Impact of Alignment Point Distance and Posture on SPAAM Calibration of Optical See-Through Head-Mounted Displays

Kenneth R. Moser*
Marxent Labs LLC

Mohammed Safayet Arefin†
Mississippi State University

J. Edward Swan II‡
Mississippi State University

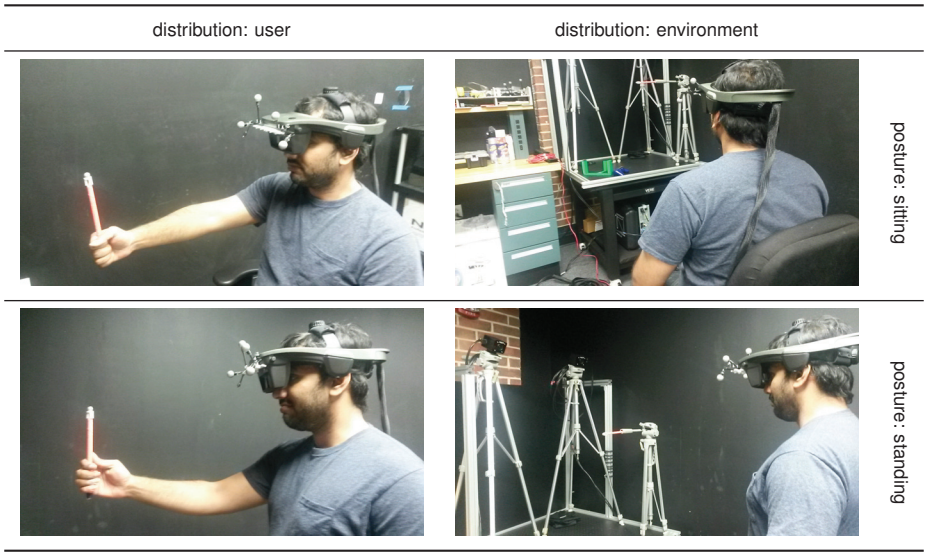


Figure 1: The effect of alignment point distance and posture on SPAAM calibration was examined. Alignment points were distributed at user-centric, reaching distances (left column), and environment-centric, room-scale distances (right column). A sitting posture (top row) and a standing posture (bottom row) were examined. A control condition was also examined, where the participant was replaced with a tripod and camera (Figure 3).

Abstract

The use of Optical See-Through (OST) technology for presenting Augmented Reality (AR) experiences is becoming more common. However, OST-AR displays require a calibration procedure, in order to determine the location of the user's eyes. Currently, the predominantly cited manual calibration technique is the Single Point Active Alignment Method (SPAAM). However, with the SPAAM technique, there remains uncertainty about the causes of poor calibration results. This paper reports an experiment which examined the influence of two factors on SPAAM accuracy and precision: alignment point distribution, and user posture. Alignment point distribution is examined at user-centered reaching distances, 0.15 to 0.3 meters, as well as environment-centered room-scale distances, 0.5 to 2.0 meters. User posture likely contributes to misalignment error, and is examined at the levels of sitting and standing. In addition, a control condition replaces the user with a rigidly-mounted camera, and mounts the OST display on a precisely-adjustable tripod. The experiment finds that user-centric distributions are more accurate than environment-centric distributions, and, somewhat surprisingly, that the user's posture has no effect. The control condition replicates these findings. The implication is that alignment point distribution is the predominant mode for induction of calibration error for SPAAM calibration procedures.

*e-mail: moserk@acm.org
†e-mail: arefin@acm.org
‡e-mail: swan@acm.org

This is an author version preprint. The final version is available as: Kenneth R. Moser, Mohammed Safayet Arefin, J. Edward Swan II, "Impact of Alignment Point Distance and Posture on SPAAM Calibration of Optical See-Through Head-Mounted Displays", *IEEE International Symposium on Mixed and Augmented Reality* (ISMAR 2018), Munich, Germany, October 16–20, 2018, pages 21–30.

Index Terms: Augmented reality—Optical-see through—Head mounted display—Single point active alignment method (SPAAM)

1 Introduction

Augmented Reality (AR) experiences, using Optical See-Through (OST) head-worn and head-mounted display (HMD) technologies, are growing in prominence, largely due to the increase in lower cost commercially available devices, such as the Epson Moverio BT-300, Microsoft Hololens, and Meta 2. AR systems using OST hardware provide a distinct perceptual difference compared to the more common Video See-Through (VST) paradigm. VST displays merge AR elements with a video feed of the world captured through a camera, sometimes mounted over or within proximity to the user's own eye. In contrast, OST displays use semi-transparent mediums to optically combine AR items directly into the user's view of the world. Because the user observes the world directly, there is no perceptual shift in viewpoint or field of view, as is the case with VST systems, where the cameras are not mounted to co-align with the user's gaze or are not designed to mitigate perceptual shifts, as orthoscopic [31, 33] and quasi-orthoscopic [7, 8, 18] VST-HMDs are able to do. The perceptual benefits of OST devices, however, come with the expense of additional calibration difficulty.

Calibration, stated simply as modeling the view parameters of a display for proper rendering of computer-generated items, is well studied for VST hardware, because the view of the world comes directly from a digital imaging camera. Numerous methods and techniques have been developed for explicitly determining the properties of camera devices [35, 37]. Calibration of an OST HMD, however, is not as easily performed, because the view of the world is accessible only to the user themselves, and cannot be extracted for direct processing.

Early calibration methods for OST HMDs employed boresighting procedures [6, 19], in which the user's head is completely restricted from movement, and the viewing frustum through the display is estimated, based on data collected from user feedback in the form of visual alignments between illuminated points on the display and visible points of interest in the world. These rigid procedures were eventually superseded by alternative techniques, which placed fewer restrictions on user head movement during alignment. The Single Point Active Alignment Method (SPAAM), presented by Tuceryan and Navab [36], has emerged as the most commonly-cited alignment-based manual calibration method.

The popularity of SPAAM is due in large part to the high degree of movement available to the user during the screen-world alignment data collection process. However, this increased mobility naturally incurs the potential for greater misalignment error, due to head instability and postural sway [1]. Subsequent variations on the original SPAAM procedure have sought to reduce or ameliorate misalignment error, while also reducing the overall calibration time, as well as the physical demands placed on the user.

The distribution of alignment points, and by association, the amount of movement needed to be taken by the user between each correspondence pair, can be typically categorized as either *user-centric* or *environment-centric* (Figure 1). A user-centric alignment distribution uses 3D points that fall within arms' reach, or the near visual field of the user. User-centric alignment benefits the user, because less physical exertion is required to move between each point, with the added potential for the entire calibration process to be performed while seated. In contrast, environment-centric calibration uses points distributed over several meters, or the medium visual field of the user. Though requiring more effort, environment-centric distributions allow for greater variance in position and coverage of the tracking space, which potentially aids in mitigating alignment error introduced from poor viewing angles, and reduces degenerate calibration results arising from excessive co-planar points. While the efficacy of each distribution scheme in producing consistently accurate calibration results has been, to an extent, investigated independently, there has yet to be a formal experiment directly focused on determining if either alignment scheme inherently produces a superior level of accuracy and precision in calibration results. Additionally, the number of alignment points required to ensure a predictable level of calibration success has yet to be explicitly examined for each of the two alignment distribution categories.

The experiment presented in this work is the first study expressly designed to compare and contrast the expected accuracy and precision of SPAAM calibration results, performed using both user-centric and environment-centric alignment point distributions. The significance of the impact of postural sway on calibration results is also considered, through comparison of seated and standing calibration results. A control condition, in which the user is replaced by a rigidly mounted camera and tripod system, is also employed to provide baseline calibration results for each alignment point distribution. Therefore, the experimental design crosses 2 levels of alignment point *distribution* (user, environment) with 3 levels of *posture* (sitting, standing, camera).

A formal description of OST HMD calibration is provided in the next section, along with a discussion of relevant studies, examining not only the user and usability aspects of SPAAM and similar variants, but also alternative error mitigation and data collection schemes. A detailed description of the experimental setup and procedures employed for performing the repeated calibrations and collecting alignment data for this study is then provided, followed by the presentation of the experimental results. The accuracy and precision of calibrations performed under each condition is examined, using the produced extrinsic eye location estimates as the common metric of comparison. The stability and variance of results is also provided for each condition at increasing alignment counts, up to a total of 50 screen-world alignments per calibration set. A complete discussion of the results, with comparison to prior investigations, is then made, concluding with thoughts on the impact of this work on AR systems employing current OST HMD technology, and directions for further research endeavors on refinement and user evaluations of OST calibration.

2 Background and Related Work

2.1 OST HMD Calibration Parameters

A simple video camera, as well as the scene camera used to perspectively render virtual geometry in modern graphics pipelines, can be modeled after an off-axis pinhole camera system [30], which can be further simplified into an imaging plane and a focal point or camera center, as shown in Figure 2b. The parameters of the pinhole camera model are expressed mathematically as an intrinsic matrix $\mathtt{K} \in \mathbb{R}^{3\times3}$, defined as:

$$^{C}\mathtt{K} = \begin{bmatrix} f_u & \tau & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where f_u and f_v describe the focal distances between the imaging plane and the camera center, c_u and c_v are the coordinates of the center of the imaging plane relative to the screen origin, and τ is the skew of the imaging plane axes. Application of the intrinsic matrix is used to transform a 3D point, with coordinate $\mathbf{p}$ relative to the camera center, into a corresponding 2D point $\mathbf{s}$ on the imaging plane:

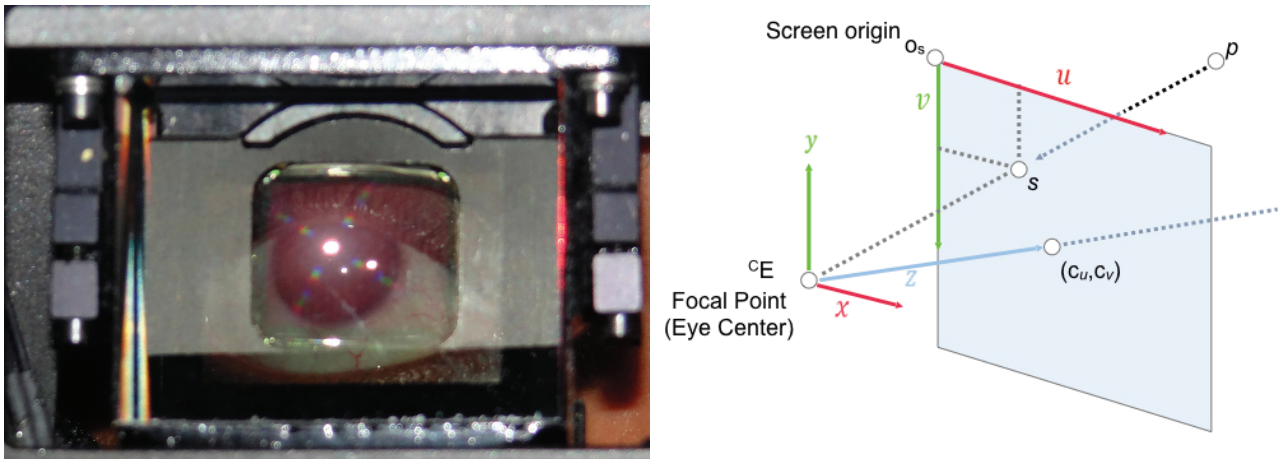
$$\mathbf{s} = {}^{C}\mathtt{K}\mathbf{p}, \tag{2}$$

This relationship, when performed in homogeneous coordinates, is valid up to a scale factor, as illustrated in Figure 2b.

The physical *camera* system created by the user's eye and optical combiner of an OST HMD, depicted in Figure 2a, parallels, in an overly simplified reduction, the pinhole camera model. The imaging plane, in the physical HMD, corresponds to the visible 2D display screen observed by the user through the optical combiner element. The camera center, naturally, refers to the focal point, or nodal point in more complex models, of the user's eye [16], referred to in this work as simply the eye center or ${}^{C}\mathtt{E}$. The 3D position and rotation of the eye center relative to the HMD's coordinate frame, *H*, that is, the translation and rotational offset of ${}^{C}\mathtt{E}$ relative to the 6DOF pose of the HMD inside a larger tracking system, can be described by a rotation matrix ${}^{H}_{E}\mathtt{R} \in \mathbb{R}^{3\times3}$, and translation vector ${}^{H}_{E}\mathbf{t} \in \mathbb{R}^{3}$. These two components together form the extrinsic matrix, which combines with the intrinsic parameters to produce a new 3×4 projection matrix ${}^{H}_{E}\mathtt{P}$:

$${}^{H}_{E}\mathtt{P} = {}^{C}\mathtt{K}\begin{bmatrix} {}^{H}_{E}\mathtt{R} & {}^{H}_{E}\mathbf{t} \end{bmatrix} \in \mathbb{R}^{3\times4}. \tag{3}$$

OST HMD calibration procedures, therefore, either attempt to estimate the values of the ${}^{H}_{E}\mathtt{P}$ matrix directly, 11 independent values defined up to a scale factor, or they independently determine the intrinsic and extrinsic matrices. A more thorough overview of the eye–screen system created by OST HMDs is provided by Grubert et al. [11], which also includes an exhaustive examination of the work surrounding OST HMD calibration methods and evaluation metrics and strategies. The following sections will highlight the most common calibration methods discussed in the literature, emphasizing those works focused on the predominant error metric for manual calibration, user alignment-error, and the usage of user-centric vs environment-centric alignment data distributions.



(a) (b)

Figure 2: Illustrations of the viewing system created by (a) the user's eye and HMD display screen, and (b) an off-axis pinhole camera model.

2.2 Two Stage and Automatic Calibration

The Display Relative Calibration (DRC) presented by Owen et al. [27] follows a two-step calibration process, in which the intrinsic properties of the display are first determined off-line, using an imaging camera set behind the display screen. Then, the extrinsic pose of the user's eye within the display is estimated at run-time, through a manual process during which the user performs a series of screen-world alignments.Makibuchi et al. [21] similarly mirror this two-stage approach, beginning with off-line intrinsic measurement followed by an active user alignment, except they utilize the alignment points to solve the perspective-n-point problem, which optimizes both the display and extrinsic parameters together.

More recently, proposed extensions of these two-step techniques attempt to replace the manual extrinsic estimations with automatic operations. Itoh et al. [14] utilize a video camera rigidly mounted to the HMD to estimate the 6DOF pose of the user's eye, using a 3D iris detection method from Swirski [32], and a localization process by Nitschke et al. [25]. This Interaction-Free Display Calibration (INDICA) procedure has been shown to produce calibration results comparable to manual methods [22]. Similarly, the Corneal Imaging Calibration (CIC) method, developed by Plopski et al. [28], estimates eye locations based on the observed distortion of fiducial patterns visibly reflected off the user's cornea. Further enhancements of CIC attempt to relax the constraints of the required camera system [29]. Although the eventual goal of these automatic methods is to accommodate systematic run-time error, such as displacement of the HMD on the user's head, to date neither of these automatic methods have been fully implemented into a dynamically updating calibration system. Additionally, while these automatic calibration procedures remove the need for manual alignments, making them user friendly, they are not inherently applicable to current commercially available OST devices, which are not factory-equipped with eye-tracking or eye-imaging cameras. This leaves the non-trivial task of implementing these systems up to the researcher or system developer desiring to use them.

2.3 SPAAM OST HMD Calibration

The SPAAM calibration process, described more thoroughly in Tuceryan and Navab [36], does not separately determine the intrinsic and extrinsic properties. Instead of a two-step approach, only the user alignment data is collected in a single process, through which the full 11 parameters, plus scale factor, of the 3×4 projection matrix, ${}^{H}_{E}\mathrm{P}$, are estimated together, using a direct linear transformation. Since no additional hardware is required to acquire the alignment data, beyond the OST HMD and tracking system used to localize the user within the AR space, SPAAM calibration can be readily adapted for use with all existing commercially and industrially available OST hardware. However, deriving the full projection matrix solely from alignment data makes the results of SPAAM highly sensitive to user misalignment and systemic tracking errors.

2.3.1 Reducing User Misalignment Error

Tang et al. [34] shows that the robustness of SPAAM calibration, with regard to misalignment error, can be increased by varying the distances between the user and the alignment point. Axholt et al. [4] shows that varying the alignment distance using a magic square distribution provides further resiliency. Additional studies by Axholt [3] also examined the impact that involuntary postural sway has on the ability of a user to produce stable screen-world alignments. Their findings show that the visual load of the user significantly influences the amount of observable sway, with the most user instability occurring when the eyes are closed.

Maier et al. [20] consider the data recording process as a source of inducing misalignment. They investigate several user signaling strategies, including verbal commands, button input, and hold and wait. Their findings showed that the hold and wait input method, in which the user indicated an alignment by holding a stationary pose for a set period of time, produced the best calibration results. Moser et al. [24] employed the hold and wait strategy in their study investigating misalignment resulting from a lack of contextual information about the proper alignment screen point. They examined several on-screen reticle styles for performing alignments against a finger tip, as well as a rod-like stylus. Calibrations using the finger alignment consistently produced less accurate and consistent results, compared to the stylus alignment calibrations. They conclude that users are more easily able to infer where the tip of the stylus is located for alignment, compared to determining where the alignment point of the tip of a finger is expected to be.

2.3.2 Reducing the Alignment Count

Variations of the SPAAM calibration have also been proposed that attempt to reduce the overall workload of the user, and thus increase the inherent usability of the technique in general. Grubert et al. [12] propose a Multi Point Alignment Method (MPAAM), in which the user performs several simultaneous alignments. While their new procedure was shown to significantly speed up the calibration process, the results also showed significantly higher errors. Fuhrmann et al. [9] propose an alternative alignment approach, in which only 8

alignments per eye are required to estimate the corners of the viewing frustum, with subsequent calibrations needing only 2 additional alignment points, for re-estimating the user's eye location within the display. Genc et al. [10] proposes a Stereo-SPAAM calibration, in which the views of both eyes are calibrated simultaneously. This method is, of course, applicable only to binocular OST HMDs, though stereo calibration has been shown to produce better calibration results compared to monocular counterparts [23]. Jun and Kim [17] also propose a calibration method for stereo OST-HMDs equipped with a depth camera. Their method employs a simplified HMD-eye model, and solves for the extrinsic location of the depth camera and both the inter-pupillary distance, and the location, of the user's eyes. They claim that this method is able to perform a full calibration with only 10 alignment points.

2.3.3 User-Centric Vs Environment-Centric Alignment Distributions

The user workload during SPAAM calibration is not only dependent on the number of alignment points, but also the amount of movement needed to move to, and perform the alignment with, each point. As noted previously, calibration robustness has been shown to increase when alignments are distributed over a varying range of distances. Numerous user study investigations have been performed considering SPAAM calibrations using alignment points distributed over several meters [4, 14, 22]. Quality metrics in these user studies include reprojection error, and estimates of the user's eye locations derived from the projection matrix result. The calibration results from these environment-centric distribution schemes show that calibration quality fluctuates greatly, with user eye estimates often varying by several centimeters across multiple calibrations by the same user.

Other work, such O'Laughlin [26] and Moser et al. [24], utilize alignment points distributed within arm's length of the user. The calibration results provided by Moser et al. [24] indicate that the user-centric alignment distribution strategy is able to produce far more repeatable results, with user eye estimates varying by less than a centimeter across multiple calibrations by the same user. However, there has yet to be a formal experiment conducted to confirm if user-centric alignment point distributions do in fact produce consistently more stable calibration results, when compared to environment-centric distributions.

3 Experimental Design

The presented experiment had two objectives: (1) formally compare the expected accuracy and precision of SPAAM calibrations, performed using both user-centric and environment-centric alignment distributions, and (2) examine the effect of involuntary user motion, due to postural sway, by contrasting calibration results produced by alignments taken by a seated user against calibration results produced by alignments taken by a standing user. A control condition was also examined, in which an RGB camera was rigidly mounted within the HMD. The control condition supplied baseline accuracy and precision values, and, by definition, was free of postural sway. This resulted in a 2 (distribution: user, environment) $\times$ 3 (posture: sitting, standing, camera) = 6 condition experiment (Table 1).

3.1 Hardware and Software

The display used for the experiment was an NVIS ST50 OST HMD, which is a binocular display capable of producing stereo images, with a resolution of 1280×1024 at each eye. The field of view for each eyepiece is stated by the manufacturer to be 40° horizontal $\times$ 32° vertical, with a spatial resolution of 1.88 arcmins/pixel. The display optical combiner contains a collimating lens used to adjust the accommodative, or focal, demand of the display to approximately 3 meters in front of the user. Graphics were rendered to the display via dual HDMI connections.

Table 1: Experimental design. 2 *distributions* (user, environment) $\times$ 3 *postures* (sitting, standing, camera) were examined, yeilding 6 experimental conditions. The *human* participant used stereo SPAAM to localize the left and right eye locations; the *camera* used mono SPAAM to localize the left eye location. 20 *calibrations* were collected for each condition, and 50 *alignments* were used for each calibration.

dist.: user	*dist.: environment*	
human	human	*posture: sitting*
human	human	*posture: standing*
camera	camera	*posture: camera*

At run time, an ART Trackpack dual-IR camera system was used to measure the 6DOF pose of the HMD, as well as the location of the physical alignment points. The Trackpack cameras were rigidly mounted to separate tripods, which were in turn affixed to an optical workbench. Calibration of the Trackpack system was performed according to the manufacturer's instructions, using version 2.10.0 of the accompanying DTrack2 software. The pose of the HMD was determined, relative to a constellation of four retro-reflective spherical markers affixed to the front-top of the HMD, using a custom 3D printed mount, as shown in Figure 3a.

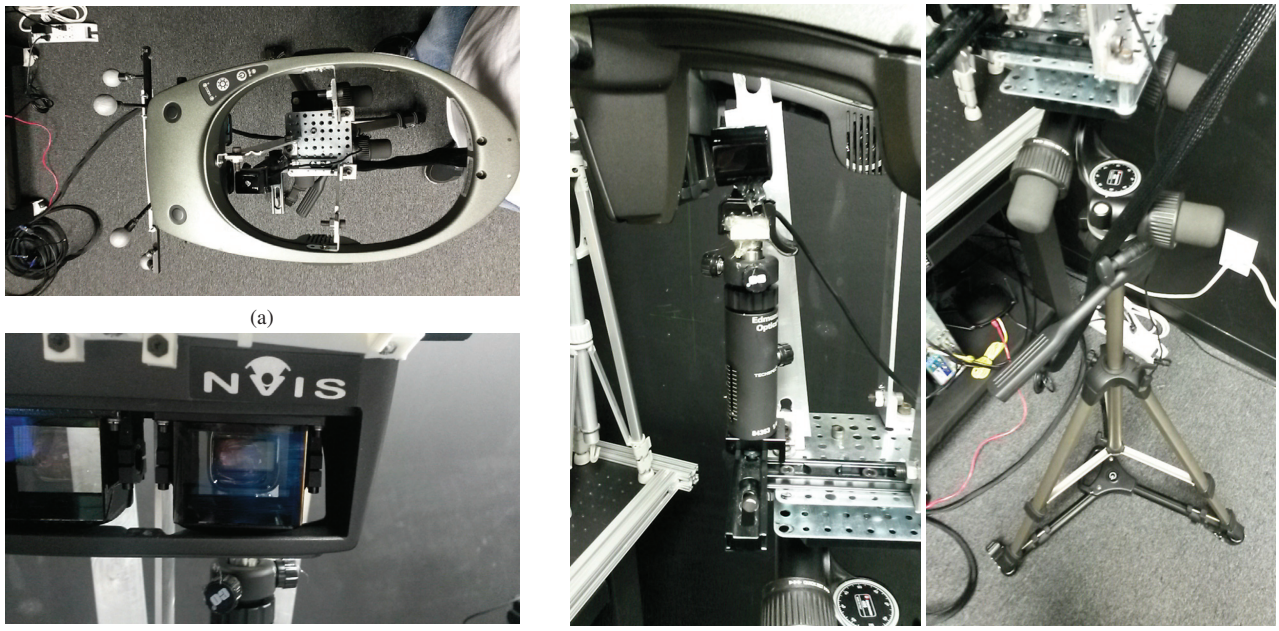
The control condition setup was constructed by rigidly mounting the HMD to a tripod system, equipped with a professional gear head assembly, allowing sub-degree rotational precision adjustment of the HMD. A Microsoft Lifecam HD-600 webcam, with a resolution of 1280×720 at 30fps, was mounted behind the left optical combiner element of the display, using an optical railing system. The railing assembly allowed the Lifecam to be adjusted in 4 DOF: vertical, horizontal, lateral, and rotated in yaw, to provide a view through the HMD screen at the approximate location of a user's eye. Video from the camera was captured through a USB 2.0 connection. Figs. 3b, c, and d show the entire HMD and camera mounting system.

The application used to control the rendering and interaction of virtual content, as well as record the user alignment data, was written in C++, utilizing an OpenGL-based rendering pipeline. The software, and all hardware connections, were driven by an Alienware m18 laptop, with an i7-4700MQ 2.4GHz processor, 16 GB RAM, and running Windows 7 x64.

3.2 SPAAM Calibration Procedure

A standard manual SPAAM calibration procedure was employed. As recommended by Hartley and Zisserman [13], normalization of the 2D screen and 3D world points was also incorporated into the procedure. During the calibration, the participant was provided an on-screen reticle, and was tasked with aligning the center of the reticle with the center of a physical target point (Figure 4). The target point for all conditions was taken to be the center of a retro-reflective sphere, 6mm in diameter, attached to the end of a cylindrical rod. During the alignment procedure, the 3D position of the sphere was actively measured by the ART tracking system, and was used, in combination with the HMD pose, to determine the head-relative coordinate of the target.

The on-screen reticle, used for the 4 conditions collected by the human participant (Table 1), was modeled after the nonius reticle design for stereoscopic calibration presented by Moser and Swan [23]. The nonius reticle was chosen due to the trends seen in Moser and Swan [23] and Moser et al. [24], suggesting that a reticle shown in stereopsis not only promotes stereo fusion at a desired depth, but also exhibits improved calibration results over single eye monocular alignments. Therefore, the stereo reticle was employed to promote the most accurate alignments in both the user-centric and



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

Figure 3: Rigidly-mounted camera system used for the control condition. (a) Top view of the HMD, with mounted retro-reflective markers. (b) View of the HMD optical combiner, with the camera visible behind. (c) Side view of the camera, mounted to adjustable optical rails. (d) View of tripod and gear-head assembly.

environment-centric conditions. The nonius reticle was a solid cross, with on-screen dimensions of 64×64 pixels, and a line thickness of 3 pixels, separated into two halves, with one half shown to each eye. The left eye received the right and bottom lines of the cross, while the right eye received the left and top lines. The on-screen location of each half was programmatically adjusted, so that the user was able to fuse the two halves into a solid cross-hair that appears to be floating in front of them at a certain distance. The on-screen offsets were different for each of the user-centric and environment-centric conditions, and are discussed further within the respective sections below.

Because the control condition used a monocular RGB camera, the nonius reticle was not employed for that condition; instead, a single solid cross was displayed on the HMD (Figure 4). While the nonius reticle was provided for the human participant, affording best-case alignment ability despite the presence of user postural sway, for the control condition the rigidity of the tripod apparatus (Figure 3) effectively mitigated systemic error from alignment motion. Likewise, the alignment distances of the camera apparatus could be explicitly measured and regulated, which removed the need for stereo computer vision-based readings, leaving a monocular reticle as the best and most concise choice.

3.3 User-Centric Calibration Procedure

The non-control, user-centric calibration condition proceeded by presenting the nonius reticle as previously described, positioned in each eye to induce stereopsis and the perception of the cross in depth. The binocular placement of the reticle was modified for each alignment, so that the perceived depth extended in front of the user, between 0.15 and 0.3 meters, or approximately arm's length. The distance change at each alignment was driven by a magic square distribution, as recommended by Axholt et al. [4]. At the start of each alignment, the color of the cross was presented in red.

The retro-reflective target point was held by the user, by means of the attached cylindrical tube. The user then moved, and fixed their gaze upon, the target point, until the nonius cross-fused into a solid

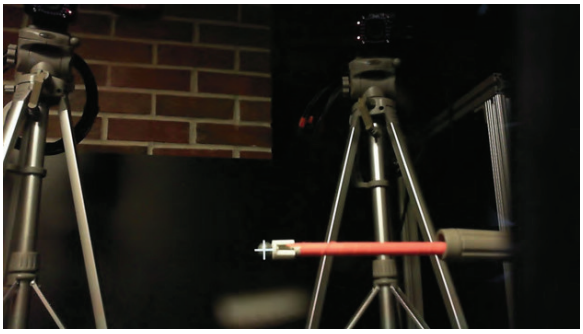


Figure 4: View through the HMD of a control condition alignment, showing the on-screen reticle aligned to the target point.

image, with the center of the perceived cross co-aligning with the center of the retro-reflective sphere. The tracked position of both the HMD and the target point were used to determine when the user's motion of both items had decreased to less than a centimeter/second for three seconds. When this happened, the reticle turned yellow, indicating to the user that data was being recorded. While the reticle was yellow, the user had 3 additional seconds to refine the alignment, and remained still until the final data capture was taken, which consisted of the 3D location of the alignment point relative to the HMD, and the on-screen pixel location of each cross half. If the user's motion exceeded 1 centimeter/second during this phase, the data was not captured, and the cross returned to red until the conditions were met. Once the data was collected for an alignment point, the next set of nonius halves was displayed in red, and the procedure repeated.

An identical set of alignment points was used for the user-centric distribution during calibrations performed by the user while seated, and calibrations performed while standing. The previously described

process did not otherwise alter. Figure 1, left column shows An example of the user-centric calibration is shown in

3.4 Environment-Centric Calibration Procedure

The non-control, environment-centric calibration condition proceeded by presenting the nonius reticle as previously described, positioned in each eye to induce stereopsis and the perception of the cross in depth. The distance between the participant and the physical marker varied between 0.5 and 2.0 meters, by the user taking steps forward or backward, or by adjusting the location of the chair while seated. The amount of distance varied between consecutive alignments was also derived from a magic square distribution, with the distances marked along the ground on a measured tape. The target point itself was affixed to a tripod and adjusted to the approximate height of the user.

The alignment process proceeded in a similar manner to the user-centric distribution process. At the start of each alignment, the color of the cross was presented in red, to indicate to the user to begin the alignment. Since the retro-reflective target point was not held by the user, an alternative means of completing the alignment was employed. The user was guided to stand at the approximate distance along the measuring tape, as prescribed by the magic square pattern, and to affix their gaze upon the target point. The user was then allowed to use a hand-held controller to independently adjust the on-screen location of each half of the nonius cross, until the fused reticle image was perceived at the approximate distance to, and co-aligned with, the center of the target point. To begin the data recording process, the user then pressed a button on the hand-held controller. The reticle then turned yellow, to indicate to the user to remain still, while data was recorded. After 3 seconds, the final data capture was taken, consisting of the 3D location of the alignment point relative to the HMD, and the on-screen pixel location of each cross half. Once the data was collected for an alignment point, the next set of nonius halves was displayed in red, and the procedure repeated.

An identical set of alignment points was used for the environment-centric distribution during calibrations performed by the user while seated, and calibrations performed while standing. The previously described process did not otherwise alter. Figure 1, right column shows an example of the environment-centric calibration.

3.5 Control Condition Calibration Procedure

The control condition utilized the camera and HMD tripod mounting system discussed previously (Figure 3). Identical sets of distances, 0.15 to 0.3 meters for the user-centric, and 0.5 to 2.0 meters for the environment-centric, were used, in order to provide comparable calibration measures for both sets of alignment distances. During this condition, the monocular cross-hair, previously described, was utilized, and the view from the webcam was referenced, in order to adjust the orientation of the HMD to align the cross with the physical target point. The retro-reflective target point was rigidly mounted to a tripod, as in the environment-centric non-control condition. The magic square distance pattern for each alignment distribution was marked on the measuring tape, and for each alignment, the tripod was manually moved, and the HMD-camera setup carefully adjusted, to ensure alignment between the center of the reticle and the target point, to within 3 pixels of visual accuracy. Although the process of adjusting the HMD rig was performed manually, the alignment precision was still expected to far exceed that possible from a human-performed calibration, because the postural and head motion from a user were expected to cause significant amounts of pixel deviation beyond that attainable from the control apparatus.

3.6 Participant

All calibration data, with the exclusion of the control condition, were recorded from repeated trails by a single expert user, the first author of this paper. As shown in Table 1, the participant completed 20 user-centric and 20 environment-centric calibrations, in both a sitting and standing posture, for both the left and right eye, resulting in $20 \times 2 \times 2 \times 2 = 160$ total calibrations. 50 alignments were used to complete each calibration set, yielding a total of $50 \times 160 = 8000$ alignment points. The user's maximum inter-pupillary distance was measured to be approximately 62 mm. The control condition utilized 20 calibrations, using the user-centric and environment-centric distance ranges, for the left eye, for $20 \times 2 = 40$ additional calibration results, with $50 \times 40 = 2000$ alignment points. Therefore, a total of 10,000 alignment points was collected.

During the calibration study, the human participant took approximately 10 minutes to complete all the alignments for a calibration set, requiring a total of approximately 1600 minutes = 26.67 hours. The control condition, which required moving the tripod and adjusting the camera, required approximately 20 minutes per calibration set, for a total of approximately 800 minutes = 13.33 hours. Therefore, the total time required to collect the data was approximately 1600 + 800 minutes = 2400 minutes = 40 hours. The data was collected over a period of several weeks.

The primary objective of this study was to compare the effect of alignment point distribution, and posture, on eye location accuracy and precision (Table 1). Therefore, restricting the calibration data to repeated measures from an expert user, extremely knowledgeable with the procedure, removed the potential for errors resulting from the subjective abilities of multiple participants, and allowed more stable and consistent results to be obtained. The same strategy has been employed in similar studies [14, 15, 24]. In addition, recruiting multiple participants to spend 40, or more, hours performing SPAAM alignments would be, to say the least, very challenging.

Therefore, this study can be considered an engineering study, which involved a human in the loop, as opposed to a user study. A user study, such as those surveyed by Grubert et al. [11], would allow the additional examination of which calibration methods could be best used by different participants.

4 Experimental Results

The quality of the calibration results, in terms of accuracy and precision, was analyzed using the standard objective metric of estimated user-eye location, employed in numerous previous studies [4, 14, 15, 22, 24, 28]. This value describes the 3D translational position of the user's eye within the HMD relative coordinate frame, and effectively equates to the extrinsic parameters of the projection matrix calculated by SPAAM. Using the inter-pupillary distance of the user, a bounding volume of where the eye was expected to be located can be formed, enabling the estimated eye location from the calibration results to represent ground truth, which allows an analysis of accuracy. The combined estimated eye locations, determined within each condition, were then used to determine a median eye estimate. The Euclidean distance between each individual eye estimate and the median eye position calculated from the calibration result was utilized to provide a measure of precision for each calibration condition. An examination of the convergence, or trend over increasing alignment count, of these metrics was also determined.

4.1 Estimated 3D Eye Location

Figure 5 plots the eye locations for all experimental conditions. Through visual inspection, it is evident that the user-centric point distributions produced eye estimate values that were far more accurate than the environment-centric distributions, especially along the depth (Y) axis. In addition, the user-centric point distributions, compared to the environment-centric point distributions, are more tightly clustered and precise, again especially along the depth (Y) axis. This is true for both the sitting and standing postures, as well as for the control condition: posture made little difference. The results from 25 alignments are about as good as those from 50 alignments.

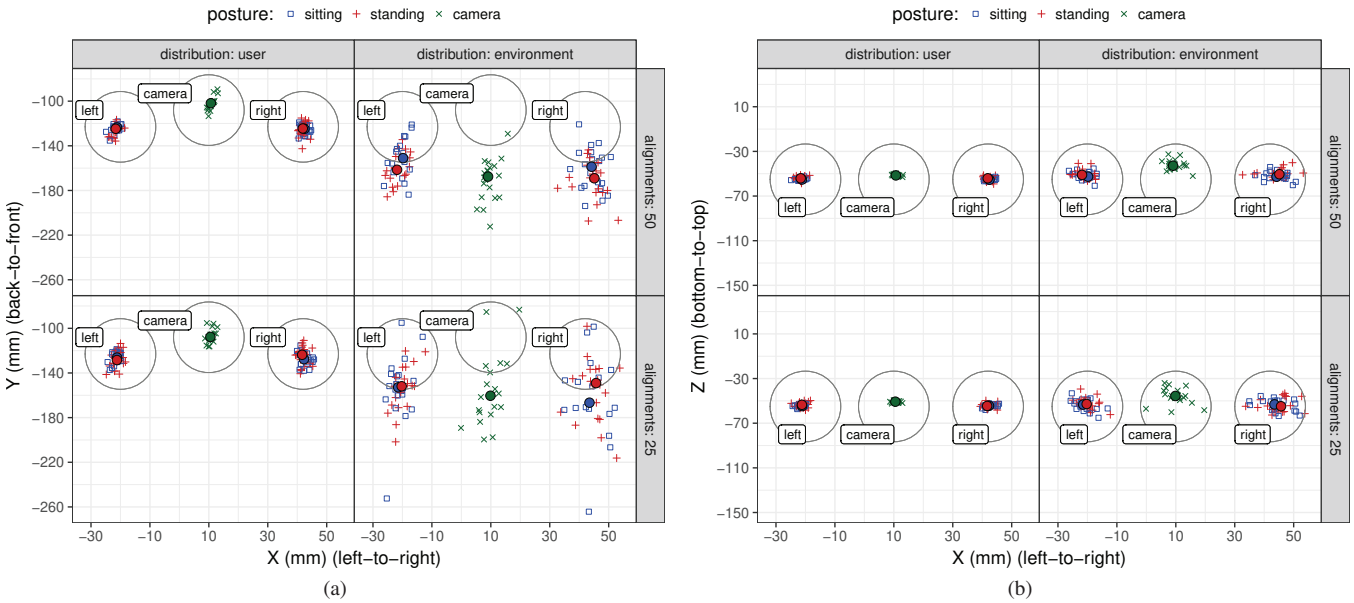


Figure 5: Estimated locations of the left eye, right eye, and camera, relative to the tracker constellation mounted on the AR display. (a) View of the XY plane, as if the reader is standing above the observer and looking at the top of their head: the observer is looking along the $+Y$ axis, towards the top of the plot. (b) View of the XZ plane, as if the reader is standing behind the observer and looking at the back of their head: the observer is looking along the $+Y$ axis, which goes into the page. On both plots, the circles show the estimated ground-truth location of the observer's left and right eyes; the circles are 24 mm in diameter, the approximate axial length of the schematic eye [5]. Another circle, of the same diameter, shows the location of the camera; for clarity, the camera locations are offset to the right by 28 mm, and would otherwise overlay the left eye. The color and shape of the points indicate *posture*: sitting, standing, or camera; the left-right columns indicate point *distribution*: user-centric, or environment-centric; and the top-bottom rows indicate the number of SPAAM *alignments* that produced each point: 50, or 25. Both plots show $N = 400$ points. For each condition, the colored circle is the median eye location. Accuracy was much higher for the user-centric distributions, especially along the depth (Y) axis. 25 alignments gave results that were as good as 50 alignments. Posture makes no difference.

4.2 Median Eye Location

Since the precise location of the user's eye was not known, a direct offset error value could not be determined. However, an alternative precision metric is the calculated distance to the median location within each result cluster (the colored circles in Figure 5). Figure 6 provides plots for distance to group medians for each calibration condition. Visual inspection indicates that the user-centric distributions produce noticeably less variation, compared to the environment-centric alignment conditions. Comparing the results after 25 alignments with those after 50 alignments shows minimal difference with increasing alignment count for the user-centric distributions, but suggests continuing improvements for the environment-centric distributions. Also, the different postures show minimal differences.

For the precision results, repeated-measures analysis of variance (ANOVA) tests were performed, in order to verify the significance, or lack there of, between conditions. The ANOVA model uses calibration (1 to 20) as the random factor, and distribution (user, environment), posture (sitting, standing, camera), and alignment (50, 25) as fixed factors that vary within each calibration. There was a significant main effect of distribution on distance to the median eye position $(F(1,19) = 44.0, p < 0.001)$, and a significant interaction between distribution and alignment $(F(1,19) = 5.7, p = 0.028)$, which also shows up as a main effect of alignment $(F(1,19) = 12.4, p = 0.002)$. There were no other main effects or interactions (all F's < 1.2). These ANOVA results conform to what is apparent in Figure 6 by visual inspection.

4.3 Increasing Alignment Count

The final metric utilized in this analysis is a comparison of the convergence, or trend, of the calibration results with increasing alignment count. This measure indicates the threshold of alignments at which the maximum calibration gains are expected to be achieved. While it is possible to produce an alignment trend graph for every metric utilized thus far, this analysis focuses on the change in distance to the median eye location values. Figure 7 provides the distance to the median value for each condition, over all 50 alignments. It is important to note, however, that no results are attainable from the direct linear transformation for a SPAAM solution until a minimum of 6 alignments have been conducted. Figure 7 begins at alignment 9, which is among the first values where reasonable eye location estimates are achieved.

Figure 7 shows that for the user-centric distributions, there is little to no improvement after 25 alignments, suggesting that 25 is a reasonable limit. For the environment-centric distributions, after alignment 25 there is some improvement in the maximum distance, but no real improvement in the median distance. The posture, whether sitting, standing, or camera, makes no difference.

5 Discussion

We begin our discussion by comparing the wide variance of eye location estimates in the depth dimension (Y in Figure 5a), for the environment-centric conditions, against those reported in previous work. In Axholt et al. [4], Itoh and Klinker [14], and Moser et al. [22], the same large variance in eye estimates along the depth dimension was shown to occur. It is possible this large variance may

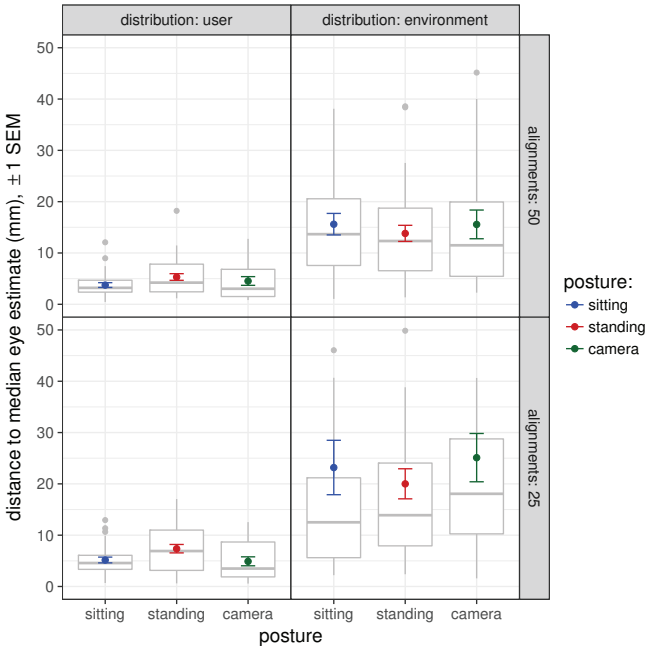


Figure 6: The distribution of the distance to the median eye estimate, for each experimental condition. The color and position of the means and error bars indicate *posture*: sitting, standing, or camera; the left-right columns indicate point *distribution*: user-centric, or environment-centric; and the top-bottom rows indicate the number of SPAAM *alignments* that produced each distribution: 50, or 25. Both plots summarize $N = 400$ points. Precision is much higher for the user-centric distributions, while posture makes little difference.

be erroneously attributed to the presumed influence of user alignment error at the larger distances. However, the control condition results, Figure 5a, exhibit the same variance in depth. Given that the user-centric calibrations for all conditions lacked the large depth variation, it can be concluded that the distribution of the alignment points themselves, and not the user alignment error, is the primary factor that contributes to calibration error. This is the primary finding of the experiment.

However, this does not mean that user alignment error plays no role in degrading calibration results. Moser et al. [24] shows that a slight depth variance does occur in user-centric alignments, when the physical target point is not obvious or contextually clear to the user, incurring a degree of misalignment. However, the non-ambiguous user-centric conditions, including those employed in this study, show significantly more consistent results. In addition, there was a lack of significant differences between the standing, sitting, and camera postures; even though standing should result in more postural sway than sitting [2], and the camera should be free of postural sway. This further supports the claim that error due to involuntary user motion contributes only a minor effect to the overall calibration outcome.

Another possible source of the alignment differences between the user-centric and environment-centric conditions is the use of the stereoscopic alignment reticle, coupled with the accommodation / convergence mismatch of the display. However, the aforementioned studies, Axholt et al. [4], Itoh and Klinker [14], and Moser et al. [22], do not employ a stereoscopic alignment procedure, but use fully monocular systems. Therefore, the strong correlation between the environment-centric results in this study, and those prior works, suggests that the use of the nonius reticle can be rejected as a primary source of influence on calibration results. Further evidence for this can be taken from the control condition, which also employed a monocular camera alignment reticle. The control condition results, again, closely match both prior studies, and the current study, for the environment-centric calibration results, and similarly match the user-centric results from this present work.

A limitation of the current work is that, even though the user-centric distributions yielded highly repeatable SPAAM calibration results, it has yet to be shown if the registration of the AR content using the SPAAM projection matrix maintains the same level of accuracy, especially when viewing AR content in the near, medium, and far visual fields. A subjective user-study evaluation, such as that conducted by Moser at al. [22], is needed. We believe, though, that since the rendering pipeline is based on the pinhole camera model, which requires a stable camera center point, the projection matrix produced by user-centric calibration will be suitable for rendering AR content at any distance.

The plot of eye estimate variance over alignments, Figure 7, also agrees with assertions made by Axholt [4] that 25 alignment points provides the optimal balance between diminished returns from further alignments, and a reasonable user workload. A steady-state of median variance, for both the user-centric and environment-centric alignment conditions, is achieved after approximately 25 alignments, for both the calibrations performed by the expert user, and those obtained from the control condition.

While 25 alignments may still seem to be an excessive number, the level of precision attainable from user-centric points may yield sufficient results to preclude re-calibration, through reuse of a previous calibration result. Moser et al. [22] indicated that registration error was not significantly impacted by reusing a previous calibration result between user sessions with the HMD. Their results also only considered environment-centric alignment calibrations. The results here suggest that a similar result would be found for user-centric alignment calibrations as well.

6 Conclusion and Future Work

This work has presented the first formal comparison of the impact of user-centric and environment-centric alignment point distributions on the consistency and variance of SPAAM calibrations for OST HMDs. Our experimental results also include values for each alignment type, taken via a control condition, in which user misalignment due to involuntary head and postural sway is eliminated, by replacing the user with a rigidly mounted camera within the HMD itself. The results of both the control condition and the user-performed calibrations show that user-centric alignment points, that is, physical alignment points presented within arm's length of the user, yield significantly more accurate and consistent results, compared to the more common environment-centric alignment point distribution scheme. Our results also show that no significant gains in calibration accuracy are achieved with alignment counts greater than 25.

These results, while empirically agreeing with findings and recommendations from previous studies, also discredits the common notion that user misalignment error is the predominately degrading factor for SPAAM calibration. Our results, in fact, show quite the contrary, that even within the control condition, the distribution of alignment points in space is the most significant determinant of the expected level of accuracy for SPAAM.

Future investigations are still needed to perform subjective user-study evaluations for comparing the perceived registration error of AR content displayed at near, medium, and far visual field distances, using results from both user-centric and environment-centric SPAAM variants. Similarly, the suitability of reusing user-centric calibration results between HMD sessions has yet to be directly examined, even though results from Moser et al. [22] suggest that repositioning of the HMD between uses has minimal impact on calibration usage. However, Moser et al.'s results may be highly dependent upon a display's optical components. Likewise, the results presented within this current work are derived from measures using

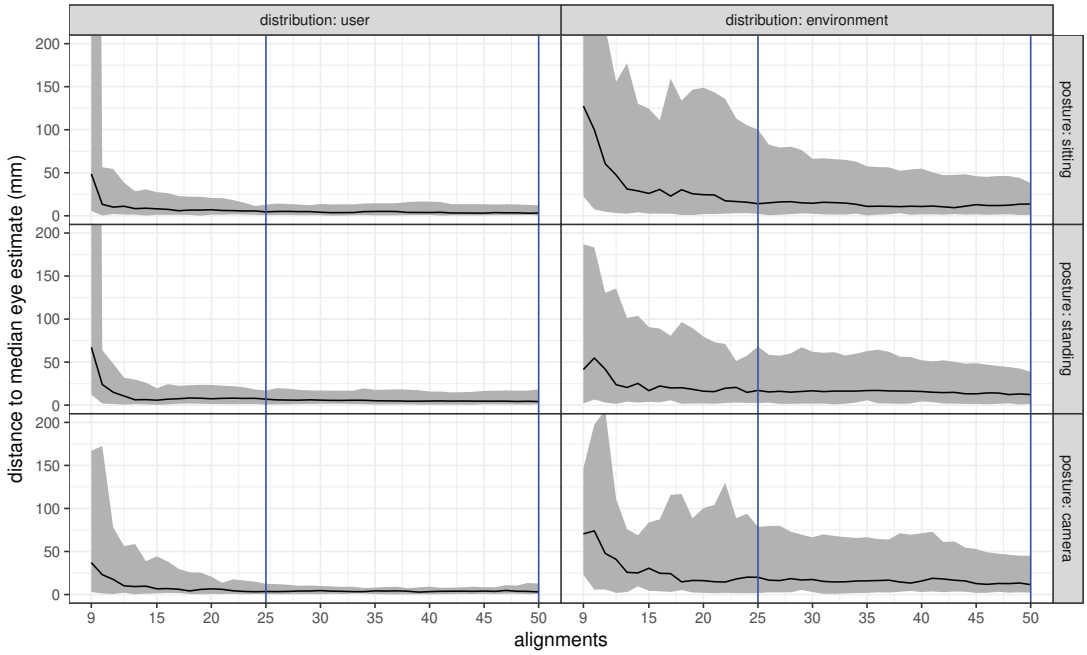


Figure 7: The distribution of the distance to the median eye estimate, as a function of alignment number, from alignment 9 to alignment 50. For each alignment number, the background polygon shows the range, from the minimum to the maximum distance, while the black line gives the median distance. The blue vertical lines indicate alignment numbers 25 and 50, as shown in Figures 5 and 6. The left-right columns indicate point *distribution*: user-centric, or environment-centric; and the top-bottom rows indicate the *posture*: sitting, standing, and camera. This graph summarizes $N = 8400$ data points (alignments 1–8 have been removed). For the user-centric distributions, alignment number 25 is as good as alignment 50, while for the environment-centric distributions, there is still some improvement in the maximum distance, but not the median, beyond 25 alignments. Posture makes no difference.

only the NVIS ST-50 display. Still, prior work has often shown SPAAM to be a viable calibration method across a number of OST-HMD types [11], and therefore we expect the same trends presented within this work to persist independently of display parameters, and as previously stated, to be predominantly dependent upon alignment distance and distribution. Future studies would also benefit from the use of integrated eye-tracking cameras, in order to obtain ground-truth eye location estimates of the actual user's eye relative to the display screen, which could then be directly compared against the estimated locations derived from the SPAAM results.

Finally, while this work has shown that physical alignments within arm's length yield superior calibration results, when compared to alignments distributed over the medium visual field, the experimental design does not conclusively show that arm's length distances are the optimal alignment distances for calibration. Further studies, with a similar design, are needed to test more specific alignment distribution ranges, in order to map the expected effect on calibration results for alignments distributed at various distance intervals, both between those used in this study, and those set closer to the user, perhaps even within only a few centimeters of the user's face.

Acknowledgments

This material is based upon work supported by fellowships provided by the NASA Mississippi Space Grant Consortium and Japan Society for the Promotion of Science (JSPS) through the East Asia and Pacific Summer Institutes Fellowship, award IIA-141477, to Kenneth R. Moser, and the National Science Foundation, under awards IIS-1018413 and IIS-1320909, to J. E. Swan II. This work was performed at the Center for Advanced Vehicular Systems, Mississippi State University.

References

[1] M. Axholt. *Pinhole Camera Calibration in the Presence of Human Noise*. PhD thesis, Linköping University, Norrköping, Sweden, 2011.

[2] M. Axholt, S. Peterson, and S. R. Ellis. User boresight calibration precision for large-format head-up displays. In *ACM symposium on virtual reality software and technology*, pp. 141–148. ACM, Bordeaux, France, 2008.

[3] M. Axholt, S. D. Peterson, and S. R. Ellis. Visual alignment precision in optical see-through ar displays: Implications for potential accuracy. In *ACM and IEEE Virtual Reality International Conference (ISMAR)*, 2009.

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

[5] A. G. Bennett and R. B. Rabbetts. Proposals for New Reduced and Schematic Eyes. *Ophthalmic Physiological Optics*, 9(2):228–230, Apr. 1989.

[6] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, vol. 2, pp. 659–669. IEEE, 1992.

[7] F. Cutolo, C. Freschi, S. Mascioli, P. D. Parchi, M. Ferrari, and V. Ferrari. Robust and accurate algorithm for wearable stereoscopic augmented reality with three indistinguishable markers. *Electronics*, 5(3), 2016. article number 59.

[8] F. Cutolo, P. D. Parchi, and V. Ferrari. Video see through ar head-mounted display for medical procedures. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 393–396, Sept 2014.

[9] A. Fuhrmann, D. Schmalstieg, and W. Purgathofer. Fast calibration for augmented reality. In *ACM Symposium on Virtual Reality Software*

and Technology, pp. 166–167. ACM, 1999.

[10] Y. Genc, F. Sauer, F. Wenzel, M. Tuceryan, and N. Navab. Optical see-through hmd calibration: a stereo method validated with a video see-through system. In *IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pp. 165–174. IEEE, Munich, Germany, 2000.

[11] J. Grubert, Y. Itoh, K. R. Moser, and J. E. Swan II. A survey of calibration methods for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018.

[12] J. Grubert, J. Tuemle, R. Mecke, and M. Schenk. Comparative user study of two see-through calibration methods. *VR*, 10:269–270, 2010.

[13] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. *Robotica*, 23(2):271–271, 2005.

[14] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 75–82. IEEE, 2014.

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

[16] J. A. Jones, D. Edewaard, R. A. Tyrrell, and L. F. Hodges. A schematic eye for virtual environments. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 221–230. IEEE, 2016.

[17] H. Jun and G. Kim. A calibration method for optical see-through head-mounted displays with a depth camera. In *IEEE Virtual Reality (IEEE VR)*, pp. 103–111. IEEE, 2016.

[18] M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya. A stereoscopic video see-through augmented reality system based on real-time vision-based registration. In *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, pp. 255–262, 2000.

[19] G. Klinker, D. Stricker, and D. Reiners. Augmented reality: a balance act between high quality and real-time constraints. *Mixed Reality–Merging Real and Virtual Worlds, Ohmsha & Springer Verlag*, pp. 325–346, 1999.

[20] P. Maier, A. Dey, C. A. Waechter, C. Sandor, M. Tönnis, and G. Klinker. An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 267–268. IEEE, 2011.

[21] N. Makibuchi, H. Kato, and A. Yoneyama. Vision-based robust calibration for optical see-through head-mounted displays. In *IEEE International Conference on Image Processing (ICIP)*, pp. 2177–2181. IEEE, 2013.

[22] 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*, 21:491–500, 2015.

[23] K. R. Moser and J. E. Swan II. Improved spaam robustness through stereo calibration. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 200–201. IEEE Computer Society, 2015.

[24] K. R. Moser and J. E. Swan II. Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 159–167. IEEE, 2016.

[25] C. Nitschke, A. Nakazawa, and H. Takemura. Corneal imaging revisited: An overview of corneal reflection analysis and applications. *Information and Media Technologies*, 8(2):389–406, 2013.

[26] M. O'Loughlin and C. Sandor. *User-Centric Calibration for Optical See-Through Augmented Reality*. PhD thesis, Master thesis, 2013.

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

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

[29] A. Plopski, J. Orlosky, Y. Itoh, C. Nitschke, K. Kiyokawa, and G. Klinker. Automated spatial calibration of hmd systems with unconstrained eye-cameras. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 94–99. IEEE, 2016.

[30] J. P. Rolland. Wide-angle, off-axis, see-through head-mounted display. *Optical Engineering-Bellinghan-International Society for Optical Engineering*, 39(7):1760–1767, 2000.

[31] A. State, K. P. Keller, and H. Fuchs. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 28–31, Oct 2005.

[32] L. Świrski, A. Bulling, and N. Dodgson. Robust real-time pupil tracking in highly off-axis images. In *Symposium on Eye Tracking Research and Applications*, pp. 173–176. ACM, 2012.

[33] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi. Development of a stereo video see-through hmd for ar systems. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pp. 68–77, 2000.

[34] A. Tang, J. Zhou, and C. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 161. IEEE Computer Society, 2003.

[35] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.

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

[37] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.