_t, y_t)). The matrix $\mathbf{S}_{\mathrm{P}}$ is the location and pose of the camera, estimated by calibration, in absolute tracker coordinates. However, we needed to know these parameters in coordinates relative to the tracked HMD position, so that the correct image could be rendered for any HMD position or orientation. We thus computed $\mathbf{D}$ as the single transform between HMD tracked centre and the HMD display:

$$\mathbf{D} = \mathbf{S}_{\mathrm{P}}\mathbf{S}_{\mathrm{T}}^{-1} \tag{7}$$

This simplicity arises from $\mathbf{S}_{\mathrm{P}}$ and $\mathbf{S}_{\mathrm{T}}$ being in the same coordinate frame. We now have a projection matrix (of the HMD display) and a modelling matrix which can be used directly in a 3D programming language such as OpenGL by post-multiplying it with the modelling matrix from the tracker:

```
// Switch to intrinsic (projection) matrix mode.
glMatrixMode(GL_PROJECTION);
// Load intrinsic matrix, P.
glLoadMatrix(P);
// Switch to extrinsic (modelling) matrix.
glMatrixMode(GL_MODELVIEW);
// Load HMD_to_optic_centre transform.
glLoadMatrix(D);
// Incorporate tracker transform.
glMultMatrix(S_T);
```

This procedure was repeated for both displays in the binocular HMD, with each display calibrated independently. In our experience, using this method we have found no need to perform an explicit stereo calibration.

Calibration accuracy was quantified as the root-mean-square reprojection error (in pixels) measured for all marker positions in the trajectory. It may seem counter-intuitive to use reprojections as an error measure here, since a non-see-through HMD has no real-world image in which to make such reprojections. The important point, of course, is that the camera did not move between capturing the HMD chequerboard and the corresponding marker trajectories and, thus, each camera pixel corresponded to the same ray irrespective of whether the HMD was present or not.

3. Results

We collected 4 trajectories using the method described above, each consisting of at least $8,000$ samples. We then physically moved the HMD to a new location within the tracked volume and collected another 4 trajectories. We repeated this procedure until we had acquired 6 sets of 4 trajectories for one HMD display. The importance of moving the HMD to new locations is two-fold. Firstly, the spatial relationship between the HMD and the camera will change each time – the effects of this change will be discussed below. Secondly, the relocation demonstrates that the calibration method works irrespective of HMD position, orientation, angle of inclination, etc.

We first present a typical calibrated solution from one trajectory. Figure 4 shows the solution for the trajectory shown in figure 3. Here, the 3D marker data ($\mathbf{X}$) is reprojected through the calibrated projection parameters to produce a new set of 2D pixel locations, $\mathbf{y}$ (shown as circles in figure 4). Plotting $\mathbf{y}$ together with $\mathbf{x}$ (the original marker centroids captured by the camera, shown as plus-signs) reveals the close coincidence between the two. Incorrect estimates of the camera parameters would result in a spatial offset between $\mathbf{y}$ and $\mathbf{x}$. No such systematic offset is evident in figure 4. The trajectory was gathered over a wide range of distances and yet the errors are low across the whole trajectory, demonstrating the applicability of the calibration over a large working volume. The reprojection error for this example is 0.89 pixels. For the nVis SX111 HMD, with a calibrated left display of 75° horizontal field of view, this reprojection error represents ≈ 3.13 arcmin (where each pixel subtends ≈ 3.5 arcmin). Note that the calibrated horizontal field of view is close to the manufacturer's specification of 76° but using the latter figure would result in a systematic error, greatest at the edges of the display, of over 2 pixels.

For any single trajectory, the presence of noise (and possibly other idiosyncratic aspects of the measurement apparatus) will inevitably lead to errors in the estimation of projection parameters. Such incorrect estimation will be evident in higher reprojection errors of *other* trajectories captured under identical conditions (i.e. when the camera
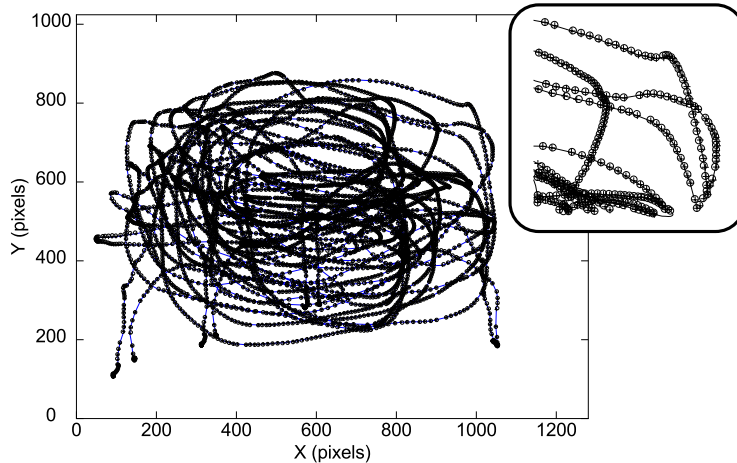
8

Figure 4: Plot of calibrated solution (circles) for the trajectory shown in figure 3 (plus-signs). The inset shows a magnified section, revealing the accurate alignment of the reprojected data. The reprojection error in this case is 0.89 pixels.

is not moved or adjusted). We used the extra 3 trajectories captured for each camera/HMD position to test this by calculating the reprojection errors for the remaining trajectories with that calibration. Figure 5 shows reprojection errors for the 4 trajectories within each data set when they were tested with the calibration generated from the first trajectory in the first data set (trajectory 1 in Figure 5a). It is clear that the original calibration generalized well to the other trajectories from the same camera/HMD position, with reprojection errors remaining at about 0.91 ± 0.12 (mean $\pm$ standard deviation) pixels for the novel trajectories. This indicates that the calibration was not over-specialized or otherwise influenced by measurement noise in the tracker and image processing coordinates.

Distinctly, we could also test generalization across the other 20 trajectories from the other 5 camera positions. Figure 5b–f illustrates the consequences of testing the original calibration (from figure 5a, trajectory 1) but now using trajectories that were collected with the camera in a slightly different location with respect to the HMD display. Each panel shows results for the camera in a different location. It is clear that these trajectories result in a larger reprojection error (up to 9 pixels). The different trajectories *within* each panel (i.e. taken from the same camera location) all have similar reprojection errors but, in each case, they confirm that the calibration is an inappropriate one.

We examined all 24 calibrations tested with each of the 24 trajectories (576 combinations, not shown). We found that it was always the case that the reprojection errors for the trajectories viewed from the same location as the trajectory used to calibrate the HMD (i.e. cases equivalent to those shown in figure 5a) were low compared to those tested with trajectories obtained from a different camera location. Specifically, generalisation within the same camera position had reprojection errors of 0.98 ± 0.01 pixels; for generalisation to other trajectories taken from a different location as the one used for calibration, reprojection errors were on average 6.71 ± 3.05 pixels.

Clearly, the effect of physically moving the headset to a different region of the tracked volume necessarily involves the camera being moved in relation to the headset and, critically, altering the spatial relationship between the camera's optic centre and principal ray relative to that of the HMD display. However, the relationship between the two optic
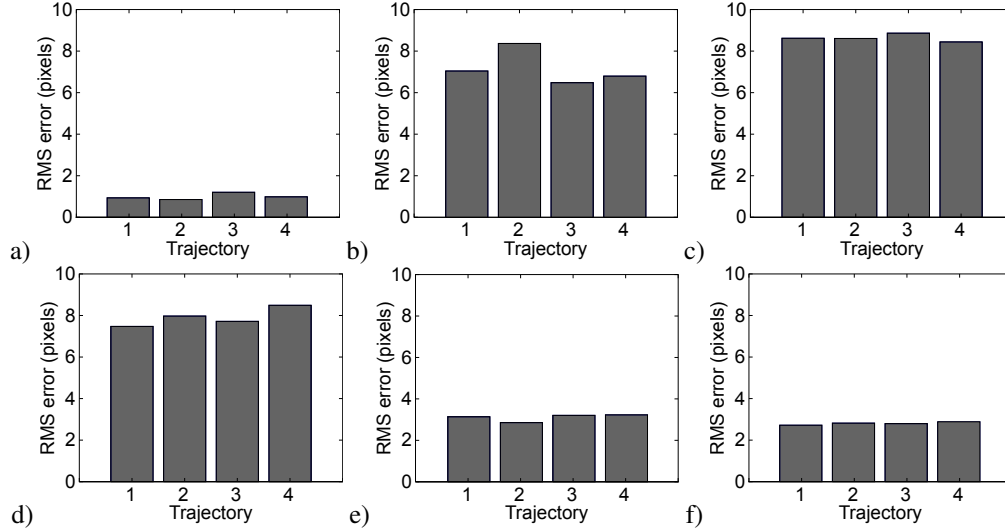
Figure 5: Generalisation of a calibration. a) Column 1 shows the reprojection error for a calibration generated from ≈ 8000 samples of marker location. Three other similar trajectories were captured and columns 2–4 show the re-projection error when these trajectories were tested using the calibration computed from the first trajectory. b) – f) show the reprojection error when trajectories recorded from the five other camera locations were tested, again using the calibration computed from the first trajectory.

centres is not unbounded – the constraint that the camera must be positioned to capture as much of the HMD grid image as possible means that it will lie within a small region whose centre is close to the HMD display's true optic centre. Further, the camera placement with respect to the HMD display is likely to mimic that of a human user wearing the same HMD: each occasion on which they re-position the HMD will yield a different alignment, but constrained by the requirement to obtain a clear image. Figure 5b–f give an indication of the errors one might expect when moving the eye relative to the calibrated optic centre. In the Discussion, we consider the consequences of such head movements for both see-through and non-see-through HMDs.

We next considered the number of samples that are required in order to obtain an accurate calibration. Because each sample quintuple of $\left[\mathbf{x}_t^{\text{CAM}} \mathbf{X}_t \right]$ is entirely independent of the others, we were able to subsample each of the trajectories obtained to see how calibration accuracy varied with the number of samples used. Figure 6 shows how reprojection error changed as more samples were made available to the calibration procedure. For each number of samples, a new calibration was generated using that number of samples from one particular trajectory. The reprojection errors for these calibrations are shown by the crosses: these are about 0.5 pixels for the smallest number of samples and rise to about 0.9 pixels for the case when all $8,000$ samples in the trajectory were used. The other 23 points plotted for a given sample size show how the calibration generalises to other trajectories, i.e. the reprojection errors are shown when other (whole) trajectories are tested using the calibration generated from the (sampled) first trajectory.

For the smallest number of samples tested (10), the algorithm over-fits the data, as is evident from the fact that there is a low reprojection error on the training data but high reprojection errors on the test trajectories. Between 10 and 1000 samples, the reprojection error on the trajectory used to generate the calibration actually deteriorates but this is accompanied by an improvement in the reprojection errors for other trajectories taken from the same camera
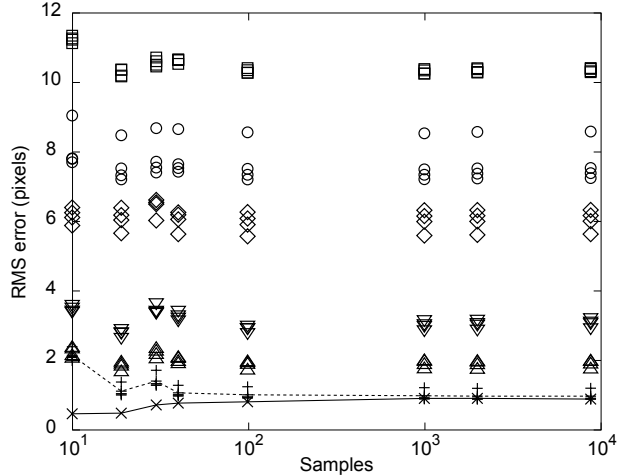
Figure 6: Calibration errors as a function of number of samples used from the tracked object trajectory. The crosses (×) show reprojection errors for a single trajectory. These are plotted against the sample size, with samples picked randomly from the ≈ 8000 locations of the tracked marker. The plus-signs (+) show how well this calibration generalised to the three different trajectories captured with the same camera position (c.f. figure 5a). The open symbols show reprojection errors when this calibration was tested against trajectories taken from the three alternate camera locations (c.f. figure 5b–f).

location (plus signs), as would be expected if the calibration is converging on the correct solution. Above 1000 samples there was no clear improvement in calibration generalization for other trajectories captured from the same camera location, even when all 4 trajectories are combined (over $32,000$ samples, reprojection error of 0.9 pixels). Trajectories taken from other camera locations lead to a worse calibration, as we discussed in relation to figure 5 but, unlike the trajectories taken from the same camera location, the reprojection errors are relatively unaffected by sample size. Of course, the main point is to determine the number of samples that would be required to achieve a reasonable calibration. From figure 6, and confirmed from similar calibrations, it would appear that $1,000$ samples is reasonable.

Given that the ultimate aim is to recover the 11 projection parameters of the display, it is instructive to plot the change in these as the number of samples increases and reprojection errors drop. Figure 7 illustrates the X and Z translation components of the optic centre plotted as a function of the number of samples used in the calibration for the 4 trajectories from the HMD position shown in figure 6. It can be seen that, when the sample number is low, the estimates of the optic centre are scattered around the location of the best estimate, obtained with $32,000$ samples. This reduction in scatter is accompanied by a relatively modest fall in the reprojection error, from ≈ 2.0 pixels for 10 samples to ≈ 1.0 pixels for $32,000$ samples. The examples illustrate the advantage of an automatic, camera-based method over those that rely on human judgements of alignment, such as SPAAM (Tuceryan et al., 2002), which are inevitably limited in the number of samples that can be obtained.

Finally, we repeated the estimation of projection parameters including a model of non-linear (radial and tangential) distortions (Heikkilä and Silvén, 1997). The extra 5 parameters were included in the simplex minimization, and the image corrections were applied to the pixel coordinates obtained in equation 6. The high density of image data points obtained by our method means that non-linear distortion parameters should be well estimated. For the SX111 HMD
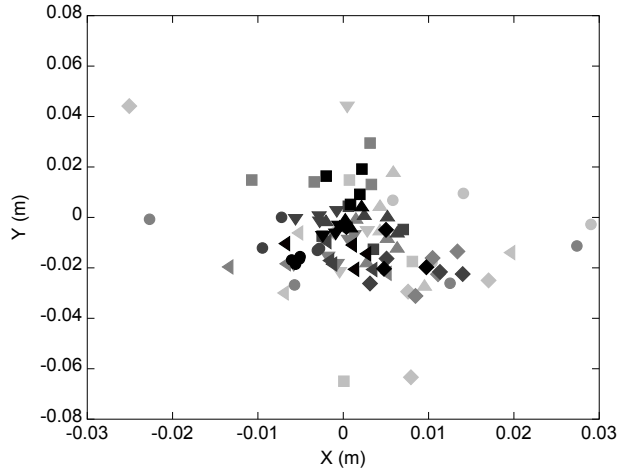
Figure 7: Two components of the HMD optic centre locations, estimated for different numbers of samples for all trajectories from the 6 HMD positions. Light grey symbols represent the results when the calibration is based on only 7 samples; mid grey: 10 samples; dark grey: 20 samples; black: 100 samples. As the number of samples increases, the estimated optic centres cluster more closely around the estimate obtained using the combined trajectories for each HMD position ($\approx 32,000$ samples, origin of plot).

used here, the non-linear parameters were small, and reprojection errors were reduced, on average, by 7%.

4. Discussion

In addition to the objective measures of calibration accuracy described above, it is important to consider the implications for humans wearing the HMD. For example, the optic centre of the wearer will never align perfectly with that found by the calibration procedure. As we saw in figure 5, movement of the camera relative to the HMD results in larger reprojection errors. Such movement is comparable to that of a human observer wearing a HMD – each usage would change the relative position of their binocular optic centres in relation to the HMD. The issue is complicated by the optics used in HMDs, many of which use approximations to collimated optics in order to relax the constraint of positioning the observer's eyes directly in line with the display's exit pupils. As a result, some shifts in the position of the eyes' optic centres relative to the exit pupils still give rise to a clear, focussed image.

The importance of small shifts in the optic centre of the user relative to the HMD display optic centre depends on whether the display is see-through or not. For a see-through display, translation of the user's optic centre will result in parallax between features drawn on the HMD screen and real objects viewed through the screen. The extent of this parallax is particularly evident in the large rise in reprojection errors for figure 5b–d relative to those in figure 5a. We do not know the range of optic centre translations that participants tolerate relative to the HMD but it is likely to be a very much smaller range in augmented reality headsets than in a non-see-through headset because human observers are exquisitely sensitive to relative offsets in the alignment of visual features.

The subjective impression that observers obtain in a non-see-through headset is quite different from an optical see-through headset. If, for example, the simulated inter-ocular separation is changed on a headset while an observer is wearing it, the observer will generally not notice much of a change in the perception of distance, size or stability of

12

objects in the scene despite the fact that, in a see-through display, such changes in simulated inter-ocular separation would cause very noticeable parallax between the real and virtual scenes. It is interesting to speculate about the reasons for this difference. Briefly, it probably implies that observers are not reconstructing the virtual scene when they view a scene with a non-see-through headset (or, indeed, when they view an ordinary scene). The argument is that there is no consistent interpretation of a static scene and fixed camera calibration parameters that could explain that scene so, if they are perceiving a stable scene, they must be doing something other than reconstruction. We have made a similar argument in relation to experiments on an expanding virtual scene (which observers perceive to be stable) (Glennerster et al., 2006; Svarverud et al., 2010). In that case, there *is* a stable interpretation of the images the observer receives but only if they are prepared to accept wildly inaccurate estimates of the optic centre locations, including both inter-ocular separation and translation of the head. More likely, in both the expanding room and the case of manually changing the inter-pupillary distance on a HMD, the reason that the world appears to remain stable is that the visual system is remarkably prone to accepting that this is the case.

A consequence of this difference is that the calibration technique we describe here is probably sufficient, in general, when using a non-see-through HMD, but a more stringent or adaptive method is likely to be required to obtain a highly accurate solution when using an optical-see-through HMD. Broadly, to calibrate more adaptively, two alternatives seem to be available. One possibility is that while the observer is wearing the see-through HMD they must adjust the position of the headset and the optics using a visual alignment procedure until the optic centre of each eye is in the 'correct' location, where 'correct' means the optic centre for which the HMD is calibrated. This may, in practice, be the simplest solution. An alternative method might be to allow the observer to adjust the HMD until they are comfortable with the view and then try different, precomputed calibrations. They could use an alignment method against a calibration rig to judge which calibration gives the best alignment relative to fixed markers in the scene. The calibration would need to be performed each time the user wears the HMD. This approach has some similarities to the SPAAM procedure described earlier (Tuceryan et al., 2002), but may not be suitable for small tracked volumes. It may also be helpful to obtain an independent estimate of the location of the optic centre relative to the headset in order to choose the best pre-computed calibration. Sousa et al. (2010) have described how this can be achieved using a long, tracked, hand-held tube containing cross-hairs at each end which the user repositions until the cross-hairs are aligned. When repeated from several different orientations, these 'rays' constrain the estimated location of the optic centre. Nevertheless, it is worth remembering that the human eye does not rotate around it's optic centre, and thus any changes in gaze will change the spatial relationship between the calibration's and observer's optic centres. Until optical-see-through HMDs are able to dynamically adapt to gaze direction, it may not be possible to obtain a truly 'perfect' alignment of real and virtual worlds.

5. Conclusion

We have presented here an HMD calibration method for non-see-through HMDs, which are the most common type of HMDs used in virtual reality applications. Our method, based on our earlier optical see-through HMDs calibration work (Gilson et al., 2008), provides a quick, reliable and robust method to calibrate each display of an HMD. Unlike

existing calibration methods (Tuceryan et al., 2002; Owen et al., 2004), ours does not require error-prone human measurements and provides an objective measure of calibration accuracy.

Acknowledgement

References

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

Gilson SJ, Fitzgibbon AW, Glennerster A. Spatial calibration of an optical see-through head-mounted display. J of Neurosci Meth 2008;173(1):140 –6.

Glennerster A, Tcheang L, Gilson SJ, Fitzgibbon AW, Parker AJ. Humans ignore motion and stereo cues in favour of a fictional stable world. Curr Bio 2006;16:428–43.

Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge University Press, UK, 2001.

Heikkilä J, Silvén O. A four-step camera calibration procedure with implicit image correction. In: IEEE Comp. Vis. Pat. Recog. San Juan, Peurto Rico; 1997. p. 1106–112.

Kuhl SA, Thompson WB, Creem-Regehr SH. Hmd calibration and its effects on distance judgments. ACM Trans Appl Percept 2009;6:19:1–19:20.

Lagarias J, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim 1998;9:112–47.

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

Mon-Williams M, Wann JP, Rushton S. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. Ophthal & Physiol Opt 1993;13:387–91.

Owen CB, Zhou J, Tang A, Xiao F. Display-relative calibration for optical see-through head-mounted displays. In: Proc. IEEE ACM Int. Symp. Mixed & Aug. Real. (ISMAR). 2004. .

Regan C. An investigation into nausea and other side-effects of head-coupled immersive virtual reality. Virt Real 1995;1:17–31. 10.1007/BF02009710.

Rinalducci E, Mapes D, Cinq-Mars SG, Higgins KE. Determining the field of view in HMDs - A psychophysical method. Presence-Teleop Virt 1996;5(3):353–6.

Sousa R, Brenner E, Smeets JBJ. A new binocular cue for absolute distance: Disparity relative to the most distant structure. Vision Res 2010;50:1786–92.

Svarverud E, Gilson SJ, Glennerster A. Cue combination for 3D location judgements. J Vis 2010;10(1).

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

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

List of Figures