ICTAE-108
Parallel Real-Time 3d Object Tracking and Structure System Architecture

Changwoo Cho
Division of Internet & Multimedia Engineering, Korea University, Seoul, Korea

Kyunam Cho, Changsung Jeong*
School of electric electrical engineering, Korea University, Seoul, Korea
csjeong@korea.ac.kr

Abstract
We propose a real-time keyframe-based tracking and modeling of the object automatically in unknown environments, while PTAM algorithm has previously been attempted in a small AR workspace. Our contribution splits each surface of the object into six threads in detecting stage, processes the tracking and modeling in parallel threads on the desktop. After detecting the target object, a thread deals with the task of object tracking using ROI in tracking stage, while the other one models the object using the sequential structure from motion method which allows the use of computationally expensive batch optimization techniques not associated with real-time operation. Therefore, we propose a system to let the user models the object very quickly. The user simply has to save and reuse the 3D model of the object next time. The system can be tracked in real-time, while all process is only processed in ROI.

Keyword: Object tracking, Structure from motion, Region of interesting

1. Introduction
Object tracking is a very popular research topic in computer vision as Augmented Reality (AR), games and applications which requires fast and robust object recognition and tracking technique. AR is defined as a real-time indirect or direct view of a real world environment that has been augmented by adding virtual computer generated information to it. AR is both interactive and registered in 3D as well as combining real and virtual objects [1]. In previous AR technical using 3D object tracking, the 3D model information utilizes the modeling software such as AUTO-CAD, 3D MAX for recognizing the object. There is a problem to take too much time by using the modeling software to model the object. Also, conventional 3D model information depends on the surrounding environment and it is limited to represent the 3D model of the various forms that exist in the real world.

In this paper, we propose three steps in our algorithm: First, we propose multiple hand-operated and automatic methods using hand-operated stereo technical and previous information in order to recognize target object. Second, we present real-time 3D object tracking method using natural features [2]. Third, we structure the target object for robust object recognition and tracking.

In order to, we attempt to implement an ROI-based (region of interest based) PTAM system. Our system offers three main challenges. The challenge comes from the preview object model which derives a model of its robustness. The lack of processing power is significant when the task of recognizing and tracking object is done. The last challenge can model the object, when it restarts the system and recognizes the object.

2. Related Work

2.1 3D Tracking

We have seen that the essential or fundamental matrix encapsulates the geometric constraint relating pairs of views [3]. Next, we will turn our attention to solving the structure from motion problem for the arbitrary number of views. The last stage is usually BA (bundle adjustment) which is used iteratively to refine the parameters of the structure and motion by the minimization of an appropriate cost function.

2.2 Structure from Motion

Sequential structure from motion method is the sequential algorithm of the most popular. It works by incorporating successive views one at a time. While each view is registered, a partial reconstruction is extended by computing the positions of all 3D points which are visible in more than one view using triangulation [4]. A suitable initialization method is typically obtained from the fundamental matrix calculated from first two sequential views. There are epipolar constraints. Therefore, one possibility is to exploit the two-view epipolar geometry that relates each view to its predecessor [5]. An alternative of resection is to determine the pose of each additional view using already-reconstructed 3D points [6, 7]. As we have seen, 6 or more 3D-2D correspondences allow linear solution for the twelve elements of the projection matrix. Another alternative is merging partial reconstruction. It is to merge partial reconstructions using corresponding 3D points [8, 9]. Typically, reconstructions of two or three views are obtained using the pairs or triplets adjacent image; and then they are merged using corresponding 3D-points.

3. System Architecture

3.1 Overview

Figure 1 is an overview of proposed algorithm architecture. Our algorithm consists of three steps: detecting stage, tracking stage and structure stage. In detecting stage, when there is not detecting the target object in the input frame, we should do feature extraction and matching in the all frames without using ROI. If we should find the

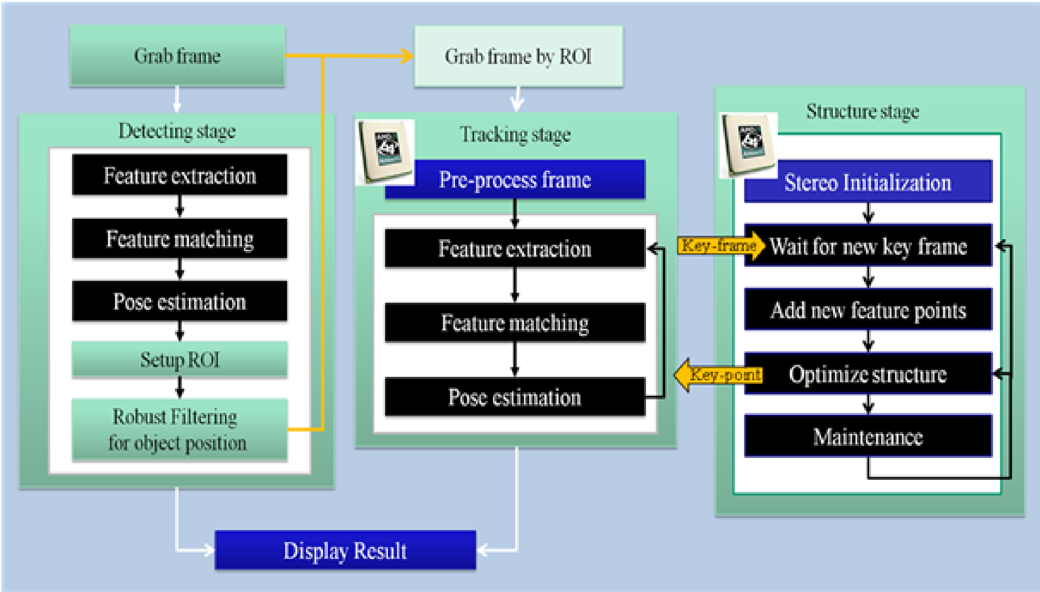

Fig. 1. Overview of system architecture and stages

object, stage is shifted to tracking stage after setup ROI and robust filtering for object position. In tracking stage, it runs SURF-based object tracking. Modeling stage should model the object in ROI. It is a way similar to PTAM [10].

3.2 Feature Extraction and Matching Steps

In feature extraction step, we extract the feature points from grabbing frame by using ROI and SURF. Then, we find matching points by comparing feature points of the object using a k-d tree in feature matching step, providing good quality and fast processing speed [11, 12].

3.3 Pose Estimation Step

In pose estimation step, we estimate the object's pose information from matched points about x, y, z positions by using RANSAC-PnP algorithm [13, 14]. Perspective-n- Problem (PnP) algorithm determines the camera position and orientation with respect to a scene object and supports pose estimation for non-planar surfaces, while 30 homography supports pose estimation for only the planar object.

3.4 Setup ROI and Robust Filtering Steps

In these steps, we exploit the region of interest (ROI) for fast object tracking and modeling in our algorithm. ROI is defined by 1.5 times the size of detecting object. It reduces wrong feature matching result. ROI is set by using object size and position in input images which are predicted from Kalman filter of object pose [15, 16].

In our filtering system, the Kalman filter is to use measurements observed over time

Prediction

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$$

Correction

$$\tilde{y}_k = z_k - H_k \hat{x}_{k|k-1}$$

$$S_k = H_k P_{k|k-1} H_k^T + R_k$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}$$

Fig. 2. Kalman filter equations

which may contain noise Therefore, the production values that tend to be closer to the true values of the measurements [15, 17]. Kalman filter equations consist of prediction and correction ones as shown below, where xk is a state vector, zk a measurement vector, Fk a system matrix and Hk a measurement matrix as shown in figure 2.

3.5 Tracking Stage

This section describes the operation of the SURF-based tracking system with the initialize of the structured object`s 3D model. This system receives the image from crabbed frame and retains a real-time pose estimation by ROI. At every frame of ROI, our system performs the following procedure:

1)The frame is crabbed from the detecting stage by ROI, and the pose information is generated from a prior stage.
2)Each frame store two pyramid levels of grayscale images: level one is the original image, and another one is level two at 80*60 pixels.
3)Matching points are projected into the frame according to the frame`s prior pose estimation.
4)The less number of the feature points is searched for in the level two of the image.
5)If the feature points are not matched in (4), its system is back to the detecting stage. Or it turns to (6).
6)The larger number of the feature points is searched for in the level one of the image.
7)The results of pose estimation are updated and transmit the key-frame to modeling stage.

3.6 Key-frame and Key-point

The object`s structured model initially is generated by first two key-frames from detection stage. And then, the new Key - frame is found in ROI of the tracking stage. The key- frame is added whenever the tracking quality is good; the distance of two matched key-points must exceed fifty pixels; and the region don`t away the ROI size.

The key-points are the feature points from all key-frames. They are used for modeling the object model in the modeling stage. Each key-point requires depth information using triangulation which can compute the 3D points from their measured image positions in two or more frames [4]. If the key-points is added, our system adjusts the positions of all key-points using bundle adjustment.

3.7 Modeling Stage

Modeling stage should model the object in ROI which is a way similar to PTAM. This stage is the process by which the 3D point of the object modeling is built from the results of pose estimation. Structure-building occurs in two distinct steps: First, an initial model is built using a stereo technique [18]. After this, the object model is continually refined and expanded by the modeling thread as the key-frames are added by the tracking system. The operation of the modeling stage is illustrated in Figure 2. This stage is individually described in [7] which called mapping section.

4. Implementation

Our system described above was implemented on a desktop PC with an Intel Core i5-2500K 3.30GHz processor running Windows7. Program is developed by C++ using Visual Studio 2010. The libraries are used the libCVD, TooN and OpenCV. We advanced our system to a below approach. First, feature matching can be parallelized in detecting stage. Therefore, we create the six threads which is the object has six surfaces. Each thread is assigned to each surface. And then, we create the two threads about tracking and modeling threads. For our experiments, we made the cube object as shown in figure 3. The object is registered previously while 3D point information of the cube.

5. Performance Evaluation

For our experiments, we made the cube object as shown in figure 3. The object is registered previously while 3D point information of the cube.

Figure 4 compares object tracking speed of our method to the other algorithms. The frame-rate of the tracking speed shows faster processing speed about 60.61fps than the others. The processing time of the object modeling depends on the number of key-frames and key-points in modeling stage of our system, and it is not associated with real-time object tracking.

Fig. 3. Sample target object

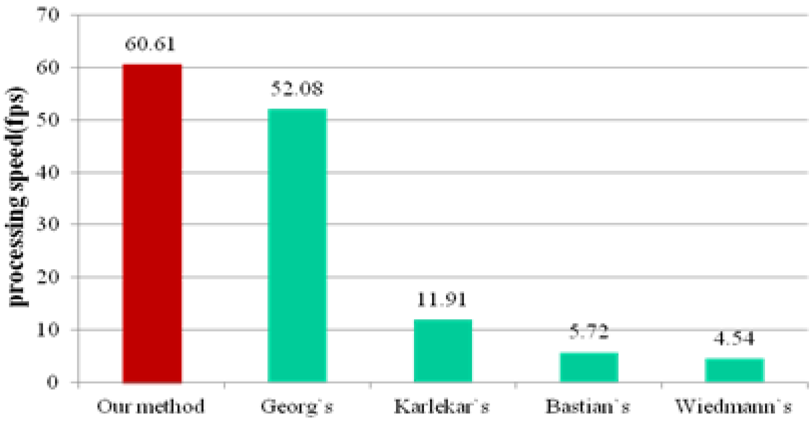

Fig. 4. Comparison of tracking speed with other algorithms

6. Conclusion and Feature

We propose a real-time key-frame-based tracking and modeling of the object automatically in unknown environments, while PTAM algorithm has previously been attempted in a small AR workspace. Our contribution splits each surface of the object into six threads in detecting stage, process tracking and modeling in parallel threads on the desktop. A thread deals with the task of robustly object tracking using ROI in tracking stage, while the other one models the object using sequential modeling from motion method which allows the use of computationally expensive batch optimization techniques not associated with real-time operation. Therefore, we propose a system to let the user models the object very quickly. The user simply has to save and reuse the 3D model of the object next time. The system can be tracked in real-time, while all process is only processed in ROI.

However, the ROI of our approach still does not have enough optimize to represent the range of the object. SURF algorithm is very slow in OpenCV. We can replace it to another algorithm or use of GPGPU for speed-up in detecting stage. Also, we can consider operate the modeling process in cloud environments which spend more time on bundle adjustment.

References

[1] G. Klein, “Visual Tracking for Augmented Reality”, University of Cambridge, 2006

[2] H. Kato, M. Billinghurst, “Marker Tracking and HMD Calibration for a videobased Augmented Reality Conferencing System”, In Proceedings of the 2nd International Workshop on Augmented Reality, 1999.

[3] O. D. Faugeras. “Three Dimensional Computer Vision: A Geometric View-point”, MIT Press, Boston, 1993.

[4] R. I. Hartley and P. Sturm. “Triangulation”, In American Image Under-standing Workshop, pages 957–966, 1994.

[5] H. Aanæs, “Methods for structure from motion”, IMM, Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, 2003

[6] P. A. Beardsley, P. Torr, and A. Zisserman. “3D model acquisition from extended image sequences”, In European Conference on Computer Vision (ECCV’96), pages 683–695, 1996.

[7] P. A. Beardsley, A. Zisserman, and D. Murray. “Sequential updating of projective and affine structure from motion”, International Journal of Computer Vision, 23 (3): 235–259, 1997.

[8] O. D. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, and I. Zoghlami. “3-D reconstruction of urban scenes from image sequences”,. Computer Vision and Image Understanding, 69 (3): 292– 309, 1998.

[9] A. W. Fitzgibbon and A. Zisserman. “Automatic camera recovery for closed or open image sequences”, In European Conference on Computer Vision (ECCV’98), pages 311–326, 1998.

[10] G. Klein, “Visual Tracking for Augmented Reality”, University of Cambridge, 2006

[11] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, D. Schmalstieg, “Pose Tracking from Natural Features on Mobile Phones”, ISMAR '08 Proceedings, 2008.

[12] P. Azad, T. Asfour, R. Dillmann, “Combining Harris Interest Points and the SIFT Descriptor for Fast Scale-Invariant Object Recognition”, International Conference on Intelligent Robots and Systems, 2009.

[13] M. A. Fischler, R. C. Bolles. “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, Communications of the ACM, Vol. 24, No. 6, pp. 381-395, 1981.

[14] V. Lepetit, F. Moreno-Noguer, P. Fua, “E PnP: An Accurate O (n) Solution to the PnP Problem”, International Journal of Computer Vision, Vol. 81, No. 2, 2009.

[15] R.E, Kalman, “A new approach to linear filtering and prediction problems:, Journal of Basic Engineering Vol. 82, 1, pp. 35-45, 2008.

[16] N.J, Gordon, “Nobel Approach to nonlinear/non-Gaussian Bayesian state estimation”, IEEE Proceedings F on Radar and Signal Processing Vol.140, 2, pp. 107-113, 2009.
[17] M.I. Ribeiro, “Kalman and Extended Kalman Filters: Concept, Derivation and Properties, Institute for Systems and Robotics”, Instituto Superior Tecnico, AV. Rovisco Pais, 1, February 2004.
[18] P. Fua. “A parallel stereo algorithm that produce dense depth maps and preserves image features”, Machine Vision Applications, pp. 35–49, 1993.

Acknowledgement

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1001) supervised by the NIPA (National IT Industry Promotion Agency), and Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the MSIP (NRF-2013043678).