Natural Feature Tracking for Extendible Robust Augmented Realities

Jun Park, Suya You, Ulrich Neumann

Computer Science Department
Integrated Media Systems Center
University Of Southern California
Los Angeles, California 90089-2561
junp@usc.edu, suyay@graphics.usc.edu, uneumann@usc.edu
Phone: (213) 740-4238 Fax: (213) 740-7285

Abstract

Vision-based tracking systems for augmented reality often require that artificial fiducials be placed in the scene. In this paper we utilize our approach for robust detection and tracking of natural features such as textures or corners. The tracked natural features are automatically calibrated to the fiducials that are used to initialize and facilitate normal tracking. Once calibrated, the natural features are used to extend the system's tracking range and to stabilize the tracked pose against occlusions and noise. The emphasis of this paper is the integration of natural feature tracking with fiducial tracking to increase the range and robustness of vision-based augmented reality tracking.

1. Introduction

Many systems have been developed to track the six-degree of freedom (6DOF) pose of an object (or a person) relative to a fixed coordinate-frame in the environment [FOXL98, GHAZ95, MEYE92, STAT96, WELC97]. These tracking systems involve a variety of sensing technologies, each with unique strengths and weaknesses, to measure a world coordinate pose, as required for virtual and augmented reality applications.

However, a large class of AR applications require annotation on objects whose positions in a room or the world may vary freely without impact on the AR media linked to them (for example, AR applications in manufacturing, maintenance, and training [CAUD92, FEIN93]). A more appropriate tracking approach for these mobile applications is one that is based on viewing the object itself [NEUM96, REKI97, SHAR97, UENO95]. Many of these AR systems depend on artificial fiducials (also called landmarks) or a prior-known model data to perform the dynamic alignment between a real and a virtual camera. These approaches are appropriate in situations where known and recognizable features are continually in view. However, it becomes difficult to ensure that this viewing constraint is satisfied in general application settings where off-screen fiducials or occlusion of fiducials may be difficult to prevent. In these cases, continued pose tracking depends on the utilization of any and all naturally occurring scene features. In this

paper, an approach is described to extend and stabilize AR tracking by using natural features.

The emphasis is the unique combination of components and the architecture of our system. To the best of our knowledge, the idea of combining uncalibrated natural features with fiducials in an AR tracking system is unique.

2. Operation Model

A sequence of how the system operates is described as follows. A user moves a camera that views the scene and the system continuously tracks the fiducials as well as any natural features detected in the images. The camera pose, computed from the reference fiducials (i.e., by fiducial-based tracking), is used to calibrate the natural scene features as they are tracked over an extended set of frames. Once calibrated, the natural scene features are used as tracking primitives to compute pose in the absence of reference fiducials (i.e., by natural feature-based tracking). Consequently, our system tracks robustly over a wider range of views and under conditions where some or all of the reference fiducials are occluded.

Our system extends the use of a Kalman filter or a 3D-based recursive filter to compute structure from motion for calibrating tracked natural features [NEUM98b]. The detection and tracking of natural features is performed by a novel adaptive motion estimation approach [NEUM98a]. We select the most reliable natural features based on a tracking confidence metric and track these features in a multi-stage process that includes feedback of feature tracking confidence. The natural feature tracking is still an order of magnitude too slow (~1Hz) for real-time applications. For this paper, real-time video sequences are recorded and processed automatically (no user intervention) and then reassembled into a video sequence. We are currently investigating dedicated hardware and DSP processor implementations of the algorithms.

2.1 Fiducial-based Tracking

Fiducials have the advantage that they can be designed to maximize the performance of AR system to detect and distinguish between them, they can be inexpensive, and they can be placed arbitrarily on objects. Our fiducial design is a colored circle or triangle [NEUM96], but other designs such as concentric circles or coded squares are equally valid [STAT96, CHO98]. Detection strategies are often dependent on the characteristics of the fiducials [REKI97, UENO95]. Currently we are working on more robust and intensity-tolerant fiducial detection system using fuzzy and rule-based algorithm [CHO98].

Computing correspondences between 2D measurements and 3D features in fiducial database is hard in the general case [UENO95], and trivial if the fiducial types are unique for each element in the database. Because only a small number of initial reference fiducials are needed for this work of fiducial-based tracking, we used unique types of fiducials to solve 2D-3D correspondence problem.

In computing camera pose, three or more corresponded fiducials are needed [FISC81]. The approach we use has known instabilities in certain poses, and in general provides multiple solutions (two or four) from a fourth-degree polynomial. Methods have been proposed to select the most likely solution [SHAR97]. We weight several tests to rank the possible pose solutions in terms of their apparent correctness [NEUM98b].

2.2 Natural Feature-based Tracking

Our novel approach of natural feature-based tracking is to combine natural feature tracking and new point position estimation. It utilizes naturally occurring scene features to compute camera pose on unprepared environments.

The system detects natural features from the initial image and tracks in the following images, providing the image coordinates of the natural features for each image frame. These image coordinates of natural features, combined with camera pose calculated by fiducial-based tracking, are used to calibrate the natural features by new point position estimation process.

The correspondence of natural features between images is solved by feature tracking algorithm as explained in section 3. The estimated 3D positions and their corresponding image coordinates of the natural features can be used to calculate camera pose in case part or all of the fiducials are occluded or undetectable.

Our contribution is the unique combination of natural feature tracking and new point position estimation to track the camera pose in an uncalibrated environment. Starting from a small set of calibrated fiducials, the camera tracking range can be extended dynamically and automatically. The user can also interact with the new environment, adding or modifying the virtual objects and real objects, because our tracking system is dynamically extendible to regions without calibrated features.

One of the issues related with natural feature-based tracking is the strategy of selecting natural features to be used for computing camera pose. Currently we have the strategy of dynamically choosing four features closest to the each corner of the image. This results in evenly distributed feature sets. We also plan to involve the uncertainties of 3D positions and 2D variances of the features.

3. Natural Feature Tracking

To perform detection and tracking of naturally occurring features, we developed a novel motion tracking approach [NEUM98a]. The system integrates three main motion analysis functions, e.g., feature selection, motion tracking, and estimate verification, in a closed-loop cooperative manner to deal with the complex natural imaging conditions.

In the feature selection module, two kinds of tracking features (points and regions) are selected and evaluated for their suitability for reliable tracking and motion estimation. The selection and evaluation processes are based on a tracking evaluation function that measures the suitability of features and the confidence of

tracking fed back from tracking module. Once selected, the features are ranked according to their evaluation values and then fed into the tracking module.

The tracking method we employed is a robust multi-stage optical flow estimation approach [NEUM98a]. For every region, an estimated motion is computed and fit to a model. Verification and evaluation are imposed to measure the confidence of the estimation and the model fit. If the estimation error is large, the motion estimate will be refined until the estimation error converges or the region is discarded as unreliable for tracking. In order to handle local geometric distortions due to large view variations and long sequence tracking, the translation model and the affine model are used for tracking point and region features, respectively. These models are the basis of the motion verification and evaluation processing. In regions, for example, the optical-flow motion estimate is computed and fit to an affine model that is used to warp the region into a confidence evaluation frame. The confidence evaluation frame is compared with the true target image to obtain a measure of tracking error.

Motion error feedback is an essential component of the architecture for robust tracking. The error information is fed back to the tracking module for motion correction and to the feature detection module for continuous feature re-evaluation. This re-evaluation of features keeps the tracking system working in an "optimum" state by selecting and maintaining the most reliable features. Although two different verification strategies are used for the two types of tracked features (points and regions) and their motion models (translation and affine), they both generate an evaluation frame that measures the estimation residual.

The closed-loop stabilization of the tracking system is inspired by the use of feedback for correcting errors in a non-linear control system. The process acts as "selection-hypothesis-verification-correction" strategy that make it possible to discriminate between good and poor estimation features, and maximizes the quality of the final motion estimation.

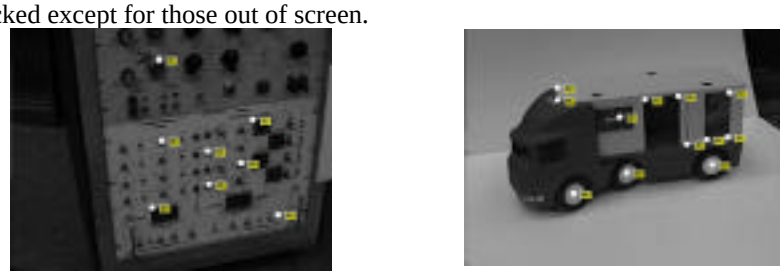
4. New Point Position Estimation

Two different recursive filters have been designed and tested to estimate 3D positions of new points based on the camera pose and measurements of the feature image coordinates [NEUM98b]. The results of synthetic and real data showed that both filters converged and were stable. The EKF (Extended Kalman Filter) is known to have good characteristics under certain conditions, however the RAC (Recursive Average of Covariances) filter gives comparable results, and it is simpler, operating completely in 3D-world space with 3D lines as measurements. The RAC filter approach eliminates the linearization processes required in the EKF with Jacobian matrices. More details about EKF can be found in references such as [MAYB79, MEND95].

The intersection[1] of two lines connecting the camera positions and the feature locations in the image creates the initial estimate of the 3D position of the feature. The intersection threshold is a design parameter and depends on the fiducial design or the natural scene and objects. These initial estimates often have quite good accuracy especially when the angles between the two lines are big enough (e.g. > 5 degree).

5. Results and Discussion

Two models (a rack and a truck) have been used for experiments. The image streams generated by a camera were directly digitized using a video editing system. The sampling rate was 15Hz resulting in about 250 frames from a 17 second video sequence for a rack model. For the truck model, the sampling rate was 25Hz resulting in about 340 frames of images from a 14 second video sequence. Figure 1 shows the result of natural feature tracking. In both rack and truck models, twenty features were detected in the first frame, twelve out of which were selected for tracking, rejecting others for being too close to fiducials or the already selected features. Even in later frames, all the features were accurately tracked except for those out of screen.



Rack model: 250th frame Truck model: 451st frame

Fig. 1 Results of Feature Detection and Tracking

Figure 2 shows the results of new point position estimation. Each chart indicates the convergence of X, Y, and Z coordinates of 3D positions of the natural features. They converged fast (at about 90th frame, i.e., in 6 seconds) and were stable after the convergence. It is noticeable that the initial estimates of Z coordinates were less accurate than X and Y coordinates, which is intuitive.

Figure 3 shows the results of camera pose tracking and virtual object/annotation overlay. The bigger dark circles indicate the projections of the estimated 3D positions of the natural features. The smaller bright circles indicate detected fiducials or the feature measurements resulted from natural feature tracking. The bright crosses indicate fiducials or natural features that were used for tracking. In cases where fewer than three fiducials were detected, 4 features (either tracked

[1] Since two lines in 3D space may not actually intersect, the point midway between the points of closest approach is used as the intersection.

natural features or detected fiducials) close to each corner of the image were selected to compute camera pose.

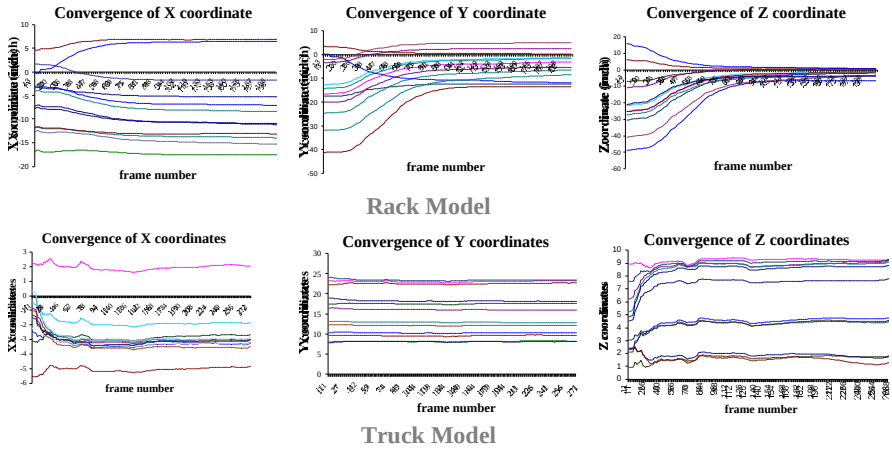
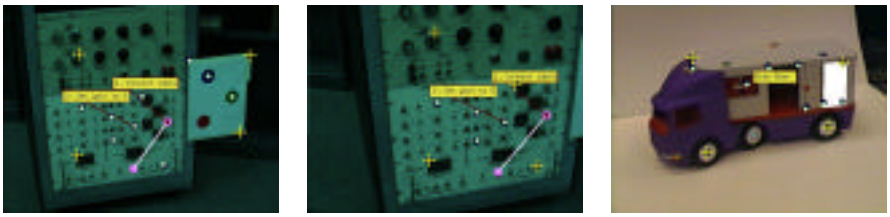


Fig. 2 Convergence of 3D positions of Natural Features



124th frame | 249th frame | 349th frame

Left and middle: Rack model- virtual annotations and a virtual line

Right: Truck model- a virtual annotation and a virtual white rectangle that covers a texture

Fig. 3 Result of Virtual Object/Annotation Overlay

The experiments of the current work were done in off-line but automatically. Most of the computation time went to the natural feature tracking, which is still an order of magnitude too slow (~1Hz) for real-time applications. Dedicated hardware and DSP processor implementations of the algorithms will hopefully allow on-line and real-time applications. We plan to set up a strategy reflecting the uncertainties of 3D positions and 2D variances of the features to choose features with more accurate 3D positions and 2D image coordinates. We also hope to build a more robust pose calculation algorithm, which adapts multiple features and is more tolerant to numerical instability. This is a position paper. A more detailed description about the work can be found in [PARK98].

6. References

[CAUD92] T. P. Caudell, D. M. Mizell, "Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes," Proceedings of the Hawaii International Conference on Systems Sciences, pp. 659-669, 1992

[CHO98] Y.K. Cho, J.Lee, and U. Neumann, "A Multi-ring Color Fiducial System and A Rule-Based Detection Method for Scalable Fiducial-tracking Augmented Reality", Proceedings of International Workshop on Augmented Reality, Nov. 1998

[FEIN93] S. Feiner, B. MacIntyre, D. Seligmann, "Knowledge-Based Augmented Reality," Communications of the ACM, Vol. 36, No. 7, pp 52-62, July 1993

[FISC81] Fischler, M.A., Bolles, R.C. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Graphics and Image Processing, Vol. 24, No. 6, 1981, pp. 381-395.

[FOXL98] E. Foxlin, M. Harrington, G. Pfeifer, "Constellation : A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications", Proceedings of Siggraph98, Computer Graphics, pp. 371-378

[GHAZ95] M. Ghazisadedy, D. Adamczyk, D. J. Sandlin, R. V. Kenyon, T. A. DeFanti, "Ultrasonic Calibration of a Magnetic Tracker in a Virtual Reality Space," Proceedings of VRAIS'95, pp. 179-188

[MAYB79] P.S. Maybeck, *Stochastic Models, Estimation, and Control*, Volume 1, Acamedic press, Inc., 1979

[MEND95] J.M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice Hall PTR, 1995

[MEYE92] K. Meyer, H. L. Applewhite, F. A. Biocca, "A Survey of Position Trackers," Presence: Teleoperator and Virtual Environments, Vol. 1, No. 2, pp. 173-200, 1992

[NEUM96] U. Neumann, Y. Cho, "A Self-Tracking Augmented Reality System," Proceedings of ACM Virtual Reality Software and Technology '96, pp. 109-115

[NEUM98a] U. Neumann, S. You, "Integration of Region Tracking and Optical Flow for Image Motion Estimation," to appear in proceedings of IEEE ICIP-98, Oct. 1998.

[NEUM98b] U. Neumann, J. Park, "Extendible Object-Centric Tracking for Augmented Reality", 1998 IEEE Virtual Reality Annual International Symposium

[PARK98] J. Park, U. Neumann, "Extending Augmented Reality with Natural Feature Tracking", Proceedings of SPIE Vol.3524, Telemanipulator and Telepresence Technologies V, Nov. 1998

[REKI97] J. Rekimoto, "NaviCam: A Magnifying Glass Approach to Augmented Reality," Presence: Teleoperator and Virtual Environments, Vol. 6, No. 4, pp. 399-412, August 1997

[SHAR97] R. Sharma, J. Molineros, "Computer Vision-Based Augmented Reality for Guiding Manual Assembly," Presence: Teleoperator and Virtual Environments, Vol. 6, No. 3, pp. 292-317, June 1997

[STAT96] A. State, G. Hirota, D. T. Chen, B. Garrett, M. Livingston, "Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking," Proceedings of Siggraph96, Computer Graphics, pp. 429-438

[UENO95] M. Uenohara, T. Kanade, "Vision-Based Object Registration for Real-Time Image Overlay," Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine, pp. 13-22, 1995

[WELC97] G. Welch, G. Bishop, "SCAAT: Incremental Tracking with Incomplete Information," Proceedings of Siggraph97, Computer Graphics, pp. 333-344