Interactive Visual Calibration of Volumetric Head-Tracked 3D Displays

Andrew John Wagemakers, Dylan Brodie Fafard, Ian Stavness
Dept. of Computer Science, University of Saskatchewan
110 Science Place, Saskatoon, SK, S7N 5C9, Canada
andrew.wagemakers@usask.ca, dylan.fafard@usask.ca, ian.stavness@usask.ca

ABSTRACT
Head-tracked 3D displays can provide a compelling 3D effect, but even small inaccuracies in the calibration of the participant's viewpoint to the display can disrupt the 3D illusion. We propose a novel interactive procedure for a participant to easily and accurately calibrate a head-tracked display by visually aligning patterns across a multi-screen display. Head-tracker measurements are then calibrated to these known viewpoints. We conducted a user study to evaluate the effectiveness of different visual patterns and different display shapes. We found that the easiest to align shape was the spherical display and the best calibration pattern was the combination of circles and lines. We performed a quantitative camera-based calibration of a cubic display and found visual calibration outperformed manual tuning and generated viewpoint calibrations accurate to within a degree. Our work removes the usual, burdensome step of manual calibration when using head-tracked displays and paves the way for wider adoption of this inexpensive and effective 3D display technology.

Author Keywords
Fish Tank VR; calibration; head tracking; visual perception

ACM Classification Keywords
H.5.1: Artificial, augmented, and virtual realities

INTRODUCTION
Commercial 3D displays have eluded wide spread adoption. Glasses-based stereoscopic televisions have declined in popularity and the "3D" feature is no longer offered on many new models. Virtual and augmented reality headsets, such as Oculus Rift and Microsoft HoloLens, have had a resurgence in recent years, but these systems require bulky headgear and either block out the physical world or lack the physicality of a real display surface. Head-tracked displays have the best potential for 3D display situated within a physical work or living space.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025685

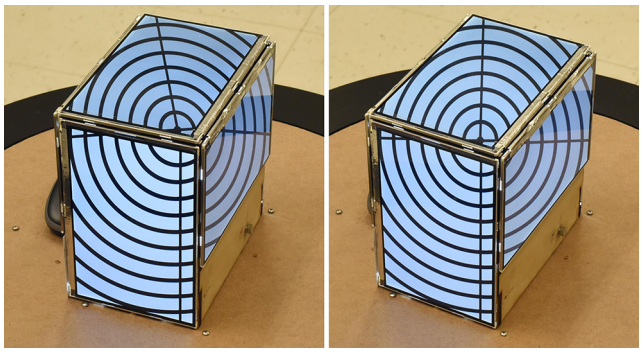

Figure 1. A calibration pattern rendered on a head-tracked cubic display before (left) and after (right) visual alignment.

To create a 3D illusion, head-tracked displays use motion parallax depth cues, which are stronger than stereoscopic cues alone [33]. Head-tracked displays naturally support work practices mixing 3D displays with traditional 2D displays. For example, in a computer-aided design scenario, orthographic views could be presented on a 2D screen while the 3D perspective view could be presented on a head-tracked 3D display. Multi-screen head-tracked displays that wrap screens around a geometric shape, such as box or sphere, provide 360° viewing around a virtual object, which enhances the motion parallax effects and improves 3D perception.

The principal drawback of head-tracked displays is that the user's viewpoint must be accurately measured in real-time with respect to each screen of the display. Even small inaccuracies in the rendered viewpoint can create visual artifacts (kinked lines, oddly floating objects, and ghosting) that can immediately and severely disrupt the 3D effect of the display (see Figure 1).

In this paper, we propose a novel user-friendly procedure to easily and accurately calibrate a head-tracked display. The procedure estimates both the location of the display and the location of user's viewpoint relative to a head tracker. It does not require any physical markers in the workspace and instead uses visual patterns rendered across a multi-screen display that are designed to appear aligned when viewed from a correct viewpoint, but to appear distorted otherwise. In this way, we can compel participants to place their viewpoint in known locations relative to the display and then measure the head-tracker error at each location to calibrate the display. We conducted a user study to evaluate different candidate visual patterns (concentric circles, a

grid, and a combination of circles and lines) on simulated spherical, cubic and box-shaped displays. We found that participants were fastest and most accurate when aligning patterns on a spherical display and that the combined circles and lines pattern was most accurately and quickly aligned across display shapes. We evaluated our procedure with a physical cubic display using a semi-automatic camera-based analysis. The analysis demonstrated that the method generates an accurate viewpoint calibration that accounts for a head-to-viewpoint offset, regardless of head-tracker orientation and invariant of depth errors.

This work makes four main contributions. First, we propose a novel way to measure viewpoint locations around a geometric display by using visual patterns. Second, we detail a mathematical approach for head-tracker calibration that estimates both the tracker-to-display and the head-to-viewpoint transformations. Third, we report the first quantitative evaluation of visual error for head-tracked displays. Fourth, we provide empirical evidence that visual pattern based calibration with a real box display is effective.

RELATED WORK

Head-tracked 3D rendering has been used with a number of previously proposed display systems. Calibration for electromagnetic tracking systems, multi-screen/projector systems and Augmented Reality systems have been extensively studied, but few have provided head-tracker calibration procedures other than manual tuning. 2D visual patterns have been proposed as fiducial markers in position tracking and calibration systems, but 3D patterns have not been well explored.

Fish Tank Virtual Reality Displays

Head-tracked 3D displays that used stereoscopic and motion parallax cues were first introduced as a practical way to view 3D data on a desktop monitor [1]. This concept, termed Fish Tank Virtual Reality (FTVR), was proposed as an alternative to Head Mounted Displays (HMDs) and immersive Virtual Reality (VR) rooms [8] of the time. The original user studies with FTVR displays demonstrated the surprising result that head-tracking provided stronger 3D cues than stereo alone [33]. An issue with early FTVR systems was the limited range of viewing angles. The user had to remain in front of the single monitor, thereby limiting their head motion, and consequently reducing the strength of motion parallax depth cues. Multi-screen FTVR displays have been developed that allow for a large variety of viewing angles, including 360 degree viewing around an enclosed volumetric display. Different configurations include: inward facing box [10], outward facing box [17, 28], tabletop [15, 7], cylinder [19] and sphere [5, 11]. Head-tracking has also been used to enhance video game experiences [21] and adapt stereoscopic display parameters [22].

Multi-screen Calibration

Many multiscreen FTVR displays use multiple projectors to illuminate a seamless geometric display surface. However, these require careful screen calibration so that overlapping projector geometry can render without visual artifacts, such as ghosting or brightness disparity. Multi-projector calibration procedures for planar [35] and curved [27, 37] display surfaces have been reported that use a camera to automatically compute accurate transformations and blending between projection regions. Box displays have used small LCD screens for a compact design, which allows them to be handheld, but has the downside of relatively thick seams between screens [28]. Seams themselves can enhance the 3D effect by providing occlusion cues, but thick seams can be obtrusive and potentially disruptive to viewing. Multiple display panels also require screen calibration, but the accuracy requirements are lower, as the screens do not overlap like projected screens. The few quantitative analyses of error sources in FTVR displays that have been reported suggest that screen calibration error contributes significantly less to overall visual error than head-tracker calibration error [8, 36]. However, little previous work has addressed or quantitatively measured head-tracker error for FTVR displays.

Tracking System Calibration

A motion tracking system is required in FTVR displays to measure the user's head position relative to the display. The tracking system must be calibrated to the display, i.e. the rigid transformation between the reference frames of the tracking system and display must be found. This is often done by tedious manual tuning, but standardized procedures have been proposed whereby the transformation is estimated from a corresponding set of tracker-measured and physically-measured 3D points. For example, Kindratenko et al. [20] used an electromagnetic (EM) tracking system to record the tracked position at several locations in the CAVE. The locations were precisely determined using a sensor mount consisting of a 1' x 1' x 0.1' wooden board with housing for a 2', 4', 6', or 8' pipe. By moving the board between marks on the floor and changing the pipe length, the sensor could be placed at a number of known locations with a precision of 0.5 cm. Other systems have used an optical system for calibration [25, 32]. EM tracking systems are often calibrated in this way because EM distortion can skew position-tracking measurements across the workspace [6, 14].

Viewpoint calibration

Although a head-tracker can be accurately calibrated, the actual viewpoint of the user is often offset from the tracker's measured "head" position. If a head-marker is worn, there will be an offset; however, even markerless trackers, such as the Kinect, result in a reported head position that is not aligned with the user's viewpoint. Most FTVR displays assume a constant offset from head-marker to the user's viewpoint [28]. Madritsch et al. [23] describe a model for locating the position of a user's eyes when viewing an FTVR display. Glasses with trackable markers are worn by the user and a constant offset is applied in the user's reference frame to locate the eyes.

Czernuszenko et al. [9] developed a method for accurate tracker calibration of a projector-based VR system which involves the user aligning physical and virtual markers at several different head positions. Physical markers are placed at known locations and virtual markers are rendered to appear at the same location. The viewer wears a pair of tracked glasses and uses a controller to align the virtual markers with the physical markers and record the offset vector at the current head position. This is repeated at any point with considerable misalignment, and an appropriate lookup table is built.

Augmented Reality Calibration

Augmented Reality (AR) displays overlay virtual imagery onto real-world environments, either by projecting onto objects or using a see-through HMD. Calibration errors of the user's viewpoint to the physical world in AR are highly noticeable and result in misalignment or ghosting of the virtual overlay. For this reason, static [3] and dynamic [4, 12] calibration schemes have been well studied in AR, including detailed quantitative error analysis [16, 24]. An advantage of HMDs is that the user's eye locations can be accurately measured using the HMD device because of its close proximity to the user's eyes [26]. The semi-transparent screen of see-through HMDs also allows them to be visually aligned similar to a rifle scope, as proposed in the SPAAM method [31]. This method involves visually overlapping a virtual marker on the HMD screen with a physical marker at a known location in the head-tracker's reference frame. Measurements with different virtual markers allow for estimation of the tracker-to-screen planar homography, and the method can be extended to binocular viewing [31].

Previous to our work, it has not been possible to perform SPAAM-style visual alignment with FTVR display screens which are looked-at rather than looked-through. Our proposed approach using visual patterns makes this possible. Our method also differs from SPAAM because it recovers both the tracker-to-display and head-to-viewpoint transformations and does not rely on a planar homography and therefore can be used with spherical displays.

Pattern-based Fiducial Markers

2D patterns have been widely used as fiducial markers in position tracking systems, including the popular AR-Toolkit [18] and patterns optimized for visual robustness [13]. Markers that generate view-dependent moiré patterns have also been used to measure position and orientation with a single camera [34]. In this work, we explore 3D patterns that can be used to guide the orientation of a participant's viewpoint; evaluate different candidate patterns; and develop a robust calibration technique that exploits these patterns.

INTERACTIVE VISUAL CALIBRATION PROCEDURE

Calibration Problem

FTVR rendering requires real-time tracking of the user's viewpoint relative to each screen of the display.

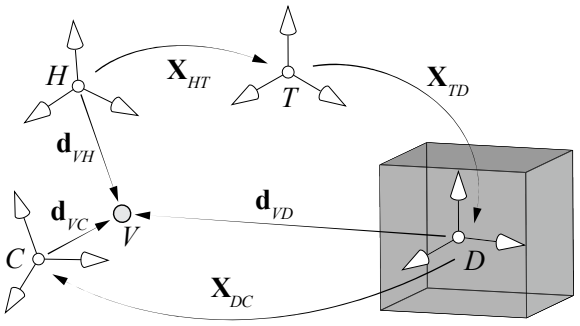

Figure 2. Coordinate frames in our calibration problem, including the display (*D*), tracking system (*T*), head (*H*), and calibration point (*C*) frames, and viewpoint position (*V*). X and d denote rigid transformations and translations, respectively.

Using a head-tracking system, the chain of transformations to accomplish this is: viewpoint-to-head, head-to-tracker, tracker-to-display, and display-to-screens (Figure 2). The tracking system measures the head-to-tracker transformation ($\mathbf{X}_{HT}$). Our proposed calibration procedure estimates two of these transformations: the rigid transformation from the tracking system to the display ($\mathbf{X}_{TD}$) and the translation between the measured head position (e.g. marker location on the head for a marker-based tracker) and the actual viewpoint of the user ($\mathbf{d}_{VH}$). For monocular viewing the user's viewpoint is their eye location, whereas for binocular viewing this is a location fused from the view of both eyes and will be person-dependent. Although $\mathbf{X}_{TD}$ is fixed for a stationary display and could potentially be measured by the tracking system [29], some tracking systems are optimized for person/head-tracking alone and manual measurement of the display location relative to the tracker (e.g. with a tape measure) can introduce significant errors. We assume that transformations between each display and each screen are calculated with existing methods for multi-screen calibration (see *Multi-screen Calibration* in Related Work).

Existing head-tracker calibration techniques involve either tedious manually tuning of these transformations or visually aligning physical and virtual markers [9]. The physical markers are placed in preset locations and the virtual markers are rendered on the surface of the display. The individual calibrating the display must move to different locations and, using a controller, translate the virtual markers until they align with the physical markers. This technique may be undesirable due to the physical markers that must be placed at precise locations.

For rendering to the correct perspective of the user, their viewpoint must be known in *D*, which can be found by solving:

$$\mathbf{d}_{VD} = \mathbf{X}_{TD}\mathbf{X}_{HT}\mathbf{d}_{VH}. \quad (1)$$

Once $\mathbf{d}_{VD}$ is known the appropriate projection matrices can be constructed. For planar screens (box, corner) this is an off-axis perspective matrix, for non-planar surfaces (sphere,

cylinder) this would require a per-pixel mapping from parametric coordinates to pixel coordinates. In the case of a 6-DoF tracking system $\mathbf{X}_{HT}$ can be measured directly, but for 3-DoF tracking system only the position of H is measured, not its orientation. In this situation, we assume the user is looking at the center of the display with their head upright. This way the vector separating the viewpoint from the head from can be solved for and applied correctly.

Proposed Procedure

We propose a visual calibration technique to find two of the necessary transformations without the use of physical markers. A pattern is rendered on the surface of the display such that it appears undistorted if viewed from C. The user must then position their head such that the pattern appears correct. This process is repeated to record correspondences of C and the orientation and position of the head-tracked point. An optimization process can be performed on these correspondences to provide accurate estimates of the display to head-tracker ($\mathbf{X}_{TD}$) transformation and the head to viewpoint ($\mathbf{d}_{VH}$) translation. For the 6-DoF head tracking, where the tracker measures $\mathbf{X}_{HT}$, the viewpoint position should be coincident with the origin of C, therefor $\mathbf{d}_{VC}$ should be zero, where $\mathbf{d}_{VC} = \mathbf{X}_{DC}\,\mathbf{X}_{TD}\,\mathbf{X}_{HT}\,\mathbf{d}_{VH}$. Thus, the following cost function can be used in the optimizer:

$$\arg\min_{\mathbf{X}_{TD},\mathbf{d}_{VH}} \sum_{i=1}^{n} \left\| \mathbf{X}^{i}_{DC}\mathbf{X}_{TD}\mathbf{X}^{i}_{HT}\mathbf{d}_{VH} \right\| \qquad (2)$$

For the 3-DoF head tracking, where the tracker measures $\mathbf{d}_{HT}$, the vector from head position to the viewpoint position ($\mathbf{d}_{VH}$) should be equal and opposite to the vector from the origin of the C to the head position, therefore $\mathbf{d}_{VH} = -\mathbf{d}_{HC}$, where $\mathbf{d}_{HC} = \mathbf{X}_{DC}\,\mathbf{X}_{TD}\,\mathbf{d}_{HT}$. Thus, the following cost function can be used in the optimizer:

$$\arg\min_{\mathbf{X}_{TD},\mathbf{d}_{VH}} \sum_{i=1}^{n} \left\| \mathbf{d}_{VH} + \mathbf{X}^{i}_{DC}\mathbf{X}_{TD}\mathbf{d}^{i}_{HT} \right\| \qquad (3)$$

In both formulations, n represents the number of recorded calibration points. Since there is minimal depth information present in the rendered visual pattern, the solver must be depth invariant. The cost functions in (2) and (3) measure error with respect to the C frames, whose z-axes are aligned to the depth direction from the display; therefore, depth errors can be removed by only considering the x and y components of the error in both cost functions. For Study #2, we solved the depth invariant form of (3) using MATLAB's nonlinear multivariable solver *fmincon*.

Verification of Proposed Procedure

A synthetic experiment was conducted as initial verification. 1000 different combinations of $\mathbf{X}_{TD}$ and $\mathbf{d}_{VH}$ were randomly generated to create a large variety of test cases. Using the inverse of the transformations, a set of 24 calibration points were transformed into viewpoints with error added to them to create synthetic, recorded positions. The error consisted of two components: normally distributed error with a mean of 1 cm in the horizontal and vertical directions, and normally distributed error with a mean of 0 cm and 11 levels of standard deviation (0 cm, 5 cm, 10 cm, 15 cm, 20 cm, 25 cm, 30 cm, 35 cm, 40 cm, 45 cm, 50 cm) in the forward direction. All 1000 x 11 conditions were run through the 3DoF version of the procedure to obtain estimates of $\mathbf{X}_{TD}$ and $\mathbf{d}_{VH}$. These estimates were used to transform the synthetic positions back into C to calculate a displacement vector. The mean magnitude of displacement vectors (excluding the forward direction) was recorded as the error for each condition. A one-way ANOVA test reported no significant effect ($F_{10,10989} = 0.0007$, $p = 1.0000$) of standard deviation of depth error on displacement error (m 1.1 cm, s.d. 0.18 cm). The maximal pairwise Hedges' g value was 0.0021. These results indicate that our proposed visual calibration technique is accurate and invariant to a broad range of depth errors.

STUDY #1: EVALUATION OF CALIBRATION PATTERNS AND DISPLAY SHAPES

We conducted a study to assess users' performance as they visually aligned 3D patterns by rotating a 3D rendered shape on a computer screen using a computer mouse. This study was intended to simulate the calibration procedure and to allow for an assessment of different patterns and display shapes.

Apparatus

A Unity application was developed for the visual alignment task that recorded accuracy and completion time. Mouse input was used to pan the camera around an object at a fixed radius. Patterns were texture mapped onto the object such that they appeared undistorted from a set camera location (goal location). The texture mapping was accomplished by placing a camera at the goal location oriented toward the center of the object. Each fragment color was determined by sampling the 2D pattern at the same location as it appeared on this camera's screen.

Experimental Conditions

Shapes. Three virtual shapes were used in the study (*Sphere*, *Cube*, *Box*) as shown in Figure 3. These shapes were chosen because spherical and cubic shapes are commonly used for FTVR displays. *Box* represents the dimensions of our current cubic display constructed from LCD screens (with seams between each screen) and *Cube* represents a cubic display with no seams that could be achieved with rear projection screens.

Patterns. Three different patterns were chosen to be texture mapped onto the shapes (*Grid*, *Circles*, *Combo*). These patterns were chosen because they contain two distinct visual cues. *Grid* contains numerous parallel lines which will appear straight and smooth when viewed from the correct viewpoint. *Circles* contains a sequence of concentric circles that are equal thickness when viewed from the correct viewpoint, like a bullseye pattern. *Combo* has cues that are present in both the *Grid* and *Circles*.

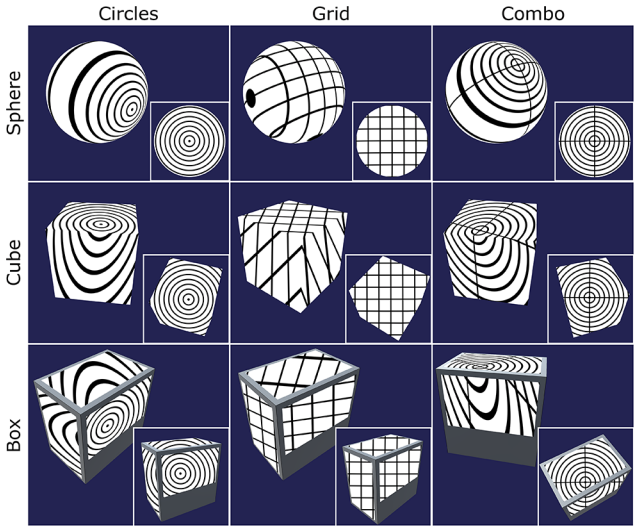

Figure 3. *Shape* (rows) and *Pattern* (columns) conditions for Study #1 showing the unaligned pattern at the start of the trial and rotated to the aligned pattern (inset).

Participants & Procedure

Eighteen participants were recruited from a local university (14 males, 4 females). In each trial, the participant used the mouse to rotate the camera around the shape until the pattern on the shape no longer appeared distorted. Once the participant was satisfied with the alignment they clicked the left mouse button. Between each trial the camera would reset to its default location and the goal location would change. Alignment accuracy was measured as the angle in degrees between the recorded camera location and the goal camera location.

Before starting the study trials, participants completed nine training trials (one per pattern/shape combination) to familiarize themselves with the controls and to better understand the possible distortions.

Participants completed trials using every combination of the three shapes and three patterns. The conditions were rotated between participants to account for possible sequencing effects. For each combination, 24 unique goal locations, equally divided into two blocks, were used. The order of the goal locations was randomly generated for each combination. Participants observed each of the patterns on a single shape before changing shapes.

After each pattern, participants filled out a NASA-TLX questionnaire; between each shape, they also answered summary questions comparing the different patterns. Participants were informed that both their accuracy and completion time would be recorded for each trial but that they should focus more on accuracy.

Design and Hypotheses

The study used a 3x3 within-participant RM-ANOVA with factors *Shape* (Sphere, Cube, Box), and *Pattern* (Circles, Grid, Combo). Dependent measures were completion time, and alignment error. We expect *Cube* to result in a higher accuracy than *Sphere* because small distortions would be more noticeable due to the sharp edges around the shape. We also expect that the Combo would be most accurate as it has both visual cues, but that Circle would be faster as it has less visual information to process. Hypotheses were:

H1. Mean error will be lowest for *Cube*.

H2. Mean error will be lowest for *Combo*.

H3. Mean completion time will be lowest for *Sphere*.

H4. Mean completion time will be lowest for *Circles*.

Results

Accuracy

The mean alignment error across conditions is shown in Figure 4. A two-way RM-ANOVA test over all trials with *Shape* and *Pattern* showed that they have a main effect on accuracy of the task ($F_{2,3879}$= 177.81, p< 0.01; $F_{2,3879}$ = 6.15, p= 0.0021). The Tukey-HSD multiple comparison post-hoc test showed that participants were significantly more accurate with *Sphere* than both *Cube* (p < 0.01) and *Box* (p < 0.01). It also showed that *Cube* was significantly more accurate than *Box* (p < 0.001). We therefore reject **H1**.

The test also found participants were significantly more accurate using *Combo* than *Circles* (p = 0.0013), but *Combo* and *Circles* were not significantly different than *Grid* (p = 0.1357, p = 0.2485). We therefore reject **H2**.

Time

Mean alignment error across conditions is shown in Figure 5. A two-way RM-ANOVA over *Shape* and *Pattern* showed that there was a main effect in task completion time ($F_{2,3879}$= 311.53, p< 0.01; $F_{2,3879}$ = 83.9, p < 0.01). The Tukey-HSD multiple comparison post-hoc test showed that *Sphere* is significantly faster than both *Cube* (p < 0.01) and *Box* (p < 0.01). It also showed that *Cube* was significantly faster than *Box* (p < 0.01). We therefore accept **H3**.

The post-hoc analysis also found that *Combo* and *Circles* are significantly faster than *Grid* (p < 0.01, p < 0.01), but not significantly different from each other (p = 0.409). We therefore reject **H4**.

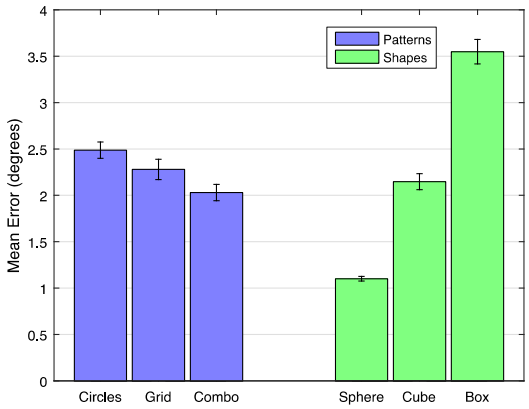

Figure 4. Mean alignment error across *Pattern* and *Shape*.

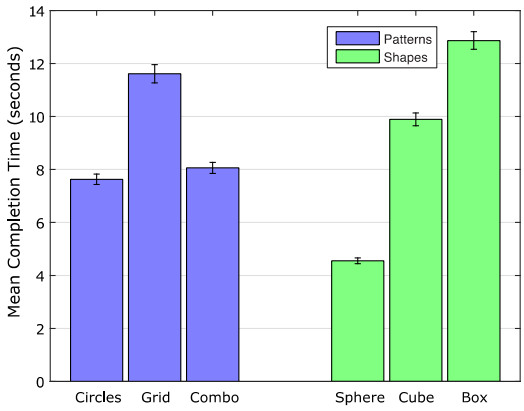

Figure 5. Mean completion time across *Pattern* and *Shape*.

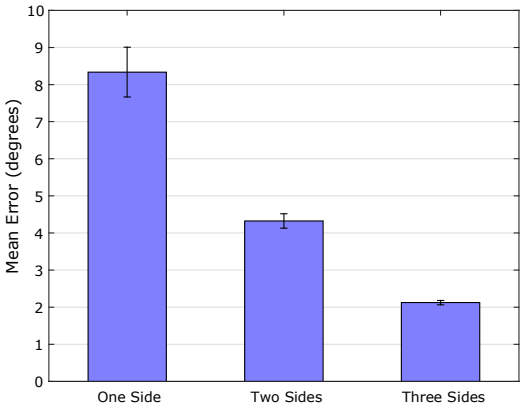

Figure 7. Mean alignment error for different number of sides visible at target viewpoint with cube and box shapes.

Questionnaire

NASA-TLX Questionnaire responses were similar across *Pattern* conditions, but varied more considerably for the *Shape* conditions (Figure 6). None of the shapes had considerably different levels of pace ($p = 0.2613$). Participants felt as though *Box* was more physically demanding than *Sphere* ($p = 0.0034$), but there was no considerable difference between *Sphere* and *Cube* ($p = 0.0667$), or between *Cube* and *Box* ($p = 0.5689$). *Box* was more mentally demanding than both *Cube* ($p = 0.0034$) and *Sphere* ($p < 0.01$), as well as *Cube* was more mentally demanding than *Sphere* ($p < 0.01$). Participants felt more successful with *Sphere* than *Box* ($p < 0.01$), but there was no considerable difference between *Sphere* and *Cube* ($p = 0.0231$) or *Cube* and *Box* ($p = 0.0162$). *Box* scored higher in Difficulty than both *Sphere* ($p < 0.01$) and *Cube*. *Cube* scored higher than *Sphere* ($p < 0.0001$), but there was no considerable difference between *Box* and *Cube* ($p = 0.03$). Participants felt more insecure with *Box* than *Sphere* ($p = 0.0001$), but there was no considerable difference between *Box* and *Cube* ($p = 0.0598$) or between *Cube* and *Sphere* ($p = 0.183$).

Sides visibility

For the *Cube* and *Box* shapes, the number of sides visible (*one*, *two*, or *three*) from every goal location was also recorded for each trial. Mean alignment error for trials with different numbers of sides visible is shown in Figure 7.

A one-way RM-ANOVA found that *three* resulted in significantly better accuracy than *two* ($p < 0.01$) and *two* resulted in significantly better accuracy than *one* ($p < 0.01$).

Discussion

The study found a considerable difference in both accuracy and completion time when different shapes were used. *Sphere* achieved the lowest mean error as well as the lowest mean completion time. This is a surprising and strong result as it breaks the typical speed vs. accuracy trade-off found in most pointing or alignment tasks. A possible reason for this is that the pattern distortion is much more gradual on a spherical display which could make it easier to determine where the camera should be moved. Performance with the *Cube* shape was worse than we expected. Although the cube edges provided strong visual cues (kinked lines at the edge between screens), the distortion effect may have been difficult to interpret by participants and required more visual processing effort and time. *Box* had the worst performance, which indicates that the seams between screens made pattern alignment significantly more difficult. This is an important design consideration as cubic displays constructed with LCD screens have seams, while rear-projection cubic displays can be made seamless.

The results show that there is a considerable difference in both accuracy and completion time when different patterns were used. *Combo* resulted in the highest accuracy, but was only considerably more accurate than *Circles*.

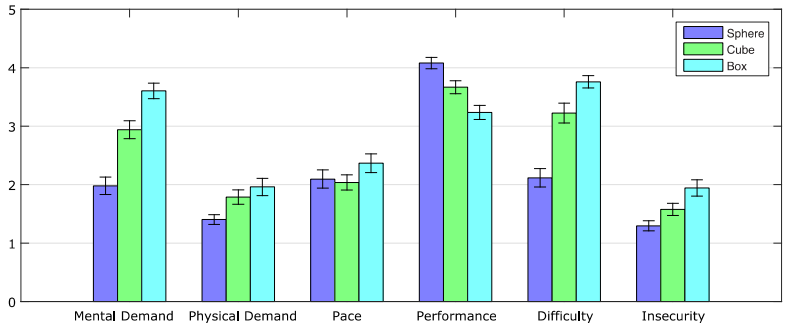

Figure 6. NASA-TLX questionnaire responses by Shape.

This could indicate that straight lines provide additional visual cues that aid in correcting distortion. However, given that *Grid* was observed to be considerably slower than the other two patterns, lines by themselves may not give enough visual cues to accurately align the pattern. *Circles* and *Combo* naturally provide a better sense of direction than *Grid* because the bullseye pattern has a defined center. Thus, the participants could have had a more accurate initial guess at how to correct the distortion using *Circles* and *Combo* making them faster.

The results for cubic displays showed that it is critical to provide three-screen views for pattern alignment. Therefore, we chose viewpoint locations for the analysis in Study #2 near the corners of the display so that three sides of the cubic display would be visible.

STUDY #2: A CAMERA-BASED ANALYSIS OF VISUAL CALIBRATION ON A PHYSICAL CUBIC DISPLAY

The aim of our second study was to assess the performance of our proposed pattern-based visual calibration technique by comparing it to a manual tuning technique. For all conditions, a marker was rigidly attached to a camera and the goal was to calibrate to the camera's viewpoint. This setup was equivalent to monocular viewing and permitted quantitative measures of calibration accuracy since test photographs could be taken from the exact viewpoint used for calibration.

Experimental Conditions

AlignmentDoF refers to how aligned the tracker and display spaces were at the start of the calibration trial. In *3DoF,* all the major axes are aligned. In *4DoF,* one of the major axes is aligned (up-axis in this case). In *6DoF,* none of the major axes are aligned.

OffsetMagnitude is the displacement, excluding the forward-looking dimension, of the head-to-viewpoint translation.

Pattern-based calibration is our proposed calibration technique of using a pattern to guide the viewer to a known position as described in the *Calibration Procedure* section.

Manual calibration is a real-time, interactive process whereby the viewer inspects a perspective-corrected scene for visual alignment errors and attempts to reduce the noted errors by using a keyboard to modify the transformation $\mathbf{X}_{TD}$. Depending on the *AlignmentDoF*, the user can either modify only the translation (*3DoF*), the translation and one rotation (*4DoF*), or the translation and all three rotations (*6DoF*).

Procedure

Based on the results of Study #1, the *Combo* pattern was chosen for *Pattern-based* calibrations and the calibration positions were situated so that three screens of the display were visible. The number of calibration points (positions to which the pattern will be rendered) was limited to 24 to balance time and accuracy. They were randomly generated in clusters of six around the top four corners of the display and presented in a clockwise order to reduce backtracking of the user. These calibration points were used for all *Pattern-based* calibrations to acquire the corresponding point set data needed for our optimization algorithm. All *Manual* calibrations require an initial transformation to begin the calibration process. The initial transformations we used emulated a best guess by adding a randomly generated translational error of 50 cm and rotational error of 10° to a ground truth transformation.

To compare calibration techniques in a variety of orientations, two experts, experienced in manually tuning FTVR displays, performed *Pattern-based* and *Manual* calibrations through all levels of *AlignmentDoF*. To demonstrate the effect of viewpoint to head-tracked offset, a single expert performed *Pattern-based* and *Manual* calibrations with the highest level of alignment (*3DoF*) through all levels of *OffsetMagnitude*.

Once the calibrations were completed, the display was set to render a perspective-corrected grid pattern. The camera was placed on a tripod and photographs were taken of each successful calibration from eight different locations. The photographs were run through a semi-automatic image processing and analysis script that used a Hough transform to find lines in the images. Lines were paired with their corresponding lines in adjacent screens and the angle at which they met was recorded. A perfect calibration would result in parallel lines and an angle of 0°. Figure 8 illustrates two representative calibration test images.

Calibration time for *Pattern-based* was measured from the first recorded point to the last recorded point. Calibration time for *Manual* was combined from two sources: the time it took to setup the alignment condition, and the time between the start of the first modification and the time at which the participant voiced their completion.

Participants and Apparatus

Two experts in visual calibration techniques with multi-screen FTVR displays were recruited to participate in this study. The study was conducted on a 5-screen (7-inch screens, 1280x800 resolution) box FTVR display.

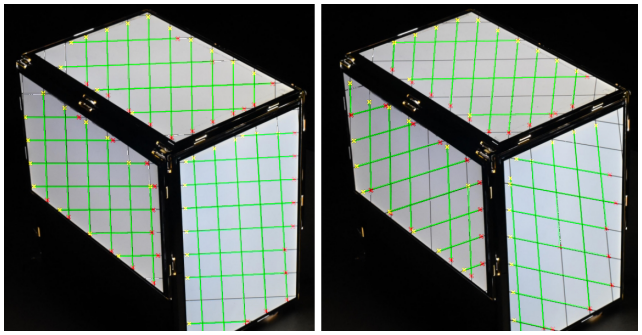

Figure 8. Pattern-based (left, 0.9°) vs. manually tuned (right, 6.3°) calibration error calculated as the average angle between auto-detected grid lines (green) across screens.

We used an OptiTrack (Natural Point, Inc.) 6DoF optical tracking system to perform head tracking, but only positional (3DoF) data was recorded. The display software was written in the Unity game engine and recorded all 3D experimental data. The optimization algorithm and photograph analysis script were implemented in MATLAB. Time was recorded using a stopwatch. All photographs and calibrations were performed using a Nikon D750 with a Nikon 50mm f/1.4 lens.

Design and Hypotheses

The study used a 2x3 within-participants one-way RM-ANOVA with factors *CalibrationTechnique* (Pattern-based, Manual), and *AlignmentDoF* (3DoF, 4DoF, and 6DoF), as well as a 2x3 within-participant one-way RM-ANOVA with factors *CalibrationTechnique* (Pattern-based, Manual) and *OffsetMagnitude* (5 cm, 10 cm, and 15 cm). Dependent measures were calibration time and errors per calibration. Hypotheses were:

H1. *Pattern-based* calibration will be at least as accurate as *Manual* calibration with no head-to-viewpoint translation.

H2. *Pattern-based* calibration will be faster than *Manual* calibration.

H3. *3DoF* will be the most accurate for *Manual* calibration.

H4. *Pattern-based* calibration will be more accurate than *Manual* calibration when there is a head-to-viewpoint translation.

H5. As *OffsetMagnitude* increases, *Manual* calibration will be less accurate.

Results

Errors

Mean calibration errors for varying *AlignmentDoF* and *OffsetMagnitude* are shown in Figures 9 and 10. A one-way RM-ANOVA showed an effect of *CalibrationTechnique* on *AlignmentDoF* ($F_{4,35}$=11.12, p<0.01). A Tukey-HSD multiple comparison post-hoc test reported a significant (p<0.01) difference between *Manual 3DoF* (m 1.24, s.d. 0.58) and *Manual 4DoF* (m 2.54, s.d 0.90). We therefore accept **H3**.

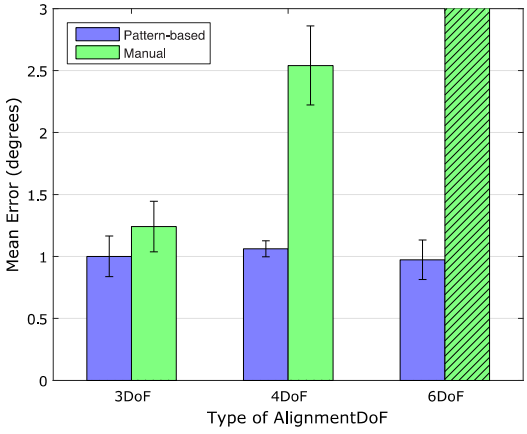

Figure 9. Mean calibration error vs. number of degrees-of-freedom to align for pattern-based versus manual procedures.

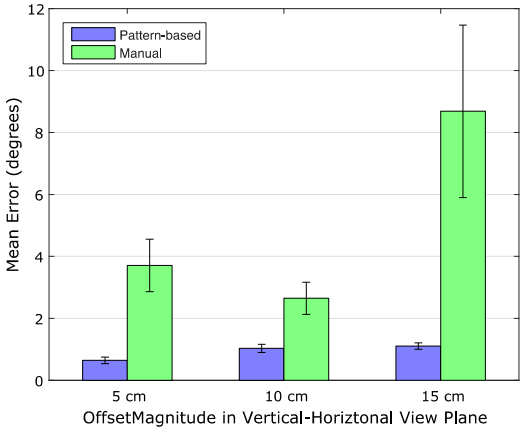

Figure 10. Mean calibration error vs. head-to-viewpoint offset magnitude for pattern-based versus manual procedures.

The post-hoc test also reported that the only significant difference (p<0.01) observed of *CalibrationTechniques* over *AlignmentDoF* is between *Pattern-based 4DoF* (m 1.06, s.d 0.18) and *Manual 4DoF* (m 2.54, s.d. 0.90). We therefore accept **H1**.

A one-way RM-ANOVA showed an effect of *CalibrationTechnique* on *OffsetMagnitude* ($F_{5,42}$=6.29, p<0.01). A Tukey-HSD multiple comparison post-hoc test reported a significant difference (p<0.01) between *Manual 15 cm* (m 8.70, s.d. 7.90) and *Pattern-based 15 cm* (m 1.10, s.d. 0.29), *Pattern-based 10 cm* (m 1.03, s.d 0.36), and *Pattern-based 5cm* (m 0.644, s.d. 0.30). We therefore accept **H4**. It also showed a significant (p<0.01) effect of *OffsetMagnitude* on *Manual* with *Manual 15 cm* error (m 8.70, s.d. 7.90) being much larger than both *Manual 10 cm* (m 2.65, s.d. 1.50) and *Manual 5cm* (m 3.71, s.d. 2.40). We therefore accept **H5**.

Times

A one-way RM-ANOVA showed no significant results for completion time. We therefore reject **H2**. Across all successful trials, we recorded a mean completion time of 2 minutes and 24 seconds and standard deviation of 57 seconds.

Discussion

We were successful in implementing a robust, accurate, and fast pattern-based visual calibration algorithm. It showed consistent and accurate results in all the tested scenarios, while not significantly increasing the amount of time a calibration takes. The quick completion times for both calibration techniques can be partly attributed to the expert's familiarity with the techniques, and the ease of re-defining the head-tracker space using OptiTrack.

Both participants spent a considerable amount of time (> 10 minutes) attempting a *Manual* calibration in the *6DoF* alignment case, but they were unsuccessful in achieving a workable transformation. The axes were so misaligned that a camera-based analysis of the error was not possible. Our inability to measure the error in this case is represented in

Figure 9 as a hatched bar extending beyond the chart. We tested a range of *AlignmentDoF* conditions to demonstrate the robustness our *Pattern-based* calibration, even though a *6DoF* case is unlikely, and avoidable if the head-tracker or display can be moved easily. The *4DoF* case is the most likely to be seen in practice, as it is quite easy to align one of the axes. For example, one could align both the head tracker and display to the floor, thus aligning at least one of their axes.

A box display was used for testing the calibration procedure because it is a challenging calibration problem. The shape of the display allows it to be viewed from almost any location, which makes calibration errors more noticeable. Our cubic display has seams between each screen, which breaks up the pattern making calibration more challenging. We also informally tested a corner-type FTVR display similar to [10] and the visual calibration scheme worked equally well. As future work, we intend to test the calibration procedure with a spherical display since the results of Study #1 suggest that a simulated spherical display is well-suited to pattern alignment.

We used a camera, instead of the participants' eyes, to allow for a quantitative analysis of calibration errors. The camera also removes the binocular disparity that would be present with participant viewing [30], which is problematic for our current cubic display that does not have stereo screens. When calibrating with both eyes, the user would choose the best visual alignment, which should naturally account for eye dominance. However, since the pattern would be viewed from two viewpoints it would not appear perfectly aligned as it does for a monocular viewpoint and we would expect this to introduce a degree of subjective error. As future work, we plan to test the calibration procedure with participants performing visual calibration of a monocular display with binocular viewing and providing a subjective evaluation of the resulting visual error.

Another limitation of our study is that calibrations were performed by experts with a high degree of familiarity with the technique. A future study is needed to test if our method achieves similar accuracy when performed by novices.

Our results show that a head-to-viewpoint translation has a large effect on the accuracy of a calibration if it is left unaccounted for. We solved this problem without requiring the tracking system to measure the orientation of the participant's head, so that our technique would have fewer requirements. Thus, it would be accessible to a larger variety of tracking systems, including those that only provide 3D position data and not head/gaze direction.

Unlike previously reported line-of-sight methods [9] or see-through HMD approaches [31], our approach does not require any additional physical apparatus attached to the display or head-mounted screens that the user looks through. This makes our method much simpler to implement, but at the cost of having poor depth accuracy at each measured viewpoint. We solve this with our depth-invariant calibration formulation: each individual viewpoint measurement ignores tracker errors along the ray of the viewpoint, and then we use multiple measurements surrounding the display to reconstruct a full rigid transformation between the display and tracker. For this reason, *Pattern-based* calibration would not work for single-screen FTVR displays with limited viewing angles, but our interest is primarily in volumetric FTVR displays that provide a stronger 3D effect and have better potential for use in work and entertainment applications.

CONCLUSIONS

Although many types of head-tracked 3D displays have been developed, all of which rely on accurate viewpoint-to-display tracking, there are few generalized calibration techniques for them. This study presented a new pattern-based calibration technique for multi-screen head-tracked displays that uses 3D patterns to guide the viewer into known viewpoint locations. We evaluated the performance of several different patterns coupled with several different simulated display types. We found that the type of display has a larger impact on both accuracy and time than the type of pattern used and that a combined pattern of circles and lines was fastest and most accurate. We tested our proposed calibration procedure through quantitative analysis of visual error with a camera and a physical cubic display. Our results demonstrate that our pattern-based visual calibration technique can effectively solve for the display and viewpoint transformations required for a convincing, artifact-free 3D illusion with a head-tracked cubic display.

ACKNOWLEDGMENTS

This research was funded by the Natural Science and Engineering Research Council of Canada. We would like to thank Sid Fels, Qian Zhou, and Carl Gutwin for their helpful suggestions and feedback.

REFERENCES

1. Kevin W Arthur, Kellogg S Booth, and Colin Ware. 1993. Evaluating 3d task performance for fish tank virtual worlds. ACM *Transactions on Information Systems* (TOIS) 11, 3 (1993), 239-265.
2. K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-D point sets. *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), 698-700.
3. Ronald Azuma and Gary Bishop. 1994. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, 197-204.
4. Michael Bajura and Ulrich Neumann. 1995. Dynamic Registration Correction in Video-Based Augmented Reality Systems. *IEEE Computer Graphics and Applications* 15, 5 (1995), 52-60.

5. John Bolton, Kibum Kim, and Roel Vertegaal. 2011. SnowGlobe: a spherical fish-tank VR display. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1159-1164.
6. Christoph W Borst. 2004. Tracker calibration using tetrahedral mesh and tricubic spline models of warp. In *Virtual Reality, 2004. Proceedings. IEEE*, 19-26.
7. Luca Cosmo, Andrea Albarelli, Filippo Bergamasco, and Andrea Torsello. 2014. Design and Evaluation of a Viewer-Dependent Stereoscopic Display. In *ICPR*. 2861-2866.
8. Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. 1993. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 135-142.
9. Marek Czernuszenko, Daniel Sandin, Thomas DeFanti. 1998. Line of sight method for tracker calibration in projection-based VR systems. In *Proceedings of 2nd international immersive projection technology workshop*. Citeseer, 11-12.
10. Tom JP Djajadiningrat, Gerda JF Smets, and Kees CJ Overbeeke. 1997. Cubby: a multiscreen movement parallax display for direct manual manipulation. *Displays* 17, 3 (1997), 191-197.
11. Fernando Ferreira, Marcio Cabral, Olavo Belloc, Gregor Miller, Celso Kurashima, Roseli de Deus Lopes, Ian Stavness, Junia Anacleto, Marcelo Zuffo, and Sydney Fels. 2014. Spheree: a 3D perspective-corrected interactive spherical scalable display. In *ACM SIGGRAPH 2014 Posters*. ACM, 86.
12. Anton Fuhrmann, Dieter Schmalstieg, and Werner Purgathofer. 1999. Fast calibration for augmented reality. *In Proceedings of the ACM symposium on Virtual reality software and technology*. ACM, 166-167.
13. S Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47, 6 (2014), 2280-2292.
14. John G Hagedorn, Steven G Satterfield, John T Kelso, Whitney Austin, Judith E Terrill, and Adele P Peskin. 2007. Correction of location and orientation errors in electromagnetic motion tracking. *Presence: Teleoperators and Virtual Environments* 16,4 (2007): 352-366.
15. Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. 2012. HoloDesk: direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2421-2430.
16. Richard L Holloway. 1997. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments* 6,4 (1997), 413-432.
17. Masahiko Inami. 1997. Media3: the virtual hologram. In *ACM SIGGRAPH 97 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH '97*. ACM, 107.
18. Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*. IEEE, 1999.
19. Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. TeleHuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2531-2540.
20. Volodymyr V Kindratenko. 2000. A survey of electromagnetic position tracker calibration techniques. *Virtual Reality* 5, 3 (2000), 169-182.
21. Arun Kulshreshth and Joseph J LaViola, Jr. 2013. Evaluating performance benefits of head tracking in modern video games. In *Proceedings of the 1st symposium on Spatial user interaction*. ACM, 53-60.
22. Arun Kulshreshth and Joseph J LaViola, Jr. 2016. Dynamic Stereoscopic 3D Parameter Adjustment for Enhanced Depth Discrimination. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 177-187.
23. Franz Madritsch, Franz Leberl, and Michael Gervautz. 1996. Camera based beacon tracking: accuracy and applications. In *VRST*, Vol. 96. 101-108.
24. Kenneth Moser, Yuta Itoh, Kohei Oshima, J Edward Swan, Gudrn Klinker, and Christian Sandor. 2015. Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique. *IEEE transactions on visualization and computer graphics* 21, 4 (2015), 491-500.
25. Kazuhisa Nakada, Masahiko Nakamoto, Yoshinobu Sato, Kozo Konishi, Makoto Hashizume, and Shinishi Tamura. 2003. A rapid method for magnetic tracker calibration using a magneto-optic hybrid tracker. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 285-293.
26. Alexander Plopski, Yuta Itoh, Christian Nitschke, Kiyoshi Kiyokawa, Gudrun Klinker, and Haru Takemura. Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics* 21, 4 (2015), 481-490.

27. Behzad Sajadi, and Aditi Majumder. 2011. Automatic Registration of Multi-Projector Domes Using a Single Uncalibrated Camera. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1161-1170.
28. Ian Stavness, Billy Lam, and Sidney Fels. 2010. pCubee: a perspective-corrected handheld cubic display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1381-1390.
29. Yichen Tang, Billy Lam, Ian Stavness, and Sidney Fels. 2011. Kinect-based augmented reality projection with perspective correction. In *ACM SIGGRAPH 2011 Posters*. ACM, 79.
30. Robert J Teather and Wolfgang Stuerzlinger. 2013. Pointing at 3d target projections with one-eyed and stereo cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 159-168.
31. Mihran Tuceryan and Nassir Navab. 2000. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on.* IEEE, 149-158.
32. Andrejs Vorozcovs, Wolfgang Stürzlinger, Andrew Hogue, and Robert S Allison. 2006. The hedgehog: a novel optical tracking method for spatially immersive displays. *Presence* 15, 1 (2016), 108-121.
33. Colin Ware, Kevin Arthur, and Kellogg S Booth. 1993. Fish tank virtual reality. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. ACM, 37-42.
34. Joshua T Weinhandl, Brian SR Armstring, Todd P Kusik, Robb T Barows, and Kristian M O'Connor. 2010. Validation of a single camera three-dimensional motion tracking system. *Journal of biomechanics* 43, 7 (2010), 1437-1440.
35. Björn Wöldecke, Dionysios Marinos, and Christian Geiger. 2014. Flexible Registration of Multiple Displays. In *Proceedings of The International Symposium on Pervasive Displays*. ACM, 86.
36. Qian Zhou, Gregor Miller, Kai Wu, Ian Stavness, and Sidney Fels. 2016. Analysis and Practical Minimization of Registration Error in a Spherical Fish Tank Virtual Reality System. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, to appear.
37. Qian Zhou, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. 2017. Automatic Calibration of a Multiple-Projector Spherical Fish Tank VR Display. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WAVC)*, to appear.