Real-Time Geometric Registration Using Feature Landmark Database for Augmented Reality Applications

Takafumi Taketomi, Tomokazu Sato and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology,
Takayama 8916-5, Ikoma, Nara, 630-0192, Japan

ABSTRACT

In the field of augmented reality, it is important to solve a geometric registration problem between real and virtual worlds. To solve this problem, many kinds of image based online camera parameter estimation methods have been proposed. As one of these methods, we have been proposed a feature landmark based camera parameter estimation method. In this method, extrinsic camera parameters are estimated from corresponding landmarks and image features. Although the method can work in large and complex environments, our previous method cannot work in real-time due to high computational cost in matching process. Additionally, initial camera parameters for the first frame must be given manually. In this study, we realize real-time and manual-initialization free camera parameter estimation based on feature landmark database. To reduce the computational cost of the matching process, the number of matching candidates is reduced by using priorities of landmarks that are determined from previously captured video sequences. Initial camera parameter for the first frame is determined by a voting scheme for the target space using matching candidates. To demonstrate the effectiveness of the proposed method, applications of landmark based real-time camera parameter estimation are demonstrated in outdoor environments.

Keywords: landmark database, natural feature, augmented reality, camera parameter estimation

1. INTRODUCTION

Augmented reality technique (AR) is applicable to various tasks such as human navigation[1–3] , work support system[4,5] and assistance in education[6] . In these AR applications, it is important to solve a geometric registration problem between real and virtual worlds. To solve the problem, many kinds of image based online camera parameter estimation methods have been proposed.

Image based camera parameter estimation methods can be classified into two groups. One is visual SLAM based method[7,8] that can estimate camera parameters without a pre-knowledge of a target environment. In this method, database construction and camera parameter estimation are carried out simultaneously. Although this method can work in unknown environment, absolute camera position and posture cannot be acquired. Thus, this method cannot be used for the position dependent tasks like navigation where guide information is located in global coordinate.

The other uses some kinds of databases that contain pre-knowledge of the target environments such as natural feature landmarks[9] and 3-D models[10,11] . In this approach, absolute camera position and posture can be acquired. Feature landmark database can be constructed automatically even in a complex environment by using the structure from motion (SFM) for omni-directional camera[12] . On the other hand, 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[9] . However, landmark based camera parameter estimation could not work in real-time because pattern matching process in this method was computationally expensive to achieve illumination and view direction independent pattern matching. Additionally, initial camera parameters for the first frame must be given manually.

In this study, in order to realize real-time and manual-initialization free camera parameter estimation using a feature landmark database, we reduce the computational cost for matching process, and an automatic initial camera parameter estimation method is introduced into the proposed framework. The cost for matching process is reduced by the following ideas: (1) Tentative camera parameters are estimated to limit the range of search for matching candidates by landmark

---

Further author information:
T.T,T.S,N.Y: E-mail: {takafumi-t, tomoka-s, yokoya}@is.naist.jp

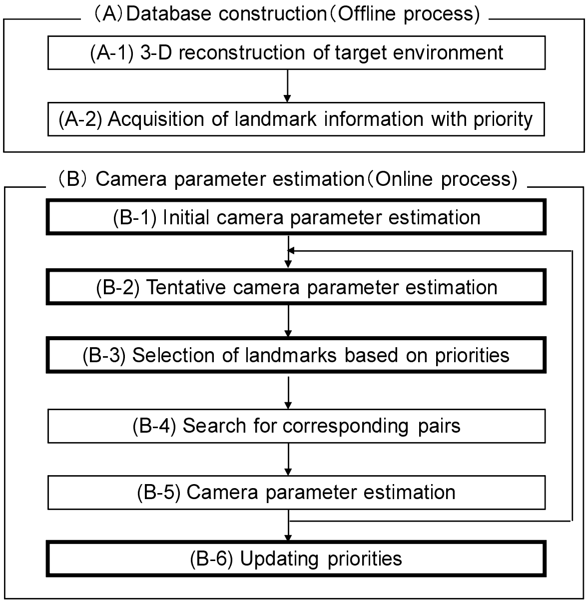


Figure 1. Flow diagram of extrinsic camera parameter estimation using landmark database.

tracking between successive image frames. (2) Priorities are associated with landmarks to select a smaller number of landmarks using previously captured video sequences. Initial camera parameters for the first frame are determined by a voting scheme for the target space using matching candidates. Figure 1 shows the flow diagram of the proposed method. In this figure, thick square indicates the new or improved process. Landmark based camera parameter estimation method is composed of database construction in offline process and camera parameter estimation in online process.

2. GEOMETRIC REGISTRATION USING FEATURE LANDMARK DATABASE

2.1 Database construction

This section describes database construction in the offline process (A). Feature landmark database consists of a number of landmarks as shown in Figure 2. In the offline process, first, an omni-directional video sequence is captured in the target environment. 3D coordinates of feature points in the omni-directional image sequence are then obtained by structure-from-motion (A-1). Finally, landmark information is registered to the database (A-2). In this research, priorities are newly associated with landmarks and characteristic scale and SIFT feature vectors are registered to the database to estimate initial camera parameters.

(A-1) 3-D reconstruction of target environment
First, the target environment is taken as omni-directional image sequences. Next, natural feature points are detected and tracked using Harris interest operator[13] . 3-D coordinates of feature points and camera parameters of the omnidirectional camera are estimated by SFM. In this SFM process, feature points of known 3-D positions[12] or absolute positions measured by GPS[14] are used as a reference of absolute position and posture. In this process, we can obtain extrinsic camera parameters of the omniderectional camera and 3-D coordinates of feature points in absolute coordinate system.

(A-2) Acquisition of landmark information
Landmark information is obtained from the result of 3-D reconstruction. Each landmark retains (a)3-D coordinate, (b)viewpoint dependent information, and (c)priority of landmark.

(a) 3-D coordinates of landmarks
To estimate extrinsic camera parameters 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,

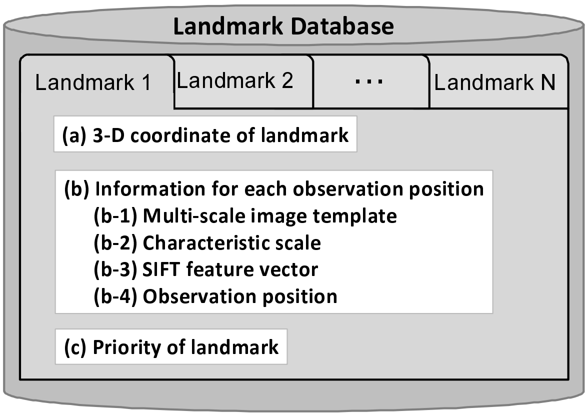


Figure 2. Elements of landmark database.

multi-scale image templates of landmark are generated. In this study, characteristic scale which is determined by using Harris-Laplacian[15] (b-2) and SIFT feature vector which is calculated by using SIFT-descriptor[16] (b-3) are newly registered to the database to estimate initial camera parameter. Additionally, positions of omni-directional camera (b-4), from which image templates are captured, are registered to select observable landmarks in online process (B-3).

(c) Priorities of landmarks
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 follows:

$$P_i = \frac{E_i}{D_i}, \tag{1}$$

where, 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 videos to determine the priorities before the system is used by users.

2.2 Geometric registration

This section describes a camera parameter estimation method in the online process (B). As shown in Figure 1, first, initial camera parameters are 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 parameters for the first frame of an input are assumed to be given by landmark based camera parameter estimation method for a still image input[17] . First, characteristic scale and SIFT feature vector of natural feature points in the input image are calculated. These features are used to search for corresponding pairs between natural feature points and landmarks. Next, outliers are rejected by considering the consistency of observable positions of landmarks. Finally, extrinsic camera parameters are estimated by solving PnP problem[18] .

(B-2) Tentative camera parameter estimation
Tentative camera parameters are 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 done by a simple SSD based tracker using image templates whose center is located at the position of the landmark in previous frame; that is matching candidates are limited to natural feature points within a fixed window W_1 centered at the position of the landmark in previous frame. After landmark tracking, tentative camera parameters $\hat{\boldsymbol{M}}$ are estimated by solving PnP problem[18] using tracked landmarks. To remove outliers, LMedS estimator[19] is employed in this process.

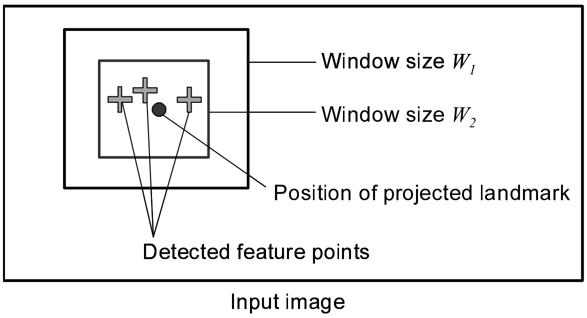


Figure 3. Search range of corresponding point.

(B-3) Landmark selection based on priorities
In this process, landmarks visible from current camera position are selected from the database by using estimated tentative camera parameters 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 will be tested in the next matching process can be reduced.

(B-4) Search for corresponding pairs
Camera parameters of the current frame are estimated by using for corresponding pairs of landmarks and image features. To determine these correspondences, first, landmarks selected from the database are projected onto the image plane using tentative camera parameter $\hat{\boldsymbol{M}}$ as follows:

$$\begin{bmatrix} a_j u_j & a_j v_j & a_j \end{bmatrix}^T = \hat{\boldsymbol{M}} \begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^T, \tag{2}$$

where, (x_j, y_j, z_j) represents 3-D coordinate of landmark j, (u_j, v_j) represents 2-D position of landmark j in the input image and a_j represents the depth of the landmark j in the camera coordinate system. Corresponding landmarks and feature points are then searched within a fixed window W_2 whose center is located at (u_j, v_j) as illustrated in Figure 3. In this process, window size can be smaller than that for the process (B-2) because camera parameter is roughly known as $\hat{\boldsymbol{M}}$. As a result, the number of feature points for matching candidates can be reduced.

(B-5) Camera parameter estimation
Camera parameters are 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-6) Updating priorities
After finishing camera parameter estimation process, priorities of landmarks are updated by considering the estimated result. The priority P_i of the landmark i is updated as follows:

$$P_i = \frac{E_{iold} + E_{inew}}{D_{iold} + D_{inew}}, \tag{3}$$

where E and D represent the frequency that is described in section 2. Subscripts $inew$ and $iold$ for these frequencies denote the result for the current and the past camera parameter estimation, respectively.

3. EFFECT OF COMPUTATIONAL COST REDUCTION

In this section, the effect of computational cost reduction by the previous method[9] in matching process is discussed. Computational cost C_{new} in matching process for the proposed method can be represented as the sum of C_{track} for tentative camera parameter estimation and C_{proj} for determination of corresponding landmarks and feature points as follows.

$$C_{new} = C_{track} + C_{proj}. \tag{4}$$

The cost C_{track} is lower than C_{proj} because illumination and view direction independent pattern matching is not needed in the landmark tracking process. By using tentative camera parameter estimated by tracking landmarks, the number of

Figure 4. Sampled images taken by omni-directional multi-camera system.

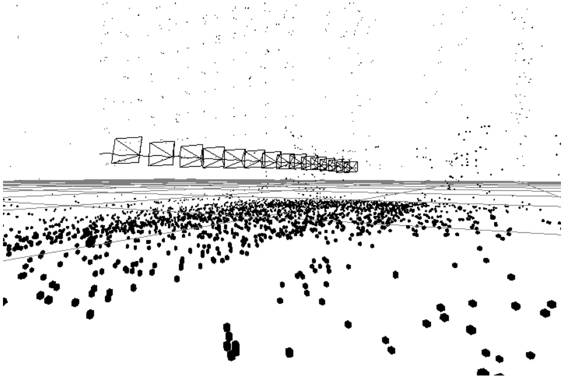


Figure 5. Result of 3-D reconstruction.

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 number of landmarks selected from the database in the previous method[9] , N_{prior} represents the number of landmarks selected from the database based on priorities and $N_{track}(N_{track} \leq N_{prior})$ represents the number of landmarks used to estimate tentative camera parameters. Resultingly, computational cost C_{proj} in the proposed method is derived as follows:

$$C_{proj} = \frac{(N_{prior} - N_{track})}{N} \frac{S_2}{S_1} C_{prev}, \tag{5}$$

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. (5) due to the cost for the overhead in the iteration process.

4. EXPERIMENTS

To show the effectiveness of the proposed method, first, the computational cost is compared with the original landmark based method[9] . 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 4 shows sampled images used for database construction. 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 5.

Table 1. Parameters in experiment.

	Oe's method[9]	Proposed method
Window size W_1 (pixel)	-	120×60
Window size W_2 (pixel)	120×60	20×20
Training video	-	Three video sequences
Initial value of priorities	-	$1/2$

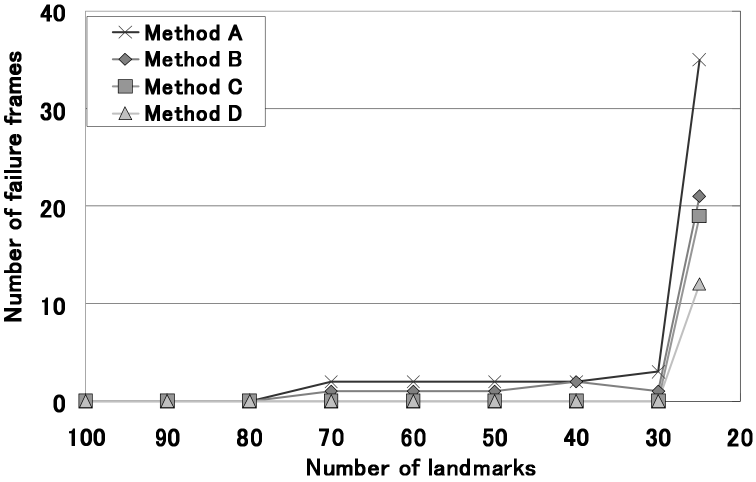


Figure 6. Relation between number of landmarks and failure frames.

After database construction, three training videos 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. In this experiment, initial camera parameter estimation process needs about 60 seconds. Thus, camera is fixed in the target environment until initial camera parameter estimation is completed. Each parameter in the online process was set as shown in Table 1.

4.1 Comparison of computational cost

To show the effect of computational cost reduction, we compared the following four methods.

Method A: Previous method[9]

Method B: Proposed method without landmark selection based on priorities

Method C: Proposed method without tentative camera parameter estimation

Method D: Proposed method

In this experiment, first, in order to determine the number of landmarks to be selected, we compared the rate of estimation failure. Next,computational cost is compared for each method.

Figure 6 shows the number of failure frames for various number of selected landmarks in process (B-3). In this experiment, we deemed the result to be a failure when the number of corresponding pairs are less than 6. The methods A and B which did not use priorities of landmarks failed to estimate camera parameter when the number of landmarks was 70 or less. The methods C and D which use priorities of landmarks did not fail when the number of landmarks was more than

Table 2. Comparison of processing time for one frame (ms).

Method	A	B	C	D
Process (B-2)	-	26	-	21
Process (B-3)	12	3	2	1
Process (B-4)	316	51	131	15
Process (B-5)	61	16	16	17
Overhead	4	4	4	5
Total cost	329	100	153	59

Table 3. Comparison of accuracy.

Method	A	B	C	D
Average position error (mm)	360	257	231	256
Std.dev. position error (mm)	528	137	204	181
Average posture error (deg.)	0.84	0.95	1.13	0.91
Std.dev. posture error (deg.)	0.71	1.20	1.16	0.91
Average re-projection error (pixel)	2.5	2.3	2.1	1.8

30. From these results, we determine the number of landmarks as 80 for the methods A and B and 30 for the methods C and D. Table 2 shows processing time for each method when we used a laptop PC (CPU: Core2Extreme 2.93GHz, Memory: 2GB). In the method D, by estimating tentative camera parameter and selecting landmarks with high priorities, the total computational cost was about 6 times cheaper than the method A. As a result, the proposed method can work in real-time. Although computational cost in the matching process (B-4) is ideally over 48 times cheaper than that of the method A from Eq. (5), actually it was 21 times because the cost for the overhead exists. Table 3 shows accuracy of each method. From this result, the methods B, C and D can reduce computational cost without increasing estimation error.

4.2 AR applications using feature landmark database

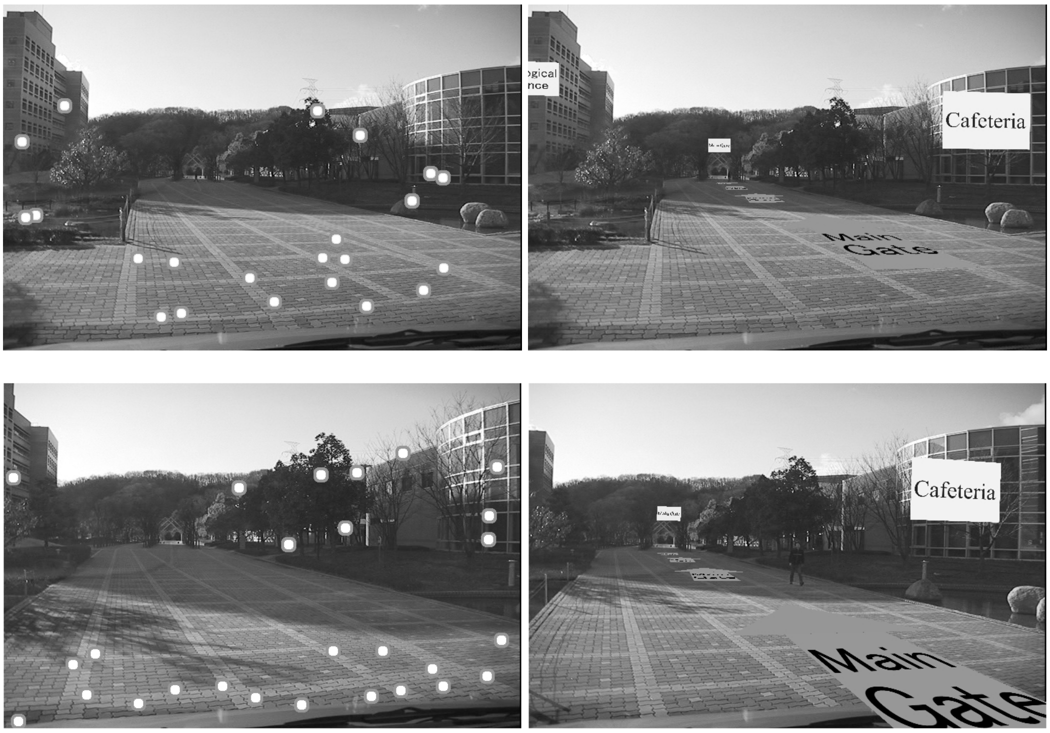
Figure 7 shows examples of AR applications using the proposed geometric registration method. In this figure, circles on the left images indicate landmarks which are used for camera parameter estimation. Figure 7(a) shows the AR car navigation. The proposed method can estimate car position and posture more accurately and more frequently than a standard GPS-based systems and we can realize highly accurate geometric registration for AR. Figure 7(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 successfully in such a natural environment as shown in Figure 7(b).

5. CONCLUSION

In this paper, we have proposed a real-time camera parameter estimation method, which is required for AR applications, by reducing matching pairs of feature points and landmarks. The number of feature points are reduced by estimating tentative camera parameters. The number of landmarks are reduced by using priorities of landmarks. Additionally, initial camera parameters can be estimated automatically by implementing the camera parameter estimation method for a still image. In the proposed method, camera parameter can be estimated in large and natural outdoor environments.

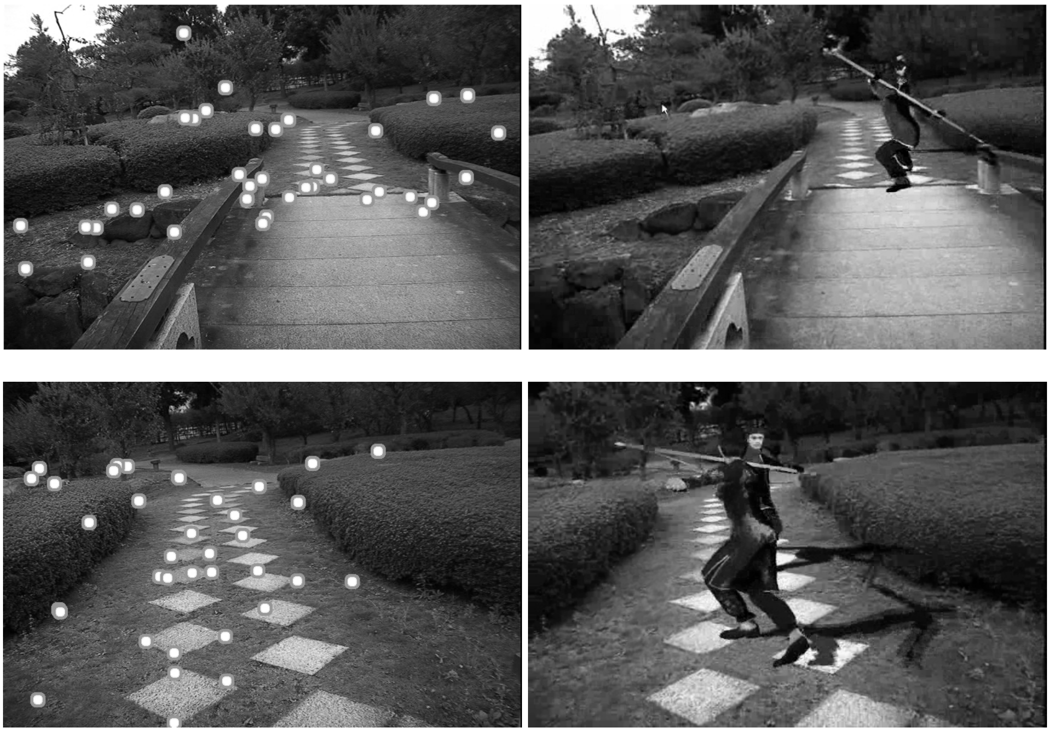
ACKNOWLEDGMENTS

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.



Detected landmarks Overlaid annotations

(a)Navigation by augmented reality



Detected landmarks Overlaid CG actors

(b)MR Pre-visualization

Figure 7. Examples of applications.

REFERENCES

[1] P. Daehne and J. Karigiannis: “Archeoguide: System architecture of a mobile outdoor augmented reality system,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 263–264, 2002.

[2] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. Int. Symp. on Wearable Computers, pp. 21–23, 2003.

[3] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov and S. Lieberknecht: “An augmented reality museum guide,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 103–106, 2008.

[4] M. Shin, B. S. Kim and J. Park: “AR storyboard: An augmented reality based storyboard authoring tool,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 98–99, 2005.

[5] G. Thomas: “Mixed reality techniques for TV and their application for on-set and pre-visualization in film production,” DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking, pp. 31–36, 2005.

[6] H. Kaufmann and A. Dunser: “Summary of usability evaluations of an educational augmented reality application,” Proc. Int. Conf. on Human-Computer Interaction, pp. 22–27, 2007.

[7] B. Williams, G. Klein and I. Reid: “Real-time SLAM relocalisation,” Proc. Int. Conf. on Computer Vision, 2007.

[8] G. Klein and D. Murray: “Parallel tracking and mapping for small AR workspaces,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 225–234, 2007.

[9] M. Oe, T. Sato and N. Yokoya: “Estimating camera position and posture by using feature landmark database,” Proc. Scandinavian Conf. on Image Analysis, pp. 171–181, 2005.

[10] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: “Fully automated and stable registration for augmented reality applications,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 93–102, 2003.

[11] L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3D camera tracking,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 48–57, 2004.

[12] T. Sato, S. Ikeda and N. Yokoya: “Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system,” Proc. European Conf. on Computer Vision, Vol. 2, pp. 326–340, 2004.

[13] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.

[14] S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: “Construction of feature landmark database using omnidirectional videos and GPS positions,” Proc. Int. Conf. on 3-D Digital Imaging and Modeling, pp. 249–256, 2007.

[15] K. Mikolajczyk and C. Schmid: “Scale & affine invariant interest point detectors,” Int. J. of Computer Vision, Vo. 60, No. 1, pp. 63–86, 2004.

[16] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int. J. of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.

[17] 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.

[18] R. Klette, K. Schluns and A. koschan Eds: Computer Vision: Three-dimensional Data from Image, 1998.

[19] P. J. Rousseeuw: “Least median of squares regression,” J. of the American Statistical Association, Vo. 79, pp. 871–880, 1984.