ISMAR03 in Tokyo, Japan, pp.342-343 (2003)

Fiducial-less 3-D Object Tracking in AR Systems Based on the Integration of Top-down and Bottom-up Approaches and Automatic Database Addition

Takashi Okuma*+, Takeshi Kurata#+, and Katsuhiko Sakaue+
*Columbia University #University of Washington
+National Institute of Advanced Industrial Science and Technology
takashi-okuma@aist.go.jp, kurata@ieee.org, k.sakaue@aist.go.jp

Abstract

We propose a novel fiducial-less 3-D object tracking method. Our method consists of three components: 1) bottom-up approach (BUA), 2) top-down approach (TDA), and 3) automatic database addition (ADA). An experimental result shows an accuracy and robustness of our method.

1. Introduction

Fiducial-less tracking methods are one of the most important issues in the field of Augmented Reality (AR) [1-4]. They determine geometric relations between a real object and users' viewing position using images. AR systems use the geometric relations to register real and virtual objects with each other. These methods don't need to set any sensing devices nor fiducial markers in the environment, so that they are appropriate for video see-through AR systems and wearable AR systems [3-5].
These methods require a correspondence between known points and local features in the image, so they often use local feature detection/matching techniques. In Ref. [1], Fischler and Bolles mentioned local feature detectors make often two types of errors: classification errors and measurement errors. These two types of errors occur in situations that lighting condition and/or viewing position changes drastically. In addition, classification errors occur when some local features are occluded. Therefore, the tracking method needs not only to use techniques for accurate local feature detection/matching, but to use techniques for parameter estimation with local feature matching result that includes some classification errors and measurement errors. These methods can be classified by how they estimate the parameters using such local feature matching results.

2. Related Works

We classify fiducial-less tracking methods into two types:
- Methods based on the bottom-up approach (BUA),
- Methods based on the top-down approach (TDA).

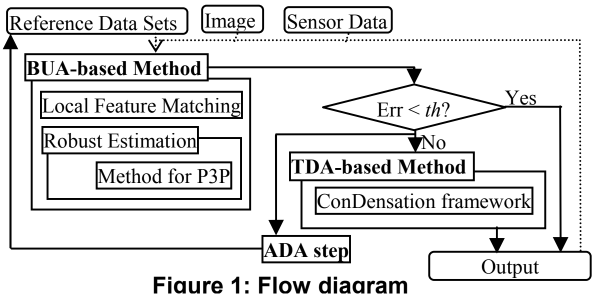

Figure 1: Flow diagram

BUA-based methods calculate 6 parameters that express the geometric relation using known 3-D position of reference points and detected 2-D position of them. Robust estimation techniques are useful to calculate appropriate parameters form the local feature matching results, which often includes the two types of errors. Tracking methods based on RANSAC [1][2] show good experimental results. However, they don't deal with change of local feature appearance that causes drifts.
On the other hand, TDA-based methods predict the six parameters of each frame using history of the parameters, motion-models, etc. Then, they search six parameters that have the smallest error around the predicted parameters. Random sampling makes calculation cost for the search reduce. We applied ConDensation framework [6] to acquiring the 6 parameters in Ref. [4]. We used a simple criterion to evaluate each sample, so that tracking easily failed when an object to be tracked moved quickly.

3. Proposed Method

Our proposed method is based on a hybrid framework of BUA-based method and TDA-based method. Figure 1 shows a flow diagram of the proposed method. In the proposed method, a reference data set of an object being tracked is composed of three components: 1) an image which the object appears in, 2) image-coordinate values of reference points in the reference image, and 3) object-coordinate values of the reference point. An appearance of local features changes with the geometric relation between the object and the viewpoint, so that one reference data set cannot be used for matching with large

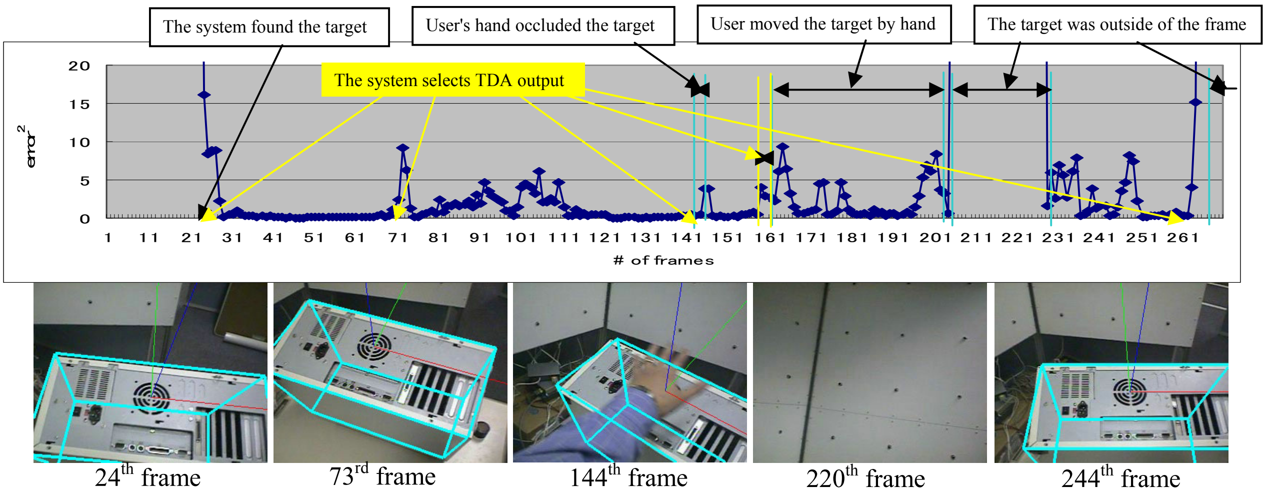

Figure 2: error value and output stills of the experiment.

area. Therefore, multiple reference data sets are required to track the object in large area. To simplify the preparation of data sets, we integrate Automatic Database Addition (ADA) step into our tracking method. The ADA step automatically adds a new data set when the proposed method is tracking the object.

The proposed method mainly outputs the result of the BUA-based method, because the BUA-based method acquires accurate results with small calculation cost. We use an LMedS framework as a robust estimation in BUA. We also use the Finsterwalder's method to solve the P3P problem because Ref. [8] shows its calculation error is smaller than other P3P methods. The Lucas-Kanade local feature tracking method [7] is used for local feature matching. To prevent drift of local features, local feature detector uses reference point data stored in advance rather than the tracking result of previous frame.

When output of the local feature matching includes classification errors more than more 50% of the number of features, the BUA-based method estimates inaccurate parameters. In this case, our method starts TDA-based estimation [4]. This works mainly when large part of an object being tracked is occluded. This method also uses an inertial orientation sensor because sensor fusion techniques help fiducial-less tracking methods when input image blurs because of quick movement [9]. The sensor output is used to predict position of local feature in the BUA and to predict 3 rotation parameters in the TDA.

We evaluated our method with a real data sequence. We prepared a single data set in advance. Figure 1 shows the error values as well as some synthesized images. The proposed method was able to continuously track the real object even though the user's hand occluded the object being tracked. The processing time per each frame was, on average, 33 ms when only the BUA-based method was used, and about 100 ms when both the BUA-based and TDA-based method were used (Intel dual Xeon 1.7GHz).

Acknowledgement

This work is supported in part by JSPS Postdoctoral Fellowships for Research Abroad (H14 and H15).

References

[1] Fischler, M. A. et al., "Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Comm. of the ACM, Vol.24, No.6, pp.381-395, 1981.

[2] Simon, G. et al., "Reconstructing while registering: A novel approach for markerless augmented reality, In Proc. ISMAR02, pp.285-294, 2002.

[3] Okuma, T. et al., "Real-Time Camera Parameter Estimation for 3-D Annotation on a Wearable Vision System," IEICE Trans. Inf. Syst., Vol.E84-E, No. 12, pp.1668-1675, 2001.

[4] Okuma, T. et al., "VizWear-3D: A Wearable 3-D Annotation System Based on 3-D Object Tracking using a Condensation Algorithm," In Proc. VR2002, pp.295-296, 2002.

[5] Weavy web site, http://www.is.aist.go.jp/weavy/.

[6] Isard, M. A., "Visual Motion Analysis by Probabilistic Propagation of Conditional Density," Ph.D. thesis, Department of Engineering Science, University of Oxford, 1998.

[7] Lucas, B.D. et al., "An Iterative Image Registration Technique with an Application to Stereo Vision," Proc. DARPA Image Understanding Workshop, pp.121-130, 1981

[8] Haralick, R.M. et al., "Analysis and solutions of the three point perspective pose estimation problem," Proc. CVPR '91, pp.592-598, 1991.

[9] Yokokoji, Y. et al., "Accurate Image Overlay on Video See-through HMDs Using Vision and Accelerometers," Proc. IEEE VR2000, PP.247-254, 2000.