Initializing Markerless Tracking Using a Simple Hand Gesture

Taehee Lee* Tobias Höllerer†

Four Eyes Laboratory, Department of Computer Science
University of California, Santa Barbara, California 93106 USA

ABSTRACT

We introduce a technique to establish a coordinate system for augmented reality (AR) on table-top environments. A user's hand is tracked and the fingertips on the outstretched hand are detected, providing a camera pose estimation relative to the hand. As a user places the hand on the surface of a table-top environment, the hand's coordinate system is propagated to the environment, detecting distinctive image features in the scene. The features are tracked fast and robustly using optical flow. In this way, a new tabletop AR environment is set up without having to carry a marker or a sophisticated tracking system to the environment itself. We also demonstrate a proof-of-concept application for establishing a table-top AR environment and recognizing a scene when detecting its features.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; I.4.8 [Image Processing and Computer Vision]: Scene Analysis

Keywords: position and orientation tracking technology, vision-based registration and tracking, interaction techniques for MR/AR

1 INTRODUCTION

In recent years, mobile computing devices have been widely deployed to customers, including cellphones, PDAs, small laptops and emerging wearable computers. As the need for mobility is growing larger, devices are becoming smaller and easier to carry with a user. In order to assist users with Augmented Reality (AR) wherever they visit, the AR system also needs to be easily accessible anywhere. Although there are many systems available to start AR in a prepared environment, we want to lower the barrier to initiate AR systems so that users can easily experience AR anywhere [5]. Considering a mobile user entering a new work place such as an office desk environment, it is necessary to enable the user to start using AR without spending much effort on setting up the environment.

In this paper, we present a technique to establish a global coordinate system for a table-top AR environment by estimating a camera pose relative to a user's outstretched hand. Handy AR [9] has shown that a user's bare hand can replace a cardboard marker [8][2] for local coordinate systems. We introduce Handy AR in our work as providing an initial camera pose estimation with scale information, as well as a user interface for interactions. The coordinate system from the user's hand is propagated to the table-top AR environment, detecting distinctive image features of the scene. The features are tracked using a hybrid tracking mechanism by combining distinctive features and optical flow. The distinctive features are used to recognize a scene, providing a scheme to continue a stabilized AR experience in different places.

*e-mail: taehee@cs.ucsb.edu
†e-mail: holl@cs.ucsb.edu

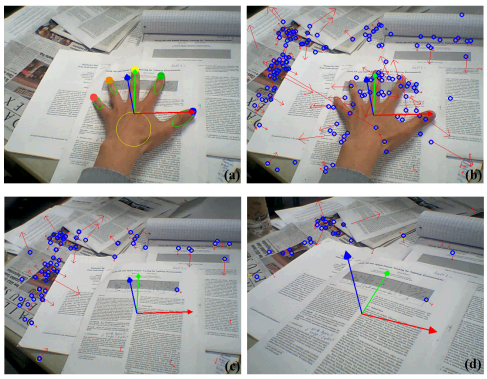

Figure 1: Establishing a coordinate system using (a) Handy AR, and (b) propagating it to the scene as detecting features. The camera pose is estimated from the features (c) after moving the hand out of the scene and (d) from different angles.

2 METHOD DESCRIPTION

We establish a coordinate system for a planar table-top AR environment using fingertip tracking, and introduce a hybrid feature tracking method that combines distinctive invariant feature detection and optical flow-based fast feature tracking. A camera pose is estimated from the tracked features relative to the 3D points that are extracted from the scene while establishing the coordinate system for the environment.

2.1 Handy AR

Handy AR [9] is used for estimating a 6 *degree-of-freedom* camera pose, replacing a cardboard marker with a user's outstretched hand. The hand is segmented by a skin-color-based classifier [7] with an adaptively learned skin color histogram, and it is tracked over frames being initially the largest blob in the view. Fingertips are located on the contour of the hand by finding points that have high curvature values and are accurately fitted into ellipses, being tracked over frames and used for estimating a camera pose based on the measured locations of the fingertips relative to each other. Five tracked fingertips provide more than the minimum number of point correspondences for a pose estimation algorithm [11]. The estimated camera pose is then used as an initial pose for further 3D scene acquisition and camera tracking as shown in Fig 1a. In this way, users do not have to carry any marker or tracking device for AR with them.

2.2 Distinctive Feature Detection and Tracking

Distinctive invariant features [10], also known as SIFT features, are extracted from a captured video frame. From the camera pose estimated using the Handy AR system as described in the previous section, the SIFT features are unprojected to a plane that is parallel to the hand, resulting in 3D locations of the features in a new world coordinate system. Given a captured video frame, newly detected SIFT features are matched to a reference frame's features, providing point correspondences for estimating a camera pose [11]. In order to more accurately match the SIFT features, the RANSAC algorithm [3] is used, eliminating outliers for pose estimation. This

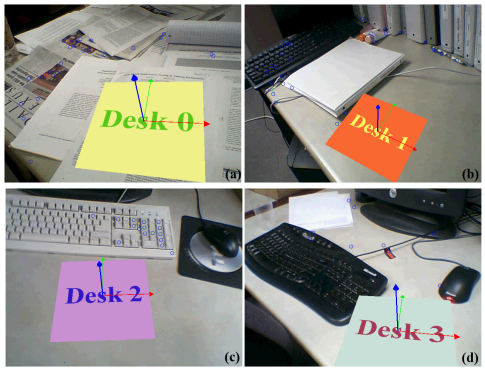

Figure 2: Snapshots of recognizing a scene. Each desk is indicated with an annotation.

also removes features that are actually not on the same working plane, allowing the features to maintain the planarity constraint we assume.

Given two frames, the interest points in one frame are tracked with regard to the other frame by iteratively computing optical flow using image pyramids [6], providing a fast tracking mechanism. The interest points to track are selected from detected SIFT features in a multi-threaded manner. SIFT detection runs at about 1.3 frames per second(fps), and optical flow at 29.8fps. The new features are added as new interest points to track if they are more than a threshold distance apart from any other currently tracked features. Incorrectly tracked features are dropped by the RANSAC algorithm.

2.3 Propagating an Established Coordinate System

Once an initial camera pose is estimated by Handy AR, the coordinate system based on the hand is propagated to the global coordinate system of a table-top or desk-top AR environment as the user places the hand on top of the surface plane. Fig 1 shows the steps of propagating the coordinate system. The camera pose from Handy AR is used to unproject the detected SIFT features to the plane (see Section 2.2), calculating the world coordinates of the features. Tracking the features for each frame, the camera pose is estimated from feature point correspondences. Thus, the user may move the hand out of the scene and start interactions in the AR environment. Note that the area that the hand is covering in Fig 1b has several SIFT features, but after moving the hand out as in Fig 1c, those features on the hand are not detected any more.

3 Results

We tested our implementation of hybrid feature tracking with regard to the speed and robustness of the system. The results show that our method is useful for establishing a coordinate system for table-top environments and is robustly tracking a camera pose for real-time AR applications, recovering from tracking failures, removing accumulated drift errors. We have tested the system using a smaller search scale space than existing implementations [4] of SIFT [10] features, while running our hybrid feature tracking method in a multi-threaded framework for real-time AR.

We have implemented a proof-of-concept application for table-top augmented reality environments using Handy AR and our hybrid feature tracking method. A user establishes coordinate systems for several desks separately as initial steps for each space. The distinctive features are stored persistently, noting their world coordinates. Fig 2 shows that a scene is recognized successfully and its annotations are registered according to the coordinate system that the user established previously. This enables users to personalize their tabletop AR system while moving into new locations as long as their coordinate systems are set up once with Handy AR.

4 Conclusions and Future Work

We introduced a method for establishing a coordinate system for tabletop AR environments using our Handy AR system. By detecting and tracking distinctive image features in the scene while propagating the coordinate system, the camera pose can be estimated from the tracked feature point correspondences with the planar surface of a table-top environment. Our experiments show that this can be used for setting up an AR environment without requiring much effort from users at a new location. The distinctive features used for estimating a camera pose are also effective to recognize a scene so that users can continue the AR interaction when moving between different locations.

We encountered interesting problems while experimenting with propagating a hand's coordinate system to the environment. As non-planar objects and the features on them may break the planarity assumption, reconstructing 3D structure of the scene is a challenging topic. From the initial camera pose from Handy AR, we can stabilize a reference frame. Subsequently a user could extend the camera tracking space of the scene by detecting landmark features incrementally [1] while moving the camera, assisting the system interactively.

5 Acknowledgement

This research was supported in part by the Korea Science and Engineering Foundation Grant (#2005-215-D00316), by a research contract with the Korea Institute of Science and Technology (KIST) through the Tangible Space Initiative project, by NSF grant #IIS-0635492, and by the NSF IGERT in Interactive Digital Multimedia grant #DGE-0221713.

References

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, June 2007.

[2] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 590–596. IEEE Computer Society, 2005.

[3] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[4] R. Hess. SIFT feature detector. http://web.engr.oregonstate.edu/ hess/ June, 2007.

[5] T. Höllerer, J. Wither, and S. DiVerdi. *Anywhere Augmentation: Towards Mobile Augmented Reality in Unprepared Environments*. Lecture Notes in Geoinformation and Cartography. Springer Verlag, 2007.

[6] Intel Corporation. Open Source Computer Vision Library reference manual. December 2000.

[7] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. In *CVPR*, pages 1274–1280. IEEE Computer Society, 1999.

[8] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, October 1999.

[9] T. Lee and T. Höllerer. Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. Submitted to: *International Symposium on Wearable Computers*, 2007 (under review).

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] Z. Y. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.