Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 2016

MARKER TRACKING FOR VIDEO-BASED AUGMENTED REALITY

YAN-FEI GAO[1], HENG-YOU WANG[1], XIAO-NING BIAN[2]

1. College of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China,
2. Department of Mathematics and Statistics, Dalhousie University, Nova Scotia B3H 4R2, Canada
E-MAIL: gaoyf@bucea.edu.cn

Abstract:

In augmented reality (AR) systems, virtual objects, graphics or texts are added to the real scene to improve the person's perception. Virtual and real information must be aligned strictly in the combination of scenes. The key of augmented reality is 3D registration technology which is based on tracking the artificial marker in the scene effectively. Using linear programming theory, this paper presents a tracking method which can recognize the vertices of any convex polygon in augmented reality system since most of the markers are squares. The transformation matrix of camera can be estimated at the aid of coordinates of the four vertices.

Keywords:

Augmented reality; 3D registration; Linear programming; Marker tracking

1. Introduction

Augmented reality is developed with the progress of computer graphics, computer vision, multi sensor technology and human-computer interaction technology. It is defined as a system in which a real environment is augmented by adding virtual figures to it for the aim that users can have more abundant experiences in augmented reality system than in pure virtual one [1]. It has broad application prospects in many fields such as education, medical industry, military, entertainment and so on. Nowadays, numerous augmented reality systems have been developed all over the world but yet it is in its initial phases.

The key issue for AR is tracking and registration: the getting of the external parameters of camera. Virtual objects must be aligned with the real world properly and dynamically, otherwise, the expected consistency of the virtual and real scene will not be achieved. Failure of geometric consistency will lead to jitter or drift of the virtual objects，which can be perceived by users easily.

AR system can be classified into two groups according to its tracking method [2].

One is based on the sensor tracking technology, in which position and rotation parameters of the camera are determined by the data coming from GPS, compass and accelerometer. To achieve precise registration, expensive trackers and sensors are needed, the cost of such system is usually very high. In order to improve the tracking accuracy of the cheap sensor, IPhone is the first device to introduce "electronic compass" by integrating acceleration meter with gyroscope and magnetometer [3].

The other one is based on the technology of computer vision. In this vision-base AR. natural feature or a pre-placed marker is tracked by means of digital image processing and computer vision technology. Park et al. proposed to estimate the position of the camera by moving the camera to extend the new natural features in the real scene [4].A set of application system based on structure from motion is constructed by Mooser [5].Since Davison and others proposed Location and Mapping Simultaneous (SLAM), SLAM method has been widely considered in the field of augmented reality [6].

This kind of registration methods are mostly based on tracking feature points and matching strategy. The extraction of feature and matching require a lot of computing time, this will cause some negative impacts on registration speed and accuracy, and even cause registration errors. Artificial markers may seem obtrusive in some cases, while it can offer more robust and faster registration [7], [8]. The 3D registration technology based on mark is being perfected. ARTag developed by Mark Fiala at National Research Institute of Canada and ARToolkit developed by Hirokazo Kato at Hiroshima City University of Japan and Maintained by the University of Washington are the most famous secondary development function libs for marker based AR. They are designed to write AR applications quickly. Many modern libraries are derived from them. They are used in various commercial applications [9], [10].

The working principle of marker-based AR can be described as follows [10], [11]:

1. The live feed video of the real scene is captured by image capturing module, and the video stream is sent to computer.

2. The processing module searches through the frames

978-1-5090-0390-7/16/$31.00 ©2016 IEEE

for the pre-placed marks which in most cases are quadrilateral.

3. If an image with specific shape is found, it will be matched with the standard images of the markers after normalized. If the matching result is true, then it means the pre-placed markers are found in the scene.

4. The position and the direction of the camera can be calculated by the vertices of the quadrilateral which is found, this step is named 3D registration. The relationships between different coordinate systems are shown in Figure 1. Origin of the camera coordinate system is the optical center of the camera, origin of the word coordinate system is the center of the marker.

5. The virtual figure is drawn on the same positions as the marker's according to the position and direction of the camera.

6. The final output consists of the virtual figures and the image of real world.

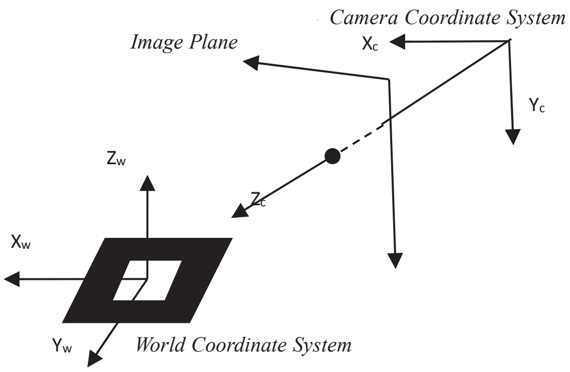

FIGURE 1 The relationship between the Coordinate System

Given that:

$$M_1 = \begin{bmatrix} \alpha_x & 0 & a_0 \\ 0 & \alpha_y & b_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$\alpha_x, \alpha_y, a_0, b_0, f$ are the intrinsic parameters of camera, and related to the internal structure of the camera only, these internal parameters can be obtained by calibration technology which is proposed by Zhengyou Zhang [12]. M_1 is named the internal parameter matrix of camera.

$$M_2 = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$$

R is a 3 x 3 orthogonal matrix; T is a 3 x 1 translation vector, they are called the external parameters of the camera since R is the rotation component and T is the translation component of the external parameter matrix of camera.

The projection transformation project a three-dimensional point onto a two-dimensional image plane. Formula (1) shows the relationship between a 3D point $[X_W, Y_W, Z_W]$ and its image $[u, v]$:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{1}$$

$[X_W, Y_W, Z_W, 1]$ and $[u, v, 1]$ are the augmented vectors for $[X_W, Y_W, Z_W]$ and $[u, v]$ respectively, Zc is an arbitrary scale factor. The process of solving T is called 3D registration and is the most important task of an augmented reality system.

Markers used in the marker-based AR for 3D registration are mostly black and white and planar. Seldom of them are chromatic or three-dimensional. Markers used by ARToolKit and ARTag are the most famous. The typical types of markers used in ARToolKit and ARTag are showed in Figure 2 and Figure 3 [11] [13].This kinds of markers are squares consisting of a thick black border and two-dimensional graphics within their white internal regions. The usage of the black and white color can help to identify marker from complex background in grabbed frame easily. The internal region marks the mark's identifier.

FIGURE 2 ARToolKit markers

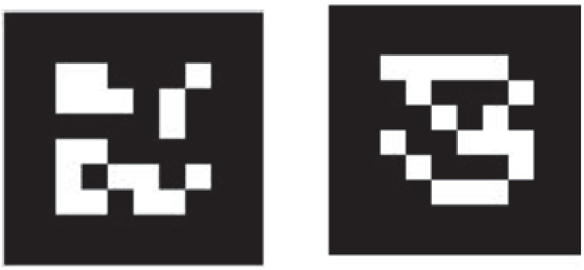

FIGURE 3 ARTag markers

The camera model is a pinhole imaging model, that is, the perspective projection model. According to the projective geometry, the shapes of markers in real scene and image plane are different. When the square markers are projected onto image plane, the projection often induces deformation. Still in this model, the projection of a convex polygon is convex one.

In this paper, a cluster of linear functions are constructed for 3d registration, the convex polygons which is the projection of the marker are took as the feasible region for these functions, the coordinate of the vertices of the polygon can be gotten by solving the optimal solutions of the functions. The transformation matrix of camera can be estimated at the aid of coordinates if the polygon itself is a convex one.

2. The theoretical basis of the algorithm

Linear programming (LP) is a problem of optimizing a linear objective function that is subjected to linear constraints. It can be expressed as following in standard form:

$$\max f = \sum_{i=0}^{n} c_i x_i \quad (1)$$

$$\begin{cases} \sum_{i=1}^{n} a_{ij} x_i \leq (=\text{or} \geq) b_j & j = 1, 2 \ldots m \quad (2) \\ x_i \geq 0 & (3) \end{cases}$$

Expression (1) is called the objective function and inequalities (2) and (3) are constraints. Inequalities in (3) are called the no negativity constraints.

f is the amount of overall resource that is limited;

c_j represents increase of f which result from each unit increase of x_i;

x_i is named as the decision variable, a set of such variables which satisfy the constraints is a feasible solution;

a_{ij} is the quantity of resource j consumed by each unit of activity i.

b_j is the amount of resource j.

The feasible region of LP is a convex polygon The feasible region consists of all feasible solutions. The region includes all points on or in the interior of polygon. An optimal solution for a LP is a point with the largest objective function value in the feasible region. For a minimization problem, objective function will obtain the minimum value at the optimal solution. If the feasible region is convex, optimum solution will be the vertices of the polyhedron.

3. Algorithm design

After the image edge is detected, the edge points will be fitted to the contour.

The algorithm can be designed as follows:

Select a set of linear objective functions $f_i(x, y) = a_i x + b_i y$, a_i, b_i should be selected properly, so that the slopes change from negative infinity to positive infinity uniformly as shown in Figure 4.

All points on the contour are traversed, and each of the points is substituted into the objective functions. It can be considered as a vertex if the point is the optimal solution of a certain number of functions as shown in Figure 5. If four vertices are found on a contour, the contour will become one of the potential convex polygons.

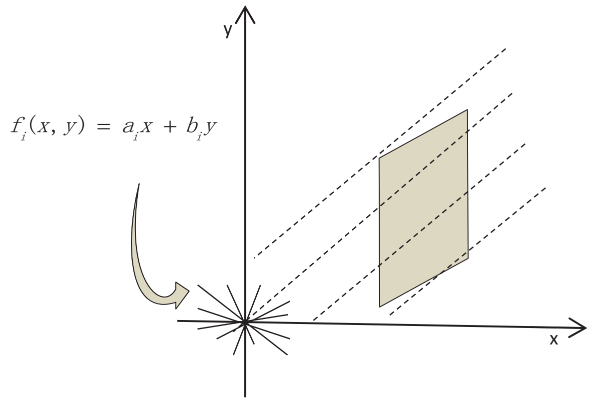

FIGURE 4 a set of linear objective functions and their feasible region

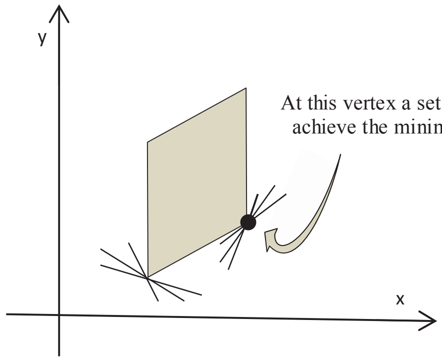

FIGURE 5 Vertices at which a set of functions achieve the minimum values

If all points between any two vertices adjacent on the potential polygon are located on the contour, the polygon will be regarded as a convex one. That is, a convex quadrilateral is found. As the result of projection, square markers will be distorted in the image plane. Before matching with the marker given, the quadrilateral should be normalized. Given that the four corners acquired after detecting grabbed frame of a potential marker are $(x_i, y_i), i=1,2,3,4$, four corners of the square in the real scene corresponding to (x_i, y_i) are $(x_i', y_i'), i=1,2,3,4$, homography matrix can be calculated by solving equation:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = H \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \tag{4}$$

The points in internal region of quadrilateral could be unwrapped using H. After that, unwrapping image are used to match templates in matching method.

Suppose $A'(x_{A'}, y_{A'})$, $B'(x_{B'}, y_{B'})$, $C'(x_{C'}, y_{C'})$, $D'(x_{D'}, y_{D'})$ are vertices of quadrilateral which is the image of the mark ABCD with sides of l and:

$$\overrightarrow{OA} = k_1\overrightarrow{OA'} \qquad \overrightarrow{OB} = k_2\overrightarrow{OB'}, \qquad \overrightarrow{OC} = k_3\overrightarrow{OC'} \qquad \overrightarrow{OD} = k_4\overrightarrow{OD'} \tag{5}$$

O is the origin of camera coordinate system. k_1, k_2, k_3, k_4 are ratio coefficients which can be gotten by solving equation (6) .

The 3D coordinate of the four vertexes can be gotten by substituting k_1, k_2, k_3, k_4 into formula (5). With these data, the transformation matrix from the world to camera coordinate system can be deduced. The matrix is known as the external parameter matrix of the camera.

$$\begin{cases} k_4 x_{D'} - k_1 x_{A'} = k_3 x_{C'} - k_2 x_{B'} \\ k_4 y_{D'} - k_1 y_{A'} = k_3 y_{C'} - k_2 y_{B'} \\ k_4 f - k_1 f = k_3 f - k_1 f \\ \sqrt{(k_2 x_{B'} - k_1 x_{A'})^2 + (k_2 y_{B'} - k_1 y_{A'})^2 + (k_2 f - k_2 f)^2} = l \end{cases} \tag{6}$$

4. The experimental results

Open CV and OpenGL library are used in the VC environment to implement the algorithm. The marker size is 8mm*8mm, Canny method is used to find image edges. In Figure 6, the marker is tracked correctly, for better observation, two sets of parallel lines were made over the four vertices of the convex polygon which is found. Since ARToolKit is the most popular open source library in c and c++ for AR. The external parameters of camera achieved by the proposed method and that of ARToolKit are both showed in Table 1. The result shows that the method used here has the same excellent tracking performance as ARToolKit. Figure 7 shows the final result that a virtual cube is "augmented" to a real desk, while Figure 8 shows two virtual cubes added to the real scene. The expected consistency of the virtual and real figures is achieved well.

FIGURE 6 A marker tracked

TABLE 1 The comparison of external parameters of camera

	The new method	ARToolKit	relative variation
Displacement vector（mm）	-15.2	-14.1	0.71%
	-41.2	-40.8	0.98%
	-399.5	-397	0.63%
Rotation angle（degree）	1.79	1.77	1.13%
	1.21	1.21	0.00%
	0.73	0.75	2.67%

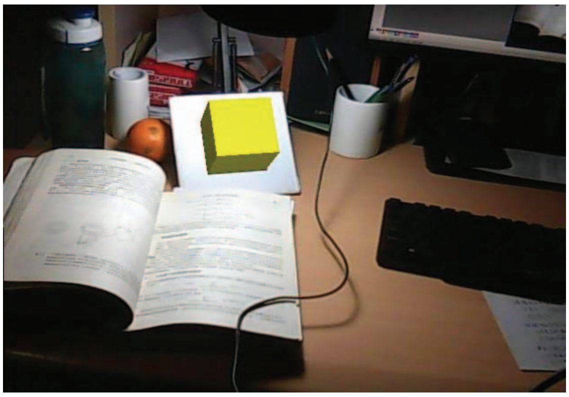

FIGURE 7 Overlay of a virtual cube

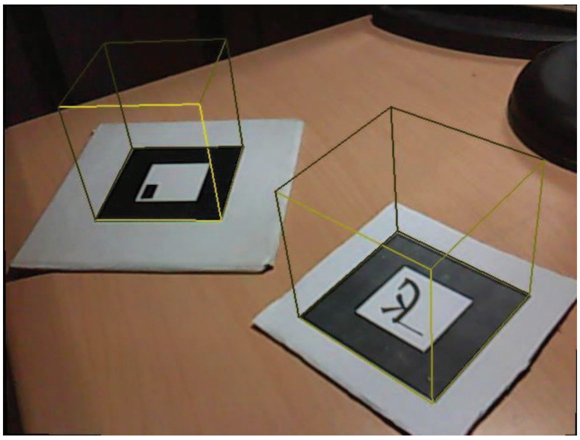

FIGURE 8 Two virtual cubes added to real scene

5. Conclusion

This paper presents a tracking method which can recognize the vertices of any convex polygon in augmented reality system since most of the markers are squares. The transformation matrix of camera can be estimated at the aid of coordinates of the four vertices. Stable, accurate and real-time properties are obtained as the algorithm is able to withstand large range of motion and fast motion speed of camera.

6. Acknowledgements

This paper is supported by National Natural Science Foundation of China (61502024), Science and Technology Plan of Beijing Municipal Education Commission (Z09117,SQKM201610016009).

References

[1] M. Billinghurst, S. Weghorst, T. Furness, "Shared Space: An Augmented Reality Approach for Computer Supported Cooperative Work", Virtual Reality Vol. 3, No. 1, pp. 25-36, Mar.1998.

[2] Liang Z M, Yuan Y, Liang J Y, "A Survey on Augmented Reality". Journal of Image & Graphics, Vol. 9, No. 7, pp.767-773, July 2004.

[3] Zhou Zhong, Zhou Yi, Xiao Jiang Jian, "Survey on Augmented Virtual environment and Augmented Reality", Vol45, No 2, pp. 157-180, 2015

[4] Park J，You S，Neumann U. "Natural feature tracking for extendible robust augmented realities", Proc．Int．Workshop on Augmented Reality．USA pp.209-217 , 1999

[5] Mooser J，You S，Neumann U，"Applying robust structure from motion to markerless augmented reality[C]", Applications of Computer Vision，2009 Workshop on．IEEE，2009：1-8

[6] Davison AJ，Reid I D，Molton N D， "MonoSLAM：Real—time single camera SLAM[J]". Pattern Analysis and Machine Intelligence，IEEE Transactions on Pattern Analysis and Machine Intelligence，Vol.29,No. 6,pp. 1052—1067,June,2007.

[7] Anuroop Katiyar, Karan Kalra and Chetan Garg, "Marker Based Augmented Reality", Advances in Computer Science and Information Technology (ACSIT), Vol. 2, No. 5, pp. 441-445, April-June, 2015.

[8] Kato, H. and M. Billinghurst. "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System." 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR99), San Francisco, CA , pp.85 – 94,1999.

[9] Daniel Wagner, Istvan Barakonyi, "Augmented reality kanji learning". The Second IEEE and ACM International Symposium on mixed and augmented reality, pp.335- 336, Oct. 2003.

[10] Kato H. "MagicCup: a tangible interface for virtual objects manipulation in table top augmented reality". Augmented Reality Toolkit Workshop, 2003. IEEE International, Japan, pp.75-76, July 2003.

[11] http://www.hitl.washington.edu/artoolkit/

[12] Zhengyou Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", Proceedings of International Conference on Computer Vision. Corfu，Greece, pp.666-673,1999.

[13] M.Fiala, "ARTag, a fiducial marker system using digital techniques". Computer Vision and Pattern Recognition, pp.590 -596, Vol.2, 2005.