

Performance and Sensitivity Analysis of INDICA : Interaction-Free Display Calibration for Optical See-Through Head-Mounted Displays

Yuta Itoh*

Gudrun Klinker†

Technische Universität München

ABSTRACT

An issue in AR applications with Optical See-Through Head-Mounted Display (OST-HMD) is to correctly project 3D information to the current viewpoint of the user. Manual calibration methods give the projection as a black box which explains observed 2D-3D relationships well (Fig. 1). Recently, we have proposed an INteraction-free DIisplay CALibration method (INDICA) for OST-HMD, utilizing camera-based eye tracking[7]. It reformulates the projection in two ways: a black box with an actual eye model (Recycle Setup), and a combination of an explicit display model and an eye model (Full Setup). Although we have shown the former performs more stably than a repeated SPAAM calibration, we could not yet prove whether the same holds for the Full Setup. More importantly, it is still unclear how the error in the calibration parameters affects the final results. Thus, the users can not know how accurately they need to estimate each parameter in practice. We provide: (1) the fact that the Full Setup performs as accurately as the Recycle Setup under a marker-based display calibration, (2) an error sensitivity analysis for both SPAAM and INDICA over the on-/offline parameters, and (3) an investigation of the theoretical sensitivity on an OST-HMD justified by the real measurements.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 INTRODUCTION

A crucial issue in AR applications using OST-HMDs is to render 3D information from the current viewpoint of the user, – and, more particularly, according to the user’s eye position, relative to a not quite stable HMD pose on the user’s head. Such rendering requires 2D-3D projection matrix from the world to a screen. Manual calibration methods like Single Point Active Alignment Method (SPAAM)[12] find a projection which explains manually-collected 2D-3D alignments best. Thus the projection is in a black box, ignores spatial relationship of a display and an eye(Fig. 1a).

We have recently proposed INteraction-free DIisplay CALibration (INDICA) method for OST-HMDs utilizing camera-based eye tracking [7]. INDICA generates the projection w.r.t the user’s current eyeball position by combining the tracked eye position online. Depending on predetermined offline parameters, the method has two setups that require either: a partial set of display parameters in combination with a previous calibration result (Fig. 1b, Recycle Setup), or a full set of all display parameters (Fig. 1c, Full Setup).

The two setups represent the same, yet their interpretations are quite different: Recycle Setup (Fig. 1b) updates the black box from a manual method using a measured eye position. The box can be given by a previously performed SPAAM calibration or a camera-based HMD calibration such as in [4]. On the other hand, Full

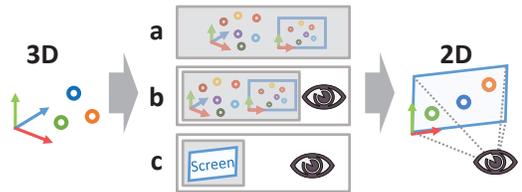


Figure 1: Interpretation of projection *black boxes* from different calibration methods: (a) SPAAM, (b) Recycle Setup, and (c) Full Setup.

Setup formulates the system as a combination of an explicit, display model and eye model (Fig. 1c). The setup requires an extra offline display calibration. Both setups have their pros and cons in practice, thus users would choose either setup over the other depending on the application (esp. means and convenience of performing the required calibrations). Therefore, evaluating and comparing both setups are valuable for the future use of INDICA.

In our previous work, we have compared our Recycle Setup with SPAAM calibrations in repeated experiments, and were able to demonstrate that the Recycle Setup performs more stably than SPAAM in estimating accurate 3D eye positions. However, we could not yet show whether the same holds for the Full Setup, which would model the system more accurately. More importantly, it is still unclear how estimation errors of the online/offline parameters affect the calibration performance. Thus, it is still unclear how accurately each of the offline parameters must be determined in practice. Therefore, it is helpful to understand the sensitivity of the overall HMD calibration accuracy to imprecision of individual parameters such that the calibration process can be designed to place high priority on the most sensitive parameters.

This paper evaluates the Full Setup by employing a marker-based offline display-parameter calibration. It confirms that the method performs as accurately as SPAAM and Recycle Setup. More importantly, this paper reports on a theoretical analysis of the calibration sensitivity of both setups as well as SPAAM with respect to the various display calibration parameters, based on real observations. The analysis allows us to reason about the display calibration accuracy for each method and provides insight into designing a suitably optimized OST-HMD system.

2 RELATED WORK

A detailed discussion of OST-HMD calibrations is in our paper [7].

Display Parameter Estimation Several research groups work on the display-parameter estimation. The offline step of Display-Relative Calibration (DRC) [11] estimates the display parameters through a standard vision-based calibration including first-order radial distortion. Gilson et al. [3, 4] employ Tsai’s method for estimating a camera frustum of an OST-HMD combined with an outside-in tracking system. Lee et al. [9] extended DRC to estimate higher-order radial distortion and showed that coefficients up to the 2nd order were actually effective.

Sensitivity Analysis to Calibration Errors Holloway [6] provide a thorough end-to-end error analysis for AR applications with an OST-HMD system. The author’s work includes a mathematical

*e-mail: itoh@in.tum.de

†e-mail: klinker@in.tum.de

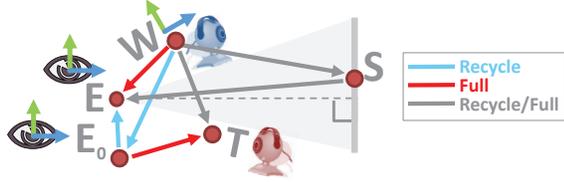


Figure 2: Spatial relationships of the virtual camera, illustrating the relevant internal coordinate systems of a screen S with an eye tracking camera E , a world camera W , and the user's eye E (or E_0).

model of the system and an evaluation which confirms the model by taking measurements by a real system. Axholt et al. [1] modeled user-dependent noise for a SPAAM calibration and observed that the noise manifests itself as a poorly estimated eyepoint, primarily along the line of sight, both in simulation and real measurements.

3 METHOD

3.1 Interaction-Free Calibration in a Nutshell

Fig. 2 shows the various coordinates to be defined as part of the display calibration process. Calibrating an OST-HMD means to estimate a 3-by-4 projection matrix $\mathbf{P}_{WE}(\mathbf{t}_{WE}) := \mathbf{K}_E [\mathbf{R}_{WE} \quad \mathbf{t}_{WE}]$ of a virtual camera defined by the OST-HMD and an eye (detail in [7]). The intrinsic matrix \mathbf{K}_E has two representations:

$$\mathbf{K}_E = \begin{bmatrix} \alpha_x & & c_x \\ & \alpha_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} z_{SE} & -x_{SE} \\ & z_{SE} & -y_{SE} \\ & & 1 \end{bmatrix} \quad (1)$$

$$= \mathbf{K}_{E_0} \begin{bmatrix} 1 + z_{EE_0}/z_{SE} & & -x_{EE_0}/z_{SE} \\ & 1 + z_{EE_0}/z_{SE} & -y_{EE_0}/z_{SE} \\ & & 1 \end{bmatrix}, \quad (2)$$

where $\mathbf{t}_{SE} = [x_{SE}, y_{SE}, z_{SE}]^T$, $\mathbf{t}_{E_0E} = [x_{EE_0}, y_{EE_0}, z_{EE_0}]^T$. $\mathbf{a} := [\alpha_x, \alpha_y]^T$ is a scaling factor that converts 3D points on the screen to pixel points. $c_x := (w-1)/2$ and $c_y := (h-1)/2$ define the image center with the pixel width w and height h . \mathbf{K}_{E_0} is the intrinsic matrix of another virtual camera defined by the old eye position E_0 .

Eq. (1) does not rely on the old eye position \mathbf{t}_{wE_0} . Instead, it requires the display pose \mathbf{t}_{wS} and the scaling vector \mathbf{a} . On the other hand, eq. (2) does not rely on these parameters, except for $[\mathbf{t}_{wS}]_z$ since $\mathbf{t}_{SE} = \mathbf{t}_{WE} - \mathbf{t}_{wS}$, and it reuses the old intrinsic matrix \mathbf{K}_{E_0} . Both cases require $(\mathbf{R}_{WE}, \mathbf{t}_{WE})$, the pose between the world and the eye. We call calibration with eq. (1) as Full Setup, and with eq. (2) as Recycle Setup.

Let T be the coordinates of an eye tracker rigidly mounted on the OST-HMD, then $\mathbf{t}_{WE} = \mathbf{R}_{WS} \mathbf{R}_{WT}^T (\mathbf{t}_{WT} - \mathbf{t}_{ET})$ (eq. (6) in [7]). In the previous work, we computed a 2D corneal limbus for estimating \mathbf{t}_{ET} through the Canny edge detector. Instead, our current implementation uses the Line Segment Detector by Gioi et al. [13].

3.2 Display Parameter Calibration

Our approach is similar to the work by Owen et al. [11] which reconstructs 3D shape of a virtual screen via the triangulation. They build a calibration jig for an HMD to obtain the pose of an calibration camera which captures the virtual screen of the HMD. Our method is modified in two-ways: we model the virtual screen as a 3D plane, and employ an inside-out marker tracking to obtain the calibration-camera poses. The following describes the calibration procedure:

Step 1: Place an OST-HMD so that a calibration camera observes a calibration pattern displayed on the virtual screen S . *Step 2:* Capture the pattern by the calibration camera, and capture a square marker M by a world camera W . *Step 3:* Remove the HMD carefully without touching the calibration camera, and capture the

marker by the calibration camera directly. *Step 4:* Repeat the step 1 to 3 N_C times.

At the step 1, a real black sheet is placed in front of the HMD so that the calibration camera can see the pattern clearly. The position of the calibration camera is changed at every iteration of the step 1. After the above procedure, one obtains poses between the world camera and each calibration camera C_k as $(\mathbf{R}_{C_kW}, \mathbf{t}_{C_kW})$. Ordinary camera calibration technique gives the virtual screen poses $\{(\mathbf{R}_{SC_k}, \mathbf{t}_{SC_k})\}_k$ up to a common scale factor s . This definition of s assumes that the scale factors $\{\alpha_x, \alpha_y\}$ are represented by a common factor α . We use this assumption through this paper. Without loss of generality, the size of a checkerboard tile is set to its pixel size. Then α becomes equal to s^{-1} .

3.2.1 Linear optimization step

The rotation estimate $\widehat{\mathbf{R}}_{wS}$ can be obtained by taking the mean of $\{\mathbf{R}_{wS}^k := (\mathbf{R}_{C_kW} \mathbf{R}_{SC_k})^T\}_k$ in the quaternion space [5]. The 3D position of the virtual screen in the world coordinates W can be written as, $\mathbf{t}_{SW}^k(s) = s \mathbf{R}_{C_kW} \mathbf{t}_{SC_k} + \mathbf{t}_{C_kW}$ for the k -th camera position. Define the averaged screen position $\overline{\mathbf{t}}_{SW}(s) := N_C^{-1} \sum_{k=1}^{N_C} \mathbf{t}_{SW}^k(s)$, and define the averaged screen position $\overline{\mathbf{t}}_{SW}(s) := N_C^{-1} \sum_{j=1}^{N_C} \mathbf{R}_{C_jW} \mathbf{t}_{SC_j} - \mathbf{R}_{C_kW} \mathbf{t}_{SC_k}$, $\mathbf{b}^k := N_C^{-1} \sum_{j=1}^{N_C} \mathbf{t}_{C_jW} - \mathbf{t}_{C_kW}$. Since the positions $\{\mathbf{t}_{SW}^k\}_k$ represent the same, we obtain the following cost function over the scale factor s :

$$f(s) := N_C^{-1} \sum_{k=1}^{N_C} \|\overline{\mathbf{t}}_{SW}(s) - \mathbf{t}_{SW}^k\|^2 = N_C^{-1} \sum_k \|\mathbf{s} \mathbf{a}^k + \mathbf{b}^k\|^2. \quad (3)$$

By solving this for $f'(s) = 0$, a linear estimate of the scale \hat{s} and the estimated display translation are obtained as follows,

$$\hat{s} = -\sum_{k=1}^{N_C} (\mathbf{a}^k \cdot \mathbf{b}^k / \|\mathbf{a}^k\|^2), \quad \widehat{\mathbf{t}}_{wS} = -\widehat{\mathbf{R}}_{wS}^T \overline{\mathbf{t}}_{SW}(\hat{s}). \quad (4)$$

3.3 Sensitivity Measurement

Having introduced INDICA, an important question arises: how accurately should we estimate each parameter of the method to achieve enough registration quality for an AR application? This section proposes a formal way to answer this question by defining a sensitivity measurement to calibration errors.

OST-HMD projection matrix can be treated as a function of calibration parameters: $\mathbf{P}_{WE}(\lambda)$ where a vector λ encapsulates the parameters, e.g. $\{\mathbf{a}, \mathbf{R}_{wS}, \mathbf{R}_{wT}, \mathbf{t}_{wT}, \mathbf{t}_{ET}, \mathbf{t}_{wS}\}$ for Full Setup. In other words, λ represents a display configuration of one particular OST-HMD design.

Let \mathbf{x}_w be a 3D point in the world coordinate system W , then \mathbf{x}_w is projected to a 2D pixel \mathbf{u}_{x_w} by the projection matrix $\mathbf{P}_{WE}(\lambda)$ as $\mathbf{u}_{x_w}(\lambda) := [p/r \quad q/r]^T$, where $[p \quad q \quad r]^T := \mathbf{P}_{WE}(\lambda) [\mathbf{x}_w \quad 1]^T$. Let λ^* be true calibration parameters and $\Delta\lambda$ represents small perturbations added during a calibration procedure, then $\mathbf{u}_{x_w}(\lambda^* + \Delta\lambda)$ represents a perturbed 2D pixel.

Define the reprojection error vector $\mathbf{e}(\mathbf{x}_E) := \mathbf{u}_{x_w}(\lambda^* + \Delta\lambda) - \mathbf{u}_{x_w}(\lambda^*)$. The first-order Taylor expansion gives an approximation of the vector as $\mathbf{e}(\mathbf{x}_E) \simeq \mathbf{J}_{\lambda^*}(\mathbf{x}_E) \Delta\lambda + O(\Delta\lambda^2)$ where $\mathbf{J}_{\lambda^*}(\mathbf{x}_w) := \left(\frac{d\mathbf{u}_{x_w}}{d\lambda}\right)_{\lambda^*}$ is a Jacobian matrix of $\mathbf{e}(\mathbf{x}_E)$ evaluated at λ^* . This Jacobian determines the primary behavior of the error caused by $\Delta\lambda$, and requires the first derivative of $(p(\lambda), q(\lambda), r(\lambda))$ only:

$$\frac{d\mathbf{u}_{x_w}}{d\lambda} = \frac{d}{d\lambda} \begin{bmatrix} p/r \\ q/r \end{bmatrix} = \begin{bmatrix} (p'r - pr')/r^2 \\ (q'r - qr')/r^2 \end{bmatrix}. \quad (5)$$

Although the calculation of eq. (5) is straightforward for most of the parameters in λ , rotation $\{\mathbf{R}_{wS}, \mathbf{R}_{wT}\}$ still need care due to their implicit parametrization. In a similar manner described in [8], we treat the little change of a rotation as an infinitesimal rotation in Lie algebra, which is expressed by a vector $\omega := [\omega_1, \omega_2, \omega_3]^T$ as



Figure 3: The OST-HMD setup used through the evaluations. The images contain annotations of the coordinate systems.

$\Delta \mathbf{R} := [\omega]_{\times} \mathbf{R}$, where $[\cdot]_{\times}$ is the skew-symmetric matrix operator. For example, $\frac{dP(\mathbf{R})}{d\mathbf{R}}$ can be computed by $= \frac{dP(\Delta \mathbf{R})}{d\omega}$. Higher-order terms of the rotation parameters such as $\omega_k * \omega_{k'}$ are treated as zero.

Now, averaging $\mathbf{J}_{\lambda^*}(\mathbf{x}_w)$ over the 3D point set $\{\mathbf{x}_w\}$ seems to behave as a sensitivity measurement. However, remember that \mathbf{x}_w itself is dependent on some display parameters. Thus the defined Jacobian does not take the same input set given different display configurations. Instead of \mathbf{x}_w , consider a 3D point \mathbf{x}_E in E with a polar coordinate representation: $\mathbf{x}_E(\delta, \theta, \varphi) := \delta \begin{bmatrix} \sin \theta \cos \varphi & \sin \theta \sin \varphi & \cos \theta \end{bmatrix}^T$, $\theta \in \Theta$, $\varphi \in \Phi$, $\delta \in L$, then $\mathbf{x}_w = \mathbf{R}_{WS}^T(\mathbf{x}_E - \mathbf{t}_{wE})$. Taking the mean of $\mathbf{J}_{\lambda^*}(\mathbf{x}_w)$ over the polar coordinate domain $\{L, \Theta, \Phi\}$ gives an expected error sensitivity measurement:

$$E[\mathbf{J}_{\lambda^*}(\mathbf{x}_w)] = (1/V) \int_{L, \Theta, \Phi} \mathbf{J}_{\lambda^*}(\mathbf{x}_w) d\delta d\theta d\varphi, \quad (6)$$

where V is the volume of the 3D space defined by $\mathbf{x}_E(\theta \in \Theta, \varphi \in \Phi, \delta \in L)$. Finally, by taking the sample mean of eq. (6), we obtain our sensitivity measurement:

$$\overline{\mathbf{J}_{\lambda^*}} := (1/N) \sum_{\delta, \theta, \varphi} \mathbf{J}_{\lambda^*}(\mathbf{R}_{WS}^T(\mathbf{x}_E(\delta, \theta, \varphi) - \mathbf{t}_{wE})), \quad (7)$$

where N is the number of 3D points sampled.

In summary, given a true OST-HMD configuration λ^* and a 3D space V in which an AR application needs to visualize AR contents, $\overline{\mathbf{J}_{\lambda^*}}$ gives a prediction of the sensitivity of each calibration parameter to calibration errors. Each column of $\overline{\mathbf{J}_{\lambda^*}}$ represents the sensitivity of a parameter with a different unit (e.g. scale, rotation and translation). For the sake of intuitive understanding, we convert $\overline{\mathbf{J}_{\lambda^*}}$ so that each calibration parameter has a scalar sensitivity measurement. For instance, let $\begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \end{bmatrix}$ be a 2-by-3 sub-matrix of $\overline{\mathbf{J}_{\lambda^*}}$ correspond to \mathbf{t}_{wT} , i.e. $\frac{d\mathbf{u}_{wT}}{d\mathbf{t}_{wT}}$. We define the scalar representation of the submatrix as $e_{\mathbf{t}_{wT}} := (\|\mathbf{e}_x\| + \|\mathbf{e}_y\| + \|\mathbf{e}_z\|)/3$. Other scalar measurements are defined in the same manner such as e_a , $e_{\mathbf{R}_{WS}}$, and so on. To compensate the difference of the units in the measurements, each measurement should be scaled properly during comparisons as explained later in the experiment section.

4 TECHNICAL SETUP

4.1 Hardware setup

We have built an OST-HMD system equipped with an eye tracker as described below and in Fig. 3. We use nVisor ST60 from NVIS—an OST-HMD with 1280x1080 resolution. The left-eye display is used for the current setup. An outward looking camera, Logitech C200, serves as the world camera W . For the eye tracker T , a PlayStation Eye camera is used. These cameras provide 640x480-pixel video and are attached to the HMD.

The position of the tracker is chosen to be at the bottom of the left display lens of the HMD. CL Eye Platform SDK¹ is used to capture images from the eye camera. The default focal length of its varifocal lens is manually adjusted and fixed to a suitable length.

¹<http://codelaboratories.com/products/eye/sdk/>

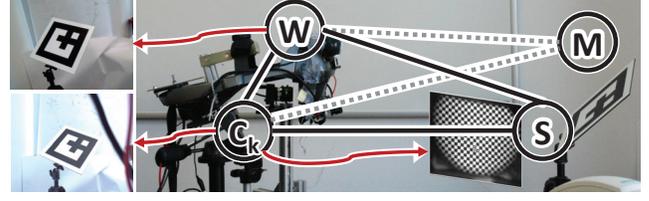


Figure 4: Display calibration setup for calibrating $\{\mathbf{a}, \mathbf{R}_{wT}, \mathbf{t}_{wS}\}$. (right) Spatial relationship with virtual screen. The screen is intentionally drawn in the image right for the schematic drawing. (left column) Sample images captured by the cameras.

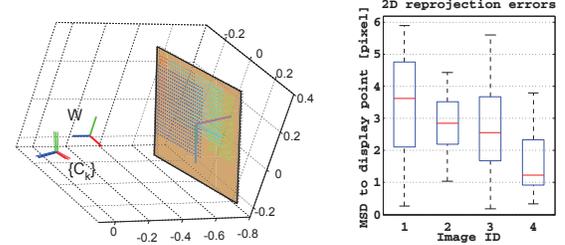


Figure 5: Display calibration result. (left) The estimated virtual screen plane and 3D grid points. (right) Reprojection error of each grid points per image. The error is relatively high compared to ordinary camera calibrations that yield subpixel errors in general.

4.2 System Calibration

To apply Full/Recycle setup to an OST-HMD system, such as the one described above, we have to precalibrate the system such that the display parameters become known. We conduct the marker-based display-parameter calibration as explained in Sec. 3.2

Fig. 4 shows our calibration setup. For the calibration camera, we used iDS’s UI-1240ML-C-HQ, an industrial camera which provides 1280x1024 color image, together with an 8-mm C-mount lens. World, tracker, and calibration cameras are calibrated beforehand by an open-source MATLAB toolbox² with printed checkerboard patterns. The poses $\{\mathbf{R}_{C_k W}, \mathbf{t}_{C_k W}\}$ were estimated via the marker coordinates M . The toolbox computes $(\mathbf{R}_{sc_k}, \mathbf{t}_{sc_k})$ up to scale. Fig. 5 left shows the calibration result. The reprojection error plot in Fig. 5 right shows relatively high error variances compared to standard camera calibrations which yield a sub-pixel accuracy in practice. Nevertheless, we will show later in the experiment that this calibration quality was sufficient to achieve required accuracy.

Calibration of $\{\mathbf{K}_{E_0}, \mathbf{t}_{wE_0}\}$ for Recycle Setup are described in [7].

5 EXPERIMENT

5.1 Design of the Test Process

The test process mostly follows the one in [7]. The main difference is that its data acquisition part is refined so that each data block can be collected individually. We have evaluated the performance of the interaction-free method (*Full/Recycle* Setup) compared to SPAAM (*training-error* condition) and to Degraded SPAAM (*test-error* condition). Fig. 7 shows an overview of the process.

5.1.1 Data acquisition

Prior to the evaluation, we acquired a series of data sets. Each set consists of 20 2D-3D point correspondences, with each 3D world point having been manually aligned to a 2D point on the screen (aka SPAAM, Fig. 6 left). The 3D points were distributed across an area of about 105x66x121 cm³ (width, height, depth) centered

²http://www.vision.caltech.edu/bouguetj/calib_doc/

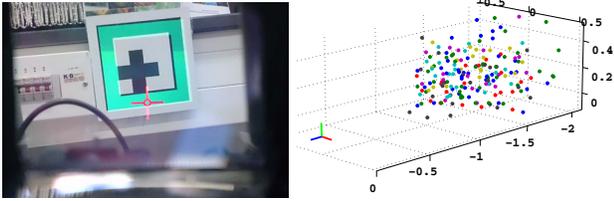


Figure 6: Data acquisition: (left) User’s view during SPAAM calibration. A virtual 2D red crosshair (2D point) will be matched to the black square marker (3D point). The green frame is a virtual image overlaid for checking the SPAAM quality. (right) Measured 3D points in meter in different colors for different blocks.

around position $(-1, 16, -149)$ [cm] relative to the operator (Fig. 6 right). Each 3D point set was also ensured to distribute well in depth for stable SPAAM calibration [1]. During this process, we also recorded at least 30 eye images per 2D-3D point correspondence. Fig. 7 (a) illustrates the step in form of a pink and a green box. The 2D-3D correspondences formed the basis for a SPAAM-based estimation of the projection matrix (blue box). The eye images were used to compute a series of 3D eye positions (orange box). We call such a data collection session a *block*.

A total of $N(=9)$ data collection sessions were performed (Fig. 7 (a)). During each session, the HMD was kept as stably as possible on the user’s head. After each session, the HMD was taken off from the head and put back on to simulate a degraded calibration situation. These blocks form the ground-truth (GT) data which are the basis for subsequent evaluations of the three evaluation conditions.

5.1.2 Data evaluation process

Training-error evaluation: For each block among N blocks, a SPAAM calibration is conducted and its quality is evaluated on the same block by using the procedure described in section 5.2. At the end, a total of $N(=9)$ error measurement sets were obtained. Fig. 7 (b) shows the procedure of this evaluation.

Test-error evaluation: One block is chosen for the SPAAM calibration and the calibration is tested against the rest of blocks –simulating the Degraded SPAAM condition in which a user continues using the same initial display calibration after the display was moved. This yields $N(N-1)$ sets of error measurements. Fig. 7 (c) shows this evaluation procedure.

Data acquisition for Full/Recycle methods (Fig. 7 (d)) is same as in [7]. Note that our experiment design is similar to [7], yet is more concise and strict. In the previous design, two sequences of consecutive blocks were recorded and the head-display position was assumed to be the same among the blocks in the same sequence. Then, SPAAM setup is evaluated *between* those blocks. Since the assumption is not exactly valid due to the head movement during a SPAAM, evaluation between any two blocks should be treated as Degraded SPAAM setup rather than SPAAM. As same as [7], our evaluation has two error measurements: 2D reprojection error and 3D eye position.

5.2 Performance Analysis

5.2.1 Comparison of 2D projection error:

Fig. 8 summarizes the result. The boxplot of the reprojection error (Fig. 8a top) shows that SPAAM setup achieves the best accuracy. This is expected since the setup learns a projection from a dataset and tested on the same. On the other hand, the other three methods show the almost same average error. For comparison, one might consider applying a statistical testing immediately. Before doing that, we analyze the histograms of the error (Fig. 8c).

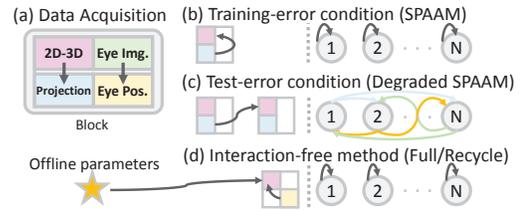


Figure 7: Overview of the experiment: (a) data acquisition, (b) training-error condition, (c) test-error condition, and (d) Full-/Recycle-setup conditions. Arrows between block nodes represent each evaluation: the source node is used for computing a projection by each method, and the destination node for evaluating the projection.

The error histogram of Degraded SPAAM gives inhomogeneous distribution somewhat similar to the Chi-square distribution. The reason can be explained by considering the re-wearing process during the calibrations. Every time an operator takes the OST-HMD off and on again, most of the time the display was set to almost the same position and few times to the position which is very different from the others. On the other hand, the histograms of **INDICA** give more homogeneous distributions with lower variances. Thus, when an OST-HMD is in a long-term use, **INDICA** is more reliable since the homogeneous property can upper bound the error range by the variance of the distributions. Overall, **INDICA** can be considered to be more stable than the Degraded SPAAM.

Furthermore, the correlation graph in Fig. 8a bottom gives another insight. The graph shows correlations between the GT points and the 2D reprojection error vectors –vectors from GT points to their corresponding reprojected 2D points. SPAAM has almost no correlation as DLT method computes an estimate which minimizes the error variance, which means that SPAAM tends to produce a projection over-fit to a given observation.

Degraded SPAAM holds some correlation, this is also understandable since different display positions on different head positions create constant bias errors. The correlation in Recycle Setup is even smaller, this implies that the method has achieved as good accuracy as it can under the combined use of SPAAM. Since Recycle Setup is relying on another projection matrix from SPAAM, the error might also reflect the *test error* of SPAAM that the method would actually achieve with other datasets taken in the same setup.

Notably, Full Setup has huge correlation while maintaining comparable calibration accuracy. This indicates that Full Setup still contains a bias error somewhere in the calibration procedure, thus has a room to further increase the calibration accuracy.

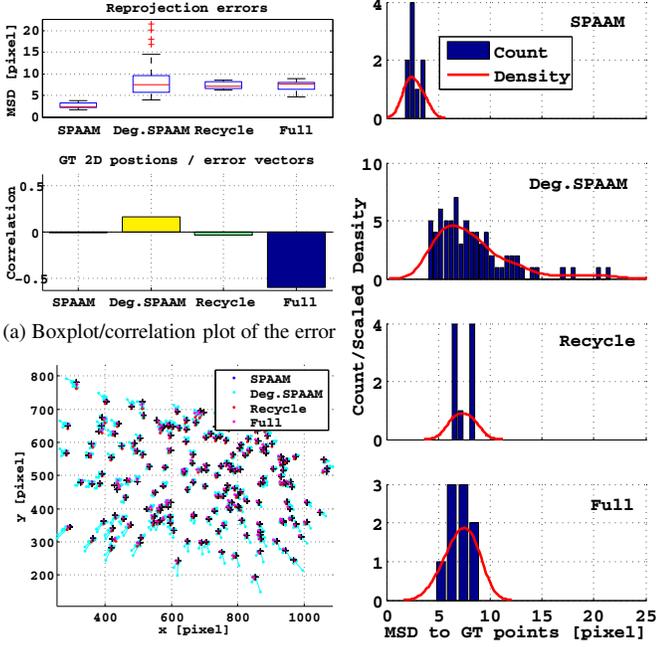
5.2.2 Comparison of 3D eye positions:

As reported in [1, 2, 7], eye position estimates by SPAAM tend to have a large variance in z-axis, typically the viewing direction of an eye. Fig. 9 shows the estimated eye positions (\mathbf{t}_{EW}) in the world coordinates. It shows the similar tendency for SPAAM results while **INDICA** gives quite stable estimates as similar to [7]. In the next section, We will provide a reasoning why SPAAM has this error tendency. In short, this is because the z-axis does not impact on the reprojection error as strong as x and y axes do.

There is a shift between the mean y position of SPAAM and that of **INDICA** (Fig. 9 right). This implies that the eye position estimates have a bias error in either or both methods.

5.3 Sensitivity Analysis

For sensitivity analysis of the obtained calibration parameters, we follow a general approach described at the section 3.4 in [6]: We deliberately add perturbations to each calibration parameter, and recompute the calibration error for each perturbation to observe how the errors propagate to the reprojection error –the errors of most



(a) Boxplot/correlation plot of the error (b) Reprojected/GT 2D points (c) Error histograms and distributions.

Figure 8: Comparison of 2D projection errors. (a) The high correlation seen at Full Setup suggests the existence of bias errors in the calibration procedure (this correlation is indeed observable in (b)). (c) Each distribution is normalized with the area of corresponding histograms for the visualization. The distribution of Degraded SPAAM has a gentle error tail toward error-increasing direction while the other three methods' distributions do not. The density estimations were done by applying the kernel density estimation.

concern for users. We treat the calibration parameters and the eye position estimated during the experiment as λ^*

As defined in eq. (7), the sensitivity measurements for Full Setup: $\{e_X\}_X$, $X \in \{\mathbf{a}, \mathbf{R}_{WS}, \mathbf{R}_{WT}, \mathbf{t}_{WT}, \mathbf{t}_{ET}, \mathbf{t}_{WS}\}$ are computed for the given calibration parameters for the error prediction. In Full Setup, λ excludes display pixel center $\{c_x, c_y\}$ since the center is determined by a known display image resolution (and S is defined at this center). Since the unit of the measurement is in the form of [pixel/Y], the measurements are scaled depending on the units of related display parameters as the following: rotation ($\mathbf{R}_{WS}, \mathbf{R}_{WT}$) by 1 with $Y=[\text{deg}]$, translation ($\mathbf{t}_{WT}, \mathbf{t}_{ET}, \mathbf{t}_{WS}$) by 0.01 with $Y=[\text{m}]$, and pixel scaling \mathbf{a} by 10 with $Y=[\text{pixel/m}]$.

The sensitivity measurements for Recycle Setup, $X \in \{\mathbf{c}, \mathbf{f}_{E_0}, [\mathbf{t}_{WS}]_z, \mathbf{t}_{WT}, \mathbf{t}_{ET}, \mathbf{t}_{WE_0}, \mathbf{R}_{WE_0}, \mathbf{R}_{WT}\}$, and the SPAAM setup, $X \in \{\mathbf{c}, \mathbf{f}_E, \mathbf{t}_{WE}, \mathbf{R}_{WE}\}$, can be obtained in the same manner by setting $\mathbf{P}_{WE}(\lambda)$ with eq. (2) or to $\mathbf{K}_E[\mathbf{R}_{WS}, \mathbf{t}_{WE}]$ respectively. \mathbf{f}_{E_0} and \mathbf{f}_E are focal-length vectors of an old and a new projection matrix respectively, and are scaled by 10 [pixel/m]. \mathbf{c} is the image center vector and scaled by 10 [pixel/pixel], namely e_c becomes constant for both SPAAM, and Full/Recycle Setup. For computing the sensitivity measurements of each method, one of the block in Sec. 5.1.1 is used. For Full/Recycle Setup, an eye position estimate from the block is also used to compute the sensitivity of \mathbf{t}_{ET} . The distribution of 3D points are chosen in the range of the 3D GT dataset.

Fig. 10 shows the analysis result. The upper row shows predicted reprojection errors for each calibration parameter, and the lower shows errors actually observed. For SPAAM, we explicitly visualized each axis of \mathbf{t}_{WE} for more detailed analysis. Overall, one can see that the predictions coincide with the observed errors. The figure provides several insights about the three methods including:

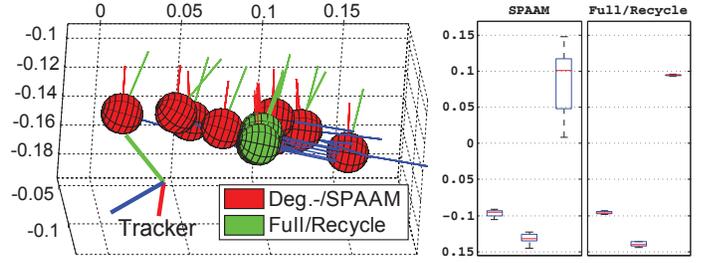


Figure 9: Comparison of 3D eye positions. (left) Visualization of estimated eye positions $\mathbf{t}_{EW}[\text{m}]$. Eyeballs are drawn with the radius used in the experiment. Orientations of the eyeballs show that of the screen (\mathbf{R}_{SW}). (right) \mathbf{t}_{EW} w.r.t xyz axes in [m]. z-axis estimates from SPAAM are instable.

(1) SPAAM is relatively insensitive to the estimation error of z-axis of the eye position \mathbf{t}_{WE} compared to that of the other axes. In other words, SPAAM tends to estimate a projection which has bigger variance in the z-axis direction. The rotation \mathbf{R}_{WE} is a dominant parameter in SPAAM, thus if SPAAM gives an accurate projection, decomposed \mathbf{R}_{WE} would be quite reliable. In turn, if Degraded SPAAM is used, HMD should be designed so that users can put it on the same orientation. (2) Both for Recycle/Full Setup, the eye position estimate \mathbf{t}_{ET} is in the second dominant parameter group. (3) Recycle setup is sensitive to old eye poses while not to the display parameter ($[\mathbf{t}_{WS}]_z$). Thus once an accurate old eye pose (and projection) is given by other methods such as SPAAM, $[\mathbf{t}_{WS}]_z$ does not require strict accuracy. This can be the reason why the evaluation of this setup by [7] worked well with a rough $[\mathbf{t}_{WS}]_z$ which was measured by hand. (4) Full Setup is especially sensitive to the calibration quality of the virtual screen (S) relative to the world. This supports that the quality of our marker-based display calibration was well enough to be compared to SPAAM methods.

5.4 Discussion

Throughout the experiment, Full (and Recycle) Setup achieved more stable and comparably accurate calibration quality against Degraded SPAAM. The analysis of the projection errors tells that Degraded SPAAM, a setup where a user compromises on a default or old calibration setting, is not a preferable solution for the long-term use of OST-HMDs; it is hard to guarantee maximum error bound and is not easy to predict how worse it can be. From the correlation analysis, the Recycle Setup seems to have achieved the ideal accuracy given that the partial calibration parameters are given by SPAAM. Thus, replacing SPAAM with camera-based methods, e.g. [4], is a possible direction further improving the performance of the setup. On the one hand, Full Setup still shows potential to achieve better performance once the following error source is identified and eliminated.

It is still unclear why the Full Setup has: (a) huge correlation to the GT points and yet has (b) small reprojection error. We suspect the eye tracking as the cause because of two reasons observed: the existence of the offset in the 3D eye position estimates for (a) (Fig. 9), and the low error sensitivity of eye position found in the sensitivity analysis for (b) (Fig. 10). As mentioned in the discussion in [7], the source of the offset can be due to the discrepancy between the eye model and the real eyeball. One can explore these issues by, e.g., installing the eye tracker in a different configuration to see the change of correlation, improving the tracking method, and so on.

The two setups have a clear difference in the number of parameters to be estimated. By recalling eq. (1) and (2), one can derive that Recycle/Full Setup yield 16/19 DoF respectively despite the fact that they represent the same projection matrix (Fig. 1). Recycle Setup aims to model the system more concisely and Full Setup does exactly. Each setup requires different precalibration proce-

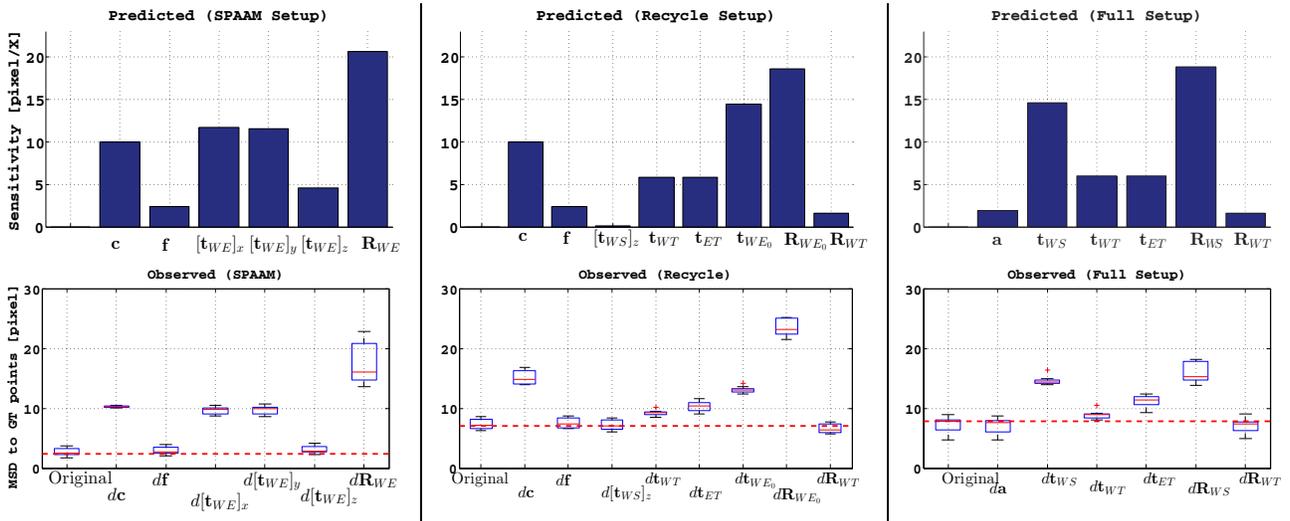


Figure 10: Sensitivity analysis against calibration errors. (top row) Predicted errors (bottom row) Observed errors. 'Original' are the original errors without perturbations, and red dotted-lines are baselines drawn at the mean values of the original errors. One can see that the predicted errors coincide with observations from the real datasets, and vice versa.

dures with different intricacy, thus one should consider the overall complexity of the calibration flow when applying **INDICA**.

The sensitivity analysis (Fig. 10) gave various insight about the use of the three calibration methods. However, strictly speaking, the result is valid only for the particular OST-HMD configuration we tested—an indoor setup with a configuration where a world camera is set on the top of the HMD and an eye tracker on the bottom. Different OST-HMDs do yield radically different configurations, and different AR applications (and FoV of HMDs) do yield different 3D point space of interest. Thus the result itself might not directly be applied to quite different scenario such as outdoor setups or HMDs with large FoV, yet one can conduct their own analysis based on our formulation once they identified their current configuration or have HMD design at hand.

The proposed sensitivity-analysis framework has potential to impact on designing an optimal OST-HMD configuration. Searching the display parameter domain with the sensitivity measurement might give the optimal configuration for a certain application. For instance, our informal investigation shows that Full Setup becomes less sensitive to the eye-position error when the eye tracker and the world camera are *straightly aligned on the eye axis*, which requires half mirror optics[10]. Contrarily, this setup becomes more sensitive to the virtual screen pose error. This setup would be preferable when an OSD-HMD can be finely calibrated in a factory, then used by variety of people with less accurate eye tracking.

6 CONCLUSION

We conduct intensive analysis of the interaction-free OST-HMD calibration method. The evaluation demonstrates the Full Setup performs as accurately as the Recycle Setup under the use of a marker-based display calibration. Furthermore, we formulates an error sensitivity analysis for both SPAAM and the interaction-free method by deriving the Jacobian of reprojection error over eye positions and display parameters. The analysis formulation is then investigated on an HMD with justification of the theory by the real measurements, which brings various insight including: high sensitivity of the virtual screen parameters, middle sensitivity of the eye position, the reasoning of SPAAM's error tendency etc.

ACKNOWLEDGMENTS

The authors wish to thank all members of the Fachgebiet Augmented Reality of TU Munich and the EDUSAFE project. The

project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under PITN-GA-2012-316919-EDUSAFE.

REFERENCES

- [1] M. Axholt, M. Skoglund, S. D. Peterson, M. D. Cooper, T. B. Schön, 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 the Human Factors and Ergonomics Society Annual Meeting*, volume 54, pages 2427–2431. SAGE Publications, 2010.
- [2] 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 *Proceedings of VR*, pages 27–34. IEEE, 2011.
- [3] S. Gilson and A. Glennerster. High fidelity immersive virtual reality. *Virtual reality-human computer interaction*, 2012.
- [4] S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster. Spatial calibration of an optical see-through head-mounted display. *Journal of neuroscience methods*, 173(1):140–146, 2008.
- [5] C. Gramkow. On averaging rotations. *Journal of Mathematical Imaging and Vision*, 15(1-2):7–16, 2001.
- [6] R. L. Holloway. Registration error analysis for augmented reality. *Presence*.
- [7] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *IEEE 3DUI*, pages 75–82, march 2014.
- [8] K. Kanatani and Y. Sugaya. Bundle adjustment for 3-d reconstruction: Implementation and evaluation. *Memoirs of the Faculty of Engineering, Okayama University*, 45:1–9, 2011.
- [9] S. Lee and H. Hua. A robust camera-based method for optical distortion calibration of head-mounted displays. In *Virtual Reality Conference (VR)*, 2013 IEEE, pages 27–30. IEEE, 2013.
- [10] S. MaNN. Through the glass, lightly [viewpoint]. *Technology and Society Magazine, IEEE*, 31(3):10–14, 2012.
- [11] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative calibration for optical see-through head-mounted displays. In *Proceedings of ISMAR*, pages 70–78, 2004.
- [12] M. Tuceryan and N. Navab. Single point active alignment method (spaam) for optical see-through hmd calibration for ar. In *Proceedings of ISAR*, pages 149–158. IEEE, 2000.
- [13] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: a line segment detector. *Image Processing On Line*, 2012.