Optical See-Through Head Mounted Display Direct Linear Transformation Calibration Robustness in the Presence of User Alignment Noise

Magnus Axholt, Martin A. Skoglund, Stephen D. Peterson, Matthew D. Cooper, Thomas B. Schön, Fredrik Gustafsson, Anders Ynnerman, Stephen R. Ellis

Division of Automatic Control

E-mail: magnus.axholt@liu.se, ms@isy.liu.se, stepe@itn.liu.se, matco@itn.liu.se, schon@isy.liu.se, fredrik@isy.liu.se, anders.ynnerman@liu.se, stephen.r.ellis@nasa.gov
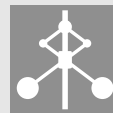
27th September 2010

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Abstract

The correct spatial registration between virtual and real objects in optical see-through augmented reality implies accurate estimates of the user's eyepoint relative to the location and orientation of the display surface. A common approach is to estimate the display parameters through a calibration procedure involving a subjective alignment exercise. Human postural sway and targeting precision contribute to imprecise alignments, which in turn adversely affect the display parameter estimation resulting in registration errors between virtual and real objects. The technique commonly used has its origin in computer vision, and calibrates stationary cameras using hundreds of correspondence points collected instantaneously in one video frame where precision is limited only by pixel quantization and image blur. Subsequently the input noise level is several order of magnitudes greater when a human operator manually collects correspondence points one by one. This paper investigates the effect of human alignment noise on view parameter estimation in an optical see-through head mounted display to determine how well a standard camera calibration method performs at greater noise levels than documented in computer vision literature. Through Monte-Carlo simulations we show that it is particularly difficult to estimate the user's eyepoint in depth, but that a greater distribution of correspondence points in depth help mitigate the effects of human alignment noise.

Optical See-Through Head Mounted Display Direct Linear Transformation Calibration Robustness in the Presence of User Alignment Noise

Magnus Axholt[1], Martin Skoglund[2], Stephen D. Peterson[1], Matthew D. Cooper[1],
Thomas B. Schön[2], Fredrik Gustafsson[2], Anders Ynnerman[1], Stephen. R. Ellis[3]

[1]Department of Science and Technology, Linköping University, Sweden
[2] Division of Automatic Control, Linköping University, Sweden
[3] Human Systems Interface Division, NASA Ames Research Center, CA, USA

The correct spatial registration between virtual and real objects in optical see-through augmented reality implies accurate estimates of the user's eyepoint relative to the location and orientation of the display surface. A common approach is to estimate the display parameters through a calibration procedure involving a subjective alignment exercise. Human postural sway and targeting precision contribute to imprecise alignments, which in turn adversely affect the display parameter estimation resulting in registration errors between virtual and real objects. The technique commonly used has its origin in computer vision, and calibrates stationary cameras using hundreds of correspondence points collected instantaneously in one video frame where precision is limited only by pixel quantization and image blur. Subsequently the input noise level is several order of magnitudes greater when a human operator manually collects correspondence points one by one. This paper investigates the effect of human alignment noise on view parameter estimation in an optical see-through head mounted display to determine how well a standard camera calibration method performs at greater noise levels than documented in computer vision literature. Through Monte-Carlo simulations we show that it is particularly difficult to estimate the user's eyepoint in depth, but that a greater distribution of correspondence points in depth help mitigate the effects of human alignment noise.

INTRODUCTION

Augmented Reality (AR) is a technique by which computer generated signals synthesize impressions that are made to coexist with the surrounding real world as perceived by the user. Human smell, taste, touch and hearing can all be augmented, but most often AR refers to human vision being overlaid with information otherwise not readily available to the user. Head-mounted display (HMD) techniques by which the human vision may be augmented are commonly divided into video see-through (VST) and optical see-through (OST) devices (Rolland & Fuchs, 2000). Correct calibration of these devices is not only important on an application level, ensuring that e.g. data labels are presented at correct locations, but also on a system level to enable display techniques such as stereoscopy to function properly (Livingston et al, 2006) (Cakmakci & Rolland, 2006), and ultimately ensuring that the user does not suffer from discomfort or injury (Stanney et al, 1998). Thus, calibration methodology is an important research area vital to AR.

In this paper we study the effect of human alignment noise on parameter estimation variability using a standard camera calibration procedure of an OST HMD. Our main findings are:

1. Eyepoint estimation along the user's line of sight is the most sensitive to noise of all calibration parameters.

2. The distribution of correspondence points in depth is a very influential variable. Simulation results show that it takes 81 manual alignments to estimate user's eyepoint with ±0.035 m precision in 50% of the calibration attempts at representative noise levels when correspondence points are distributed over ±0.1 m. However, if the correspondence points are distributed over ±0.5 m instead, only 9 alignments are needed for the same precision.

3. Parameter estimation variability can be modeled linearly as a function of noise for all calibration parameters.

RELATED WORK AND PROBLEM STATEMENT

Being inherently based on a camera, VST devices are generally considered easier to calibrate due to the straightforward application of camera calibration techniques found in e.g. computer vision literature (Abdel-Aziz & Karara, 1971), (Haralick, 1989), (Tsai, 1987), (Zhang 2000), (Hartley & Zisserman, 2000). Using a pinhole camera model as an analogy for the human eye, similar calibration techniques have been suggested for OST devices as well (McGarrity & Tuceryan, 1999), (Tuceryan et al, 2000), (Genc et al, 2002), (Tang et al, 2003), (Owen et al, 2004), (Gilson et al, 2008).

Either as a part of their calibration procedure (Owen et al, 2004), (Gilson et al, 2008), or as a part of their evaluation process (McGarrity & Tuceryan, 1999), (Tuceryan et al, 2000), previous researchers have performed the calibration of an OST HMD on stationary platforms where the calibration

procedure was not affected by the human alignment noise characteristic of postural sway and head rotation. Furthermore, in the cases where quantitative calibration quality is reported (Genc et al, 2002), (Tang et al, 2003), or otherwise illustrated as lateral registration error (McGarrity & Tuceryan, 1999), (Tuceryan et al, 2000), no information on head orientation is given. Therefore, this paper sets out to answer three questions: **1)** What is the effect of human alignment noise on OST HMD calibration? **2)** How does parameter variance, e.g. uncertainty in the eyepoint position, affect the resulting calibration quality? **3)** Which is the most influential parameter?

Human Operator Limitations

To calibrate an OST HMD according to a standard camera calibration procedure the human operator must manually and subjectively align at least six landmarks with their corresponding pixel coordinates on the transparent screen through what is known as a boresight exercise. This presents challenges in terms of **a)** simultaneous correspondence point acquisition, **b)** human alignment precision, and **c)** the use of assisting technology.

a) In the case of VST HMD, a video frame provide a "snapshot" in which hundreds of correspondence points can be collected simultaneously, but with OST a human operator will inevitably move between each alignment, thereby preventing the boresight lines from converging into a single eyepoint in space. A solution to this challenge is to use only one landmark, but reference it in the head coordinate system according to the Single Point Alignment Algorithm (SPAAM) (Tuceryan & Navab, 2000). Since the tracker sensor now serves as the origin, the user can collect an arbitrary number of correspondence points moving freely in between alignments as the boresight lines will all converge at the end of vector **t** (9).

b) Human alignment precision in an OST HMD is predominantly dependent on head rotation precision which has been reported to be 0.13° (Nicholson, 1966), 0.9° (Verona, 1978), and 0.04° (Wells & Griffins, 1987) standard deviation. Most recently (Axholt et al, 2009) reported 0.25° precision for 12 standing subjects using an OST HMD with VGA resolution through 28° by 37° FOV, which converts to 4.3 px misalignment. However, when the effect of head translation was removed, subjects exhibited sub-pixel alignment precision on the order of human visual acuity (0.016°, 0.2 px). Hence head rotations are thought to compensate postural sway. Unfortunately the translation compensation is only possible in dynamic modeling using time series and is not an option for the static standard camera calibration procedure.

c) (McGarrity & Tuceryan, 1999), (Tuceryan et al., 2002), (Genc et al., 2002), and (Tang et al.,2003) adopt a calibration procedure relying on the user to subjectively construct alignments. In the presence of noise, calibration quality generally improves with the number of correspondence points (Chen et al., 1994). However, due to human fatigue it is reasonable to believe that there exists an optimal calibration quality as a function of the number of correspondence points

versus deteriorating user alignment due to fatigue over the time it takes to construct the alignments. To reduce the effect of noise in "time-consuming and error-prone human measurements" Gilson et al. (Gilson et al, 2008) replaces the user's eye with a camera during calibration to estimate the parameters of the OST HMD with techniques similar to that of VST calibration. The subsequent evaluation is however made with the same camera, and not with a human eye, and therefore does not illustrate the effects of mismatching camera and eye position. Owen et al. (Owen et al, 2004) also use camera aided calibration and addresses the challenge of switching camera for human eye by dividing the calibration procedure in two phases, one for intrinsic and one for extrinsic parameters. These two works rely on the fact that the intrinsic parameters only need to be estimated once, and do not change between user sessions. This fact is unfortunately only true for an ordinary camera with a rigid camera house, but not for an OST HMD as the location of the eye, after it has replaced the camera, does not necessarily coincide with the apex of the frustum defined by the intrinsic parameters. Thus focal length, **f**, and principal point ($\mathbf{p_x}$, $\mathbf{p_y}$) also need to be adjusted between each user session. Genc et al. (Genc et al, 2002) addresses this fact, but chooses to correct either principal point or tracker-eye offset **t**, not both. Thus it seems as the technique of separating extrinsic and intrinsic parameters during calibration is not a viable path for reducing the number of alignments made by a human operator using an OST HMD. A more promising approach is to study the design of the calibration environment and in particular the configuration of correspondence points: For a fixed configuration there exists an optimal number of correspondence points (Chen et al., 1994) and for a variable configuration, setups with points at the perimeter of the calibrated volume yield better results (Challis & Kerwin, 1992).

Computer Models

To be able to correctly merge the real and the virtual world during user interaction with a dynamic scene, an AR system maintains a computer model to represent the location of real
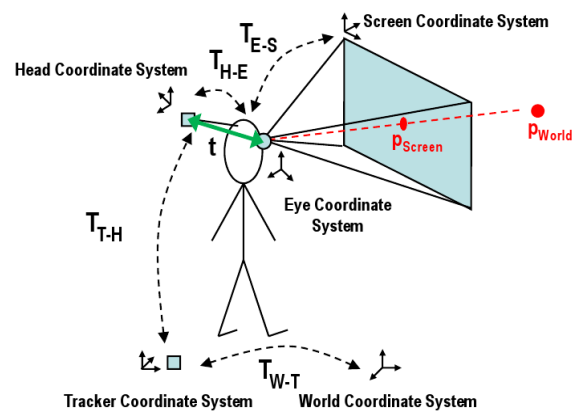

Figure 1: Transformations in an AR system.

and virtual objects. The spatial relationships are normally modeled using linear transformation matrices. As 4-by-4 matrices, they can be aggregated through multiplication to symbolize the traversal through local coordinate systems and

so describe the exact location of surrounding objects relative to the user's eye. In Figure 1 T_{W-T} illustrates where the tracker transmitter is located in the world. In this model we assume that the tracker and world coordinate systems coincide, thus $T_{W-T} = I$. As the user moves, the tracker continuously reports the position and orientation of the head-worn tracker sensor relative the tracker transmitter through T_{T-H}. The tracker sensor and the display are assumed to be rigidly mounted to a helmet which in turn is assumed to perfectly follow the user's head. This means that both T_{H-E} and T_{E-S} are static. The green arrow illustrates the tracker-eye offset t.

At the eyepoint, the user's view is traditionally modeled as a pinhole camera frustum. Assuming no radial distortion in the optics, this camera subsystem can be modeled as two matrices holding extrinsic and intrinsic camera parameters which conveniently can be multiplied into the matrix aggregation. By aggregating the static matrices $\{T_{H-E}, T_{E-S}\}$ separately the calibration procedure becomes the task of populating the elements of the aggregated static matrix T_{CAL}, see eq. 1-3, instead of determining each measurement individually. This is preferable since some of the measurements needed for an accurate calibration model are hard to obtain directly. In the case of an OST display the offset t between the tracker sensor and the eye is an example of such a measurement.

$$p_{screen} = T_{W-T}T_{T-H}T_{H-E}T_{E-S}p_{world} \quad (1)$$

$$T_{W-T} = I, T_{H-E}T_{E-S} = T_{CAL} \quad (2)$$

$$p_{screen} = T_{T-H}\,T_{CAL}p_{world} \quad (3)$$

Standard Camera Calibration Procedure

Common calibration procedures usually spring from camera resectioning (Hartley & Zisserman, 2000), camera pose estimation (Haralick, 1989), and direct linear transform (DLT) problems (Abdel-Aziz & Karara, 1971), in which the relationship between landmarks of known locations in the surrounding real world, p_{world}, and points of known pixel coordinates on the screen, p_{screen}, are used to determine a 3-by-4 camera matrix T_{CAL}, see eq. 4. This corresponding point data can be expressed as a system of homogeneous linear equations, eq. 6, in which x is a vector of the elements in matrix T_{CAL}, and A is the result the matrix multiplication in eq. 5 when the perspective divide, w, has been substituted, see e.g. Appendix A in (Sutherland, 1974) for details. The positions in p_{world} and p_{screen} are usually normalized to a common order of magnitude. By conditioning the matrix A this reduces the effect of noise (Hartley, 1997) (Wan & Xu, 1996). The minimum number of correspondence points depends on how the degrees of freedom (DOF) of the calibration model have been parameterized. For the 12 entries in T_{CAL} at least six points are needed, but in practical applications more points are usually gathered to further mitigate the effect of noise. This prompts the use of singular value decomposition (SVD) by which A is factored into two bases (U,V) and a diagonal matrix (Σ), see eq. 7. The eigenvalue calculations in SVD effectively perform a least square approximation. Thus the 12th and last column in the base matrix V, that by convention corresponds to the smallest singular value in Σ, can be interpreted as the calibration matrix T_{CAL}, see eq. 8, which projects landmark coordinates onto screen coordinates with the smallest residual between screen points and corresponding landmarks as seen by the user.

$$p_{screen} = Tp_{world} \quad (4)$$

$$w[u\ v\ 1]^T = \begin{bmatrix} T_{1,1} & T_{1,2} & T_{1,3} & T_{1,4} \\ T_{2,1} & T_{2,2} & T_{2,3} & T_{2,4} \\ T_{3,1} & T_{3,2} & T_{3,3} & T_{3,4} \end{bmatrix} [x\ y\ z\ 1]^T \quad (5)$$

$$Ax = 0 \quad (6)$$

$$A = U\Sigma V^T \quad (7)$$

$$T_{CAL} = \begin{bmatrix} V_{1,12} & V_{4,12} & V_{7,12} & V_{10,12} \\ V_{2,12} & V_{5,12} & V_{8,12} & V_{11,12} \\ V_{3,12} & V_{6,12} & V_{9,12} & V_{12,12} \end{bmatrix} \quad (8)$$

At this point, some calibration procedures adjust for non-linear lens effects by using T_{CAL} as initial values for a Levenberg-Marquardt (LM) optimization procedure further refining T (Tsai, 1987) (Zhang 2000). Known measurements can be used as soft parameter constraints for the LM, and further robustness to noise can be provided by weighting the optimization cost function to decrease the effect of outliers (Hartley & Zisserman, 2000).

The matrix T_{CAL} can be divided further into extrinsic, $R|t$, and intrinsic, K, camera parameters with RQ-decomposition using Givens rotations, see eq. 9. At this stage the offset from the tracker sensor origin to the center point of the eye is accessible through t, and R describes the rotation of the screen. K gives focal length, α, which in turn holds the distance to the screen (in meters), f, if the pixel ratio, m (pixels per meter), is known, see eq. 11. With known screen resolution, this information also gives the theoretical FOV. The practical FOV is, however, dependent on the eye remaining inside the exit pupil defined by the Lagrange Invariant (Cakmakci & Rolland, 2006), as well as being inside the aperture stop, i.e. that the eye aligns with the frustum apex defined by the the principal point, (p_x, p_y) and focal length (α_x, α_y). s denotes skew, i.e. perpendicularity of display surface axes, and is 0 for most normal cameras (Hartley & Zisserman, 2000).

$$T_{CAL} = K[R|t] \quad (9)$$

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} R = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (10)$$

$$\alpha = m * f \quad (11)$$

METHODS

Modeling and Simulation

Using MATLAB R2006b a right-handed frustum object with negative z view vector was built to roughly model a Kaiser ProView 50ST with 640 by 480 px (VGA) resolution,

28° by 37° FOV, centered principal point, projection plane located 0.05 m in front of the user's left eye, and the camera origin located stationary in the world origin. Correspondence points were located 2 meters in front of the user's left eye, such that they were evenly spaced over the FOV, but at randomized depth to avoid coplanar correspondences. The independent variables of the simulation where: **1)** number of correspondence points {6,9,12,16,20,42,81} distributed in a grid pattern with even spacing throughout the display surface, **2)** human noise distribution {fixed range, white noise, Gaussian} parameterized using range, **3)** human noise magnitude defined as pixel range {0,1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15} introduced as permutations of p_{screen} in random (white) direction, and **4)** random (white) distribution of correspondence points in depth with range ±0.1-1.0 m with 0.1 m increments. The simulation was run with 1,000 iterations per combination of independent variables to collect T_{CAL}.

RESULTS

Parameter Estimation Variability

The simulation results in Table 1 show that human noise introduced during the boresight exercise mainly manifests itself as a poorly estimated eyepoint, primarily along the line of sight, t_z. This effect is visible reading Figure 2 horizontally. The simulation also confirmed previous findings, that variance in the estimated parameters due to poor boresighting can be mitigated through the use of additional correspondence points, but more importantly it also showed that increasing the range of correspondence points in depth greatly improves the parameter estimation: E.g. Table 1 shows that 9 correspondence points distributed over ±0.5 m perform equally well as 81 points over ±0.1 m in the presence of human alignment noise of 5 px range. The improvement due to depth distribution is visible reading Figure 2 vertically.

Figure 2: Variability in the eyepoint estimation as a function of Gaussian alignment noise and distribution of 20 correspondence points in depth. Blue boxes denote interquartile ranges (IQR). Red + signs denote outliers > 1.5 IQR from distribution median.

Simulation results also showed that variance in all of the estimated camera parameters increases linearly as a function of human alignment noise (correlation $r^2 > 0.99$) for all three human noise models in the range of 1-15 px using 12-81 evenly distributed grid correspondence points. For the noise interval 0-1 px and for 6-9 correspondence points the relationship between variance and noise was found non-linear. As some HMD model the virtual image plane at optical infinity (> 2 m), the relationship was also tested for screen distance set to 3.0 m. It was found that the linear relationship between parameter variance and noise is independent of screen distance, but dependent on screen resolution and FOV. The reason for this is that noise is defined as pixel range and resolution and FOV changes the physical size of a pixel. Lastly we found that calibration parameters exhibited increasingly greater variance when human noise was modeled with Gaussian noise, white noise and lastly fixed range, although this was expected given the fact that range was used to parameterize distributions of different shape.

Support in Empirical Data

In a pilot study for a related experiment, not yet published, two human subjects collected 9 datasets of alignment noise over 81 grid correspondence points distributed over ±0.1 m using a Kaiser ProView 50ST with 640 by 480 px (VGA) resolution, 28° by 37° FOV. Using a bootstrap method (Efron, 1979), mean interquartile range (min, max) was estimated to 0.21 (0.09,0.45), 0.11 (0.04,0.20), 0.57 (0.33,1.02), 512 (235,1174), 512 (176,1125), 150 (101,298), 189 (60,415), 4.9 (1.7,10.1) for t_x, t_y, t_z, f_x, f_y, p_x, p_y, orientation, respectively (p=0.05). While the confidence interval is far too large to be conclusive and the number of subjects too few to be representative, these observations still suggest t_z to be the more variable parameter.

DISCUSSION AND FUTURE WORK

Traditional camera calibration has generally only been studied at relatively low noise levels (< 1 px) (Sun & Cooperstock, 2006) and not at levels like those of head rotation precision (~5 px). The simulation in this paper illustrates that human alignment noise induces parameter estimation variance primarily in the extrinsic parameters along the user's line of sight. This effect has not been observed in any of the related work on OST HMD calibration cited in this paper as their respective experiment design and validation were either camera-aided or otherwise done in the absence of such noise. It can however be mitigated, and calibration procedure can be simplified, by distributing the correspondence points over greater depth rather than simply adding more points. We are currently investigating this approach empirically by calibrating an OST HMD with correspondence points at varying depth. Future work involves modeling virtual computer graphics objects using the estimated parameters and allowing subjects to measure perceived registration error between virtual and real objects to find thresholds for acceptable calibration quality.
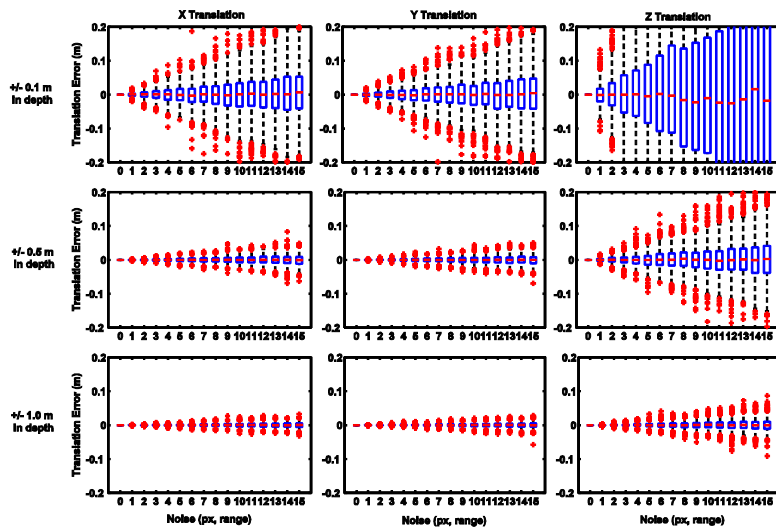
Correspondence Points	Depth distribution (m, +/-)	Human noise (px, range)	X Translation Error (m)	Y Translation Error (m)	Z Translation Error (m)	Principal Point	Focal Length	Orientation (deg.)
9	0.1	1	0,013	0,012	0,092	39,040	8,477	0,458
		5	0,070	0,057	0,464	197,895	41,489	2,260
		10	0,127	0,120	1,003	418,827	82,516	4,613
	0.5	1	0,003	0,002	0,015	6,359	3,981	0,255
		5	0,014	0,013	0,070	30,416	19,669	1,268
		10	0,027	0,026	0,151	65,345	37,360	2,546
	1.0	1	0,001	0,001	0,004	1,821	1,447	0,000
		5	0,007	0,007	0,019	9,317	7,594	0,525
		10	0,014	0,012	0,035	17,011	14,909	1,005
20	0.1	1	0,006	0,006	0,037	15,978	3,709	0,162
		5	0,034	0,030	0,172	72,723	19,588	0,958
		10	0,069	0,056	0,341	143,729	38,577	1,668
	0.5	1	0,001	0,001	0,005	2,245	1,363	0,000
		5	0,007	0,007	0,026	11,671	6,879	0,485
		10	0,014	0,013	0,050	23,608	14,325	0,989
	1.0	1	0,001	0,001	0,001	0,756	0,577	0,000
		5	0,003	0,003	0,007	3,470	2,752	0,162
		10	0,007	0,007	0,014	7,461	5,632	0,458
81	0.1	1	0,003	0,003	0,013	5,389	1,538	0,000
		5	0,014	0,014	0,070	29,495	7,804	0,324
		10	0,028	0,028	0,134	57,025	15,163	0,704
	0.5	1	0,001	0,001	0,002	0,860	0,487	0,000
		5	0,003	0,003	0,009	4,257	2,568	0,162
		10	0,006	0,006	0,020	8,944	5,280	0,362
	1.0	1	0,000	0,000	0,000	0,281	0,212	0,000
		5	0,001	0,002	0,002	1,447	1,025	0,000
		10	0,003	0,003	0,005	2,925	2,024	0,162

Table 1: Interquartile ranges for camera calibration parameters as a function of the simulation's independent variables .

REFERENCES

Adbel-Aziz, Y., Karara, H. (1971), Direct Linear Transformation from Comparator to Objecct Space Coordinates in Close-Range Photogrammetry, *ASP Symp. Close-Range Photogrammetry*, pp. 1-18.

Axholt, M., Peterson, S. D. & Ellis, S. R. (2009), Visual Alignment Precision in Optical See-Through AR Displays: Implications for Potential Accuracy, *Proceedings of the ACM/IEEE Virtual Reality International Conference*.

Cakmakci, O. & Rolland, J. (2006), Head-Worn Displays: A Review, *Journal of Display Technology*, 2(3):199-216.

Challis, J. H. & Kerwin, D. G. (1992), Accuracy Assessment and Control Point Configuration When Using the DLT for Photogrammetry, *Journal of Biomechanics*, 25(9):1053-1058.

Chen, L., Armstrong, C. W. & Raftopoulos, D. D. (1994), An Investigation on the Accuracy of Three-Dimensional Space Reconstruction Using the Direct Linear Transform Technique, *Journal of Biomechanics*, 27(4):493-500.

Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7(1):1-26.

Genc, Y., Tuceryan, M., & Navab, N. (2002), Practical Solutions for Calibration of Optical See-Through Devices, *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 169-175.

Gilson, S.J., Fitzgibbon, A.W., & Glennerster, A. (2008), Spatial Calibration of An Optical See-Through Head Mounted Display, *Journal of Neuroscience Methods, 173*, pp. 140-146.

Haralick, R. M. (1989), Pose Estimation from Corresponding Point Data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):1426-1446.

Hartley, R. I. (1997), In Defense of the Eight-Point Algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580-593.

Hartley, R. I. & Zisserman, A. (2000), *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, ISBN: 9780521540513.

Livingston, M. A., Ellis, S. R. and White, S. M. (2006), Vertical Vergence Calibration for Augmented Reality Displays,*Proceedings of the IEEE Virtual Reality Conference*, pp. 293-294.

McGarrity, E. & Tuceryan, M. (1999), A Method for Calibrating See-Through Head-Mounted Displays for AR, *Proceedings of the IEEE International Workshop on Augmented Reality*, pp. 75-84.

McGarrity, E., Genc, Y., Tuceryan, M., Owen, C. & Navab, N. (2001), A New System for Online Quantitative Evaluation of Optical See-Through Augmentation, *IEEE/ACM International Symposium on Augmented Reality*, pp. 157-166.

Nicholson, R. M. (1966), The Feasibility Helmet-Mounted Sights as A Control Device, *Human Factors*, 8:417-425.

Owen, C. B., Zhou, J. Tang, A. & Xiao, F. (2004), Display-Relative Calibration for Optical See-Through Head-Mounted Displays, *IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 70-78.

Rolland, J. P. & Fuchs, H. (2000), Optical versus Video See-Through Head-Mounted Displays in Medical Visualization, *Presence*, 9(3):287-309.

Stanney, K. M., Mourat, R. R. & Kennedy, R. S. (1998), Human Factors Isuues In Virtual Environments: A Review of the Literature, *Presence*, 7(4):327-351.

Sun, W. & Cooperstock, J. R. (2006), An Empirical Evaluation of Factors Influencing Camera Calibration Accuracy Using Three Publicly Available Techniques, *Journal of Machine Vision and Applications*, 17(1):51-67.

Sutherland, I. E. (1974), Three-Dimensional Data Input by Tablet, *Proceedings of the IEEE*, 62(4):453-461.

Tang, A., Zhou, J. & Owen, C. (2003), Evaluation of Calibration Procedures for Optical See-Through head-Mounted Displays, *IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 161-168.

Tsai, R. (1987), A Versatile Camera Calibration Technique for High-Accuarcy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Journal of Robotics and* Automation, RA-3(4):323-344.

Tuceryan, M., Greer, D. G., Whitaker, R.. T., Breen, D. E., Crampton, C., Rose, E. & Ahlers, K. H. (1995), Calibration Requirement and Procedures for a Monitor-Based Augmented Reality System, *IEEE Transaction on Visualization and Computer Graphics*, 1(3):255-273.

Tuceryan, M., Genc, Y. & Navab, N. (2000). Single Point Active Alignment Method (SPAAM) for Optical See-Through HMD Calibration for AR, *Proceedings of IEEE and ACM International Symposium on Augmented Reality*, pp. 149-158.

Verona, R. W. (1978), Head aiming/Tracking Accuracy in a Helicopter Environment, *Proceedings of Advisor Group for Aerospace Research and Development (AGARD)*, pp. 51:1-51:18, US Airforce Research Laboratory (US-AARL) P.O. Box 577, Fort Rucker, AL 36362, USA.

Wan, X. & Xu, G. (1996), Camera Parameters Estimation and Evaluation in Active Vision System, *Journal of Pattern Recognition*, 29(3):439-447.

Wells, M. J. & Griffin, M. J. (1987), A Review and Investigation of Aiming and Tracking Performance with Head-Mounted Sights, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17(2):210-221.

Zhang, Z. (2000), A Flexible New Technique for Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334.