

AR Cultural Heritage Reconstruction Based on Feature Landmark Database Constructed by Using Omnidirectional Range Sensor

Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
{takafumi-t, tomoka-s, yokoya}@is.naist.jp

Abstract. This paper describes an application of augmented reality (AR) techniques to virtual cultural heritage reconstruction on the real sites of defunct constructs. To realize AR-based cultural heritage reconstruction, extrinsic camera parameter estimation is required for geometric registration of real and virtual worlds. To estimate extrinsic camera parameters, we use a pre-constructed feature landmark database of the target environment. Conventionally, a feature landmark database has been constructed in a large-scale environment using a structure-from-motion technique for omnidirectional image sequences. However, the accuracy of estimated camera parameters is insufficient for specific applications like AR-based cultural heritage reconstruction, which needs to overlay CG objects at the position close to the user's viewpoint. This is due to the difficulty in compensation of the appearance change of close landmarks only from the sparse 3-D information obtained by structure-from-motion. In this paper, visual patterns of landmarks are compensated for by considering local shapes obtained by omnidirectional range finder to find corresponding landmarks existing close to the user. By using these landmarks with local shapes, accurate geometric registration is achieved for AR sightseeing in historic sites.

1 Introduction

AR is a technique that enhances the real world by overlaying CG objects. In this study, AR techniques are used for virtual cultural heritage reconstruction on the real site of the defunct temple in ancient Japanese capital Asuka. By using our method, visitors can see virtually reconstructed buildings by CG at the original place as shown in Figure 1. To realize AR-based cultural heritage reconstruction, geometric registration of real and virtual worlds is required; that is, real and virtual world coordinates should be aligned. In the literatures, vision-based registration methods, which result in estimating extrinsic camera parameters, are extensively investigated [1–6] because they can achieve pixel-level geometric registration. These methods can be classified into the following two groups.

One is a visual-SLAM based method [1, 2] that estimates camera parameters without pre-knowledge of target environments. In this method, database



Fig. 1. AR sightseeing in historic site.

construction and camera parameter estimation are carried out simultaneously. The problem of visual-SLAMs is that they only estimate relative camera motion. Thus, this approach cannot be used for position-dependent AR applications like navigation and landscape simulation.

The other uses some kinds of pre-knowledge of target environments such as 3-D models [3–5] and feature landmarks [6]. In this approach, camera parameters are estimated in the global coordinate system. However, a 3-D model based approach usually requires large human costs to construct 3-D models for large-scale environments. On the other hand, a feature landmark-based camera parameter estimation method [6] has been proposed. The method constructs a feature landmark database automatically by using structure-from-motion (SFM) in a large-scale environment. However, the accuracy of estimated camera parameters is insufficient for some kinds of AR applications like AR sightseeing where CG objects may be placed at the position close to the user’s viewpoint as shown in Figure 1. This is due to the difficulty of matching feature landmarks that exist close to the user’s position (close landmarks). In order to successfully compensate for the appearance change caused by the viewpoint change for close landmarks, sparse 3-D information by obtained the SFM process is not sufficient.

In this study, in order to improve the accuracy of vision-based geometric registration at the spot where CG objects of cultural heritage must be placed at the position close to the user, we newly compensate for visual patterns of landmarks using a dense depth map obtained by an omnidirectional laser range sensor. Figure 2 shows the flow diagram of the proposed vision-based registration method. The feature landmark-based geometric registration method is composed of two stages: the database construction stage in an offline process and the geometric registration stage in an online process. Although the framework of geometric registration method is basically the same as the method proposed by Taketomi et al. [6], in our method, image templates of close landmarks are compensated by considering local 3-D structure around the landmark.

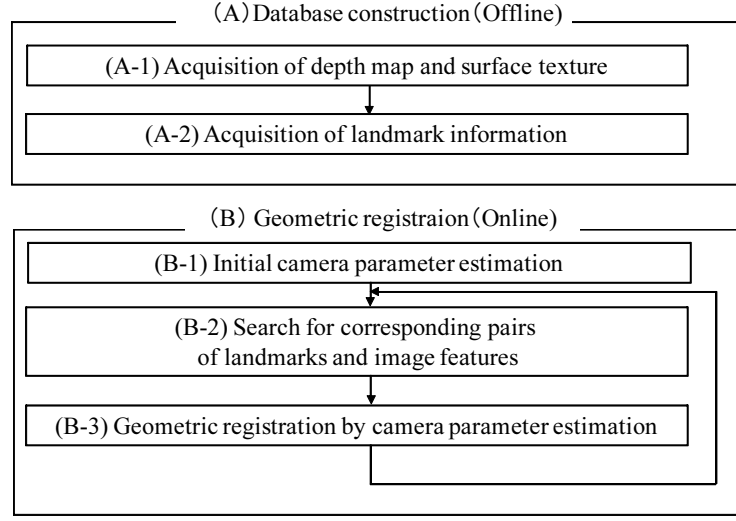


Fig. 2. Flow diagram of proposed vision-based registration.

2 Landmark Database Construction Considering Local 3-D Structure

This section describes a feature landmark database construction process in the offline stage (A) in Figure 2. The feature landmark database must be constructed for the target environment before the online camera parameter estimation (B) is started for geometric registration. In this study, to compensate for image templates of landmarks, we use dense depth information obtained by the omnidirectional laser range sensor.

2.1 Acquisition of Depth Map and Surface Texture

Range data and texture are acquired using the omnidirectional laser range sensor and the omnidirectional camera in the target environment. In this scanning process, the geometrical relationship between these sensors is calibrated and fixed in advance. In the obtained depth map, some parts including the sky area cannot be measured by the laser range sensor. If we simply mask these unmeasurable areas in the pattern matching process, the aperture problem will easily be caused, especially for landmarks that exist at the boundary of the sky and landscape. It should be noted that such landmarks often become the key points for estimating the camera posture. To avoid the aperture problem, in the proposed method, infinite depth values are set to the sky area. Concretely, the largest region where depth values are not available in the omnidirectional image is determined as the sky area.

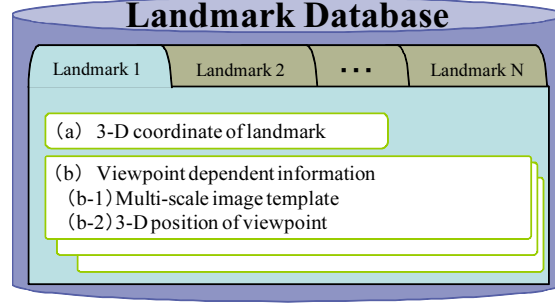


Fig. 3. Elements of landmark database.

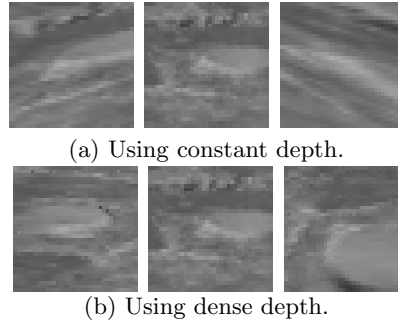


Fig. 4. Example of warped image patterns for some landmarks.

2.2 Acquisition of Landmark Information

The feature landmark database consists of a number of landmarks as shown in Figure 3. Each landmark retains (a) a 3-D coordinate and (b) viewpoint dependent information.

(a) 3-D Coordinate of Landmark: 3-D positions of landmarks are used to estimate extrinsic camera parameters in the online stage (B). For all the feature points detected by using a Harris corner detector [7] from omnidirectional images, 3-D positions of feature points (a) are determined from the depth map obtained by the omnidirectional laser range sensor. These Harris corners are then registered as landmarks.

(b) Viewpoint Dependent Information: In this process, view dependent information is generated for every grid point that is placed on the ground plane around the sensor position. Figure 4(a) shows warped image patterns of close landmark that is stored in the database as image templates by using the SFM-based method [6]. In this figure, the second column shows the generated image pattern from the original viewpoint (position of the sensor). The first column and third column show warped image patterns where the viewpoints are set five meters to the right and forward from the original viewpoint, respectively. As can be seen in Figure 4(a), in the SFM-based database construction, warped images of landmarks that exist close to the user’s position are largely distorted because

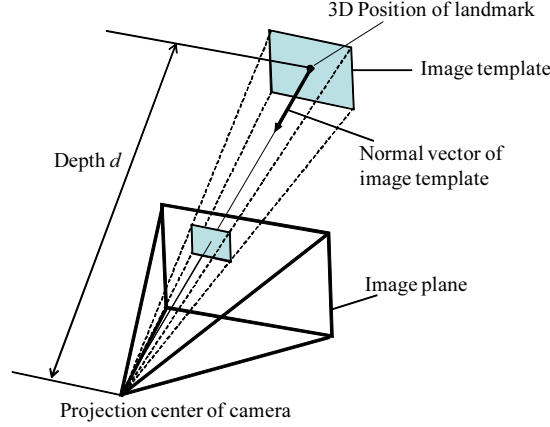


Fig. 5. Generation of image template by conventional method.

image patterns are compensated with constant depth values d acquired for the landmark by SFM as shown in Figure 5.

In the proposed method, dense 3-D data obtained by the omnidirectional laser range sensor is used to correctly compensate for image templates of landmarks. Concretely, first, depth values d_i for each pixel i on the image template are obtained from range data. Next, pixel i values on the image template is determined by projecting an omnidirectional image using these depth values d_i . In this pattern generation, occluded areas in the image template are set as masked areas in order to ignore them in the pattern matching process. Figure 4(b) shows warped images obtained by using the proposed method. It can be observed that warped images are generated without large distortion by using dense 3-D information.

3 Geometric Registration: Extrinsic Camera Parameter Estimation Using Landmark Database

This section describes the camera parameter estimation stage in the online process (B) in Figure 2 for AR geometric registration. In this process, first, initial camera position and posture are estimated. Initial camera position and posture for the first frame of the input are assumed to be given by the landmark-based camera parameter estimation method for a still image input [8] (B-1). Next, search for corresponding pairs (B-2) and geometric registration (B-3) are repeated.

Search for Corresponding Pairs: In this process, corresponding pairs of landmarks and image features are searched for in the current frame. First, landmarks used to estimate camera parameters 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 (Sum of Squared Differences) based tracker. After landmark tracking,

tentative camera parameters are determined using tracked landmarks. Image templates from the nearest viewpoint from the current camera position are then selected from the database. Finally, corresponding pairs between landmarks and image features are searched for using NCC (Normalized Cross-Correlation) with ignoring masked pixels.

Geometric Registration by Camera Parameter Estimation: After determining the corresponding pairs of landmarks and image features, extrinsic camera parameters are determined in the world coordinate system by solving the PnP problem [9] using these pairs. In order to remove outliers, the LMedS estimator [10] is employed in this process. After estimating extrinsic camera parameters, CG objects that are placed in the world coordinate system in advance are overlaid on the input image by using projection matrix computed by estimated camera parameters.

4 Experiments

To demonstrate the usefulness of the proposed method, first, the effectiveness of pattern compensation by considering local 3-D structure of the landmark is evaluated. Next, estimated camera parameters are compared with those by the SFM-based method [6].

In this experiment, the landmark database is constructed for an outdoor environment using an omnidirectional multi-camera system (Point Grey Research Ladybug2) and an omnidirectional laser rangefinder (Riegl LMS-Z360). Figure 6 shows a panoramic image and corresponding depth map used for database construction. In this experiment, the ground plane of the target environment is divided into 10×10 grid points at 1 meter intervals. To compare the accuracy of estimated camera parameters, the SFM-based feature landmark database is also constructed in the same place. For both methods, the same video image sequence (720×480 pixels, progressive scan, 15fps, 250 frames) captured in the target environment is used as the input video for the online camera parameter estimation. In this experiment, camera position and posture for the first frame are given manually.

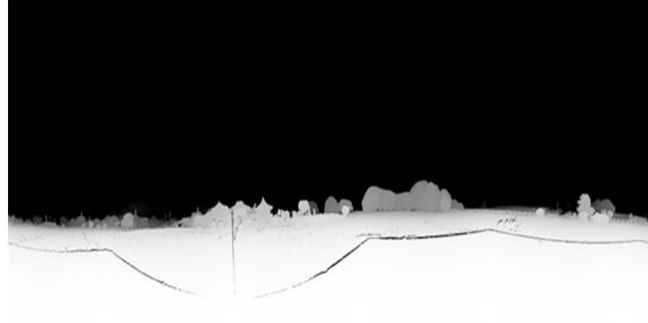
4.1 Quantitative Evaluation of Pattern Compensation

Generated image templates of landmarks by the proposed and the previous methods are quantitatively evaluated by comparing to ground truth. To show the effectiveness of pattern compensation, compensated image templates of landmarks exemplified in Figure 4 are compared with image patterns of landmarks in input images. In this experiment, viewpoints for pattern compensation are given by estimating camera parameters with manually specified correspondences of landmarks in input images.

Table 1 shows average and standard deviation of normalized cross-correlation values between compensated image templates and image patterns of landmarks in input images for 30 image templates of landmarks. In the proposed method



(a) Panoramic image taken by omnidirectional multi-camera system.



(b) Depth map taken by omnidirectional laser range sensor.

Fig. 6. Acquired omnidirectional data.**Table 1.** Comparison of normalized cross-correlation value.

	Proposed method	Previous method [6]
Average	0.63	0.47
Standard deviation	0.039	0.052

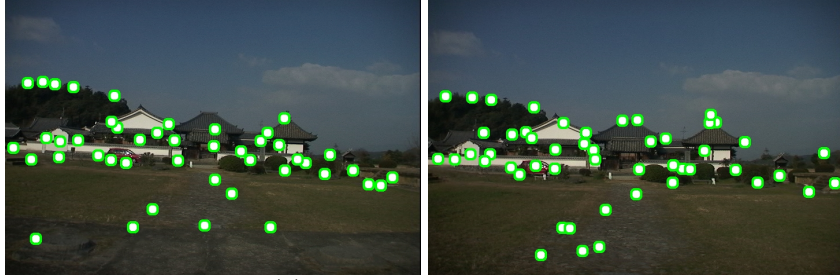
which considers dense depth information, the average normalized cross-correlation value (0.63) is higher than that of the previous method (0.47) which does not consider the local 3-D structure around the landmark. From this result, we can confirm that compensated image templates are more similar to image patterns of landmarks in input images than that of the previous method.

4.2 Quantitative Evaluation of Estimated Camera Parameters

In the second experiment, the accuracy of estimated camera parameters is quantitatively evaluated and compared with the previous method [6]. Figure 7 shows landmarks in example frames that are used for camera parameter estimation. As shown in this figure, corresponding pairs of landmarks and feature points are successfully found for the ground part in the proposed method, while the previous method could not find any corresponding landmarks at the ground part of



(a) By the previous method [6].



(b) By the proposed method.

Fig. 7. Corresponded landmarks.**Table 2.** Comparison of the accuracy

	Proposed method	Previous method [6]
Average of position error (mm)	231	342
Standard deviation of position error (mm)	107	164
Average of posture error (degree)	1.11	1.41
Standard deviation of posture error (degree)	0.52	0.46

the images. This is regarded as the effect of appropriate pattern compensation using dense 3-D information.

Table 2 shows the accuracy of each method. To evaluate the accuracy of estimated camera parameters, we create the ground truth by estimating camera parameters with manually specified correspondences of landmarks. Note that we have removed several frames in which the reprojection error of the obtained ground truth is over 1.5 pixels. From this result, the accuracy of the proposed method has been proven to be improved than that of the previous method. Figures 8 and 9 illustrate errors in position and posture, respectively. In most frames, errors of the proposed method are the same or smaller than that of the previous method.

Figure 10 shows examples of generated images using the proposed method for AR sightseeing in Asuka, Japan. Virtual objects are overlaid on the site of the old temple. We have confirmed that CG objects placed at the position close to the user's viewpoint are correctly registered.

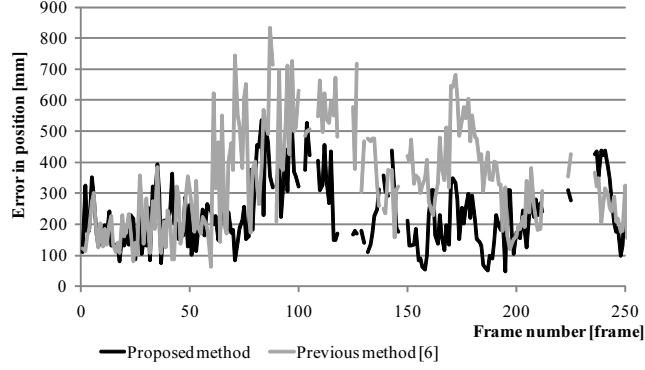


Fig. 8. Error in position for each frame

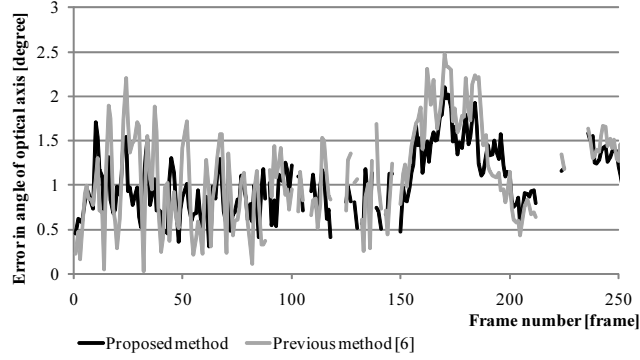


Fig. 9. Error in posture for each frame

5 Conclusion

In this paper, we have proposed a method to use dense depth information for landmark-based geometric registration for realizing AR sightseeing in the historic site. In this method, unlike other methods, the landmarks close to the user's viewpoint that effect the accuracy of geometric registration are aggressively used by compensating its visual patterns based on dense depth information acquired by using omni-directional range finder. Importance of close landmarks are validated quantitatively through the experiment. It should be noted that the proposed method is not for large-scale environments but for selected places where the accuracy of geometric registration largely depends on close landmarks. In future work, we will develop a method that uses both dense and sparse 3-D structures for efficiently constructing the database in large-scale outdoor environments.

Acknowledgement

This research is supported in part by Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Agency (JST), and the "Ambient Intelligence" project granted by the Ministry of Education, Culture, Sports, Science and Technology.

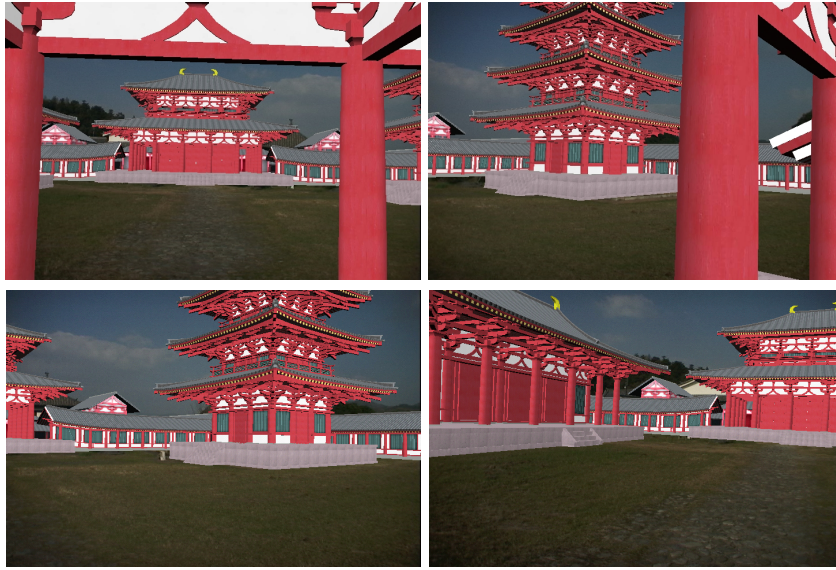


Fig. 10. User's views in AR sightseeing.

References

1. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. *Proc. Int. Symp. on Mixed and Augmented Reality* (2007) 225–234
2. Williams, B., Klein, G., Reid, I.: Real-time SLAM relocalisation. *Proc. Int. Conf. on Computer Vision* (2007)
3. Lepetit, V., Vacchetti, L., Thalmann, D., Fua, P.: Fully automated and stable registration for augmented reality applications. *Proc. Int. Symp. on Mixed and Augmented Reality* (2003) 93–102
4. Vacchetti, L., Lepetit, V., Fua, P.: Combining edge and texture information for real-time accurate 3D camera tracking. *Proc. Int. Symp. on Mixed and Augmented Reality* (2004) 48–57
5. Yang, G., Becker, J., Stewart, C.V.: Estimating the location of a camera with respect to a 3d model. *Proc. Int. Conf. on 3-D Digital Imaging and Modeling* (2007) 159–166
6. Taketomi, T., Sato, T., Yokoya, N.: Real-time camera position and posture estimation using a feature landmark database with priorities. *CD-ROM Proc. 19th IAPR Int. Conf. on Pattern Recognition* (2008)
7. Harris, C., Stephens, M.: A combined corner and edge detector. *Proc. Alvey Vision Conf.* (1988) 147–151
8. Susuki, M., Nakagawa, T., Sato, T., Yokoya, N.: 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* (2007) 124–129
9. Klette, R., Schluns, K., Koschan Eds, A.: *Computer Vision: Three-dimensional Data from Image*. (1998)
10. Rousseeuw, P.J.: Least median of squares regression. *J. of the American Statistical Association* **Vol. 79** (1984) 871–880