A tracking by detection approach for robust markerless tracking

Chunrong Yuan

Fraunhofer FIT
Schloss Birlinghoven
D–53754 Sankt Augustin
chunrong.yuan@fit.fraunhofer.de

ABSTRACT

Augmented reality needs tracking technologies to render virtual objects into the field of view of the user. The success of ARToolkit demonstrates the advantages of optical tracking systems. Recently, much research has been carried out on markerless optical tracking. As an alternative to markers and due to the potential applications in industry, markerless tracking has got much attention and acceptance. In this paper we would like to survey the current state of the art of markerless tracking, analyze the challenges, present our current work, and discuss related issues.

Keywords: markerless tracking, augmented reality, invariant feature detection, matching, pose estimation

1 INTRODUCTION

Augmented Reality (AR) requires accurate registration and visualization of virtual objects relative to the real world. Among different software and hardware technologies, computer vision based approach is regarded as one of the most promising solutions to the registration problem. Using e.g. a camera attached to a head–worn display, the viewpoint of an AR user can be registered based on the changes in the video stream captured by the camera.

Based on simple image segmentation, marker–based tracking system can perform fast detection and identification of the markers. But in some critical AR applications (e.g. service maintenance in automotive industry), it is impossible to use markers. Furthermore, marker–based tracking is very sensitive to occlusions. Tracking stops immediately even if only a tiny portion of a marker is occluded.

An alternative to markers is to use features which exist naturally in the tracking environment. Such features can be edges, corners and/or textures. One approach to markerless tracking is based on temporal registration and prediction of the features [1] [7]. While real–time performance can be achieved, manual initialization of the tracker has to be performed before tracking commences. In [1], two images of the scene are captured beforehand and the initial 3D positions of the camera have to be determined manually. In [7], the camera must be kept stable in the beginning so that manual initialization can be carried out. Besides the problem of initialization, further concerns are the drift and the reinitialization problem.

The model–based approach [4] [8] has also been applied for markerless tracking. Usually a 3D model is required so that 2D features extracted from the video frame can be matched to those belonging to the 3D model. In principle, the building of a 3D model is difficult and time–consuming, particularly for large–scale environment.

Sensor fusion based approaches [3] [5] have also been applied for markerless tracking, where other sensors are used to compensate optical tracking. Due to the introduction of additional devices into the tracking system, the alignment of the device coordinate system with that of the camera must be handled carefully.

2 CHALLENGES

Considering the factors of speed, accuracy and flexibility, the performance of the current markerless tracking approaches is still beyond what real–world AR applications demand. The stringent requirements have made the development of markerless trackers a challenging task. The user who wears the head mounted camera should be allowed to move freely, which makes the motion of the camera unpredictable. The tracking environment changes constantly and dynamically. Despite environmental changes and user invention, the 3D camera pose must be estimated, both robustly and accurately.

Accordingly, the solution to markerless tracking should satisfy a number of criteria. Among other things, it should

1. need few or at least easy offline preparation
2. capable of automatic initialization/reinitialization
3. be able to detect natural features reliably
4. have a reasonable computation cost
5. be able to provide accurate augmentation in real time
6. work in large–scale and unconstraint environment
7. be flexible and adaptive to the application needs

With these considerations, we present in what follows our solution to the markerless tracking problem.

3 MARKERLESS TRACKING BY DETECTION

In common with the well–known ARToolkit, we use the approach of tracking by detection. Only the current camera frame is used for pose estimation. Moreover, we don't require a 3D model of the environment. Inspired by [2], we developed a new local invariant feature detector, which can facilitate robust matching and accurate pose estimation.

Only limited offline preparation is required. A single image of the environment is captured as reference. From the reference image we select a plane by identifying its four corner points. According to application needs, the plane can be chosen from anywhere in the scene. Shown in Figure 1 on the left is the reference image. The rectangle whose edges and corners are drawn in blue illustrates the plane we selected.

Figure 1: The reference image and one of the tracked frames.

We use the plane for both registration and augmentation purposes. Based on the plane, a reference coordinate system can be defined. For example, the origin can be put at the center of the rectangle, with x axis pointing to the right and y axis pointing up. Having such a reference coordinate system, the camera pose can be estimated relative to it. As a consequence, our approach does not require 3D engineering of the tracking environment.

During the online tracking stage, local invariant features are extracted from the current image and matched against those features extracted from the reference image. Matching of two feature points is based on an Euclidean distance measure. Candidates of matched pairs are selected based on a global threshold. The matches are further verified by fitting an affine transform to those matched points. Those matches that do not agree with the transform are regarded as outliers and discarded. From the remaining inliners, a final transform matrix between the current and the reference image is calculated.

Having obtained the transform matrix, we can easily locate the reference rectangle in the current frame. Now the problem becomes finding the 3D camera pose with four coplanar points whose configuration is known. A homography–based method can be applied to find the rotation and translation of the camera. Since pose estimation based on a single projective image may suffer from the problem of pose ambiguity, the initial calculated pose (particularly the rotation part) can be incorrect. In order to refine the pose parameters, a more robust pose estimation method [6] is applied.

A camera image has quite a lot of noise. Due to this reason, small jitter may be observed in the estimated camera poses if the camera remains still. A simple method has been used to get rid of the jitter. We compute a difference image between the current and the last camera frame. Based on the difference image, we can detect whether the scene has changed or not. We calculate a new camera pose only when camera motion has occurred.

We have tested the approach in several experiments with different scenarios. One of the experiment was carried out before a large operation panel. For each camera image, the camera pose is calculated in real–time (20–25 frames/s on a conventional laptop). The pose matrix can be applied for annotation and augmentation purposes. Shown in Figure 1 on the right is one tracked frame with augmentation. We render into the tracked scene as augmentation the reference coordinate system. The two axes are visualized with text annotations in different colors. The visually correct augmentation indicates that the camera pose has been calculated precisely.

Figure 2: Further examples of tracking and augmentation.

Some further examples of augmented frames are shown in Figure 2. As can been seen, the augmentation remains correct despite large viewpoint changes and occlusions.

The operation panel used for our experiment is located in a corridor. A glass wall which can not be seen in the pictures stands behind the camera. Due to this reason, the illumination changes constantly. Thanks to the invariant ability of our feature detector, robust tracking and accurate augmentation have been achieved.

4 Discussions

A solution to markerless pose tracking has been presented in section 3. It satisfies most of the criteria listed in section 2 (definitely 1 to 5, partially 6 & 7). Tracking is based on the detection of salient feature points. We achieve robust matching, accurate pose estimation and real–time augmentation under environmental changes as well as unconstraint motion of the camera.

Since the approach does not need a 3D model, offline preparation is minimal. By using a tracking by detection approach and due to the robust wide–baseline matching algorithm, the system is capable of automatic initialization (actually the system reinitialize itself automatically all the time). With a high frame–rate, the tracking algorithm can be ported to run on mobile devices as well.

The approach could be extended to cope with large–scale environments by building a database of reference frames distributed evenly in the environment. With such an improvement, markerless tracking in unconstraint environment, e.g. an extended industrial plant, is possible.

We would like to integrate our solution to the markerless tracking problem into actual industrial applications. With the knowledge of the real industrial settings and specific application requirements, tracking algorithms can be improved and made adaptable to industrial needs. The improvement of tracking performance could then lead to widespread use of AR technology in industry.

References

[1] K.W. Chia, A.D. Cheok, and S.J.D. Prince. Online 6DOF augmented reality registration from natural features. In *First Int. Symposium on Mixed and Augmented Reality*, pages 305–313, Darmstadt, Germany, Oct. 2002.

[2] M. Grabner, G. Helmut, and B. Horst. Fast approximated SIFT. In *7th Asian Conference of Computer Vision*, pages 918–927, Hyderabad, India, Jan. 2006.

[3] R. Koch, K. Koeser, B. Streckel, and J.F. Evers-Senne. Markerless image-based 3d tracking for real-time augmented reality applications. In *The 7th Int. Workshop on Image analysis for multimedia interactive services*, Montreux, Switzerland, April 2005.

[4] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua. Fully automated and stable registration for augmented reality applications. In *Second Int. Symposium on Mixed and Augmented Reality*, pages 93–101, Tokyo, Japan, Oct. 2003.

[5] H. Najafi, N. Navab, and G. Klinker. Automated initialization for marker–less tracking: A sensor fusion approach. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 110–119, Airlington, VA, USA, Nov. 2004.

[6] G. Schweighofer and A. Pinz. Robust pose estimation from a planar target. In *Technical Report, TR-EMT-2005-01*, Graz University of Technology, May 2005.

[7] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *Int. Symposium on Augmented Reality*, pages 120–128, Munich, Germany, Oct. 2000.

[8] I. Skrypnyk and D.G. Lowe. Scene modeling, recognition and tracking with invariant image features. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 110–119, Airlington, VA, USA, Nov. 2004.