A complete calibration procedure for optical see-through head-mounted display

Weili Shi [a], Huamin Yang [b], Wang Wang [c], Zhengang Jiang [d]

School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China

[a]shiwl@cust.edu.cn, [b]yhm@cust.edu.cn, [c]ww1619525@163.com, [d]jiangzhengang@cust.edu.cn

Abstract. Augmented reality (AR) is a technology in which images generated by computer overlay the eyes' view of the real environment. AR can provide much help too many domains such as computer-aided surgery, repair and maintenance of complex engines, facilities modification, and interior design. OSTHMD (optical see-through head-mounted display) calibration is the key problem in AR (augmented reality) technology. In this document, we show the common model of camera and two processes of calibration. The first step is to find the rigid transformation between two coordinate systems. This step may be completed by Berthold's method which can find the relationship between two coordinate systems using measurements of the coordinates of a set of points in both coordinate systems. And the second step presents a self-calibration process of OSTHMD without having to evaluate the intrinsic and extrinsic parameters of camera separately. In the whole process of calibration, we don't need fix the position and pose of head because we can get the relationship in real-time between world coordinate system and mark coordinate system which is similar to the pose and position of OSTHMD through the first step. Trough above process, the complete calibration is accomplished.

Keywords: Augmented reality, optical see through head-mounted display, camera model, calibration, registration.

1. Introduction

Augmented reality (AR) is a technology in which images generated by computer overlay the eyes' view of the real environment. AR can provide much help too many domains such as computer-aided surgery, repair and maintenance of complex engines, facilities modification, and interior design.

In an AR system, the view of a real object will be augmented by covering computer generated graphics on this view. This requires that the generated graphics are accurately aligned with image of object in real world. That means the generated graphics must be relative to the position and pose of people's head in real time. All these ask for the calibration of OSTHMD.

With the development of research of OSTHMD, many researchers propose different methods to accomplish OSTHMD calibration. Deering came up with an integrated process of calibration which can achieve the high resolution virtual reality [2]. Janin made use of multi-points measured through manual to solve nonlinear equations with the optimization method [3]. Fuhrmann presented a simple and fast calibration scheme, which does not require additional instrumentation or complicated procedures [4]. Tuceryan proposed SPAAM method which is use-friendly and overcame the disadvantage of fussy operation appearing in others' papers [5]. They all made great contributions to calibration of OSTHMD.

In order to ensure the effective AR system, the real and computed-generated object must be accurately positioned relative to each other in real time. So this requires that the calibration must be achieved at the start of system. In our system, the calibration involves two steps. The first step is to estimate the position and pose of OSTHMD in world coordinate system. The second step is to accomplish the OSTHMD calibration.

The first step's task is to find the relationship between world coordinate system and mark coordinate system. The mark coordinate system is constructed by four marks attaching to the OSTHMD and these four marks should not be coplanar. The mark coordinate system is fixed. We can see that this relationship must be a rigid transformation with which the points in a coordinate system can transformed into another coordinate system. As we know, this transformation is consisted of a rotation and a translation.

Traditionally the second step is to ensure the intrinsic parameters and extrinsic parameters if just the head-mounted display places in the world coordinate system. But with the mark coordinate system attached to the OSTHMD, the extrinsic parameters aren't changed just as the intrinsic parameters. That is to say, we have no need to solve intrinsic parameters and extrinsic parameters separately. All points in world coordinate system can find their corresponding points in mark coordinate system. We just need to find the transformation with which the points in mark coordinate system can map into image plane coordinate system. And this transformation is steady which is a 3×4 matrix consisted of 12 parameters and will not calibrate again before next using system.

2. Calibration system

This calibration system is consisting of calibration point, OSTHMD (optical see through head mounted display), marks attached to OSTHMD and eye. Eye can be modeled by a camera model. There are four coordinate systems in this system as figure 1.

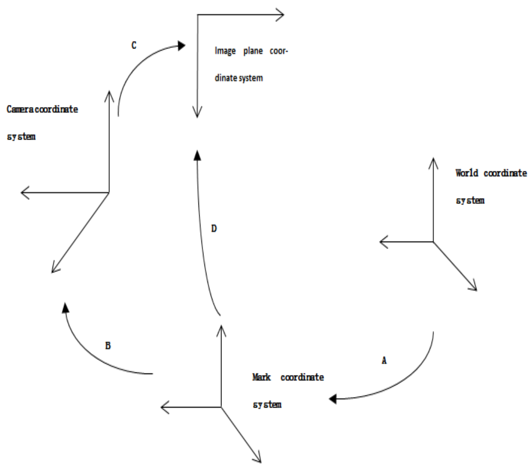

Figure 1: The coordinate systems exist in calibration system and the relationship between them

As the figure 1 showing, usual calibration wants to obtain the transformation with which the points in the world coordinate system can be mapped into 2D image coordinate. In this paper, we should make the 3D world points mapped into the mark coordinate system and then make it transformed into the image coordinate. The mark coordinate system can be constructed by attaching four marks to the OSTHMD. If we find the relationship between world coordinate system and mark coordinate system, the pose and position of head will be got and the transformation between mark coordinate system and image plane coordinate system would be fixed. That is to say, the moving of head in experiment has no effect on the transformation with which the points in mark coordinate system can map to the points in image plane coordinate system. The advantage of this method is that we can move the head freely when we calibrate the OSTHMD and also the process of calibration is simplified.

2.1Camera model

The usual model is used for the camera is the pinhole model, with which the 3D points are projected onto the 2D image plane. This model usually defines a projective imaging geometry. It is an ideal model and doesn't take the optical effects into consideration.

The camera can be modeled by a set of parameters which are divided into two categories. Some parameters belong to one category are called extrinsic parameters. Others are called intrinsic parameters. The extrinsic parameters are what define the position and pose of the camera and they are related to the world coordinate system. The intrinsic parameters define the optical properties of the camera.

The extrinsic parameters are usually represented by a rotation and a translation. These parameters usually define the pose of the camera in the world coordinate system. Let's assume a 4×4 homogeneous matrix representing the extrinsic parameters of the camera. In top left corner of the matrix there is a 3×3 matrix representing rotation and in the right there is a 3×1 vector representing translation. This rigid transformation can be written as

$$T_{pose} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The intrinsic parameters can be represented by the perspective projection and it can be written as a 3×4 projection matrix

$$T_{proj} = \begin{bmatrix} f_u & \tau & r_0 & 0 \\ 0 & f_v & c_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2)$$

So, the target of traditional camera calibration is to obtain these two matrices. In this document, we want to achieve the self-calibration of the OSTHMD. So the result we prefer to get is the product of these two matrices, we get D

$$T_{camera} = T_{proj} T_{pose} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \quad (3)$$

2.2 Transformation among coordinate systems

As figure 1 show, there are three transformations (A, B and C) that need to be estimated. A is a rigid transformation between worlds coordinate system and mark coordinate system. Because the marks are attaching to the OSTHMD, we can regard this transformation as the position and pose of OSTHMD in the world coordinate system. It can be represented by the T_{pose} in (1). ThroughT_{pose}, the points in world coordinate system can be mapped into mark coordinate system. With the change of position and pose of head, this transformation will be changed. So it is dynamic, and we should estimate it in real time. B is also a rigid transformation and it can be represented by T_{pose} too. But because the marks attached to the OSTHMD aren't moved anymore, the camera can be defined and calibrated with respect to the mark coordinate system. Therefore, this transformation is fixed and unaffected by the moving of head. C is a perspective projection, and the product of B and C which we usually called projection matrix can be represented by T_{camera} in (3). This projection matrix with respect to the mark coordinate system is what we want to obtain and if we achieve this, the OSTHMD calibration is accomplished.

To sum up, which problem we will solve is to get transformation a through tracking marks and to get transformation D through the relationship between point in the image coordinate system and the relevant point transformed by A in the world coordinate system.

3. Calibration formulation

In this section, the method which solve transformation A (the rigid transformation between two coordinate system) and D (the traditional projection matrix) will be shown. To obtain A, the method mentioned in Berthold's paper [5] is used. This method provides a closed-form solution to solve the problem of absolute orientation using unit quaternions. And it is easy to understand and use this method. To get the projection matrix D, SPAAM [1] algorithm would be a good choice. This algorithm is user-friendly and the people who without getting corresponding training can use this system conveniently.

3.1Transformation Estimation

At the beginning of this section, how to construct two coordinate systems should be illustrated. We assume that mark coordinate system coincides with world coordinate system in the beginning time. Using four balls which can be tracked with tracking device construct the mark coordinate system. All these four balls shouldn't be coplanar. And four balls' coordinates should be known when mark coordinate system coincides with world coordinate system. We design four balls' coordinates asM_0, M_1, M_2 and M_3. These coordinates represent four balls' coordinates in mark coordinate system, and their values aren't change anymore. When changing the pose and position of head, we can get four changed coordinates (W_0, W_1, W_2 and W_3) with respect to world coordinate system. First, we find the centroids $\overline{M}$ and $\overline{W}$ of the two sets of measurements in the mark coordinate system and the world coordinate system

$$\overline{M} = \frac{1}{4\sum_{i=0}^{3} \mathrm{M}_i}$$

$$\overline{W} = \frac{1}{4\sum_{i=0}^{3} \mathrm{W}_i}$$

And then, we only deal with measurements relative to the centroids. Represent the points in the new form

$$M_i' = M_i - \overline{M}$$

$$W_i' = W_i - \overline{W}$$

For each set of coordinates we compute the nine possible products

$x_{Wi}'x_{Mi}'$, $x_{Wi}'y_{Mi}'$, $x_{Wi}'z_{Mi}'$,

$y_{Wi}'x_{Mi}'$, $y_{Wi}'y_{Mi}'$, $y_{Wi}'z_{Mi}'$,

$z_{Wi}'x_{Mi}'$, $z_{Wi}'y_{Mi}'$, $z_{Wi}'z_{Mi}'$

We now compute the ten independent elements of the 4×4 symmetric matrix N by combining the sums obtained above.

Where

$S_{xx} = \sum_{i=0}^{3} x_{Wi}'x_{Mi}'$, $S_{xy} = \sum_{i=0}^{3} x_{Wi}'y_{Mi}'$,

$S_{xz} = \sum_{i=0}^{3} x_{Wi}'z_{Mi}'$,…, $S_{zz} = \sum_{i=0}^{3} z_{Wi}'z_{Mi}'$

We can get a 4×1 vector Q (q_0, q_x, q_y, q_z) which is the eigenvector corresponding to the most positive eigenvalue of matrix N.

$$\mathrm{N} = \begin{bmatrix} (S_{xx}+S_{yy}+S_{zz}) & S_{yz}-S_{zy} & S_{zx}-S_{xz} & S_{xy}-S_{yx} \\ S_{yz}-S_{zy} & (S_{xx}-S_{yy}-S_{zz}) & S_{xy}+S_{yx} & S_{zx}+S_{xz} \\ S_{zx}-S_{xz} & S_{xy}+S_{yx} & (-S_{xx}+S_{yy}-S_{zz}) & S_{yz}+S_{zy} \\ S_{xy}-S_{yx} & S_{zx}+S_{xz} & S_{yz}+S_{zy} & (-S_{xx}-S_{yy}+S_{zz}) \end{bmatrix}$$

Then, use Q to obtain the rotation matrix R with the formula

$$\mathrm{R} = \begin{bmatrix} {q_0}^2+{q_x}^2-{q_y}^2-{q_z}^2 & 2(q_x q_x + q_0 q_z) & 2(q_z q_x - q_0 q_y) \\ 2(q_x q_y - q_0 q_z) & {q_0}^2-{q_x}^2+{q_y}^2-{q_z}^2 & 2(q_y q_z + q_0 q_x) \\ 2(q_x q_z + q_0 q_y) & 2(q_y q_z - q_0 q_x) & {q_0}^2-{q_x}^2-{q_y}^2+{q_z}^2 \end{bmatrix}$$

Finally, we compute the translation T as the difference between the centroid of the coordinates in mark coordinate system and the rotated centroid of the coordinates in world coordinate system.

$$\mathrm{T} = \overline{M} - R(\overline{M})$$

So, we can estimate the transformation A through above process, and

$$\mathrm{A} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$$

This 4×4 homogeneous matrix we obtained is rigid transformation between world and mark coordinate system. And with this method, we can estimate the head's position and pose in world coordinate system in real time. That is to say, if we accomplish the calibration of OSTHMD, we can transform the points in the world into the image plane in the correct position in real time.

3.2OSTHMD Calibration

When we get the transformation A, we can transform points in world coordinate into mark coordinate system. What we should do now is to calibrate OSTHMD. That is to say, we should estimate transformation D.

We represent points in mark coordinate system and in image coordinate system with homogeneous coordinate $\mathrm{P_{M,i}} = (\mathrm{x_{M,i}}, \mathrm{y_{M,i}}, \mathrm{z_{M,i}})^{\mathrm{T}}$ and$\mathrm{P_{I,i}} = (\mathrm{u_{I,i}}, \mathrm{v_{I,i}}, \omega_{\mathrm{I,i}})^{\mathrm{T}}$. This problem can be simplified by the equation

$$\begin{pmatrix} u_i \\ v_i \\ \omega_i \end{pmatrix} = \mathrm{D} \begin{pmatrix} x_{M,i} \\ y_{M,i} \\ z_{M,i} \\ 1 \end{pmatrix} \text{ for i} = 1,2,\ldots,\mathrm{n}$$

And, we can get equations

$$u_i = a_{11}x_{M,i} + a_{12}y_{M,i} + a_{13}z_{M,i} + a_{14}$$

$$v_i = a_{21}x_{M,i} + a_{22}y_{M,i} + a_{23}z_{M,i} + a_{24}$$
$$\omega_i = a_{31}x_{M,i} + a_{32}y_{M,i} + a_{33}z_{M,i} + a_{34}$$

Let matrix D as

$$D = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

Assume that (x_i, y_i) is the image coordinate. Because$x_i = u_i/\omega_i$,$y_i = v_i/\omega_i$, we can get

$$x_i(a_{31}x_{M,i} + a_{32}y_{M,i} + a_{33}z_{M,i} + a_{34})$$
$$= a_{11}x_{M,i} + a_{12}y_{M,i} + a_{13}z_{M,i} + a_{14}$$
$$y_i(a_{31}x_{M,i} + a_{32}y_{M,i} + a_{33}z_{M,i} + a_{34})$$
$$= a_{21}x_{M,i} + a_{22}y_{M,i} + a_{23}z_{M,i} + a_{24}$$

We can put all the elements of D into a column vector$p = [a_{ij}]^T$, then we can get the equation

$$Bp = 0$$

where

$$B = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{M,i} & y_{M,i} & z_{M,i} & 1 & 0 & 0 & 0 & 0 & -x_i x_{M,i} & -x_i y_{M,i} & -x_i z_{M,i} & -x_i \\ 0 & 0 & 0 & 0 & x_{M,i} & y_{M,i} & z_{M,i} & 1 & -y_i x_{M,i} & -y_i y_{M,i} & -y_i z_{M,i} & -y_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

This matrix has 2n rows. The data of one point can construct 2 rows and 12 columns.Solving this equation can obtain transformation D. The problem is to find p to make Bp minimum. This can be done by finding the SVD of the matrix B ($B = UDV^T$). Getting the smallest singular value of matrix V, and the result is the corresponding column of this matrix.If we get the vector p, we can construct matrix D by it. And this is what we want to estimate.

Up to now, we accomplish OSTHMD calibration with relative to the mark system. As long as we complete this process, we will not calibrate again unless this system is used next time or by other people.

Because the calibration is just relative to the mark coordinate system and the points in the world coordinate system is what we want to map into image plane, the system should estimate the transformation between the world and the mark coordinate system immediately in case of the changing of head's position or pose in real time.

4. Conclusion

The code of the experiment is accomplished by C++.

Before experiment, we construct mark coordinate system through attaching 4 marks to OSTHMD and the coordinates of these four marks are convenient which can be measured manually. In the process of calibration, moving head or mouse makes the cursor aligned with the image of the calibration point. And clicking the left button of mouse, the data will be collected. And then, repeat this process by changing the position and pose of head and six sets of data should be collected.The calibration procedure can be summarized as follows:

The world coordinate system is fixed, and it is corresponding to the tracking device. The tracking device defines the origin of world coordinate system.

Attaching four balls which can be tracked by the tracking device to the optical see-through head-mounted display to construct the mark coordinate system. Assume that mark coordinate system coincide with the world coordinate system in the beginning.

A point in the world coordinate system which we called calibration point should be known at the beginning of the process. Its coordinate in world coordinate system is defined at the start. This calibration point is used to collect the calibration data. When user wants to collect the data in one position, move the mouse to align this point and click left button. The data should be collected six times at least.

All data collected is the four balls' coordinate in the world coordinate and in the image plane coordinate system. Based on the data we can transform the calibration point to the mark coordinate system through the method in 3.1. And now, we can get at least six sets of data in mark coordinate system, and also they are all different because they are corresponding to the different pose of head.

When we get all data we want, we can complete the calibration through the method mentioned in 3.2. The calibration process is completely accomplished. Notice that the user is encouraged to move his/her head in a large range, and the area of the range around the origin of world coordinate system may be better.

When use the system, the object points in world coordinate system should be transformed into mark coordinate system in real time. That is to say, we should calculate the transformation between world coordinate system and mark coordinate system in real time. The transform between mark coordinate system and image plane coordinate system is just estimated one time before the next use.

Reference

[1] Tuceryan M, Navab N. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR [C] Proceedings of the IEEE and ACM International Symposium for Augmented Reality, Munich, 2000: 149-158

[2] Deering M. High resolution virtual reality [J]. Computer Graphics, 1992, 26(2): 195-202

[3] Janin A L, Mizell D W, Caudell T P. Calibration of head-mounted displays for augmented reality applications [C] Proceedings of IEEE Virtual Reality Annual International Symposium, Seattle, 1993: 246-255

[4] Fuhrmann A, Schmalstieg D, Purgathofer W. Fast calibration for augmented reality [C] Proceedings of the ACM Symposium on Virtual Reality Software and Technology, London, 1999: 166-167

[5] Berthod k. p. Horn, “Closed-form solution of absolute orientation using unit quaternions,” Reprinted from Journal of the Optical Society of America A, Vol. 4, page 629, April 1987.

[6] D. Drascic, J. J. Grodski, P. Milgram, K. Ruffo, P. Wong, and S. Zhai. Argos: A display system for augmenting reality. In Formal video program and proc. of the Conference on Human Factors in Computing Systems (INTERCHI’93), page 521, 1993.

[7] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge based augmented real-ity. Communications of the ACM, 36(2):53–62, 1993.

[8] M. Gleicher and A. Witkin. Through-the-lens camera control. Computer Graphics, pages 331–340, July 1992.

[9] A. Janin, D. Mizell, and T. Caudell. Calibration of head mounted displays for augmented reality applications. In Proc. of the Virtual Reality Annual Internation-al Symposium (VRAIS’93), pages 246–255, 1993.

[10] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality ’99, pages 85–94, San Francisco, CA, October 20–21, 1999.

[11] Feiner, S., MacIntyre, B., and Seligmann, D. Annotating the Real World with Knowledge-Based Graphics on a See-Through Head-Mounted Dis-play. Proceedings of Graphics Interface’92, 78-85, 1992.

[12] Janin, A., Mizell, D. and Caudell, T. Calibration of head mounted dis-plays for augmented reality applications. In Proceedings of the Virtual Reality Annual International Symposium (VRAIS '93), pages 246–255, September 1993.

[13] Tsai, R.. An efficient and accurate camera calibration technique for 3D machine vision. Proceedings CVPR '86, pages 364-374, IEEE, June 1986.

[14] Whitaker, R., Crampton, C., Breen, D., Tuceryan, M., and Rose, E. Object Calibration for Augmented Reality. Proc. EUROGRAPHICS'95, pp. 15-27, 1995.