

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

Magnus Axholt<sup>1</sup>, Martin Skoglund<sup>2</sup>, Stephen D. Peterson<sup>1</sup>, Matthew D. Cooper<sup>1</sup>,  
Thomas B. Schön<sup>2</sup>, Fredrik Gustafsson<sup>2</sup>, Anders Ynnerman<sup>1</sup>, Stephen. R. Ellis<sup>3</sup>

<sup>1</sup>Department of Science and Technology, Linköping University, Sweden

<sup>2</sup>Department of Electrical Engineering, Linköping University, Sweden

<sup>3</sup>Human Systems Interface Division, NASA Ames Research Center, CA, USA

Abstract goes here...

## 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 commonly AR refers to the human vision being overlaid with information otherwise not readily available to the user. A correct calibration is 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 [SOURCE]. Thus, vital to AR, calibration methodology is an important research area. While great achievements already have been made, there are some properties in current calibration methods for augmenting vision which do not translate from its traditional use in automated cameras calibration to its use with a human operator. This paper uses a Monte Carlo simulation of a standard direct linear transformation camera calibration to investigate how user introduced head orientation noise affects the parameter estimation during a calibration procedure of an optical see-through head mounted display.

## OVERVIEW AND RELATED WORK

### Display Techniques

The display techniques by which the human vision is augmented are commonly divided into head-slaved helmet mounted displays (HMD) and stationary head-up displays (HUD). Both can be further subdivided into video see-through (VST) and optical see-through (OST) devices. In the case of VST, the user's visual impressions are relayed by a video camera where the signal is composited with a data layer before it is presented to the user on an opaque screen. VST carries several benefits in terms of ease of calibration: A frame from the camera's auxiliary video signal can be routed through a processing segmentation and corner detection procedure to simultaneously acquire hundreds of correspondence points with precision only limited by image blur and pixel quantization (Hartley & Zisserman, 2000). Said auxiliary signal can also be used to objectively estimate the calibration quality. Other benefits relate to the fact that the optics of the

eye and the camera are separated: In the monoscopic case the requirement for correct screen positioning relative to the user's eye is relaxed as the eye's view vector and the screen's center, the principal point, need not to be aligned. While harder to calibrate, OST has the benefit that it does not reduce the human vision to the capabilities of the camera in terms of acuity, dynamic range, and field of view (FOV) as the data layer is composited through an optical combiner, usually a slanted half-silvered mirror and a system of lenses. The direct optical path between the real world and the eye also provides the safety to revert to non-augmented vision in case of system failure. OST devices are however harder to calibrate for the exact opposite reasons as described above: The optical system of the screen must be carefully aligned with that of the user's eye, which is a high demand when the calibration is dependent on the user's subjective judgment in the absence of an auxiliary video signal. The paper focuses on calibration methods involving OST HMD for applications where human vision cannot be substituted with a camera view.

### Computer Models

To be able to correctly merge the real and virtual world during user interaction with a dynamic scene, the AR system maintains a computer model to represent the location of real and virtual objects. The spatial relationships are normally modeled using linear transformation matrices containing rotations as well as translations through the use of homogeneous coordinates. As 4-by-4 matrices, they can be aggregated through multiplication to symbolize the traversal through local coordinate systems to describe the exact location of surrounding objects relative to the user's eye, see Figure 1 and (9). At the eye point the user's view is traditionally modeled as a pinhole camera. Assuming no radial distortion in the optics this subsystem can be modeled as two additional matrices holding extrinsic and intrinsic camera parameters which together conveniently can be multiplied into the matrix aggregation. In the total matrix aggregation there is a subset of matrices that update dynamically due to user and object movement, and a subset that remain static where the spatial relationship between the objects is fixed. By singling out the subset of static matrices and aggregating them separately, the calibration procedure becomes the task of populating the elements of the aggregated matrix 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 between the tracker sensor and the eye is an example of such a measurement. The casing of the tracker hides the exact location of its sensor origin. Similarly the skull and lobe hides the internal workings of the eye. Therefore the calibration method must estimate this type of measurements implicitly with the help of other measurable distances in the model.

### Standard Camera Calibration Procedure

Common calibration procedures usually spring from camera decomposition (Ganapathy, 1984), 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,  $\mathbf{p}_w$ , and points of known pixel coordinates on the screen,  $\mathbf{p}_s$ , are used to determine a 3-by-4 camera matrix  $\mathbf{T}$  (1). This corresponding point data can be expressed as a system of homogeneous linear equations (3) in which  $\mathbf{x}$  is a vector of the elements in matrix  $\mathbf{T}$ , and  $\mathbf{A}$  is the result of matrix multiplication (2) when the perspective divide,  $\mathbf{w}$ , has been substituted for, see Appendix A in (Sutherland, 1974) for details. The values in  $\mathbf{p}_w$  and  $\mathbf{p}_s$  are usually normalized to a common order of magnitude to condition the matrix  $\mathbf{A}$  to reduce 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, but at least six points are needed to solve for the 12 entries in  $\mathbf{T}$ . Generally more points are gathered to further mitigate the effect of noise. This results in  $\mathbf{A}$  being non-square, thus prompting the use of a More-Penrose pseudo inverse by which  $\mathbf{A}$  is factorized into two bases ( $\mathbf{U}, \mathbf{V}$ ) and a diagonal matrix ( $\mathbf{\Sigma}$ ) using singular value composition (SVD) (4). The eigenvalue calculations in SVD effectively perform a least square approximation. Thus the last column in the base matrix  $\mathbf{V}$ , that by convention corresponds to the smallest singular value in  $\mathbf{\Sigma}$ , can be interpreted as the calibration matrix  $\mathbf{T}$  (5) that projects landmark coordinates onto screen coordinates with the smallest residual between screen points and corresponding landmarks as seen by the user.

$$\mathbf{p}_s = \mathbf{T}\mathbf{p}_w \quad (1)$$

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

$$\mathbf{A}\mathbf{x} = 0 \quad (3)$$

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (4)$$

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

At this point, some calibration procedures adjust for non-linear lens effects by using matrix  $\mathbf{T}$  as initial values for a Levenberg-Marquardt (LM) optimization procedure further

refining  $\mathbf{T}$  (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 importance of outliers (Hartley & Zisserman, 2000).

The matrix  $\mathbf{T}$  can further be divided into extrinsic,  $\mathbf{R}\mathbf{t}$ , and intrinsic,  $\mathbf{K}$ , camera parameters with RQ-decomposition using Givens rotations. At this stage the offset from the tracker sensor origin to the center point of the eye is accessible through  $\mathbf{t}$  (7), and  $\mathbf{R}$  describes the rotation of the screen.  $\mathbf{K}$  gives focal length,  $\alpha$ , which in turn holds the distance to screen (in meters),  $\mathbf{f}$ , if the pixel ratio (pixels per meter) is known (8). With knowledge of screen resolution, this information also gives the theoretical FOV. The practical FOV is however dependent on how well the eye aligns with the exit pupil defined by the principal point,  $(\mathbf{x}_0, \mathbf{y}_0)$  [SOURCE].

$$\mathbf{T} = \mathbf{K}[\mathbf{R}\mathbf{t}] \quad (6)$$

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (7)$$

$$\alpha = m * f \quad (8)$$

This procedure seems well documented and straight forward in the case of camera calibration, but how about its use with a human operator for the purpose of calibrating OST HMD?

### 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 can be said to 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 eye point in space. A solution to this challenge is to use only one landmark instead of six, 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  $\mathbf{t}$  (7).

**b)** Human alignment precision in an OST HMD is predominantly dependent on head rotation precision which has been reported to be  $0.13^\circ$  (Nicholson, 1966),  $0.9^\circ$  (Verona, 1978), and  $0.04^\circ$  (Wells & Griffins, 1987) standard deviation. Most recently Axholt et al. reported  $0.25^\circ$  precision for 12 standing subjects using an OST HMD with VGA resolution

through  $28^\circ$  by  $37^\circ$  FOV which effectively converts to an average of 4.3 pixel boresight misalignment. However, when the effect of head translation was removed, subjects exhibited sub-pixel alignment precision on the order of human visual acuity ( $0.016^\circ$ , 0.2 pixel). Hence head rotations are thought to compensate postural sway (Axholt et al, 2009), but 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) In the presence of noise, calibration quality generally improves with the number of correspondence points (Hartley & Zisserman, 2000). 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 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. However, together with Genc et al. (Genc et al, 2002), all three works rely on the fact that 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  $(x_0, y_0)$  also need to be adjusted between each user session. In conclusion, 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 a HDM OST.

## Problem Statement

Under the hypothesis that OST HMD must be calibrated by a human operator, through a process that cannot be camera aided, we wish to investigate the limits of existing calibration procedures with respect parameter estimation robustness to human alignment noise in order to understand, and possibly compensate, for its limitations..

## METHODS

### Modeling

In a first step a standard DLT camera calibration according to SPAAM (Tuceryan & Navab, 2000) was modeled using an object oriented node tree structure implemented in MATLAB R2006b to manage the relationship

between objects, transformation and projection matrices as depicted in Figure 1 and described in (9). A frustum object was set to roughly model a Kaiser ProView 50ST with VGA resolution,  $28^\circ$  by  $37^\circ$  FOV and a projection plane located 0.05 m in front of the user’s left eye. Each simulation iteration

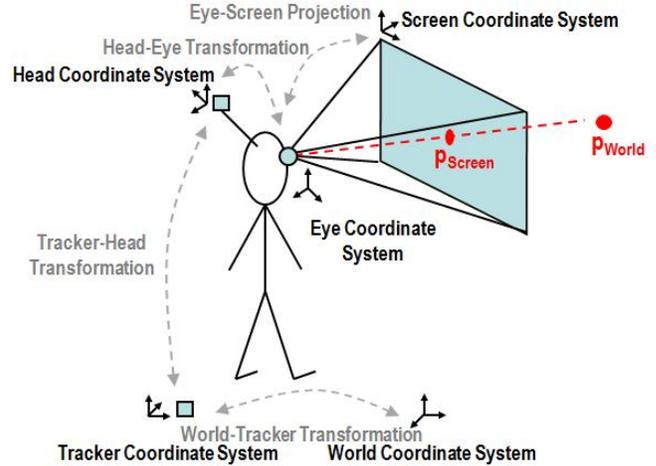


Figure 1: Objects, transformations and projections used in simulation

started by randomizing the user’s head rotation and position in front of the landmark,  $p_{world}$ . Then a Nelder-Mead (NM) optimization algorithm estimated the optimal orientation of  $T_{T-H}$  to minimize the distance between a crosshair,  $p_{screen}$ , and the landmark as projected on the screen according to (9).  $T_{H-E}$  simulates the offset between the tracker sensor and user’s eye, and thereby constitutes the extrinsic parameters of a standard camera calibration procedure. Similarly  $P_{E-S}$  is the intrinsic parameters. Since both are static when the user is wearing the HMD, they may be multiplied together into  $T_{CAL}$  (10) which corresponds to the camera calibration matrix  $T$  (1) to be estimated.

$$p_{screen} = T_{W-T} * T_{T-H} * T_{H-E} * P_{E-S} * p_{world} \quad (9)$$

$$T_{W-T} = I, T_{H-E} * P_{E-S} = T_{CAL} \quad (10)$$

$$p_{screen} = T_{T-H} * T_{CAL} * p_{world} \quad (11)$$

In a second step, the simulation was extended with a normalization procedure as described in (Hartley, 1997), and a LM optimization procedure implementing quadratic, pseudo-Huber, and Blake-Zisserman cost functions further refining  $T$  (Hartley & Zisserman, 2000). It was at this point the effect of human noise in a standard camera calibration became noticeable as the LM search for an appropriate  $T$  would fail at surprisingly low noise levels regardless of cost function, which in turn prompted further changes to the model for investigation.

Thus, in a third step, the NM optimization, which was implemented to model the human operator limitation a), using SPAAM, was excluded for the benefit of faster calculations enabling a Monte Carlo simulation. With no drawback to the final result, the frustum was thus made stationary and pointed towards not one but a collection of landmarks whose ideal projections were precomputed using (9). Also the LM step

was excluded, thereby solely studying the effect of human noise on a standard DLT camera calibration.

### Simulation

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, Weibull} parameterized using range or 99.9% probability, **3)** human noise magnitude defined as pixel range {0,0.2,0.4,0.6,0.8,1,2,3,4,5,6} introduced as permutations of  $p_{screen}$  in random (white) direction. The simulation was run with 1,000 iterations per combination of independent variables to collect the dependent variables  $T_{CAL}$  and normalized calibration error (NCE) (Weng et al, 1992).

### RESULTS

Figure 3 shows how the translation component of  $T_{H-E}$ , i.e. the estimation of the relative location between head tracker sensor and user’s eye,  $t$ , varies based on user alignment noise and number of correspondence points. Notably, the estimation error is particularly large along the user’s line of sight (z axis).

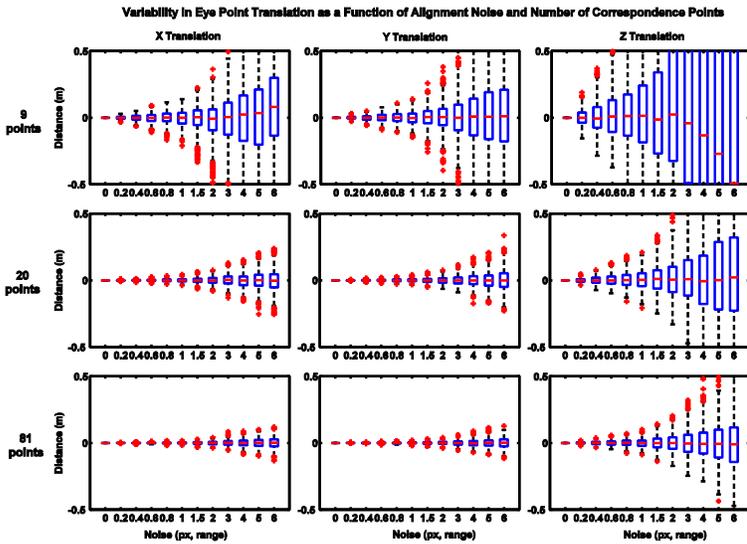


Figure 3 Data from Monte Carlo simulation illustrating how variability in eye point translation vector  $t$  increases with alignment noise. The data in the plot was collected in a right-handed coordinate system with negative  $z$  along the user’s viewing direction, using a fixed range model for 1,000 iterations. Boxes denote quartiles. Outliers, marked with red + signs, are defined as 1.5 times interquartile range.

The boxes in Figure 3 represents the parameters’ range for which there is a 50% probability of repetability. Table 1 presents the numerical values for  $T_{H-E}$  relative the four noise models. E.g. a user exhibiting a head orientation noise equivalent to six pixel fixed range would have to align 81 points for a 50% probability of the eye point being correctly estimated within 13 cm ( $=0.259/2$ ) from its true location (\*).

	X translation		Y translation		Z translation	
	20 p	81 p	20 p	81 p	20 p	81 p
<b>Fixed *</b>	0.099	0.052	0.103	0.051	0.548	0.259
<b>White</b>	0.059	0.030	0.051	0.030	0.324	0.137
<b>Gaussian</b>	0.034	0.017	0.030	0.017	0.1731	0.089
<b>Weibull</b>	0.017	0.009	0.016	0.009	0.0929	0.041

Table 1: Interquartile range (in meters) for eye point translation as a function of 20 or 81 correspondence points and four various user alignment noise models of 6 pixel range magnitude.

Similar results are found in the rotational component of

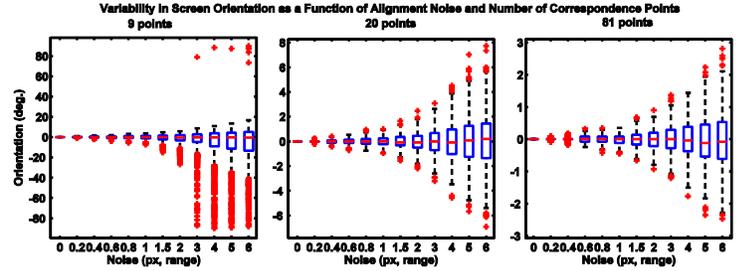


Figure 2 Variability in screen orientation  $R$  as a function of alignment noise and number of correspondence points based on a fixed range noise model for 1,000 iterations. Orientation is reported as the dot product cosine angle between the original screen orientation and estimated screen orientation after simulated calibration. Note the varying scale on the y axis.

$T_{H-E}$ , i.e. the orientation of the screen relative to user’s eye,  $R$ . Figure 2 shows that there is a 50% probability of estimating the screen rotation accurately within  $\pm 1^\circ$  if the user can correctly align 81 correspondence points within a range of six pixels. Similar to the trend in Table 1, the 50% probability accuracy is somewhat better for the other noise models.

The variability of the intrinsic parameters is shown in Table 2. Through division with the pixel ratio (8) the values the measurements in which there is a 50% probability to find the estimated frustum apex and subsequently the eye point.

	Principal Point X Focal Length X		Principal Point Y Focal Length Y	
	20 p	81 p	20 p	81 p
<b>Fixed</b>	0.013	0.007	0.014	0.007
<b>White</b>	0.008	0.004	0.008	0.003
<b>Gaussian</b>	0.004	0.002	0.004	0.002
<b>Weibull</b>	0.002	0.001	0.002	0.001

Table 2: Interquartile ranges (in meters) for the intrinsic parameters as a function of 20 or 81 correspondence points and four various user alignment noise models of 6 pixel range magnitude.

### DISCUSSION

This paper was justified because investigations on camera calibration accuracy usually models lower levels of noise than those found in a human operator (Sun & Cooperstock, 2006). To bridge the data gap between camera and human performance a controlled environment, such as a Monte Carlo simulation, was motivated.

The results showed that, even with a fair amount of alignments, a standard camera calibration procedure is unlikely to estimate user's eye point with repeatable accuracy in the presence of head rotation noise levels found in literature. However, as the predominant error is along the user's viewing direction, there may be some instances in which the performance of the standard camera calibration may be acceptable: If the eye point is erroneously estimated along the user's line of sight, the registration error will manifest itself as a scaling error as long as the user looks straight ahead. However, as the head is turned, the frustum will pivot around the erroneous eye point and introduce a lateral registration error as the frustum is shifted sideways. The current findings should be taken into account for applications with a fixed user, i.e. in vehicles and observations posts where the user explores primarily with head rotations.

To determine if alignment distribution had an effect on parameter robustness, one fixed range and three noise models were implemented and set to have range as a common parameter. While white and Gaussian noise increase their spread with increasing range, the Weibull distribution, popular for modeling cluster concentrations, shifts its mode towards zero for increasing range. This should be kept in mind when comparing the results.

#### ACKNOWLEDGEMENT

Magnus Axholt and Stephen D. Peterson were supported by PhD scholarships from the Innovative Research Programme at the EUROCONTROL Experimental Centre, Brétigny-sur-Orge, France. These authors were also supported through the NASA Grant NNA 06 CB28A to the San José State University Research Foundation. The experiment was conducted at NASA Ames Research Center in Mountain View, CA (corresponding author: Stephen R. Ellis, sellis@mail.arc.nasa.gov).

#### REFERENCES

Abdel-Aziz, Y., Karara, H. (1971), Direct Linear Transformation from Comparator to Object 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*.

Genk, 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.

Ganapathy, S. (1984), Decomposition of Transformation Matrices for Robot Vision, *Proceedings of the IEEE International Conference on Robotics and Automation*, pp.130-139.

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*, 2<sup>nd</sup> Edition, Cambridge University Press, ISBN: 9780521540513.

Milgram, P. & Kishino, A. F. (1994), Taxonomy of Mixed Reality Visual Displays, *IEICE Transactions on Information and Systems*, E77-D(12), pp. 1321-1329.

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.

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., & 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.

Sun, W. & Cooperstock, J. R. (2006), An Empirical Evaluation of Factors Influencing Camera Calibration Accuracy Using Three Publicly Available Techniques, *Machine Vision and Applications*, 17(6):51-67.

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

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

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.

Weng, J., Cohen, P. & Herniou, M. (1992), Camera Calibration with Distortion Models and Accuracy Evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965-980.

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