Computing Homography with RANSAC Algorithm:

A Novel Method of Registration

Xiaowei Li, Yue Liu, Yongtian Wang, Dayuan Yan

Department of Opto-electronic Engineering, Beijing Institute of Technology

Beijing 100081, China

Tel: (+86 10) 6891 2565

Email: {nemesis, liuyue, wyt, ydyuan}@bit.edu.cn

ABSTRACT

An AR (Augmented Reality) system can integrate computer-generated objects with the image sequences of real world scenes in either an off-line or a real-time way. Registration, or camera pose estimation, is one of the key techniques to determine its performance. The registration methods can be classified as model-based and move-matching. The former approach can accomplish relatively accurate registration results, but it requires the precise model of the scene, which is hard to be obtained. The latter approach carries out registration by computing the ego-motion of the camera. Because it does not require the prior-knowledge of the scene, its registration results sometimes turn out to be less accurate. When the model defined is as simple as a plane, a mixed method is introduced to take advantages of the virtues of the two methods mentioned above. Although unexpected objects often occlude this plane in an AR system, one can still try to detect corresponding points with a contract-expand method, while this will import erroneous correspondences. Computing homography with RANSAC algorithm is used to overcome such shortcomings. Using the robustly estimated homography resulted from RANSAC, the camera projective matrix can be recovered and thus registration is accomplished even when the markers are lost in the scene.

Keywords: AR, registration, plane tracking, **RANSAC**, homography

1 INTRODUCTION

Augmented Reality, or AR, is a new technology with which computer-generated virtual objects can be merged into the image sequences of real scene. Rather than immersing the user into an absolutely virtual world as a Virtual Reality (VR) system does, an AR system allows its users to view the real environment with the virtual objects imposed on. It is widely applied in many fields such as medical care, entertainment, manufacturing, outdoor navigation and cartoon film etc. [1]. The demand for "Seamless" combination makes registration technique one of the key issues in an AR system.

Registration, also known as pose estimation or camera tracking, is used to obtain the motion variables of a camera by hardware or computer vision analysis method, or the combination of both of them, in either a real-time or off-line way. Accurate registration result can ensure the integration of virtual objects into the image sequences.

Current vision registration methods can be classified as model based and move-matching based. When we have already defined a model of the target scene, 3D/2D feature correspondences can be extracted for every frame. Using

these correspondences, though the features would be of various kinds (points, lines, free curves, etc.), cost functions or equations based on the pose variables are setup, because these correspondences must satisfy certain camera model [1][2]. Registration is completed after minimizing the cost functions or solving the equations. The model itself would be either accurate or crude. Accurate model is relatively difficult to get, and crude model is always sufficient to use. Move-matching refers to the multiple view geometry among several adjacent frames. It generally provides the relationship of the 2D/2D feature correspondences in two or three frames, which is actually a tensor. Such tensor among more frames could be also obtained, but it seems to be too complex to use. Similarly, registration could also be achieved by minimizing a cost function or solving a set of equations. However, such kind of registration only gets a relative camera motion pose, because no explicit world coordinates is defined and it is always set to the first frame's camera coordinate as an arbitrary one.

Registration with markers is extensively used in AR systems for its accuracy and real time quality. When markers are introduced into the scene, the scene is modified. Because of this modification, a simple model of the scene can be gained. Thus, this kind of registration is actually model based.

Obviously, when markers are lost in certain frames, the registration will face failure, since no 3D/2D correspondences exist. Noting the point that in many situations, the marker is a planar board with more than four corners in it, and is always placed on planar objects in the scene, like grounds, walls, roads, etc., a registration method with markers is proposed, which can still achieve reliable result when markers are lost in certain frames.

The typical situation of the problem is that in an image sequence, the marker can be extracted in frames k-2 and k-1, and in frame k and subsequent frames k+1, k+2, $\cdots$, k+n, the marker is lost or not complete, which will resulting in erroneous registration. In frame k+n+1, the marker comes into sight again, and the original method succeeds once more. Our method is dealing with the registration in subsequence of frames k-2, k-1, $\cdots$, k+n.

This method combines the virtues of both model based and move-matching based methods. The core issues involve plane detection, tracking, and robust estimation of homograhpies of the plane between consecutive frames.

2   HOMOGRAPHY IN MULTIPLE VIEW GEOMETRY

Since homography simplifies the relationship between correspondences and its computation cost is relatively small, homograhpy is used extensively in computer vision systems and plays an important part in multiple view geometry in computer vision.

Single view geometry indicates how the world scene is imaged into the final image we see, which is also referred as camera model. A 3D/2D correspondence (x, X), where x and X are both in homogeneneous form and the world coordinate is defined on the plane 3D point X is laid on, should satisfy the following equation[3]:

$$x = PX = K[R|t] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K[r_1, r_2, t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = HX \qquad （1）$$

where P is known as projective camera matrix, K is the camera's intrinsic matrix, and R and t are the pose matrix, which

is referred as the rotation matrix and translation vector respectively. r_i is the ith column of R. H is the homography matrix relating 3D/2D correspondences on a world plane.

For a 2D/2D correspondence (x, x'), when it refers to 3D point on a plane, it will satisfy Equ. 2 [3][4]:

$$x' = Hx \quad H^{-1}x' = x \qquad (2)$$

where H is the homography matrix relating two views of planar target. Unlike fundamental matrix F in epipolar geometry, which images a point to a line, H describes a "Point-Point" imaging. H is a full rank matrix. Four point correspondences can be used to get a unique solution of H up to a scale.

3   PLANE DETECTION

In frame k-2 and k-1, the marker can be seen, and extract accurate 2D/2D correspondences can be extracted in it. With these 2D/2D correspondences, accurate homography H could be gained. Obviously, the marker would be laid on the world plane. The boundaries defined by the corners on the marker in extended to certain scale in frame k-2 and k-1, and these extended boundaries could be seen as the crude regions of the world plane in the two frames.

After performing Harris corner detection in the two plane regions and matching them in a classic way, a point correspondence set P can be obtained. Using the accurate H, every correspondence is tested and the one approximately satisfying Equ. 2 are considered as the images of 3D points lying on the same world plane as the marker does. Removing the correspondences doesn't satisfy Equ. 2, and the plane boundaries in frame k-2 and k-1 can be updated using the left ones. The final boundaries in the two frames are much more reliable.

4   PLANE TRACKING

When the markers in frame k and the subsequent frames k+1, k+2, ···, k+n are lost, the plane must be tracked to accomplish the final registration. Take the situation between frames k-1 and k for example.

First we contract the plane region boundary in frame k-1 to make sure most corners within this region are highly "likely" on the same plane. Then, Harris corner detection is performed in frame k and correspondences between the resulting corners and the corners in the plane region of frame k-1 are detected using the proximity and similarity of their intensity neighborhood. The acquired correspondence set can be denoted as P.

Surely, P has outliers. RANSAC algorithm can help the computing of homography beginning with P, known as putative correspondence set. RANSAC is a modern and robust estimation method. Computing H with RANSAC has been proved to be relatively fast and robust. The final results of this method are a set of inliers and a reliable H. [3], then boundary of the plane region in frame k-1 and k could be obtained with the set of inliers. Now, these new boundaries are extended to a larger scale again, plane detection process described in section 3 are performed using the reliable H and the updated plane regions in the two frames can be obtained. Then the plane can be tracked[6].

5 POSE COMPUTING

When reviewing Equ. 2, if the homography H relating 3D and 2D plane is already known and the camera has been calibrated beforehand, the projective matrix could be recovered as:

$$\text{H} = \text{K}[r_1, r_2, \text{t}] \qquad \text{P} = \text{K}[r_1, r_2, r_1 \times r_2, \text{t}] \tag{3}$$

where the unit orthogonality quality of rotation matrix is used. Actually, Equ. 3 is the basis for many marker based registration method[3].

Taking the situation in frame k-1 and k for example, frame k is the first frame that loses marker in the sequence. 3D/2D homograhpy **H** is gained for frame k, and robust 2D/2D homography **H'** is gained between frame k-1 and k during plane tracking procedure. The 3D/2D homography **H''** for frame k can be recovered as [3]:

$$\textbf{H''= H'H} \tag{4}$$

Then the camera projective matrix **P** for frame k can be recovered using Equ. 3. For the subsequent frames k+1, k+2, $\cdots$, k+n, the same process is done. Thus the registration can be accomplished even when the markers are lost.

6 CONCLUSION

In this paper, a registration method to be used for the situation when the markers lying on a textured world plane lost in some frames is proposed. Although using the GRIC function [7] is a new method to detect planes in the scene, the proposed method seems to be much easier to implement, and relatively fast. However, because the method computes the projective matrixes of frames losing marker based on the previous frame's registration result, there would be accumulative errors and this is what our future work will solve.

REFERENCES

[1] Ronald Azuma. Recent Advances in Augmented Reality. Computers & Graphics, November 2001.

[2] Jing Chen. Research in The Spatial Registration Technology of A Video AR System, Doctoral Dissertation, Beijing Institute of Technology, 2002.

[3] Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision, Second Edition. Cambridge University Press, 2004.

[4]Daniel F. Dementhon and Larry S. Davis. Model-based object pose in 25 lines of code. International Journal of Computer Vision, 15:123-141, 1995.4, 2000.

[5] Simon J.D. Prince, Ke Xu, and Adrian David Cheok. Augmented Reality Camera Tracking with Homographies. Computer Graphics, Vol. 22, No. 6, November/December 2002, pp.39-45.

[6] Gilles Simon and Marie-Odile Berger. Reconstructing while registering : a novel approach for markerless augmented reality. In International Symposium on Mixed and Augmented Reality - ISMAR'02, Darmstadt, Germany, September 2002.

[7] Surviving dominant planes in uncalibrated structure and motion recoveryM. Pollefeys, F. Verbiest, L. Van Gool, Surviving dominant planes in uncalibrated structure and motion recovery, A. Heyden, G. Sparr, M. Nielsen, P. Johansen (Eds.) Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Lecture Notes in Computer Science, Vol.2351, pp. 837-851.