Estimating 3D Object Coordinates from Markerless Scenes

Ki Woon Kwon[1], Sung Wook Baik[2], and Seong-Whan Lee[1,*]

[1] Korea University, Seoul 136-713, Korea
{kwkwon, swlee}@image.korea.ac.kr
[2] Sejong University, Seoul 143-747, Korea
sbaik@sejong.ac.kr

Abstract. This paper presents a novel method for estimating the coordinates of a 3D object using the four vertices of a quadrangle and the camera motion parameters. Estimation of 3D object coordinates from 2D images of video is a studied problem in augmented reality. However, most solutions are dependent on fiducial markers in video or known coordinate systems which are required with superimposition of virtual object on frames. In this paper, we begin with the fact that the rectangular objects in 3D real world are projected the perspective quadrangle onto image planes. We can estimate 3D object coordinates from 4 vertices of quadrangular objects through transformation of image coordinates. The camera motion parameters between pairs of successive frames in a sequence are calculated using epipolar geometry.

1 Introduction

An AR system should be able to [1] 1) combine real environments and computer-generated virtual objects, 2) operate virtual objects interactively with the change in the real world, and 3) align virtual graphic objects onto real environments. When a virtual object is superimposed in a reference frame, the frame should contain one plane with which a 3×3 planar homography can be found [2-5].

The homography is the transformation modeling the 2D movement of coplanar points under perspective projection. To obtain another planar homography between two consecutive frames in a sequence, different methods calculating camera motion to use multiple planes have been considered [4].

Kutulakos and Vallino proposed a system that can represent 3D graphic objects using four pairs of prior affine basis points that correspond to a sequence of images extracted from two uncalibrated affine cameras [6]. Another system involves a perspective camera model. This is more difficult to estimate the projective reconstruction from perspective views than using affine reconstruction from orthographic views [7].

In this paper, we estimate the direction of the Z-axis and the vertices of a quadrangle that is defined in a reference frame. The consecutive frames are computed for the essential stereoscopic matrix using epipolar geometry, and the estimated coordinates of the 3D object are determined from the camera motion parameters [8].

* Author for Correspondence.

V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3516, pp. 850 – 853, 2005.

2 Estimating the Coordinates of a 3D Object

The rectangle is deemed to be one side of a rectangular parallelepiped. Consequently, its X-, Y-and Z-axes are at right angles to each other. Based on this fact, the Z-axis of the rotation angle is determined via complex rotations of the X-and Y-axes in the real world. The estimated direction of the Z-axis can be used to calculate the angles among the X-, Y-and Z-axes, and these angles are used when overlaying a 3D graphic object on the frame image.

To estimate the direction of the Z-axis, the image coordinates are rotated by applying the Euler-angle to each axis. The vertices and center point of a quadrilateral which the user designates from a reference frame, are applied to Equation 1.

$$\begin{bmatrix} 1+D\cdot E & -C+A\cdot H & B+A\cdot J \\ C+D\cdot H & 1+D\cdot F & -A+D\cdot I \\ -B+D\cdot J & A+D\cdot I & 1+A\cdot G \end{bmatrix} \tag{1}$$

where

$$A = a\cdot\sin\theta,\ B = b\cdot\sin\theta,\ C = c\cdot\sin\theta,\ D = (1-\cos\theta)$$

$$E = a^2-1,\ F = b^2-1,\ G = c^2-1,\ H = ab,\ I = bc,\ J = ac$$

Then, the unit vector in the direction of the Z-axis can be calculated. Equation 1 is a transformation matrix used for rotating an object about an arbitrary axis.

3 Calculating Camera Motion Parameters

We develop another method for extracting camera motion parameters using a monoscopic system. Most of the video sequences used are pictures in which the intrinsic parameters of the camera are unknown. Therefore, the intrinsic parameters of the camera are set to a fixed skew of 0, an aspect ratio of 1, and the principle point is in the center of the quadrangle. The extrinsic parameters are calculated as an essential matrix. A pair of successive frames has the similar property with images of left and right camera. To get the cross product with vector and matrix, $\boldsymbol{S}$ is a skew symmetric matrix. And, the $\boldsymbol{R}\cdot\boldsymbol{S}$ matrix, an essential matrix is computed by Equation 2.

$$\boldsymbol{q}_r^T\cdot(\boldsymbol{R}\cdot\boldsymbol{S})\cdot\boldsymbol{q}_l = \boldsymbol{0} \tag{2}$$

The essential matrix $R\cdot S$ is computed using Equation 2, using an 8-point algorithm. In addition, E in a 3×3 matrix is computed using Equation 3 and 4, and the property of the epipolar constraint. In order to calculate the essential matrix, we expand the equation,

$$\boldsymbol{x}'^T\,\boldsymbol{E}\boldsymbol{x} = 0 \tag{3}$$

For 8 point correspondences, Equation 6 becomes

$$Ae = 0 \tag{4}$$

where

$$A = \begin{bmatrix} u'_1 u_1 & u'_1 v_1 & u'_1 & v'_1 u'_1 & v'_1 v_1 & v'_1 & u_1 & v_1 & 1 \\ u'_2 u_2 & u'_2 v_2 & u'_2 & v'_2 u'_2 & v'_2 v_2 & v'_2 & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u'_8 u_8 & u'_8 v_8 & u'_8 & v'_8 u'_8 & v'_8 v_8 & v'_8 & u_8 & v_8 & 1 \end{bmatrix}$$

and $e = (e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33})$.

In Equation 4, $(u_{1\cdots8}, v_{1\cdots8})$ and $(u'_{1\cdots8}, v'_{1\cdots8})$ are obtained points from the images of the left and right cameras, where $e_{1\cdots8}$ are the components of essential matrix E. Since Equation 4 is too time-consuming to process, $e_{1\cdots8}$ are computed using singular values decomposition (SVD).

4 Experimental Results and Analysis

The proposed method, which estimates the coordinates of a 3D object using a planar structure for video-based AR, has been tested for the camera motion information to the coordinates of a 3D object created on an image in a sequence of frames. Our method applies the camera motion parameters to the coordinates of a 3D object in the frames of a video, and compares the estimated direction of the Z-axis with the direction of the real Z-axis. Fig. 1 shows the measured difference between the estimated and real directions of the Z-axis. The accumulated error increases towards the end of sequences. Fig. 2 shows examples of superimposition of an object located at different backgrounds in video sequence.

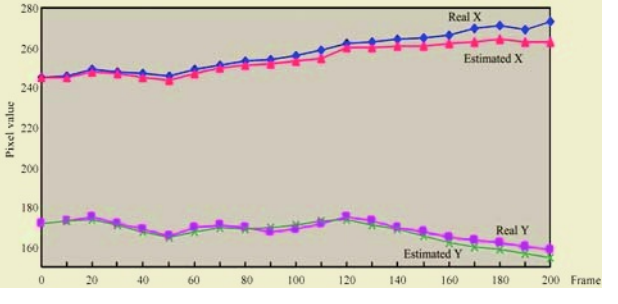

Fig. 1. Comparison for the registration between the estimated and real directions of Z-axes

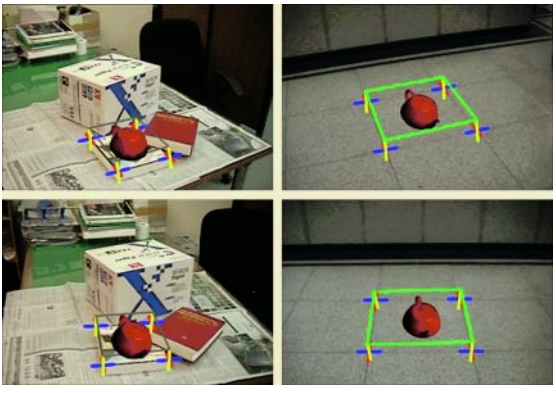

Fig. 2. Superimposition of a teapot in video sequence

References

1. Ronald T. Azuma: A Survey of Augmented Reality, Teleoperators and Virtual Environments, Vol. 6, No. 4, pp.355-385, 1997
2. Gilles Simon and Marie-Odile Berger: Estimation for Planar Structures, IEEE Computer Graphics and Applications, Vol. 22, pp.46-53, 2002
3. Simon J.D. Prince, Ke Xu and Adrian David Cheok: Augmented Reality Camera Tracking with Homographies, IEEE Computer Graphics and Applications, Vol. 22, pp.39-45, 2002
4. Gilles Simon, Andrew W. Fitzgibbon and Andrew Zisserman: Markerless Tracking using Planar Structures in the Scene, Proc. International Symp. Augmented Reality, pp.137-146, 2000
5. Peter Sturm: Algorithms for Plane-Based Pose Estimation, Proc. of the Conference on Computer Vision and Pattern Recognition, pp.1010-1017, 2000
6. Kiriakos N. Kutulakos and James R. Vallino: Calibration-Free Augmented Reality, IEEE Trans. on Visualization and Computer Graphics, Vol. 4, pp.1-20, 1998
7. Yong duek Seo and Ki Sang Hong: Calibration-Free Augmented Reality in Perspective, IEEE Trans. on Visualization and Computer Graphics, Vol. 4, No. 6, pp.346-359, 2000
8. Kumar Rakesh, Sawhney, H. Sawhney and Allen R. Hanson: 3D model acquisition from monocular image sequences, Proc. of the Conference on Computer Vision and Pattern Recognition, pp.209-215, 1992