Research on Calibration of Optical See-Through Head-Mounted Display

Hailong Yu and Guozhong Liu

Department of Instrument Science and Optoelectronic Engineering, Beijing Information Science & Technology University, No.12 east qinghexiaoying road, haidian district, Beijing, China
Email: liuguozhong@bistu.edu.cn

Abstract. In order to successfully apply augmented reality technology and provide users with immersive 3D experience of augmented reality, it is necessary to calibrate the optical see-through head-mounted display (OST HMD) of augmented reality equipment. With the help of the video tracking system, a new display calibration method was proposed, in which the data of the video tracker were taken as input and the 3D coordinates of the visual virtual model were taken as output to find the projection transformation matrix between the 3D coordinates of the virtual space and the 3D coordinates of the real space. In addition, in order to obtain the video tracker data conveniently, the 3D calibration target with checkerboard was designed. The display calibration experiment was carried out on Microsoft Hololens platform. The experimental results show that the displacement error of the reprojection of the virtual object is 2.7 cm within the active range of 80 cm × 80 cm × 100 cm of the tracker.

1. Introduction

With the progress in engineering technology, more and more augmented reality devices choose OST HMD as their output medium. In the augmented reality application system, users can see the virtual objects superimposed on the real scene when they see the real world scene. In order to achieve good augmented reality experience and application function in augmented reality application, it is necessary to calibrate the optical perspective helmet display for augmented reality devices. The purpose of the display calibration process is to calculate the transformation relationship between the real world coordinates and the virtual space coordinates. So far, display calibration is still a challenging problem.

As commercial interest in augmented reality technology has increased in recent years, new augmented reality headsets with stereoscopic display capabilities such as Hololens, MagicLeap One and Meta Two have flooded the market. All of them can provide users with 3D augmented reality experience. In order to have an immersive 3D augmented reality experience, they need to calibrate the 3D display. Augmented reality in stereoscopic display device, with the aid of video tracker designed a three dimensional space to the three-dimensional space of the new display calibration method, the real world of the coordinates of the point under the tracking camera as input data, the corresponding virtual point coordinates in the virtual space as the output data, the calculation of the real world coordinate system to the virtual space coordinate system transformation.

In the following part, firstly, the principle of OST HMD calibration technology is introduced. Secondly, the necessary equation of stereo display calibration is derived. Finally, the calibration method was tested with Hololens as the augmented reality stereo display platform, and the average displacement error of the corresponding virtual object was calculated, and the calibration method was evaluated by analyzing the error.
2. Calibration Principle of Optical See-Through Head-Mounted Display

2.1. Calibration System

In order to achieve accurate virtual and real alignment, the pose of the virtual rendering camera in the real world should be acquired first, and the position direction information obtained by the augmented reality system can be used to calculate the corresponding position displayed by the virtual object \cite{1}. The attitude of virtual rendering camera is usually measured by the tracking system, and the more accurate the information is, the higher the precision of virtual and real alignment will be. Therefore, in order to achieve accurate virtual and real alignment, the tracking system needs to be calibrated first \cite{2}. At present, most of the calibration systems of OST HMD are composed of target, head tracker and virtual rendering camera. Virtual rendering camera is an integrated imaging system composed of human eye and optical transmission helmet display. Camera video tracker is the most commonly used tracking system tracker. As shown in figure 1, it is a typical video tracker optical transmission helmet display calibration system. It includes two imaging processes, real scene and virtual object. If the virtual rendering camera composed of human eyes is precisely synchronized with human eyes, the two imaging processes will also achieve precise alignment on the optical transmission helmet display \cite{3}.

![Figure 1](image-url)

Figure 1. Coordinate system transformation of OST HMD calibration system. Where, camera coordinate system: \(F_a(X,Y,Z)\), image screen coordinate system: \(F_d(U,V)\), virtual camera coordinate system: \(F_v(X,Y,Z)\), display screen coordinate system: \(F_d(U,V)\), world coordinate system: \(F_w(X,Y,Z)\).

The origin of the world coordinate system is fixed at a vertex on the target. The homogeneous coordinates of any point in the world coordinate system are defined as \(P_w=[x_w, y_w, z_w, 1]^T\).

The transformation relation between 3D-3D coordinate system is defined as rigid transformation, while the 3D-2D optical imaging relation is perspective transformation. The rigid transformation from any point \(P_a=[x_a, y_a, z_a, 1]^T\) in coordinate system \(F_a(X,Y,Z)\) to coordinate system \(F_b(X,Y,Z)\) can be written as:

\[
P_b = T_{b \rightarrow a} P_a = \begin{bmatrix} r_{b \rightarrow a} & t_{b \rightarrow a} \\ 0 & 1 \end{bmatrix} P_a \tag{1}
\]

Where, \(r_{b \rightarrow a}\) and \(t_{b \rightarrow a}\) respectively represent the rotation part and the translation part of the transformation from coordinate system \(F_a(X,Y,Z)\) to coordinate system \(F_b(X,Y,Z)\).

The perspective transformation from any point \(P_v=[x_v, y_v, z_v, 1]^T\) in camera coordinate system \(F_v(X,Y,Z)\) to image coordinate system \(F_d(U,V)\) can be written as:

\[
\rho P_v = P_{c \rightarrow b} P_v = \begin{bmatrix} p_{c \rightarrow b} & \theta_{3 \times 1} \end{bmatrix} P_v
\]

Where, \(\theta_{3 \times 1}\) is the zero matrix of 3\(\times\)1, \(\rho\) is the proportional coefficient, and \(p_{c \rightarrow b}\) is the internal parameter of the camera, including the focal length \(f_u, f_v\) and the center point \(v_0, v_0\).

Therefore, the effective mapping from any point \(P_v=[x_v, y_v, z_v, 1]^T\) in coordinate system \(F_v(X,Y,Z)\) to coordinate system \(F_d(U,V)\) of the virtual camera is transformed into:
\[
\rho P_c = P_{c \to b} P_b = P_{c \to b} T_{b \to a} P_a = M_{c \to a} P_a
\]  
(3)

Where, \( M_{c \to a} = P_{c \to a} T_{a \to d} \) is the projective matrix between coordinate system \( F_a(X,Y,Z) \) and camera image plane coordinate system \( F_c(U,V) \).

The coordinate transformation of the helmet display calibration system of optical perspective is shown in table 1.

| Transformation | Description | Features |
|---------------|-------------|----------|
| \( T_{t \to w} \) | \( F_v(X,Y,Z) \to F_v(X,Y,Z) \) | Rigid, variation |
| \( P_{d \to v} \) | \( F_v(X,Y,Z) \to F_d(U,V) \) | Perspective, fixation |
| \( T_{v \to t} \) | \( F_i(X,Y,Z) \to F_i(X,Y,Z) \) | Rigid, fixed |
| \( T_{v \to w} \) | \( F_v(X,Y,Z) \to F_v(X,Y,Z) \) | Rigid, variation |
| \( P_{d \to v} \) | \( F_i(X,Y,Z) \to F_d(U,V) \) | Perspective, fixation |

2.2. Calibration System Calculation Model

The virtual rendering camera conversion relationship between the actual scene and the virtual scene in the OST HMD calibration system is as follows:

\[
\rho P_d = M_{d \to v} P_w = P_{d \to t} T_{t \to w} P_v
\]  
(4)

We assume that the position of the human eye and the virtual camera is always exactly aligned in the HMD calibration system. By observing figure 1, it can be found that the position of human eyes is connected with the target of the world coordinate system through the imaging process of actual scene and virtual image, and the rigid transformation of these two imaging processes should also be equal, thus:

\[
T_{t \to w} = T_{v \to t} T_{v \to w}
\]  
(5)

Combine (4) and (5) to get:

\[
\rho P_d = P_{d \to t} T_{v \to t} T_{v \to w} P_v = M_{d \to t} P_v
\]  
(6)

Where, \( P_t \) is the coordinate point in the tracker coordinate system, \( M_{d \to v} = P_{d \to t} T_{v \to w} \) is the effective projective matrix for the imaging of the target point in the tracker coordinate system on the virtual camera screen. In the effective projective matrix, \( P_{d \to v} \) contains the parameters of OST HMD display calibration \( \{f^d_x, f^d_y, u^d_0, v^d_0, \gamma^d \} \). Since the tracking camera is fixed on the augmented reality display, the fixed transformation relation between the tracking camera coordinate system and the virtual rendering coordinate system is \( T_{v \to t} \), which includes two parts of rotation and translation. \( T_{v \to w} \) can be determined by tracker calibration, so it is easy to calculate \( P_v \). Under the measurable conditions of \( P_t \) and \( P_v \), by measuring the multi-point correspondence between \( P_t \) and \( P_v \), the OST HMD calibration is transformed into the mathematical problem of \( M_{d \to v} \).

3. Calibration Method of Stereo Display

In most OST HMD calibration methods, users collect 2D-3D corresponding points and use linear transformation to calculate the mapping and transformation relationship from 3D world points to 2D screen points. In fact, most OST HMD can create visualized 3D virtual objects in front of the user’s eyes and provide access to the 3D virtual objects. From this point of view, virtual coordinates can be represented in three-dimensional coordinates of space rather than two-dimensional coordinates of the screen. Therefore, the mapping model of the stereo display calibration system becomes a registration process from 3D real space to 3D virtual space.
We assume $P_t, P_d \in \mathbb{R}^3$, and we solve the estimated mapping matrix $M_{d \leftarrow t}$ through a set of corresponding points $(P_t, P_d)_{i=1,\ldots,n}$. More specifically, $P_t$ is obtained through the tracking system measurement, and the information of the corresponding point $P_d$ is predefined and visualized on the OST HMD.

We further assume that $M_{d \leftarrow t}$ is linear, and based on the linear hypothesis, we assume that the transformation $M_{d \leftarrow t}$ we are solving for is a general case of the perspective transformation. Therefore, we use the perspective transformation model to solve the mathematical expression of the transformation of perspective:

$$
M_{d \leftarrow t} = 
\begin{bmatrix}
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{31} & m_{32} & m_{33} & m_{34} \\
m_{41} & m_{42} & m_{43} & m_{44}
\end{bmatrix}
$$

During the calibration test, the calibration target cube in the world coordinate system is aligned with the virtual cube generated by the computer. The tracking camera is calibrated to determine the rigid transformation $T_{t \leftarrow w}$ between the world coordinate system and the tracking camera coordinate system, then the homogeneous coordinate of the coordinate point in the world coordinate system in the tracking coordinate system is expressed as $P_t = [x_t, y_t, z_t, 1]^T$. The homogeneous coordinate of the corresponding image point in the virtual camera coordinate system is $P_d = [u_d, v_d, w_d, 1]^T$, then:

$$
K \begin{bmatrix} u_d \\ v_d \\ w_d \\ 1 \end{bmatrix} = M_{d \leftarrow t} \begin{bmatrix} x_t \\ y_t \\ z_t \\ 1 \end{bmatrix}
$$

Where, $K$ represents the proportionality coefficient. Use $m_{i1}^t, m_{i2}^t, m_{i3}^t, m_{i4}^t$ to represent $M_{d \leftarrow t}$, then:

$$
\begin{align*}
(m_{i1} - u_d m_{i4}) \cdot P_t &= 0 \\
(m_{i2} - v_d m_{i4}) \cdot P_t &= 0 \\
(m_{i3} - w_d m_{i4}) \cdot P_t &= 0
\end{align*}
$$

The projection matrix $M_{d \leftarrow t}$ has 15 unknown parameters in the general perspective transformation model. According to equations (8) and (9), three independent constraint equations can be determined for each reference point. For 3d perspective transformation, the number of theoretical minimum reference points is only 5. In general, the number of corresponding points will be greater than 5. In this case, it is necessary to solve the overdetermined equation of the corresponding target point. The overdetermined equation can be solved to the approximate solution through the nonlinear optimization of the least square method, so as to calculate the transformation matrix of the helmet display calibration of optical perspective.

4. Calibration Experiment and Result Analysis of Stereo Display

4.1. Stereo Display Calibration Experiment
We used Microsoft HoloLens as an optical perspective helmet stereo display for the experiment. In this case, the HoloLens embedded front-facing RGB camera as a video tracker. The alignment point is expressed in the tracking coordinate system, and its corresponding virtual point is expressed in the holographic display coordinate system. The checkerboard method was used to calibrate the Hololens front-facing camera, and the geometric transformation $T_{w \leftarrow t}$ between the world object coordinate system and the tracking camera coordinate system was determined, so as to calculate the coordinate $P_t$. 

4
of the alignment point in the tracking coordinate system. Finally, the point set \( \{ P_i, \ldots, P_t \} \) and \( \{ P_i, \ldots, P_{t+d} \} \), \( i = 1, \ldots, n \) are used for the stereo ost-hmd calibration described in section 3.

To make virtual and real alignment easier in stereo displays and provide better depth cues during alignment, we need a 3D object rather than the 2D crosscut typically used for SPAAM or the flat disk used for stereo-SPAAM. We designed a 10cm*10cm*10cm cube as the 3D calibration target, as shown in figure 2. At the same time, the corresponding virtual model was made with 3dsMax. In order to make the task of alignment easier, each face of the virtual model was painted with different colors matching the calibration target.

Before calibration, the virtual overlay cube is not properly aligned with the tracked tag, as shown in figure 3(a). First, display a virtual 3D model and try to align the virtual display model with the Angle of the real object. It is important to note that only the angular positions are measured aligned. Once it is aligned, you can see the virtual cube superimposed on the real cube in the holographic environment, as shown in figure 3(b).

Then record the image and virtual model coordinate information at this time. Next, the calibrator moves to a different angle position to complete the above operation. This process continues until 20 sets of points are collected, at which point the system notifies the user that the calibration is complete.

4.2. Analysis of Stereo Display Calibration Results

Evaluation of the optical perspective HMD calibration has been a challenge because only the wearer can see the superimposed objects produced by the calibration. To make this process more objective, some studies have used cameras rather than the user's eyes [5] and measured images taken by the "eye camera". However, the use of the camera will inevitably affect the depth information. In addition, using a camera will require accuracy to be reported in dimensionless pixels. In this experiment, the displacement error of the reprojection virtual object is used to evaluate the calibration method.

In this experiment, we used the training and testing methods studied by Ttoh et al [6] to evaluate the calibration results. Specifically, 7 additional sets of corresponding point data were collected during the calibration process, that is, 27 sets of corresponding point data were taken during the calibration. The first 20 groups of corresponding points were trained to calculate and calibrate the mapping relationship. The remaining 7 groups of corresponding points were used as test samples of the mapping matrix and were not used for the training of display calibration. Based on the display calibration mapping matrix, the average displacement error of the virtual object of the test sample is calculated, and the calibration method of the stereo display is evaluated.

The calculation formula of the mean reprojection error is:

\[
E = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(x_i - x_{im})^2 + (y_i - y_{im})^2 + (z_i - z_{im})^2}
\]  

(10)

Figure 2. 3D target calibration

Figure 3. The state of virtual and real alignment (a) virtual and real are not aligned (b) virtual and real are aligned
Where, \( x_d, y_d, z_d \) is the measured value of the virtual object and \( x_{mt}, y_{mt}, z_{mt} \) is the calculated value of the virtual object.

The experimenter wore Hololens augmented reality glasses and aligned the calibration cube and its corresponding virtual display model in the holographic display, as shown in figure 3(a) and 3(b). When the virtual and real model as shown in 3(b) is aligned, the camera is recorded. Repeat the above operation 27 times to complete the data collection of corresponding points.

Since the task of virtual and real alignment is performed by human during the experiment, it is prone to errors. When the display is calibrated with 20 sets of corresponding points, the possible outliers can be removed according to the reprojection error. In the experimental training stage, four outliers were removed according to the reprojection error, and the calibration mapping relationship was calculated by using the data of the remaining 16 alignment points. The mean and standard deviation of the reprojection error of the 16 alignment points were \((25.1\pm12.5\text{mm})\).

In the experimental test stage, the mean and standard deviation of the reprojection error of the alignment points of the 7 test samples were \((26.9\pm11.8\text{mm})\), as shown in table 2.

| Sample standard data (mm) | Sample calculation data (mm) | Displacement error (mm) |
|---------------------------|-----------------------------|------------------------|
| -375.168                  | -37.32                      | 1455.354               |
| -377.378                  | -55.2499                    | 1455.081               |
| 15.01891                  | 23.17801                    | 935.9674               |
| 4.049797                  | 12.817                      | 942.2595               |
| -67.8124                  | -681.599                    | 1016.977               |
| -82.2879                  | 192.197                     | 1000.997               |
| 227.731                   | 9.431635                    | 986.1048               |
| 35.55547                  | 1303.74                     | 29.79006               |
| 1335.029                  | 52.65612                    | 12.817                 |
| -170.67                   | 6.101996                    | 1375.605               |
| -181.888                  | -35.2796                    | 1301.248               |
| -159.115                  | -49.2503                    | 1340.078               |
| 15.01891                  | 23.17801                    | 935.9674               |
| 4.049797                  | 12.817                      | 942.2595               |
| -67.8124                  | -681.599                    | 1016.977               |
| -82.2879                  | 192.197                     | 1000.997               |
| 227.731                   | 9.431635                    | 986.1048               |
| 35.55547                  | 1303.74                     | 29.79006               |
| 1335.029                  | 52.65612                    | 12.817                 |
| -170.67                   | 6.101996                    | 1375.605               |
| -181.888                  | -35.2796                    | 1301.248               |
| -159.115                  | -49.2503                    | 1340.078               |
| 15.01891                  | 23.17801                    | 935.9674               |
| 4.049797                  | 12.817                      | 942.2595               |
| -67.8124                  | -681.599                    | 1016.977               |
| -82.2879                  | 192.197                     | 1000.997               |
| 227.731                   | 9.431635                    | 986.1048               |
| 35.55547                  | 1303.74                     | 29.79006               |
| 1335.029                  | 52.65612                    | 12.817                 |
| -170.67                   | 6.101996                    | 1375.605               |
| -181.888                  | -35.2796                    | 1301.248               |
| -159.115                  | -49.2503                    | 1340.078               |
| 15.01891                  | 23.17801                    | 935.9674               |
| 4.049797                  | 12.817                      | 942.2595               |
| -67.8124                  | -681.599                    | 1016.977               |
| -82.2879                  | 192.197                     | 1000.997               |
| 227.731                   | 9.431635                    | 986.1048               |
| 35.55547                  | 1303.74                     | 29.79006               |
| 1335.029                  | 52.65612                    | 12.817                 |
| -170.67                   | 6.101996                    | 1375.605               |
| -181.888                  | -35.2796                    | 1301.248               |
| -159.115                  | -49.2503                    | 1340.078               |
| 15.01891                  | 23.17801                    | 935.9674               |
| 4.049797                  | 12.817                      | 942.2595               |
| -67.8124                  | -681.599                    | 1016.977               |
| -82.2879                  | 192.197                     | 1000.997               |
| 227.731                   | 9.431635                    | 986.1048               |
| 35.55547                  | 1303.74                     | 29.79006               |
| 1335.029                  | 52.65612                    | 12.817                 |

The mean and standard deviation of the reprojection errors of X, Y and Z axes are \((11.8\pm6.7\text{mm})\), \((9.5\pm4.4\text{mm})\), and \((19.5\pm12.7\text{mm})\), as shown in figure 4. Where, the xy plane is perpendicular to the user view and the z-axis is parallel to the user line of sight, indicating the alignment depth.

![Figure 4. Displacement error of virtual object experiment](image)

It can be seen from the evaluation results that the depth alignment error is the largest among the three directions of the mean displacement error and plays a decisive role in the mean displacement error. This is because visually the eye is not as sensitive to depth information as the xy plane, and depth alignment is difficult. In addition, in order to prevent users from viewing the virtual model in a close distance for a long time, the virtual model of augmented reality glasses can only be seen when it is more than 0.8m away from users. Therefore, the relative distance in the depth direction is also the furthest in the calibration experiment, which further increases the depth alignment error. In addition, the monocular camera tracking system has a low accuracy in the depth direction, and the mark-based
optical tracking research also shows that the depth direction is the most susceptible to the tracking results [7].

5. Summary
In this study, an optical perspective helmet display calibration system is designed for stereo display. With the help of video tracker, the transformation relationship between the real coordinate system and the virtual space coordinate system is calculated, and the virtual object is superimposed onto the real scene accurately. Video tracker carried out the calibration system experiment in the range of 80cm×80cm×100cm. The experimental results show that the displacement error of the virtual object in the calibration range is 2.7cm. In the future, our goal is to design an asymmetric calibration target, to solve the problem of manual virtual and real alignment difficulties, in order to reduce the error of human factors in the calibration system.

6. Acknowledgments
This work was supported by Science and Technology Program of State Grid Corporation of China (5222JZ170022R).

7. References
[1] Reichelt S, Schwerdtner A, Futterer G 2010 “holographic 3-D displays-electro-holography within the grasp of commercialization” INTECH Open Access Publisher
[2] Kato H and Billinghurst M 1999 “marker tracking and HMD calibration for a video-based augmented reality conferencing system” in IEEE and ACM Intern. Workshop on Augmented Reality (IWAR) pp 85–94
[3] Tsai R Y 1987 “a versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf T V cameras and lenses [J]” IEEE Journal of Robotics and Automation vol 3 chapter 4 pp 323-34
[4] Bin L, Yongtian W, Yue L 2009 Journal of Computer-Aided Design & Computer Graphics vol 21 chapter 4 pp 432-441
[5] Makibuchi N Kato H and Yoneyama H 2013 “vision-based robust calibration for optical see-through head-mounted displays” In 20th IEEE International Conference on Image Processing (ICIP) 2177–2181
[6] Itoh Y and Klinker G 2014 “interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization” In IEEE Symposium on 3D User Interfaces 75–82
[7] M. Bauer 2007 “tracking errors in augmented reality” PhD thesis, Technical University Munich