Real-time camera position and posture estimation using a feature landmark database with priorities

Takafumi Taketomi, Tomokazu Sato and Naokazu Yokoya
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan
{*takafumi-t, tomoka-s, yokoya*}*@is.naist.jp*

Abstract

In the field of computer vision, many kinds of camera parameter estimation methods have been proposed. As one of these methods, an extrinsic camera parameter estimation method that uses pre-constructed feature landmark database has been studied. In this method, extrinsic camera parameters of video images are estimated from correspondences between landmarks and image features. Although this method can work in a large outdoor environment, its computational cost in matching process is expensive and it cannot work in real-time. In this paper, to achieve real-time camera parameter estimation, the number of matching candidates are reduced by using priorities of landmarks that are determined from previously captured video sequences.

1. Introduction

Camera parameter estimation is very important in various vision-based applications. Some of these applications like augmented reality (AR), human navigation, and self-localization of robots and automobiles need absolute camera position and posture and the estimation process should be done in real-time. In this paper, we propose an extrinsic camera parameter estimation method for such applications.

In the field of computer vision, many kinds of image based online camera parameter estimation methods have been proposed. These methods can be classified into two groups. One is visual SLAM based method [1, 2] that can estimate camera parameters without a pre-knowledge of target environment. In this method, database construction and camera parameter estimation are carried out simultaneously. However, absolute camera position and posture cannot be acquired. The other uses some kinds of pre-constructed databases such as 3-D models [3, 4] and feature landmarks [5]. In this approach, absolute camera position and posture estimation can be realized. However, construction of 3-D models for large and complex outdoor environments needs large human costs. Thus, we employ feature landmarks as the database that stores 3-D positions of image features and image templates [5]. Feature landmark database can be constructed automatically even in a complex environment by using the structure from motion (SFM) for omni-directional camera [6]. However, landmark based camera parameter estimation could not work in real-time because pattern matching process in this method is computationally expensive to realize illumination and view direction independent pattern matching.

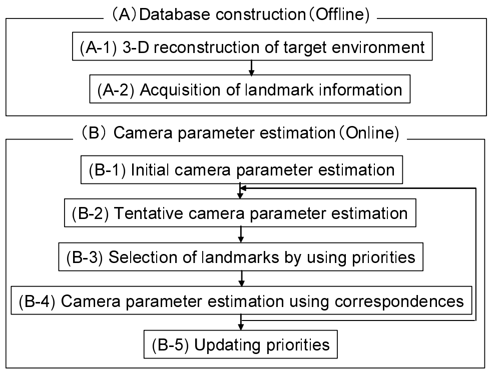


Figure 1. Flow diagram.

In this study, in order to realize real-time camera parameter estimation using landmarks, the number of matching candidates are reduced by the following ideas. (1) Reduction of feature points: Tentative camera parameters are estimated to limit the range of search using landmark tracking between successive image frames where view direction and illumination change can be ignored. (2) Reduction of landmarks: Priorities are associated with landmarks using previously captured video sequences to select a smaller number of landmarks whose matching confidences are high.

2. Camera parameter estimation using feature landmark database with priorities

Figure 1 shows the flow diagram of the proposed method. Feature landmark database must be constructed for the target environment before online camera parameter estimation. In this research, priorities are

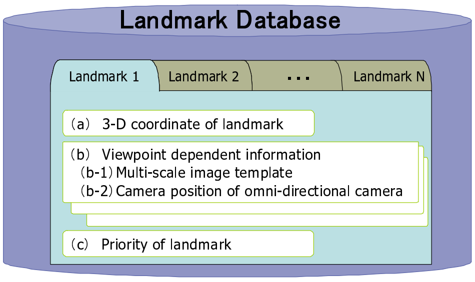


Figure 2. Elements of landmark database.

newly associated with the landmarks in the database.

2.1 Construction of landmark database

(A-1)3-D reconstruction of target environment
First, the target environment is taken as omni-directional image sequences. For the captured image sequences, SFM is applied to estimate the 3-D coordinates of feature points and camera parameters of omni-directional camera. In this SFM process, feature points of known 3-D positions [6] or absolute positions measured by GPS [7] are given as a reference of absolute position and posture.

(A-2)Acquisition of landmark information
Feature landmark database consists of a number of landmarks as shown in Figure 2. Each landmark retains (a)3-D coordinate, (b)viewpoint dependent information, and (c)priority of landmark.

(a)3-D coordinate of landmark: To estimate the camera parameter in the online process, 3-D coordinates of landmarks are registered. 3-D coordinates of landmarks are obtained by the SFM (A-1).

(b)Viewpoint dependent information: In order to deal with viewpoint dependent visual aspect changes of landmarks, for each position of omni-directional camera, multi-scale image templates of landmark (b-1) and position of omni-directional camera (b-2) from which the image template is captured are registered.

(c)Priority of landmark: Priorities are associated with landmarks to select reliable landmarks. These priorities are determined by calculating probabilities that landmarks are used in online camera parameter estimation. Priority P_i of landmark i is defined as $P_i = E_i/D_i$. Here, E_i represents the frequency that the landmark i is judged as an inlier by robust estimation in the online process and D_i represents the frequency that the landmark i is selected from the database as a matching candidate. In this study, we assume that system administrator gives several training video to determine the priorities before the system is used by users.

2.2 Camera parameter estimation

This section describes a camera parameter estimation method in the online process (B). As shown in Figure 1, first, initial camera parameter is estimated (B-1). Next, tentative camera parameter estimation (B-2), selection of landmarks with high priorities (B-3), and camera parameter estimation (B-4) are repeated. After finishing camera parameter estimation, the priorities in the database are updated based on the result of current camera parameter estimation (B-5).

(B-1)Initial camera parameter estimation
Initial camera parameter for the first frame of input is assumed to be given by landmark based camera parameter estimation method for a still image input [8]. Currently, this part does not implemented to the system and initial parameters are given manually in the experiments.

(B-2)Tentative camera parameter estimation
Tentative camera parameter is estimated by landmark tracking between successive frames. In this process, landmarks that are used to estimate camera parameter in the previous frame are selected and tracked to the current frame. In the successive frames, visual aspects of landmarks hardly change. Thus, tracking of landmarks can be realized by a simple SSD based tracker with low computational cost. After landmark tracking, tentative camera parameter is estimated by solving PnP problem [9] using tracked landmarks. To remove outliers, LMedS estimator [10] is employed in this process.

(B-3)Selection of landmark based on priorities
In this process, landmarks that are visible from current camera position are selected from the database by using estimated tentative camera parameter and geometric location of landmark. Next, top N_{prior} confident landmarks are selected based on priorities of landmarks. By using priorities of landmarks, unreliable landmarks such as repeatable texture and natural object are efficiently discarded. As a result, the number of landmarks that should be tested in the next process can be reduced.

(B-4)Camera parameter estimation using correspondences
Camera parameter of the current frame is estimated by searching corresponding pairs between selected landmarks and image features. To determine these correspondences, first, landmarks selected from the database are projected to the image plane using tentative camera parameter. Corresponding landmarks and feature points are then searched within a fixed window whose center is at the projected landmark. In this process, window size can be smaller than that for the process (B-2) because camera parameter is roughly known. As a result, the number of feature points for matching candidates can be reduced. Finally, camera parameter is estimated by solving PnP problem using corresponded pairs of landmarks and feature points. In this process, outliers are rejected by using a LMedS estimator as in (B-2).

(B-5)Updating priorities of landmarks
After finishing camera parameter estimation process, priorities of landmarks are updated by using estimated result. The priority P_i of the landmark i is updated as follows:

$$P_i = \frac{E_{iold} + E_{inew}}{D_{iold} + D_{inew}}, \quad (1)$$

where E and D represent the frequency that is described in section 2.1. Subscripts $inew$ and $iold$ for these frequency denote the current and the past camera parameter estimation, respectively.

2.3 Effect of computational cost reduction

In this section, the effect of computational cost reduction from the previous method [5] in matching process is discussed. Computational cost C_{new} in matching process for the proposed method can be represented as sum of C_{track} for tentative camera parameter estimation and C_{proj} for determination of corresponding landmarks and feature points: $C_{new} = C_{track} + C_{proj}$. The cost C_{track} is lower than C_{proj} because illumination and view direction independent pattern matching is not needed in landmark tracking process. By using tentative camera parameter estimated by tracking landmarks, the number of feature points are reduced to S_2/S_1, where S_1 and S_2 represent the size of search window in the previous and the proposed methods, respectively. The number of landmarks are also reduced to $(N_{prior} - N_{track})/N$ by selecting landmarks with high priorities. Here, $N(N \geq N_{prior})$ represents the maximum number of landmarks which are selected from the database in the previous method and $N_{track}(N_{track} \leq N_{prior})$ represents the number of landmarks which are used to estimate tentative camera parameter. Resultingly, computational cost C_{proj} in the proposed method is derived as follows:

$$C_{proj} = \frac{(N_{prior} - N_{track})}{N} \frac{S_2}{S_1} C_{prev}, \quad (2)$$

where C_{prev} is the cost of matching process in the previous method. Note that effect of computational cost reduction does not perfectly conform with Eq. (2) due to the overhead of the iteration process.

3. Experiments

To show the effectiveness of the proposed method, first, the computational cost is compared with the original landmark based method [5]. Applications of landmark based real-time camera parameter estimation are then demonstrated.

First, the landmark database is constructed for an outdoor environment using omni-directional multi-camera system (Point Grey Research Ladybug). Figure 3 shows sampled images used for database construction.

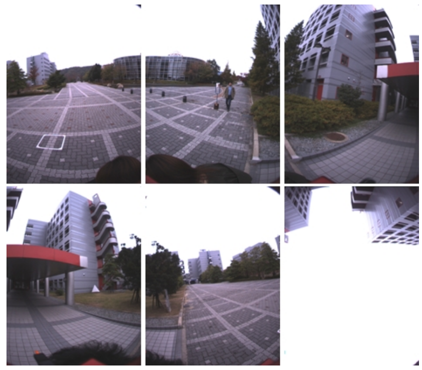


Figure 3. Sampled images taken by omni-directional multi-camera system.

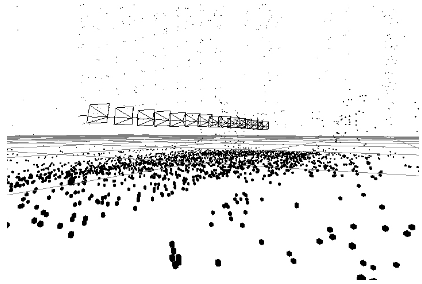


Figure 4. Result of 3-D reconstruction.

By applying the SFM process for these input image sequences, 3-D positions of image features and extrinsic parameters of omni-directional camera are estimated as shown in Figure 4. After database construction, 3 training video are taken in the target environment to compute the priorities of landmarks. To evaluate the proposed and the previous method, another video image sequence (720×480 pixels, progressive scan, 15fps, 1,000 frames) is also captured. Each parameter in the online process was set as shown in Table 1.

First, the number of landmarks to be selected from the database is determined. Figure 5 shows estimation errors in position for various number of selected landmarks. In this experiment, camera parameter estimation by the proposed method has never failed even the number of landmarks is set as 30, while camera parameter estimation by the previous method failed when the number of landmarks is 70 or less. In the proposed method, error in position is slightly increased when the number of landmarks is 50 or less. From this result, we determine the number of landmarks as 60 for the

Table 1. Parameters in experiment.

	Previous method	Proposed method
Range of search in process (B-2)	-	120×60 pixels
Range of search in process (B-4)	120×60 pixels	20×20 pixels
Training video	-	Estimation results of three sequences
Initial value of priorities	-	0.5

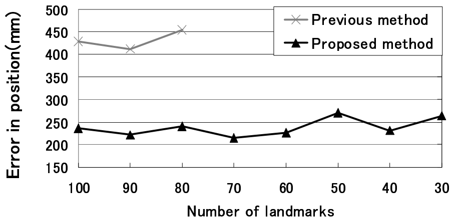


Figure 5. Relation between number of landmarks and error in position.

proposed method and 80 for the previous method. For these numbers, average estimation error in position is 233mm and 360mm for the proposed and the previous methods, respectively. In this case, the accuracy is increased even when the number of selected landmarks are decreased. It is considered as the effect of the priorities of landmarks. Table 2 shows processing time for the previous and the proposed methods. By estimating tentative camera parameter and selecting landmarks with high priorities, total computational cost is about 6 times cheaper than the previous method. As a result, the proposed method can work in real-time. Although computational cost in the matching process is ideally over 24 times cheaper than that of the previous method from Eq. (2), actually it was 20 times.

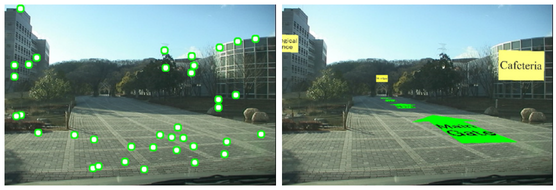
Figure 6 shows examples of AR applications using the proposed method. In this figure, circles on the left images indicate landmarks which are used for camera parameter estimation. Figure 6(a) shows the AR car navigation. The proposed method can estimate car position and posture more accurately and more frequently than standard GPS-based systems and we can realize high accurate geometric registration for AR. Figure 6(b) shows the application for pre-visualization tool for filmmaking using AR. Pre-visualization is a technique that is used for testing a camera work and an acting in the pre-production process of filmmaking. Our method has worked in such a natural environment as shown in Figure 6(b).

4. Conclusion

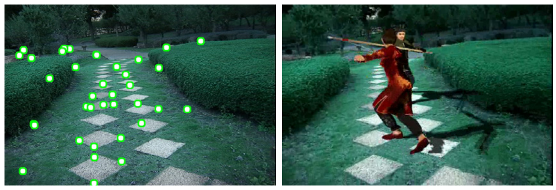
In this paper, we have proposed a real-time camera parameter estimation method by reducing matching pairs of feature points and landmarks. Number of feature points are reduced by estimating tentative camera parameters. Number of landmarks are reduced by using priorities of landmarks. In the proposed method, camera parameter can be estimated in large and natural environments.

Table 2. Comparison of processing time.

Individual step	Processing time (ms)		Ratio of processing time
	Previous method (Number of landmarks 80)	Proposed method (Number of landmarks 60)	
Tentative camera parameter estimation	-	28	-
Selection of landmarks	12	1	0.08
Matching process in (B-4)	316	15	0.05
Camera parameter estimation in (B-4)	61	17	0.28
Total (including of overhead cost)	393(2.5fps)	66(15.1fps)	0.16



Detected landmarks. Overlaid Annotations.
(a)Navigation by augmented reality.



Detected landmarks. Overlaid CG actors.
(b)Pre-visualization.

Figure 6. Examples of applications.

Acknowledgement

This research is supported by Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Agency (JST), and Strategic Information and Communications R&D Promotion Programme (SCOPE) of Ministry of Internal Affairs and Communication Japan.

References

[1] B. Williams, G. Klein and I. Reid: “Real-time SLAM relocalisation,” Proc. ICCV, 2007.
[2] G. Klein and D. Murray: “Parallel tracking and mapping for small AR workspaces,” Proc. ISMAR, pp. 225–234, 2007.
[3] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: “Fully automated and stable registration for augmented reality applications,” Proc. ISMAR, pp. 93–102, 2003.
[4] L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3D camera tracking,” Proc. ISMAR, pp. 48–57, 2004.
[5] M. Oe, T. Sato and N. Yokoya: “Estimating camera position and posture by using feature landmark database,” Proc. SCIA, pp. 171–181, 2005.
[6] T. Sato, S. Ikeda and N. Yokoya: “Extrinsic camera parameter recovery from multiple image sequences captured by an omnidirectional multi-camera system,” Proc. ECCV, Vol. 2, pp. 326–340, 2004.
[7] S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: “Construction of feature landmark database using omnidirectional videos and GPS positions,” Proc. Int. Conf. on 3DIM, pp. 249–256, 2007.
[8] M. Susuki, T. Nakagawa, T. Sato and N. Yokoya: “Extrinsic camera parameter estimation from a still image based on feature landmark database,” Proc. ACCV’07 Satellite Workshop on Multi-dimensional and Multi-view Image Processing, pp. 124–129, 2007.
[9] R. Klette, K. Schluns and A. koschan Eds: Computer vision: Three-dimensional data from image, Springer, 1998.
[10] P. J. Rousseeuw: “Least Median of Squares Regression,” J. of the American Statistical Association, 79, pp. 871–880, 1984.