Research On Marker Tracking Method for Augmented Reality

Guangjun Liao[1,2]
[1]College of Automation Science and Engineering,
South China University of Technology,
Guangzhou, China
E-mail:gjliao@yahoo.com.cn

Yueming Hu[1], Zhifu Li[1]
[2]The Faculty of Forensic Science and Technology,
GuangDong Police College
Guangzhou, China

***Abstract*—This paper focuses on the device tracking method for augmented reality. Based on stereo vision principle, marker location in the space is realized. The nonlinear model of camera, structure parameters of stereo vision system and marker's sub-pixel tracking by gradient method which influence the system's accuracy have been discussed. The problem of transformation matrix for rigid body structure in the space is resolved based on iterative closest point method. The research results can be used in the precise field for augmented reality.**

Keywords-***Augmented reality, Stereo vision, Gradient sub-pixel method, Iterative closest point algorithm***

I. BACKGROUD

Augmented reality (AR) is used to combine virtual objects with the real world. It can provide a more natural user interface for engineers and establish fascinating environment for education, virtual assembly, training and some other fields. [1,2]

Modern mobile augmented reality systems use one or more of the following tracking technologies: digital cameras and/or other optical sensors, accelerometers, GPS, gyroscopes, solid state compasses, RFID, wireless sensors. Each of these technologies has different levels of accuracy and precision. For example, the ARToolKit[3] and ARTag[4] video tracking libraries calculate the real camera position and orientation relative to physical markers in real time. But there are some limitations in application fields due to high false recognition rate and at least 4 coplanar markers for tracking device.

In this paper, based on binocular stereo vision, the device tracking system has been discussed. The paper provides a method of the interface for human machine interaction.

II. INTRODUCTION

To track the device's 3-D position and orientation through gathering the three makers' 3-D coordinates, a binocular vision system and a method for determining 3-D motion parameters of an object from binocular sequence images are introduced. Fig.1 shows the system block diagram. Fig.2 shows equipment of stereo vision system. Fig.3 shows tracking device with three markers.

Figure 1. System block diagram

Figure 2. Equipment of stereo vision system

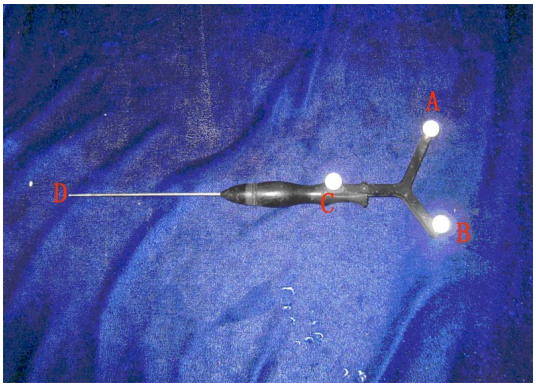

Figure 3. Tracking device with three markers

Taking the factors which largely influence the accuracy into consideration， especially lens distortion and camera interior parameters, the same model of cameras and lens are recommended. In order to capture two images at the same time, the image acquisition card with synchronous acquisition is necessary.

III. BINOCULAR STEREO VISION SYSTEM

Stereo vision consists of capturing two images of a scene, taken from different viewpoints and estimating the depth of the scene from analysing the disparity between corresponding features. This methodology finds its basis in trigonometry and triangulation and is employed by the human binocular vision system.

The algorithm principle and implementation scheme about binocular stereo vision are presented in [5,6,7]. In this paper, the factors which influence the accuracy are analysed, and the method of tracking device with three markers is discussed.

A. Camera Model

The camera model is a mathematical expression of optical imaging system. Generally speaking, there are two models for camera: linear model and nonlinear model. Combined with practical engineering, the nonlinear model of camera is utilized in which the radial distortion and tangential distortion of camera are considered. In order to improve the locating accuracy, distortion correction operation for pixel of image is necessary. Reference [8] describes the nonlinear model of camera and lens aberration in detail.

B. Structure of Binocular Cameras

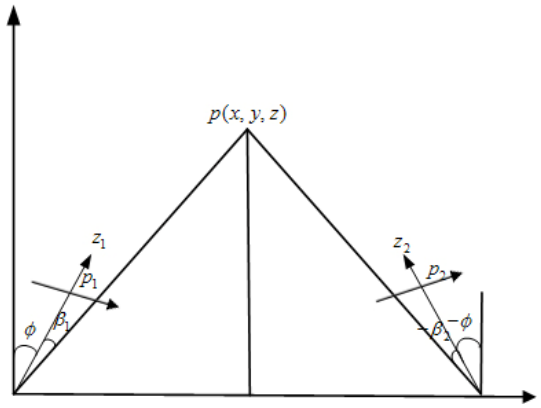

Figure 4. The relation between point $p(x,y,z)$ with in xoz plane and two cameras[9]

For simplicity, like in [9], we suppose that the two cameras are on the same plane. Fig.4 shows the relation between point $p(x,y,z)$ with in xoz plane and two cameras. As Fig.4 shows, $p_1(u_1,v_1)$ and $p_2(u_2,v_2)$ are pixel coordinates for two cameras, f_1, f_2 are focuses, β_1 and β_2 are angles between optical axis and xy plane, φ is angle between optical axis and z axis.

Therefore:

$$tg\beta_1 = -U_1 / f_1 \tag{1}$$

$$tg(\phi+\beta_1) = x/z \tag{2}$$

$$tg(\phi+\beta_2) = (B-x)/z \tag{3}$$

And:

$$z = B/[tg(\phi+\beta_1)+tg(\phi+\beta_2)] \tag{4}$$

$$x = B(1-\frac{tg(\phi+\beta_2)}{tg(\phi+\beta_1)tg(\phi+\beta_2)}) \tag{5}$$

$$y = v_1\frac{z}{f_1}\frac{\cos\beta_1}{\cos(\phi+\beta_1)} = v_2\frac{z}{f_2}\frac{\cos\beta_2}{\cos(\phi+\beta_2)} \tag{6}$$

By the error analysis, the measurement errors at the three directions for point p are:

$$dz = \pm\sqrt{\left[\frac{\partial z}{\partial u_1}du_1\right]^2+\left[\frac{\partial z}{\partial u_2}du_2\right]^2} \tag{7}$$

$$dx = \pm\sqrt{\left[\frac{\partial x}{\partial u_1}du_1\right]^2+\left[\frac{\partial x}{\partial u_2}du_2\right]^2} \tag{8}$$

$$dy = \pm\sqrt{\left[\frac{\partial y}{\partial u_1}du_1\right]^2+\left[\frac{\partial y}{\partial u_2}du_2\right]^2+\left[\frac{\partial y}{\partial v_1}dv_1\right]^2+\left[\frac{\partial y}{\partial v_2}dv_2\right]^2} \tag{9}$$

So，the total measurement error is:

$$\pm\sqrt{(dx)^2+(dy)^2+(dz)^2} \tag{10}$$

From above analysis, it can be concluded that:

- The measurement accuracy is proportional to baseline and inversely proportional to effective focal length.
- With the angle β increasing, the measurement is more accuracy, but the field of camera's view becomes narrow. 50~80 degree between two cameras is recommend by experiment.

According to the above conclusion, some methods and proposals are proposed for increasing the measure accuracy, increasing baseline length, using long focal length lens, tilting the cameras to the midline under ensuring the two cameras with sufficient field of view could increase the measure accuracy.

C. Sub-Pixel Accurate Center

Generally speaking, the image processing system considers the sampling information and the resolution for a pixel, but rarely considers the information between the sampling points. Sub-pixel algorithm is using mathematical tools to express the non-sampling information between two pixels by utilizing the continuity of image. There are some common methods about sub-pixel algorithm including: distribution-fitting based on the approximate Gaussian

distribution of intensity for marker's surface, image center of gravity method based on the image brightness, relevant operator method based on the objectives of gray similar operation near the target template and the secondary quadratic curve fitting surface.[10] However, the effectiveness of the algorithms are not good enough to meet the expectation of real-time application. For the system, to track the reflective markers, sub-pixel edge detecting at the gradient direction is proposed, which get the location of brightness of the vertex tracker based on analyzing the image's gradient direction. Fig.5 shows the schematic diagram of gradient sub-pixel method. By analyzing the point of the vector group, the markers sub-pixel center could be calculated.

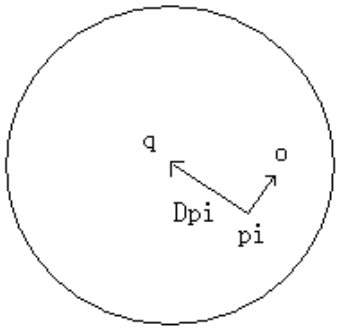

Figure 5. Schematic diagram of gradient sub-pixel method

The core idea of this algorithm is based on the observation that every vector from the center q to a point p located within a neighborhood of q is orthogonal to the image gradient at p.

$$\begin{cases} \varepsilon_i = DI_{P_i}{}^T \bullet (q - p_i) \\ \varepsilon_i \to 0 \end{cases} \tag{11}$$

Where $DI_{P_i}{}^T$ is the image gradient at the one of the points p in a neighborhood of q. The value of q is to be found through minimizing ε_i. The system equations may be set up with ε_i's set to zero.

$$\sum_i \left(DI_{P_i} \bullet DI_{P_i}{}^T \right) - \sum_i \left(DI_{P_i} \bullet DI_{P_i}{}^T \bullet p_i \right) \tag{12}$$

where the gradients are summed within a neighborhood ("search window") of q. Calling the first gradient term G and the second gradient term b, we can get:

$$q = G^{-1} \bullet b \tag{13}$$

The algorithm sets the center of the neighborhood window at this new center q and then iterates until the center keeps within a set threshold.

D. Others

To improve the robustness of the system, on the hardware design, the camera is near-infrared camera with coaxial near-infrared light; the markers are made of special reflective material. The system runs in the near-infrared range for reducing the external environment interference.

IV. RIGID TRANSFORMATIONS

All the rigid transformations we considered will be affine. This means that if we have chosen a linear coordinate system in whatever set we are looking at (a line, a plane, or space). Then the transformation $P \to P_*$ is calculated in terms of coordinate vectors x and x_* according to the formula:

$$x_* = Ax + v \tag{14}$$

Where A is a matrix and v is a vector. In 3D, for example, we require:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{15}$$

It turns out that all rigid transformations are in fact affine, but we shall not worry about that here. The matrix A is called the linear component, v is the translation component of transformation.

According to [11,12], in order to get the transformation matrix, the device tracked in the system has four non-coplanar markers at least. There is certain limitation by using device with four makers. To expand its application, the tracking device with three markers is more convenient.

To achieve the above purpose, we cite idea about function vtkLandmarkTransform of VTK[13], and use iterative closest point (ICP) algorithm to calculate the transformation matrix. ICP is an iterative algorithm for matching point-sets. [14] Consider two point-sets $A, B \subseteq R^d$ where $|A| = n$ and $|B| = m$. We are interested in a one-to-one matching function $\mu : A \to B$ that minimizes the root mean squared distance (RMSD) between A and B. Mathematically, we want to minimize the following:

$$RMSD(A, B, \mu) = \sqrt{\frac{1}{n} \sum_{a \in A} \left\| a - \mu(a) \right\|^2} \tag{16}$$

Incorporating rotation and translation into the matching, we could define:

$$\mu : A \to B, t \in R^d, R \in SO(d) \min \sum_{a \in A} \left\| Ra - t - \mu(a) \right\|^2 \tag{17}$$

Where R is the rotation matrix, t is the translation vector and $SO(d)$ is the set of special orthogonal matrices in d dimensions.

The ICP algorithm seeks to minimize the RMSD, by alternating between a matching step and a transformation step. In the matching step, given a certain rotation and translation, the optimal matching is calculated by minimizing the RMSD. In the transformation step, given a matching, the optimal rotation and translation are computed. This alternating process terminates when the matching remains unchanged in successive iterations. The following is an algorithmic description of ICP:

- Initialize R=I(the identity matrix),t=0.
- There Matching Step: Given R and t, compute optimal μ by finding $\min_{\mu} RMSD(A,B,\mu)$.
- Transformation Step: Given μ, compute optimal R and t by finding $\min_{R,t} RMSD(RA-t,B,\mu)$.
- Go to step 2 unless μ is unchanged.

In the system, we should determine the relation between three tracking markers and track's D point (Fig.3) in space. We mount another marker at D point and get the four markers' 3-d coordinates as reference data. Then, we remove the marker at D point, calculate the rotation matrix and translate vector by three markers'(A,B,C) 3-d coordinates based on ICP algorithm. Using the reference data computes the D's 3-d coordinates in real time. If the real 3-d coordinate system and 3-d coordinate system of the virtual environment unified by matrix transformation, the human-computer interaction could be realized in virtual environment.

V. EXPERIENCE

In the system, in order to reduce cost, the near-infrared camera whose brand is SONY XC-75 is remodeled, the lens are SPACECOM S4.8mm which are made in Japan. As Fig.2 shows, the baseline length is 1.2m. Fig.6 shows the inspection equipment for accuracy.

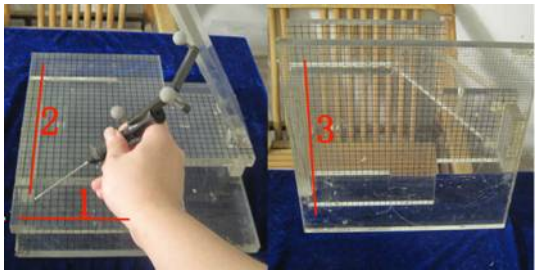

Figure 6. Inspection equipment for accuracy

As Fig.6 shows, in the equipment every grid's width and length are 1 cm. There are three orthogonal directions just as red lines 1,2,3 depicts. Firstly, we mount a marker at D point (Fig.3), capture images with four markers by the stereo vision system (Fig.2) , calculate the four markers' 3-d coordinates based on the binocular stereo vision algorithm and take them as reference data. Then, we remove the marker at D point, capture images with three makers and get the three markers' 3-d coordinates, calculate (R,t) based on ICP. Now, we could get the D's 3-d coordinates in real time according to four markers' reference data and (R,t).

In the experiment, after correcting the lens distortion, we focus on verifying the effectiveness of the sub-pixel accurate center. We use the device's D point to hit the grid's corner at different directions and length. According to data of the two points' 3-d coordinates, we could compute the length of the line segment in space. The results are given as follows:

TABLE I. EXPERIMENTAL RESULTS FOR NONUSING THE SUB-PIXEL ACCURATE CENTER

Direction	*Actural distance/mm*	*Measurement data/mm*	*Errors/mm*
1	5	5.37	+3.7
1	20	20.45	+4.5
2	5	4.46	-5.4
2	20	19.45	-5.5
3	5	5.34	+3.4
3	20	19.69	-3.1

TABLE II. EXPERIMENTAL RESULTS FOR USING THE SUB-PIXEL ACCURATE CENTER

Direction	*Actural distance/mm*	*Measurement data/mm*	*Errors/mm*
1	5	5.08	+0.08
1	20	20.12	+1.2
2	5	4.84	-1.6
2	20	19.82	-1.8
3	5	5.11	+1.1
3	20	19.92	-0.08

TABLE I shows the average accuracy is less than 8mm for nonusing the sub-pixel accurate center. TABLE II shows the average accuracy is less than 2 mm for using the sub-pixel accurate center. At direction 2, the error is bigger than others. So we can conclude that using the sub-pixel accurate center is effective.

VI. CONCLUSION

Compared with other binocular vision system, the innovation of system is:

- Consider the hardware cost, and the remodeled near-infrared camera is used.
- The design of interface is more advance for the tracking device with only 3 markers.
- Research on sub-pixel detecting at the gradient direction and improve the location accuracy.
- Based on ICP algorithm solves the rigid body transformation by three coplanar markers.

The next work is to develop 3-d interactive virtual scene by OpenGL with discussed tracking system and consider its

applications like virtual assembly, virtual maintance, virtual training and other fields.

ACKNOWLEDGMENT

The research presented abrove was supported by Major Program of National Natural Science Foundation of China (No. 60835001) and the cooperation project in industry, education and research of Guangdong province and Ministry of Education of P.R.China(No. 912220500017). The authors would like to express thanks for their financial support.

REFERENCES

[1] Ren Peng, Kang Bo.New Marker Tracking Method for Augmentation Reality.Journal of System Simulation, 2009(1),21(2):465-468(in Chinese)

[2] Wang X. Experiential mixed reality learning environments for design education. In: Proceedings of the 41st Australia and New Zealand Annual Conference of the Architectural Science Association (ANZAScA). Deadkin University, Geelong, Australia, 2007: 272-277.

[3] Chang Yong, Shi Chuang. 3D Visualization and Analysis of Spatial Information Basedon Augmented Reality. Journal of System Simulation(S1004-731X), 2007, 19(9): 1991-1995

[4] Mark Fiala. ARTag,a fiducial marker system using digital techniques.Computer Vision and Pattern Recognition, 2005, (2) :590-596 .

[5] Qiu Maolin, Ma Songde, Li Yi. Overview of Camera Calibration for Computer Vision. ACTA AUTOMATICA SINICA,2000(1),26(1):43-55(in Chinese)

[6] Wang Li, Ma Shu-yuan, Wu Ping-dong, Sun Guang-yu. Study on Camera Calibration of Vision System.(in Chinese)

[7] Radu Orghidan,Joaquim Salvi,El Mustapha Mouaddib. Modelling and Accuracy Estimation of a New Omnidirectional DepthComputation Sensor .Pattern Recognition Letters, 2006, 27 (7) :843-853 .

[8] Chatterjee C, Roychowdhury V P. Algorithms for coplanar camera calibration .Machine Vision andApplications, 2000, 12 (2) :84-97 .

[9] Zhang Jianxin,Duan Fajie,Ye Shenghua.An Optimization Structure Design of Binocular Sensor.Opto-Electronic Engineering,1996(6),23(3):12-17(in Chinese)

[10] TANG Guan-qun. Analysis and comparison of several calculation methods of beam spot center. Journal of Beijing Institute of Machinery,2009,24(1):61-64(in Chinese)

[11] D.W. Eggert,A. Lorusso,R.B. Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. Machine Vision and Applications, 1997, 9(5-6): 272～290.

[12] K. Achour ,M. Benkhelif. A new approach to 3D reconstruction without camera calibration. Pattern Recognition, 2001(12),12(34): 2467–2476.

[13] VTK Home Page .http://www.vtk.org. .

[14] P.J.Besl.N.D.McKay, IEEE Transactions on Pattern Analysis and Machine Intelligence,PP.239-256,1992.