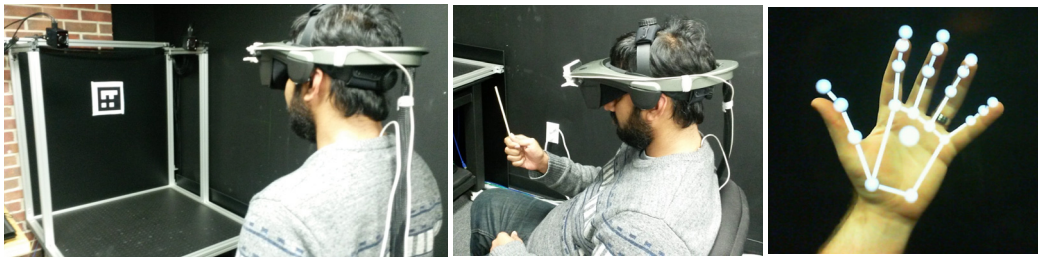
Evaluation of User-Centric Optical See-Through Head-Mounted Display Calibration Using a Leap Motion Controller

Kenneth R Moser* J. Edward Swan II†

Mississippi State University



(a) (b) (c)

Figure 1: Examples of manual OST AR calibration procedures. (a) A standard environment-centric calibration performed by aligning on-screen points with the center of a fiducial marker in the world. (b) A user-centric calibration performed by aligning on-screen points with the tip of a stylus held by the user. (c) A virtual hand overlay generated using Leap Motion tracking information.

Abstract

Advances in optical see-through head-mounted display technology have yielded a number of consumer accessible options, such as the Google Glass and Epson Moverio BT-200, and have paved the way for promising next generation hardware, including the Microsoft HoloLens and Epson Pro BT-2000. The release of consumer devices, though, has also been accompanied by an ever increasing need for standardized optical see-through display calibration procedures easily implemented and performed by researchers, developers, and novice users alike. Automatic calibration techniques offer the possibility for ubiquitous environment independent solutions, un-reliant upon user interaction. These processes, however, require the use of additional eye tracking hardware and algorithms not natively present in current display offerings. User dependent approaches, therefore, remain the only viable option for effective calibration of current generation optical see-through hardware. Inclusion of depth sensors and hand tracking cameras, promised in forthcoming consumer models, offer further potential to improve these manual methods and provide practical intuitive calibration options accessible to a wide user base.

In this work, we evaluate the accuracy and precision of manual optical see-through head-mounted display calibration performed using a Leap Motion controller. Both hand and stylus based methods for monocular and stereo procedures are examined, along with several on-screen reticle designs for improving alignment context during calibration. Our study shows, that while enhancing the context of reticles for hand based alignments does yield improved results, Leap Motion calibrations performed with a stylus offer the most accurate and consistent performance, comparable to that found in previous studies for environment-centric routines. In addition, we found that stereo calibration further improved precision in every case. We believe that our findings not only validate the potential of hand and gesture based trackers in facilitating optical see-through calibration methodologies, but also provide a suitable benchmark to help guide future efforts in standardizing calibration practices for user friendly consumer systems.

Index Terms: H.5.1 [[Information Interfaces and Presentation]: Multimedia Information Systems]: Artificial, augmented, and virtual realities—; H.5.2 [[Information Interfaces and Presentation]: User Interfaces]: Ergonomics, Evaluation/methodology, Screen design—

1 Introduction

The growing availability of consumer level Optical See-Through (OST) Head-Mounted Displays (HMDs) has produced a burgeoning market for new and innovative Augmented Reality (AR) applications intended for use by the general populace. This surge in consumer access is accompanied by a rapidly growing need for standardized calibration procedures offering intuitive and easily implemented methodologies for researchers, developers, and novice users alike. While past offerings in the consumer domain, for example the Google Glass and Epson Moverio BT-200, include standard sensory options, such as gyroscopes, accelerometers, and built-in cameras; next generation OST devices, including the Microsoft HoloLens and Epson Moverio Pro BT-2000, promise a wide variety of additional integrated and on-board sensors. Depth, gesture, and hand tracking cameras, in particular, offer unique potential for not only novel application and interface designs, but also for the creation of standard calibration practices and methodologies applicable across devices and markets.

The goal of OST HMD calibration is to properly model the user's viewpoint through the display screen. Unlike Video See-Through (VST) AR, where the user's view of the world is provided by means of an externally mounted camera, OST AR allows virtual content to be directly overlaid onto the real world from the user's own natural perspective. Since the "camera" in an OST system is the user's eye itself, it is not possible to use standard image processing techniques for calibrating the viewing frustum. Approximation methods must instead be used to estimate the user's perspective. The accuracy of these methods ultimately determines the registration quality, or correctness in location, of virtual content in relation to physical objects, such as an overlay registered to a user's hand, Figure 1 (c).

*e-mail: moserk@acm.org

†e-mail: swan@acm.org

IEEE Symposium on 3D User Interfaces 2016
19–20 March, Greenville, SC, USA
978-1-5090-0842-1/16/$31.00 ©2016 IEEE

Automatic calibration is an increasing possibility for future display systems, with promising results being shown by recently developed approaches, such as Itoh and Klinker's INteraction-Free DIsplay CAlibration (INDICA) [9] and Plopski et al.'s Corneal Imaging Calibration (CIC). These methods utilize additional camera hardware and image processing algorithms to directly measure the location of the user's eye within the head-mount, allowing dynamic calibration updates at run-time. While automatic processes fundamentally allow ubiquitous, any time any where, calibration, off the shelf commercial HMD offerings do not natively include the necessary eye-tracking hardware fundamental to these procedures, making the applicability of INDICA and similar approaches limited at the current time. As a result, calibration techniques for OST displays have traditionally relied on manual interaction methods, which estimate the viewing frustum through user feedback.

Common manual calibration techniques, such as Tuceryan et al.'s [28] Single Point Active Alignment Method (SPAAM), require users to move about their environment and visually align on-screen points to real world target locations, as demonstrated in Figure 1 (a), with the fiducial marker indicating a known world location. Providing that an adequate number of screen to world correspondence pairs are obtained, an estimation of the perspective projection, used to model the user's viewpoint, can be obtained. Traditionally, target points taken from the world are static locations or fiducial markers whose locations are measured off-line or actively tracked within the user's environment. Unfortunately, the use of external objects or patterns for alignments inherently imposes dependency on the environment restricting users to specific locations or forcing them to construct and maintain portable targets and fiducials. Nevertheless, environment agnostic calibration may be feasible through utilization of body tracking technologies, allowing the replacement of environment dependent target points with the user's own fingers and hands.

The Leap Motion controller is one commercially available solution for providing low cost hand and stylus tracking. Differing from the Microsoft Kinect and Asus Xtion, which are primarily whole body motion trackers, the Leap Motion affords close proximity gesture and hand detection for natural user interaction and interfaces in both VR and AR applications. Unpredictable tracking accuracy, particularly for occluded or oblique hand orientations, imposes some limitation on the utility of the technology. However, small size, form factor, and software versatility make the Leap Motion ideally suited for integration into consumer HMD offerings. The ability to perform SPAAM-like manual calibration using hand and finger alignments, as seen in Figure 1 (b), liberates the user from environment dependent target constraints and achieves a user-centric methodology, easily integrated into most OST AR systems, easing the burden on not only users, but also application developers alike.

In this work, we present the results of an investigation into the efficacy of the Leap Motion controller as a means for facilitating hand and stylus alignment based calibration of an OST HMD system. In contrast to a previous cursory demonstration [30], our study additionally includes an examination of accuracy and precision differences between monocular and stereo calibration variants. We also explore several reticle designs and the effect of alignment context on hand calibration results. Our analysis employs standard objective measures and metrics, including reprojection error and extrinsic eye location estimates, to compare not only the performance of each condition, but also the viability of OST calibration with Leap Motion in general. We provide a final discussion of our results and their benefit to standardizing calibration practices, and close with recommendations and remarks to guide future work.

2 Background and Related Work

Previous studies evaluating or presenting improved calibration techniques primarily focus on accuracy and time reduction measures. While these metrics are fundamental to comparisons of efficiency and robustness between methods, considerably less attention has been attributed to reducing the restrictions that calibration procedures impose on users and system developers.

2.1 Environment-Centric Calibration

Initial modalities for determining a user's perspective through an OST HMD required rigid head fixation, [4, 5, 13], while users manually adjusted the location of on-screen reticles to align with specific target points in the environment. Even though these bore-sighting approaches offer a hardware independent calibration solution, inhibition of user movement makes this methodology both tedious and impractical for consumer use. Tuceryan and Navab's SPAAM procedure [28] relaxes this restriction allowing far greater freedom of movement, while users perform screen to world alignments. A minimum of 6 alignments is required to solve for the 11 free parameters, plus a scale factor, of the 3×4 projection matrix used to describe the viewing frustum created by the user's eye and the display. Subsequent developments by Genc et al. [6] extend SPAAM to stereo OST HMDs, allowing simultaneous calibration of both the left and right eye views. Stereo SPAAM alignments utilize 3D virtual points by exploiting stereoscopic depth cues produced by binocular displays. The usability gains of SPAAM calibration, though, is counter balanced by the introduction of misalignment error caused by movement, or sway, of the user's body and head during the procedure.

Tang [26] revealed that misalignment error could be ameliorated, to an extent, by increasing the variation of screen to world alignments in depth. Axholt et al. [1] further indicates that Tang's Depth SPAAM achieves more stable results with 20–25 alignments well distributed to cover a larger region of the environment. However, related studies, [3, 14], show that error contributions from user misalignment are highly dependent on display type, user stance, and input methods. Alternative two phase calibration methods provide various means for not only reducing user interaction, but also environmental constraints as well.

Easy SPAAM [7, 16] optimizes previous calibration results by adjusting the projection matrix to an updated location of the user's eye triangulated through standard user driven 2D-3D alignments. This approach, while still bounded by the accuracy of the existing calibration data, significantly reduces the number of screen to world correspondences required and also isolates the impact of further alignment error to the new extrinsic, eye location, measures. Owen et al.'s [19] Display Relative Calibration (DRC) removes the dependency on previous calibration results by fully decoupling the eye's perspective into intrinsic, display dependent, and extrinsic, user specific, components, and implements two separate procedures for independent measure of each. The first phase, conducted off-line, directly computes display specific parameters through image processing and computer vision techniques. The second phase, similar to that of Easy SPAAM, still utilizes a small number of screen-world alignments for triangulating the extrinsic location of the user's eye relative to the display screen. Innovative, recently proposed, two step calibration methods seek to improve the second phase of the DRC approach and completely eliminate user interaction, creating a user-centric, environment independent, process as a consequence.

2.2 User-Centric Calibration

Itoh and Klinker [9] enhance the on-line component of the DRC methodology, by automating the measurements needed to estimate user eye location. Their INDICA technique employs image processing methods, similar to Swirski's iris detection [25] and Nitschke's localization algorithm [17], to estimate the 3D position of the user's eye relative to miniature cameras mounted within the HMD. Plopski et al.'s Corneal-Imaging Calibration [20] parallels this approach by using feature detection algorithms to identify the orientation

of known patterns reflected onto the user's cornea by the display screen. Comparing the distortion of the reflected features to a simplified model of the eye's structure, yields more accurate 3D localization estimates compared to direct iris detection. Both techniques, though, fully remove the need for user interaction and provide a completely autonomous process, ideally suited for consumer devices. Unfortunately, these automatic schemes require additional built-in eye-tracking cameras, and related components, not yet standard in current HMD products. Additionally, the high level of expertise needed to effectively deploy ad-hoc versions of these systems is beyond the capability of casual users. Therefore, manual interaction methods are still the most viable calibration option for current consumer OST HMD offerings.

Alternative techniques, which diverge from the two phase methodology, but maintain a user-centric approach, employ hand held tracking markers for alignment based calibration [12, 18]. Moving the marker around the field of view, mimicking the distance variation of Depth SPAAM, yields viable accuracy results and reduces user movement since the entire procedure may be conducted while seated. Even though this method offers a usable environment agnostic procedure, like other manual calibrations it too ultimately requires the possession of external alignment targets to proceed. Hand and gesture tracking technology, such as the Leap Motion controller, may be a viable option for replacing hand held markers with the hand itself.

2.3 Leap Motion Accuracy and Precision

The Leap Motion controller employs modern computer vision algorithms to process images from stereo IR cameras, facilitating near proximity hand, gesture, and stylus tracking. Research applications for the device center primarily around innovative and novel computer interfaces [21, 24, 27, 31] as well as VR interaction capabilities [22, 23]. A number of formal evaluations targeting the accuracy and precision of Leap Motion tracking [8, 11, 29] have shown that, while both hand and stylus accuracy quickly diminish at distances greater than 250mm from the device, stylus tracking to within 0.7mm or less is possible. Measurements for hand tracking showed much larger fluctuations according to the degree of hand and finger orientation and occlusion, as well as hand shape and size. Despite current tracking inconsistencies, the accessibility and current prominence in VR applications offers a strong foundation for transitioning use of Leap Motion into OST AR, and exploiting body tracking for calibration use. Our study provides the first formal evaluation of attainable calibration accuracy in OST AR systems using Leap Motion controllers.

3 Experiment Design

We construct a complete OST AR framework, combining a Leap Motion controller with a commercially available HMD. Since the Leap Motion is able to perform both hand and stylus tracking, we utilize each for performing monocular and stereo SPAAM based calibrations of our system. We employ a straight-forward manual procedure, requiring multiple alignments between on-screen reticles and either a finger or stylus. While achieving consistent alignments to a single point on a stylus is relatively intuitive, the ability of a user to maintain repeated alignments with a single point on a finger tip is far more inexact. Instead of utilizing an additional physical cap, ring, or other wearable indicator, we incorporate several variations of on-screen reticle design to provide visual context for aiding the user in properly positioning their finger during calibration alignments. We believe this approach more adequately represents what a viable consumer oriented calibration mechanism would provide.

3.1 Hardware and Software

An NVIS ST50 binocular OST HMD is used as the primary display for our study. The ST50 is capable of producing stereo images

Figure 2: (right) Our HMD and Leap Motion system. (left) A user performing stylus to screen alignments. The right-handed head relative coordinate frame, in relation to the Leap Motion, is shown. Positive X extends horizontally to the user's left, positive Y forward, and positive Z downward to the ground.

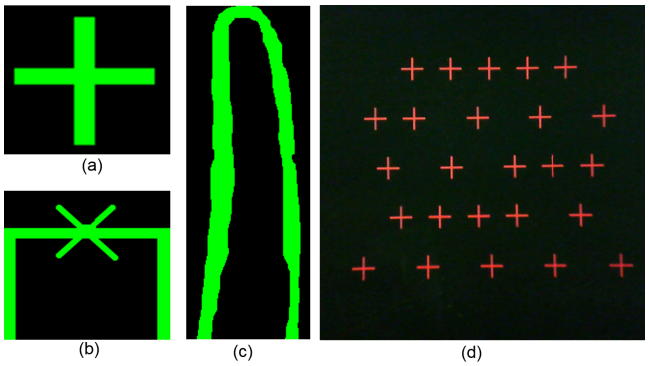


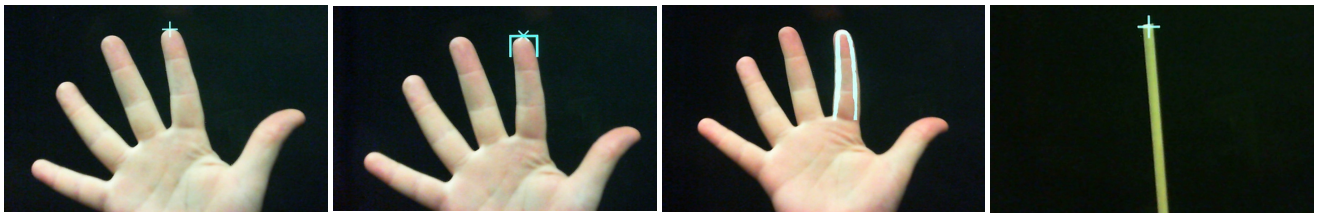
Figure 3: Illustrations of the cross (a), box (b), and finger (c) reticles used for hand alignments. (d) View through the HMD of the complete grid of alignment points for left eye calibration.

with a resolution of 1280×1024, a 40° horizontal and 32° vertical field of view, and a manufacturer specified spatial resolution of 1.88 arcmn/pxl. A custom 3D printed mount attaches the Leap Motion to the front of the display. Figure 2 shows the complete assembly and orientation of the Leap Motion tracking coordinate frame. The tracking calibration of the Leap Motion controller itself was confirmed using the resources included with the Leap Motion software. Integrating the Leap Motion tracking information into our HMD calibration software is performed using version 2.3.1.31549 of the Leap Motion SDK. An Alienware m18 laptop, i7-4700MQ 2.4GHz processor with 16 GB RAM running Windows 7 x64, is used to drive the graphics for the display, via HDMI, as well as run all necessary software applications.

3.2 Alignment Methods

Calibration of our system, by a user, is performed by aligning a sequential series of on-screen reticles to either the index finger tip on the user's right hand or to the end of a hand held stylus. As previously noted, we incorporate several reticle designs to provide a visual indication of finger orientation during alignment. However, only a single reticle type is employed for stylus alignments, since the tip in this case is more well defined.

Hand Alignments: Tracking data from the Leap Motion is able to provide the position and orientation of numerous points along the hands and arms, as long as they are within the field of view of the device. We reduce the alignment complexity of hand calibration by using only the position information for the tip of the right index finger. A more thorough specification for the location of this point, relative to the finger bones, is provided within the Leap Motion's accompanying documentation. We employ three on-screen reticle



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

Figure 4: Stylus and hand alignments for each reticle design, as seen through the HMD. (a) Cross reticle alignment. (b) Box reticle alignment. (c) Finger reticle alignment. (d) Stylus alignment.

designs to provide varying degrees of context to aid in finger positioning during calibration alignments. Perfect alignment occurs when the center of the right index finger tip coincides with the target point specified for each reticle.

Cross A simple cross-hair, Figure 3 (a), comprised of a horizontal and vertical line, is displayed with the target point located at the center of the intersection point. The on-screen dimensions of the cross are 64×64 pixels with line thickness of 3 pixels. The ubiquitous application of cross-hairs for targeting and aiming purposes makes this a natural design for alignment procedures. Figure 4 (a) illustrates an alignment between a hand and cross reticle as seen through our HMD system.

Box The design of the box reticle, Figure 3 (b), is intended to provide a pseudo cap for the user's finger. The box reticle is created from a 3 sided rectangle displayed with an 'X' placed on the upper edge intersecting the target alignment point. The on-screen dimensions of the box are 128×128 pixels with line and 'X' thickness of 10 and 5 pixels respectively. The structure of the box design is such that a user would naturally center their finger within the outlined region, improving the likely hood of consistent alignment to the finger tip. Figure 4 (b) illustrates an alignment between a hand and box reticle as seen through our HMD system.

Finger The finger reticle, Figure 3 (c), provides an anatomical finger outline onto which the user's real finger is aligned. The target point for this reticle is located at the center tip of the outline's upper edge. The on-screen dimensions of the finger is 128×384 pixels with average outline thickness of 20 pixels. The finger reticle is intended to provide the most position context of all three designs. The center portion of the reticle is not filled in order to allow the user a clear view of their finger throughout alignment. While it is possible to provide a completely solid design, the brightness of the display often inhibits the ability to clearly distinguish the location of real objects behind AR content. Figure 4 (c) illustrates an alignment between a hand and finger reticle as seen through our HMD system.

Stylus Alignments: A stylus, or tool as it is referred to in the Leap Motion documentation, is treated as a simplified cylinder object with a length, diameter, and tip located at the center point of the cylinder's end. The tip position and pointing direction of any cylindrical object of appropriate length and diameter, as specified within the Leap Motion documentation, is automatically tracked by the Leap Motion as long as an appropriate amount of the object is visible within the Leap Motion's field of view. Our stylus object is created from a 5mm diameter wooden dowel rod, approximately 20cm in length, held to allow the majority of the length to extend beyond the user's hand. The cross reticle, as described previously, is used for stylus calibrations, and perfect alignment occurs when the center of the cross coincides with the center of the stylus tip. Figure 4 (d) illustrates a stylus alignment as seen through our HMD system.

3.3 Calibration Procedure

We utilize a standard alignment based SPAAM procedure to calibrate our OST AR system. As recommended in Axholt et al. [1], a total of 25 alignments is used to generate the final calibration results. The distribution of on-screen alignment points is arranged to produce a 5×5 grid, as shown in Figure 3 (d), providing an approximately even coverage of on-screen area. In order to facilitate binocular calibration of both eyes simultaneously, the on-screen location of all reticles is shifted differently for each eye so that the fused left and right binocular reticle image will be perceived stereoscopically in depth by the user, between 200–400cm, or within arms reach. Even though the placement of on-screen reticles differs between left and right eyes, we do not change the patterns between monocular and stereo calibrations, or across calibration sets. Therefore, a consistent pattern is always shown to the left eye regardless of condition, and likewise a consistent sequence of locations is maintained for the right eye alignments regardless of condition. Both hand and stylus calibration conditions proceed in an identical manner.

A single on-screen reticle is rendered to the display screen. The user then moves their right index finger or stylus tip until it aligns as closely as possible to the target point of the reticle, as previously described. Once a sufficiently accurate alignment has been achieved, a button press, on either a keyboard or wireless controller activates recording of positional data acquired by the Leap Motion. Throughout the recording process, the color of the on-screen reticle is changed from green to yellow providing visual confirmation to the user that measurement has begun and to indicate that the alignment should be maintained until recording has ceased. The 3D finger or stylus tip location, relative to the Leap Motion coordinate frame, is measured every 100ms for 1sec resulting in 10 data points per alignment. The median X, Y, and Z position value from the 10 recording points is used as the final measure. This location estimate, along with the X, Y, screen pixel location of the reticle's target point, is saved, and the complete set of 25 world and screen correspondence pairs is combined to produce the calibration result.

Monocular calibration sets always proceed by first calibrating the left eye followed by the right without interruption. Stereo calibration sets, of course, produce both left and right results together. Additionally, the user is instructed to perform all hand alignments in an identical manner, by keeping their palm flat and facing toward the display screen with all five fingers as evenly spaced as possible. We impose this requirement to maintain a hand tracking quality as consistent as possible across conditions. All stylus alignments are also performed with the user holding the stylus in their right hand. No further restrictions were placed on the alignment procedure.

4 Analysis of Results

We recorded alignment data from repeated calibrations by a single expert user. Our primary objective is to verify the efficacy of the Leap Motion controller itself, for calibrating OST displays, and not the inherent usability or intuitiveness of our design. In this regard, repeated measures from an expert user, knowledgeable with the procedure as similarly employed by [9, 10], provide more stable results, void of subjective affects. The ordering of reticle conditions, along with monocular and stereo presentation, was randomly permuted across all conditions. Our expert completed 20 monocular and 20 stereo calibrations for each of the three hand and single stylus

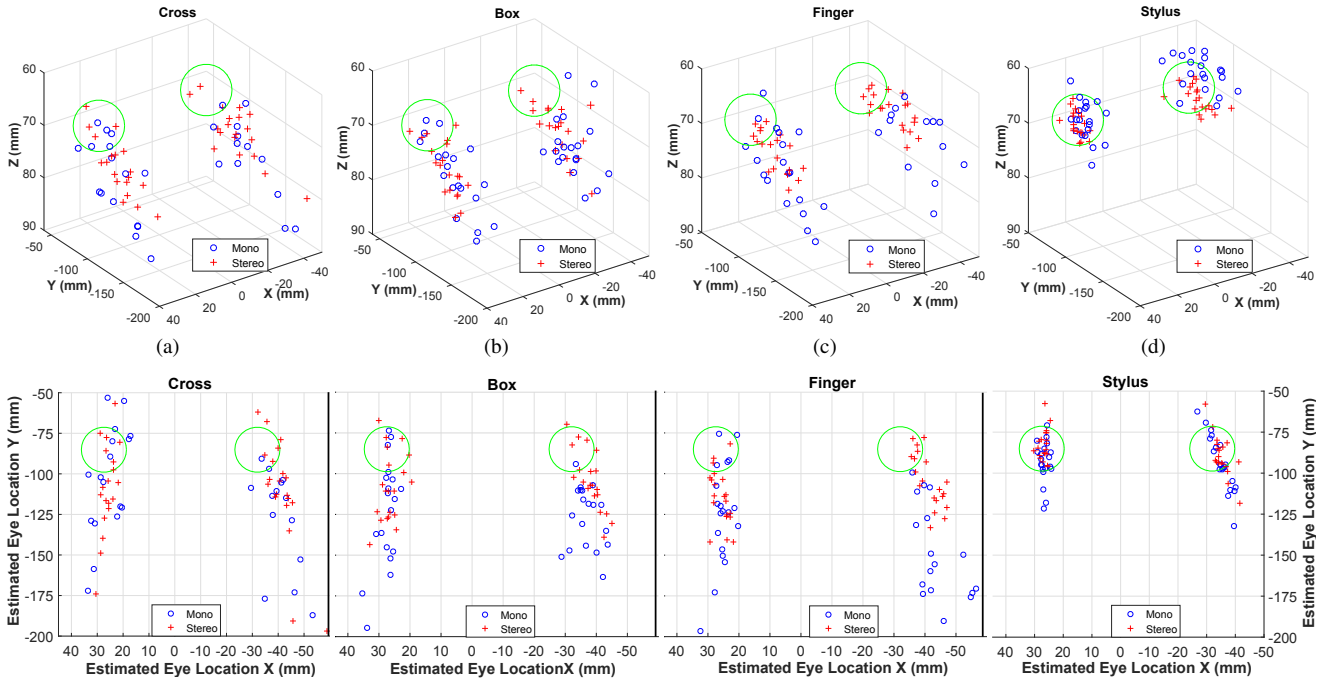


Figure 5: Estimated user eye locations relative to the Leap Motion coordinate frame. (a) Cross reticle, (b) Box reticle, (c) Finger reticle, and (d) Stylus calibration 3D position estimates. 2D eye position plots showing only X and Y estimate locations for (e) Cross reticle, (f) Box reticle, (g) Finger reticle, and (h) Stylus calibrations. In all plots, the center of the Leap Motion is at location (0, 0, 0), with monocular calibration estimates displayed in blue, stereo calibration estimates plotted in red, and green circles used to denote an area of plausible eye points.

alignment methods, resulting in $20 \times 4 \times 2 = 160$ calibrations total. Each calibration session lasted between 30 and 40 minutes, with the entire data collection task requiring an estimated 12 hours, spread over a period of two weeks. In order to decrease the amount of discomfort experienced by our expert, we limited the amount of time that the display was worn to no more than 2 hours in a day. Before beginning the data collection task, we carefully aligned the OST HMD optics for the user's interpupillary distance, 62mm. However, the user did not perform any additional optical alignments before each calibration session; instead the user simply placed the HMD upon their head and began calibrating. Although this means that the user's eyes were not always aligned with the optical centers of the HMD's monocles, we felt that this represented a more realistic use pattern for HMD calibration.

We utilize three primary metrics for determining the accuracy of calibration results for each condition. First, the estimated eye location of the user relative to the Leap Motion coordinate frame is obtained by decomposing the extrinsic component from the SPAAM calibration results. This is a standard evaluation metric for OST HMD calibration, [2, 10, 15]. Our second metric is reprojection error, calculated as the difference between the ground truth on-screen position of each reticle, used during calibration, and the screen coordinate that results from back projecting the corresponding finger or stylus 3D tip position using the calibration result for that alignment set. We additionally examine binocular disparity values, taken as the difference between the left and right eye location estimates for each monocular and stereo calibration, along each axis relative to the Leap Motion coordinate frame.

We tested dependent measures with repeated-measures analysis of variance (ANOVA). We used calibration session (1–20) as the random variable, in order to remove session-to-session variance that may have resulted from the slightly changing alignment between the user's eyes and the OST HMD optics. For each ANOVA test, if Mauchly's test indicated non-sphericity, we adjusted the p-value according to the Huynh-Feldt ε value; in such cases we report ε along with the ANOVA F-test.

4.1 Eye Location Estimates

Figure 5 provides the estimated eye locations obtained from the monocular and stereo calibration results of each alignment method. Since the actual ground truth location of the user's eye, with respect to the Leap Motion, is not directly attainable in our system, a region of plausible locations, based on measurements of the physical display screen and the user's interpupillary distance (IPD) of 62mm, is presented within each plot. The 3D eye positions, relative to the Leap Motion coordinate frame, are provided in Figure 5 (a) through (d) for the Cross, Box, Finger, and Stylus alignment method results respectively. Likewise, a top-down view, showing only the positions relative to the Leap Motion's X and Y axis, are plotted for each alignment condition in Figure 5 (e) through (h).

Visual inspection of both 3D and 2D eye location estimates show that the stylus alignment method produced the most accurate extrinsic results, in relation to plausible ground truth eye positions. The highest level of precision, for both monocular and stereo calibrations, likewise occurs for stylus alignments. Nearly all values for stylus alignments fall within plausible locations. Similar to the findings of Axholt [1] and Moser et al. [15], the primary errors occur along the depth dimension (the Y axis). The three hand alignment methods produce noticeably less uniform results.

A more quantitative measure of precision is the amount of variance, or spread, in location estimates relative to the centroid value of each related group. We obtain this value by calculating the Euclidean distance from each 3D eye estimate location to the corresponding median position for all estimates relative to each eye and alignment method grouping. Distance values for each alignment method group are provided in Figure 6. Mean and standard deviation values for the group median distances, with both left and right eye values combined, are provided in Table 1. We performed

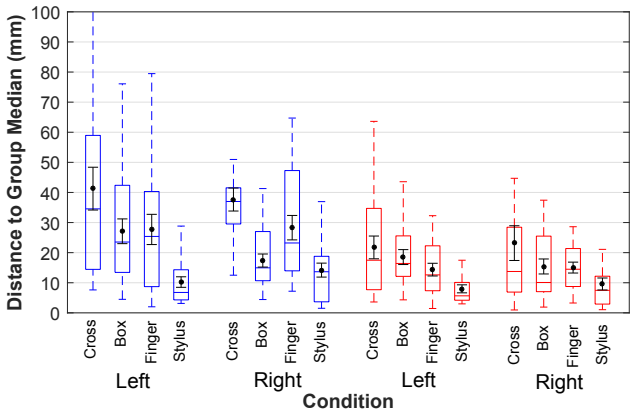


Figure 6: Distances between estimated eye positions and the median location value for monocular (blue) and stereo (red) calibrations for each alignment method. Mean and ±1 SEM bars are shown for the values within each condition group.

ANOVA analysis to identify significant differences between the four alignment method conditions in relation to both monocular and stereo calibration results. ANOVA results indicate significant differences between alignment methods for monocular calibrations ($F(3,57) = 13.3, p < 0.001$). A post-hoc Tukey-Kramer honest significant difference test confirms that the cross alignment method median group distances are significantly higher compared to the remaining three methods at $p < 0.001$. No further significance between groups is found. ANOVA results for stereo calibrations, similarly indicate significant differences between alignment methods ($F(3,57) = 4.7, p = 0.01, \varepsilon = 0.68$), with post-hoc analysis confirming that the cross alignment method median group distances, for stereo calibration, are significantly higher than the finger and stylus distances at $p < 0.001$. No further significance between groups is found.

We also directly compare the monocular and stereo results within each alignment method, combining values for left and right eye for each condition. We find that cross distances significantly differ between mono and stereo calibration ($F(1,19) = 7.7, p = 0.01$). Distances to the group median for the finger alignment method also differ significantly between monocular and stereo calibrations ($F(1,19) = 15.6, p < 0.001$). Mono and stereo calibrations for the box and stylus methods do not significantly differ ($F(1,19) = 2.4, p = 0.14$ and $F(1,19) = 1.8, p = 0.20$, respectively).

4.2 Reprojection Error

Figure 7 provides a visualization to illustrate the reprojection error amounts in stereo calibrations performed using the cross and stylus alignment methods. Direction vectors, drawn in red, represent the difference in screen position, pixels, between the back projected coordinate of each alignment point and the ground truth reticle location used during calibration, drawn in blue. Figure 8 provides the complete set of reprojection error values for both mono and stereo calibrations within each of the four alignment methods. Additionally, mean and standard deviation values for the error in each condition, with both left and right eye groupings combined, are provided in Table 1.

Visual inspection of the reprojection error values of Figure 8 clearly shows a noticeably lower occurrence of error in stylus calibration results compared to the three finger alignment methods. Repeated measures ANOVA for monocular calibration results indicates a significant difference across alignment method ($F(3,117) = 41.0, p < 0.001, \varepsilon = 0.68$). Post-hoc analysis confirms that reprojection error values for both the cross and stylus alignment method are significantly different from all other methods ($p < 0.001$). No further significance between groups were found. ANOVA for stereo calibration results, similarly, indicates a significant difference between alignment method reprojection error ($F(3,117) = 45.8, p < 0.001, \varepsilon = 0.54$), with post-hoc results showing that the cross and stylus alignment methods' reprojection error values differ significantly from all other methods ($p < 0.001$).

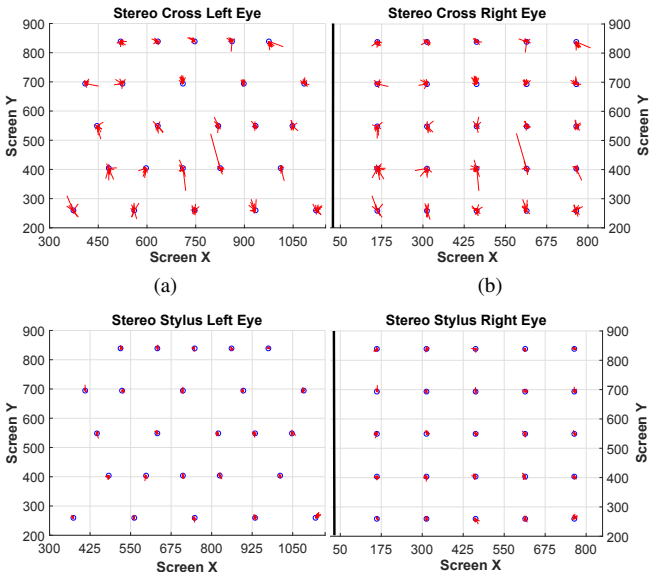


Figure 7: Reprojection error plots for stereo calibrations using the Cross reticle, (a) and (b), and Stylus, (c) and (d). Blue circles indicate ground truth reticle location during alignment. Red lines indicate direction and magnitude of reprojection error in pixels.

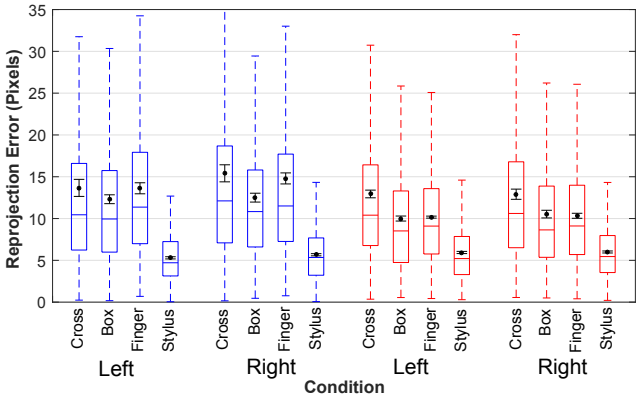
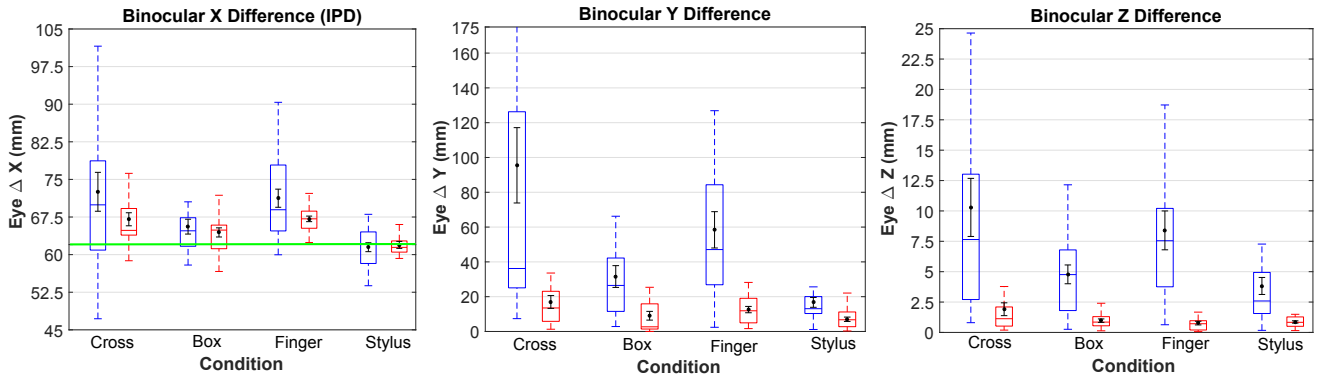


Figure 8: Reprojection error for monocular (blue) and stereo (red) calibrations of each alignment method. Mean and ±1 SEM bars are shown for the values within each condition group.

We perform additional ANOVA analysis to directly compare reprojection error results between mono and stereo calibrations of each alignment method. No significant difference is found between mono and stereo reprojection error for cross calibrations ($F(1,39) = 2.1, p = 0.15$), though a significant difference in reprojection error, between mono and stereo calibrations, does occur for the box alignment method ($F(1,39) = 8.7, p = 0.005$), as well as for the finger and stylus methods ($F(1,39) = 30.6, p < 0.001$ and $F(1,39) = 9.4, p = 0.004$, respectively).

4.3 Binocular X, Y, Z Disparity

Our final evaluation metric is the difference between the separate X, Y, and Z components of the paired left–right eye location estimates (Figure 5) from mono and stereo calibrations. These three binocular



(a) (b) (c)

Figure 9: Differences between left and right eye location estimates for monocular (blue) and stereo (red) calibration pairs. Difference between X (a) positions represents interpupillary distance (IPD). The green line indicates the measured IPD, 62mm, of the expert user. Y (b) position differences indicate forward and backward offsets and Z (c) vertical offsets between left and right eye estimates in relation to the Leap Motion coordinate frame. Mean and ± 1 SEM bars are shown for the values within each condition group.

disparity values, for eye pairs within each condition, are provided in Figure 9 (a), (b), and (c), respectively. The difference in X location between eyes provides a measure of interpupillary distance (IPD). The real IPD for the expert user is measured to be approximately 62mm, also indicated in Figure 9 (a). The physical differences in the Y, depth, and Z, vertical, offsets between the expert user's eyes are not directly measured, but are reasonably expected to be approximately 0mm. Examination of the binocular differences along all three directions clearly shows that stereo calibrations result in more accurate disparity, or offset, values compared to monocular calibration results. Table 1 provides the mean and standard deviation for disparities in each direction for all four alignment method conditions.

Repeated measures ANOVA for the monocular calibration IPD values (X axis), in Figure 9 (a), indicate a significant difference between alignment conditions ($F(3,57) = 5.4, p = 0.01, \varepsilon = 0.56$). Post-hoc results reveal that the stylus values are statistically different from both the cross and finger values ($p = 0.004$). Stereo calibration values similarly indicate a significant difference between alignment methods ($F(3,57) = 7.4, p < 0.001$) , with post-hoc results showing that stylus values differ significantly from cross and finger values ($p < 0.001$). No further significant differences are found between groups. Additional ANOVA analyses, comparing the stereo and mono values within each alignment method, reveal that only finger calibration results show a significant difference between mono and stereo IPD values ($F(1,19) = 5.8, p = 0.03$). No significant differences were found between mono and stereo IPD values for the cross, box, or stylus alignment method.

ANOVA across alignment conditions for binocular disparity along the Y axis, Figure 9 (b), shows a significant difference between monocular calibrations ($F(3,57) = 8.0, p = 0.003, \varepsilon = 0.52$), with post-hoc results confirming significant differences between the cross alignment method and both the box and stylus conditions ($p < 0.001$). Stereo calibrations also show a significant difference between calibration conditions ($F(3,57) = 3.1, p = 0.04$), with post-hoc results confirming significant differences between the cross and stylus alignment method values ($p = 0.03$). No further significant differences are found between groups. Additional ANOVA analyses reveal that mono and stereo values significantly differ within each calibration condition: cross ($F(1,19) = 12.9, p = 0.002$), box ($F(1,19) = 10.5, p = 0.004$), finger ($F(1,19) = 17.9, p < 0.001$), and stylus ($F(1,19) = 8.3, p = 0.009$).

Similar statistical results are found for binocular disparity values along the Z axis, Figure 9 (c), with mono calibrations showing a significant effect from alignment condition ($F(3,57) = 4.2, p = 0.02, \varepsilon = 0.76$), and stereo calibrations showing a trend towards being significantly affected by alignment condition ($F(3,57) = 3.5, p = 0.06, \varepsilon = 0.43$). Post-hoc results for both cases confirm a significant difference between cross and stylus groups for mono ($p = 0.012$) and stereo calibrations ($p = 0.023$). No further significant differences are found between groups. Additional ANOVA analyses reveal that mono and stereo values significantly differ within each calibration condition: cross ($F(1,19) = 12.7, p = 0.002$), box ($F(1,19) = 24.1, p < 0.001$), finger ($F(1,19) = 22.1, p < 0.001$), and stylus ($F(1,19) = 17.5, p = 0.001$).

5 Discussion

We begin our discussion with a review of the eye location error metric, starting with comparison between monocular and stereo calibration results. While our statistical analysis only showed significant differences between mono and stereo sets for the cross and finger alignment calibrations, the mean values from Table 1 show that stereo results produced consistently smaller distance to group median values, indicating that stereo calibration yields a consistently higher level of precision, over all. This can also be, somewhat, visually seen by the tighter clustering of eye points in both the 3D and 2D plots of Figure 5. A clear distinction, though, between not only the precision, but also the accuracy, can be seen across alignment method. Calibrations performed with the stylus alignment method produce lower distances to group medians and the highest level of accuracy in comparison to plausible eye locations. In fact, the Leap Motion stylus based calibrations yield far greater extrinsic estimates compared to both the finger alignments employed in our study and also compared to extrinsic values found in nearly all previous studies evaluating SPAAM procedures for environment-centric alignments, [1, 2, 10, 15]. We believe that significantly more accurate extrinsic values for stylus alignments is directly related to the tracking ability of the Leap Motion and the easily discernable tip of the stylus itself.

As previously noted, the accuracy of hand tracking, by the Leap Motion, is highly dependent on the orientation, position, and occlusion level of the fingers, palm, and hand features. This inherent systematic variability naturally leads to less consistent measures in all three of our finger based alignment methods. Conversely, the high precision in stylus tracking by the Leap Motion inherently yields more reliable results. Additionally, the presence of actual misalignment error between the target points of the on-screen reticles and the user's finger tip further increases the potential for inaccuracies and high variability in calibration results. We attempted to address the concerns about the ability of a user to perform repeated alignments to the exact finger tip, by employing the three reticle designs. Our results do show that the improved positioning context afforded by the

Table 1: Mean and standard deviation for each error metric, with both left and right eye values combined, separated by alignment method.

		Distance to Group Median Eye Location (mm)		Binocular Disparity Along X (mm)		Binocular Disparity Along Y (mm)		Binocular Disparity Along Z (mm)		Reprojection Error (pixel)	
		mean	std	mean	std	mean	std	mean	std	mean	std
Cross	mono	39.46	25.26	72.52	17.38	95.5	96.82	10.29	10.7	14.54	6.46
	stereo	22.45	21.56	67.06	5.83	16.96	16.48	1.91	2.37	12.92	3.82
Box	mono	22.23	15.32	66.57	6.63	31.58	27.99	4.79	3.46	12.41	3.37
	stereo	16.97	11.07	64.44	4.08	9.05	11.24	0.97	0.63	10.27	2.94
Finger	mono	28.01	20.17	71.26	8.09	58.42	46.67	8.4	7.18	14.21	4.12
	stereo	14.7	8.66	67.13	2.48	12.58	8.12	0.77	0.68	10.25	1.92
Stylus	mono	12.22	9.3	61.48	3.98	16.72	12.62	3.83	3.15	5.54	0.82
	stereo	8.78	7.54	61.94	3.11	7.02	5.57	0.86	0.43	5.98	0.78

box and finger, positively effect calibration results for both monocular and stereo procedures, though not to a high enough degree to be comparable to the stylus results.

The heightened performance of our stylus calibration, compared to the results from prior investigative studies, we believe is a product of the environment-centric methodology employed in those systems. The SPAAM procedures commonly employed, and evaluated, almost entirely utilize static locations or markers within the environment as alignment points. The systemic errors due to measurement inaccuracies of these alignment points is not present in our user-centric approach. Also, all calibration alignments in these previous works were performed by standing users, which, as shown by Axholt et al. [3], increases the occurrence and magnitude of misalignment error due to postural sway. The Leap Motion calibration allows alignments to be performed while seated, reducing the tendency of sway and body motion during the procedure. The reprojection error and binocular disparity results further validate the viability of Leap Motion procedures for OST HMD calibration.

The low reprojection error seen in stylus calibrations is, no doubt, a product of the more accurate extrinsic eye position estimations, with an attainable reprojection accuracy to within 5–6 pixels (Table 1), again well within the limits obtained by prior SPAAM evaluation studies. The more contextual box and finger reticles also showed improvement in reprojection error when utilized within stereo alignments, confirming the practicality for intuitive reticle designs in finger based OST calibration. The binocular IPD measures also support the utility of stylus usage for Leap Motion calibration, yielding estimates very close to the actual ground truth measure. The improvements in disparity values, in every direction, for stereo calibrations is not surprising, though, since the stereo procedure inherently forces the alignment to two screen points coupling the error across both eyes.

Even though we believe our results conclusively show that accurate calibration of an OST AR system is possible using the Leap Motion, our evaluation method could still be improved in a number of ways. We did not consider any environmental conditions, such as lighting and brightness, that may effect the tracking accuracy of the Leap Motion. Therefore, we are not able to speak to the feasibility of the method as an outdoor calibration solution. Furthermore, we imposed several restrictions on the finger calibration procedure to maintain tracking consistency, such as forcing the user to maintain an open, fingers spread, hand pose during calibration. Future studies will be needed to examine the resulting accuracy trends from calibrations allowing full freedom of hand movements. Moreover, standardization of reticle designs and alignment processes will eventually be required in order to maintain uniformity of OST calibration across devices, increasing the accessible of the practices to novice consumers.

6 Conclusion and Future Work

In this work, we have examined the use of a Leap Motion in facilitating the calibration of an OST HMD system. We utilized monocular and stereo SPAAM based manual alignment procedures for the calibration process, and examined the resulting accuracy and precision from both finger and stylus based methods performed by an expert user. Our results show that both modalities yield calibration qualities well within acceptable levels, as compared to those presented in prior SPAAM evaluation studies. Our findings, additionally, indicate that hand based calibration accuracy can be improved by using more visibly contextual reticle designs to aid in finger placement during the alignment process. However, our overall analysis indicates that the higher tracking accuracy and repeatability of stylus alignments make this the preferred method for OST calibration methods incorporated using the current version of the Leap Motion. Even though the ultimate goal for user-centric calibration is the removal of dependency on physical alignment targets, we believe that the inclusion of a storable stylus, in forthcoming consumer OST devices, is a reasonable requirement to facilitate manual calibration techniques.

While our results confirm the utility of the Leap Motion in OST calibration, our study did not consider extraneous factors, which may impact tracking performance of the device, such as lighting conditions, and degraded internal calibration of the device itself. Future studies will also be required to further refine and identify the most intuitive presentation strategies for on-screen reticles, improving the robustness and usability of manual alignment methods. The standardization of OST calibration practices will also benefit future hardware offerings, making the technology more accessible across systems and markets.

Acknowledgements

The authors wish to thank Sujan Anreddy for his assistance. Support for this work is provided in part by NSF awards IIS-1320909 and IIS-1018413 to J. E. Swan II, and a NASA Mississippi Space Grant Consortium fellowship to K. Moser.

References

[1] M. Axholt, M. A. Skoglund, S. D. O'Connell, M. Cooper, and S. R. Ellis. Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration. In *Proceedings of IEEE Virtual Reality*, pages 27–34, 2011.

[2] M. Axholt, M. A. Skoglund, S. D. OConnell, M. D. Cooper, S. R. Ellis, and A. Ynnerman. Accuracy of eyepoint estimation in optical see-through head-mounted displays using the single point active alignment method. In *IEEE Virtual Reality Conference 2012, Orange County (CA), USA*, 2011.

[3] M. Axholt, M. A. Skoglund, S. Peterson, M. D. Cooper, T. B. Schon, F. Gustafsson, A. Ynnerman, and S. R. Ellis. Optical see-through head mounted display direct linear transformation calibration robustness in the presence of user alignment noise. In *Proceedings of Human*

Factors and Ergonomics Society Annual Meeting, volume 54, pages 2427–2431, 2010.

[4] R. T. Azuma. Predictive tracking for augmented reality. *PhD thesis*, 1995.

[5] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the 25th Hawaii International Conference on Systems Sciences*, volume 2, pages 659–669, 1992.

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

[7] Y. Genc, M. Tuceryan, and N. Navab. Practical solutions for calibration of optical see-through devices. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, pages 169–175, 2002.

[8] J. Guna, G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, 14(2):3702–3720, 2014.

[9] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *Proceedings of IEEE Symposium on 3DUI*, pages 75–82, march 2014.

[10] Y. Itoh and G. Klinker. Performance and sensitivity analysis of indica: Interaction-free display calibration for optical see-through head-mounted displays. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pages 171–176, Sept. 2014.

[11] G. Jakus, J. Guna, S. Tomazic, and J. Sodnik. Evaluation of leap motion controller with a high precision optical tracking system. In M. Kurosu, editor, *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, volume 8511 of *Lecture Notes in Computer Science*, pages 254–263. Springer International Publishing, 2014.

[12] H. Kato and M. Billinghurst. 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*, pages 85–94. IEEE, 1999.

[13] G. Klinker, D. Stricker, and D. Reiners. *Augmented Reality: A Balance Act Between High Quality and Real-Time Constraints.*, pages 325–346. Ohmsha & Springer Verlag, 1999.

[14] P. Maier, A. Dey, C. A. Waechter, C. Sandor, M. Toennis, and G. Klinker. An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 267–268, October 2011.

[15] K. R. Moser, Y. Itoh, K. Oshima, J. E. Swan II, G. Klinker, and C. Sandor. Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, 21(4):491–500, 2015.

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

[17] C. Nitschke, A. Nakazawa, and H. Takemura. Corneal imaging revisited: An overview of corneal reflection analysis and applications. In *Information and Media Technologies*, volume 8, pages 389–406, 2013.

[18] M. O'Loughlin and C. Sandor. User-centric calibration for optical see-through augmented reality. *Master thesis*, 2013.

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

[20] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura. Corneal-imaging calibration for optical see-through head-mounted displays. *Visualization and Computer Graphics, IEEE Transactions on*, 21(4):481–490, 2015.

[21] L. E. Potter, J. Araullo, and L. Carter. The leap motion controller: A view on sign language. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI '13, pages 175–178, New York, NY, USA, 2013. ACM.

[22] J. Quarles. Shark punch: A virtual reality game for aquatic rehabilitation. In *Virtual Reality (VR), 2015 IEEE*, pages 375–375. IEEE, 2015.

[23] H. Regenbrecht, J. Collins, and S. Hoermann. A leap-supported, hybrid ar interface approach. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, pages 281–284. ACM, 2013.

[24] J. Sutton. Air painting with corel painter freestyle and the leap motion controller: a revolutionary new way to paint! In *ACM SIGGRAPH 2013 Studio Talks*, page 21. ACM, 2013.

[25] L. Swirski, A. Bulling, and N. Dodgson. Robust real-rime pupil tracking in highly off-axis images. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*, pages 173–176, March 2012.

[26] A. Tang, J. Zhou, and C. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 161–168, 2003.

[27] T. Travaglini, P. Swaney, K. D. Weaver, and R. Webster III. Initial experiments with the leap motion as a user interface in robotic endonasal surgery. In *Robotics and Mechatronics*, pages 171–179. Springer, 2016.

[28] M. Tuceryan and N. Navab. Single point active alignment method (spaam) for optical see-through hmd calibration for ar. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, pages 149–158, October 2000.

[29] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.

[30] C. W. Yuta Itoh, Frieder Pankratz and G. Klinker. [Demo] Calibration of Head-Mounted Finger Tracking to Optical See-Through Head Mounted Display. Demonstration at 12th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013, Adelaide, Australia, October 1-4, 2013, 2013.

[31] I. Zubrycki and G. Granosik. Using integrated vision systems: three gears and leap motion, to control a 3-finger dexterous gripper. In *Recent Advances in Automation, Robotics and Measuring Techniques*, pages 553–564. Springer, 2014.