Achieving Robust Alignment for Outdoor Mixed Reality using 3D Range Data

Masaki Inaba
The University of Tokyo
Tokyo, Japan
inaba@cvl.iis.u-tokyo.ac.jp

Atsuhiko Banno
Advanced Industrial Science and Technology
Ibaraki, Japan
atsuhiko.banno@aist.go.jp

Takeshi Oishi
The University of Tokyo
Tokyo, Japan
oishi@cvl.iis.u-tokyo.ac.jp

Katsushi Ikeuchi
The University of Tokyo
Tokyo, Japan
ki@cvl.iis.u-tokyo.ac.jp

ABSTRACT

Mixed reality (MR) technology can be applied to various applications such as architecture, advertising, and navigation systems, so the desire to utilize MR in outdoor environments has been increasing. In order to utilize MR, it is necessary to achieve alignment super imposing virtual contents in the desired position. However, because light changes continually in outdoor environments, and the appearance of real objects changes also, in some cases the previous image-based alignment methods do not work well. In this paper, a robust image-based alignment method to be used in outdoor environments is proposed. In the proposed method, the albedo of real objects is estimated using 3D shapes of these objects in advance, and the appearance is reproduced from the albedo and current light environment. The appearance of real objects and reproduced image becomes close, so a robust image-based alignment is achieved.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Artificial, augmented, and virtual realities

Keywords

Mixed Reality, Augmented Reality, Alignment, 3D Range Data, Albedo, Analysis by Synthesis

1. INTRODUCTION

Mixed reality (MR) is the technique to combine the real world and the virtual world by superimposing virtual contents onto the real world. By using MR technology, users can experience virtual contents as though they existed in the real world through a head mounted display (HMD), a smartphone, or some other device. MR technology can be applied to many fields such as education, architecture, medical treatment, advertising, navigation systems, and entertainment. So there is much research on MR [2, 3].

In recent years, considering the many possible applications, desire to utilize MR in outdoor environments as well as in indoor environments has been increasing. Applications using outdoor MR technology include virtual completion of outdoor building and virtual outdoor advertising. In addition, virtual restoration of lost cultural assets has attracted interest. By superimposing a 3D CG model restored using CAD into the real world, users can watch restored cultural assets in their actual setting, so they can feel that they are present [10, 18, 26].

One of the issues to be solved for seamless MR is geometric registration. Geometric registration is necessary to achieve alignment between a real object and a virtual object. In order to achieve this alignment, estimation of the camera position and pose is necessary. The processes to calculate absolute camera parameters and relative camera parameters are hereinafter referred to as **global alignment** and **local alignment**.

In this paper, by using a dense 3D shape of the real environment (3D model) retrieved by laser range sensor [9], a robust image-based alignment method for light changing is proposed because light changes continually in an outdoor environment. The following describes the gist of the proposed method.

First, the reflectance (albedo) of real objects is estimated from a 3D model and omnidirectional image measurement in advance [17]. Thus, a 3D model with albedo is created using texture mapping. This makes it possible to achieve robust feature-matching for light changes and to estimate the exact camera position and pose. In global alignment, the appearance is reproduced by radiosity from the 3D model with albedo and current light environment, and correspondences between the camera image and reproduced appearance are searched using natural features.

In local alignment, by projecting natural features of a previous camera frame on a 3D model, a 3D natural features map is generated in real time. Because these natural features are extracted from the camera image in the current light environment, the feature-matching between the current camera frame and the previous camera frame is easy, so camera position and pose are estimated relatively. By repeating the camera parameters estimation and natural features projection, a robust local camera position and pose tracking is possible. However, errors of camera parameters estimation are accumulated using only local alignment. So the appearance is reproduced in real time by simple shading from the 3D model with albedo and current light environment, and global realignment

VRST'12, December 10–12, 2012, Toronto, Ontario, Canada.

is tried at the same time. If the feature-matching between the current camera frame and reproduced appearance is successful, global realignment is performed, and it is possible to eliminate the accumulated error.

This paper is organized as follows. Chapter 2 introduces related research on alignment, and describes the problems of previous methods and the significance and purpose of this study. Chapter 3 describes the proposed method of robust global alignment and local alignment for light changes using a 3D model with albedo. Chapter 4 shows the effectiveness of the proposed method by evaluation experiments. Chapter 5 contains the conclusion and future issues.

2. RELATED WORK

In order to achieve alignment, it is necessary to acquire the camera position and pose, and continue to match the coordinate systems of the real world and the virtual world. There are various studies to estimate the camera position and pose, and the technologies that have attracted attention recently are the feature-based method, the 3D map-based method, and the analysis by synthesis (AbS)-based method.

2.1 Feature-based Method

Feature-based Methods use visual tracking to estimate camera pose. Diverdi et al [7] use Lucas and Kanade's optical flow [14] by Shi and Tomasi's good features [23], on the other hand Wagner et al [27] use Fast corner tracking [20]. These methods can estimate camera pose in real time. But, feature-based methods cannot estimate camera position, so they must be combined with inertial sensor, and so on [5, 21]. Further, estimation errors accumulate little by little.

2.2 3D Map-based Method

In 3D map-based methods, a 3D range map of local point features is constructed from camera images. The camera position and pose can be estimated from the 2D-3D correspondences between the camera image (real world) and the 3D features map (virtual world).

A method called PTAM (Parallel Tracking And Mapping) [11] uses visual SLAM (Simultaneous Localization And Mapping) to perform 3D range mapping of environment and camera parameter estimation at the same time by tracking local point features of the camera image. In this approach, it is possible to achieve rough local alignment without prior information using only one camera. As a method similar to PTAM, DTAM (Dense Tracking and Mapping) [15] has been proposed. PTAM adopts a stereo method using only feature points, but DTAM adopts a stereo method using all the pixels of the camera image. By creating a dense depth map, the accuracy of alignment improves. However, in these methods, since there is no prior information of the real world, the virtual contents have to be placed roughly, so exact alignment between real objects and virtual objects is difficult. In addition, errors accumulate gradually.

Taketomi et al [25] propose a method to create a 3D landmark database of local point features by calculating the 3D position using SfM (Structure from Motion) in advance. The camera position and pose estimation is carried out by feature-matching between the camera image and the 3D landmark database. A highly accurate global alignment can be achieved in real time using this method. However, if the light environment of the current camera image is very different from that of the database image due to time or weather, feature-matching is difficult because appearance of these images is different. In particular, because light changes continually in outdoor environments, this is a fatal problem.

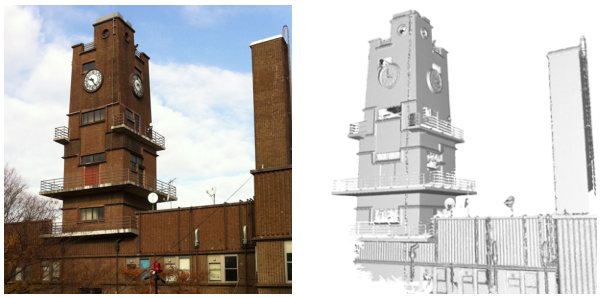


Figure 1: Real object and 3D model

2.3 Analysis by Synthesis-based Method

Analysis by synthesis-based methods use 3D CAD models of real objects. The flow of 3D model rendering, feature-matching between the rendered image and the camera image, feature projection to 3D model, and camera parameter estimation is repeated.

Reitmayr et al [19]use a textured 3D model of the surrounding environment. In this method, edges and shape of the rendered image and the camera image are extracted and matched. Then, these features are projected back to the 3D model, so 2D-3D correspondences are acquired. This method achieves sufficiently accurate alignment in real time. In [30], 3D line model is created, and edges and shape are used as features similarly. This method is more faster than the method of Reitmayr et al. But edges and shape are weak against the occlusion, and cannot be used for the objects that have few edges.

In [12, 24], local point features are used instead of edges. These methods also achieve precise alignment in real time. Local point features are robust against the occlusion, but computational cost is high. In addition, these methods do not consider the influence of the light sources. So, if the light environment changes, feature-matching cannot work well.

Schumann et al [22] use a 3D model with diffuse parameters and try to achieve tracking by AbS under the assumption that the 3D model can be rendered in the current light environment. In order to estimate camera position and pose, various local feature-based methods are compared and the similarity-based method is considered. However, these methods cannot run in real time. Further, the creation of a 3D model with diffuse parameter, the rendering of a 3D model in the current light environment, and light environment estimation are not achieved in this study.

From the above, in this research, it is proposed to achieve a markerless absolute alignment method that is robust for light changing. The proposed method is based on the AbS method and considers the effect of light sources.

3. METHOD

In order to achieve geometric registration, it is necessary to always obtain the camera position and pose. By using a color image and a 3D model acquired in advance, in addition to the input camera image, global alignment and local alignment are achieved. The color image and the 3D model acquired in advance are hereinafter referred to as the database.

If the color image and the 3D model in the database are calibrated, by searching correspondences between the input camera image and the database image, camera parameters can be estimated, and global alignment can be performed. By using an existing fea-

ture point extraction method, correspondences can be automatically extracted from the two images, but if the light environment is very different, correct feature-matching becomes difficult. In such a case, the camera parameters estimation will fail and the virtual contents cannot be superimposed on the desired position.

In the proposed method, the database image is converted to a reproduced image that has the same light environment as the input camera image does. The reflectance (albedo) of real objects is estimated from the database image and the 3D model in advance [17], and the appearance in the current light environment is reproduced from the albedo. Since the light environment of the reproduced appearance is similar to that of the input camera image, the appearance of the two images becomes close, and feature-matching becomes easy.

By mapping the natural features of the input image to the 3D model after the initial global alignment, a 3D natural features map is created in real time, so local alignment is achieved. Feature-matching is easy because the natural features are extracted from the current appearance. However, the error of the camera parameters estimation will accumulate. So appearance is reproduced by simple shading in real time, and global realignment is attempted at the same time.

3.1 Database Creation

The database comprises the color image and the 3D model of the real object (Fig. 1). The color image is obtained by shooting multiple times while changing the shutter speed using an omnidirectional camera so as to be an HDR (High Dynamic Range) image that has a wide brightness range. The 3D model is acquired by a laser range sensor. The camera parameters of the HDR image are calculated in the coordinate system of the 3D model by acquiring correspondences manually. Then, the calibration of the color image and the 3D model is performed.

3.1.1 Albedo Estimation

The albedo of real objects is estimated using the method of Okura et al [17]. Under the narrow-band camera assumption, the camera sensitivity can be approximated by the Dirac delta function, so the wavelength of the light source is constant [29]. Therefore, when k means the channel of RGB, the intensity of camera image I_k is represented by the following equation:

$$I_k = \tau_k S_k E_k \qquad (k = \{r, g, b\}) \tag{1}$$

τ_k is the camera gain, S_k is the albedo value, E_k is the illuminance value.

In order to estimate the albedo value S_k from the above equation, it is necessary to determine the illuminance value E_k on the surface of the real object. Illuminance value E_k is calculated by the radiosity method using IBL (Image-Based Lighting) [6]. IBL can be performed using the obtained omnidirectional HDR image as a light environment map with the rendering software Radiance[1], hereinafter referred to as **Radiance**. By setting the albedo value to 1.0 at the surface of the 3D model in the rendering, the illuminance image I'_k can be generated by the following equation:

$$I'_k = \tau_k E_k \tag{2}$$

Because the brightness distribution of the light source and the real object is acquired by the same omnidirectional camera, camera gain τ_k of Eq. (1) and Eq. (2) are equal. The albedo value S_k of the real object is obtained by the following equation:

$$S_k = \frac{I_k}{I'_k} \tag{3}$$

The albedo image is obtained in this way, and the 3D model with albedo is created by texture mapping.

[1]http://radsite.lbl.gov/radiance/

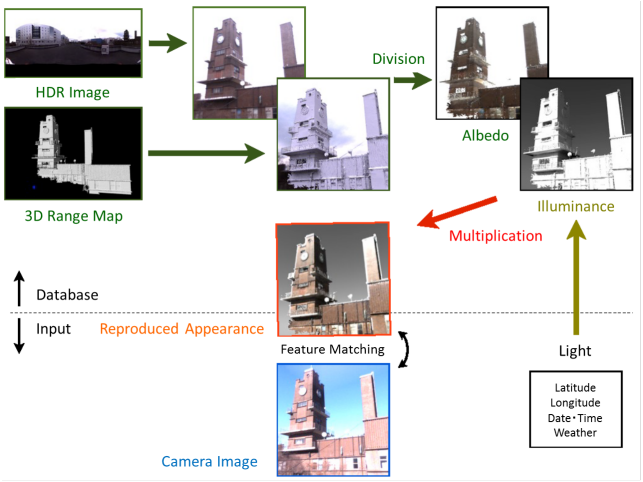


Figure 2: Process of global alignment

3.2 Global Alignment

Fig. 2 shows the process of the proposed method from appearance reproduction to global alignment.

3.2.1 Appearance Reproduction

First, given as parameters the date, time, weather, and GPS information when performing global alignment, the sky light environment is simulated using Radiance, and then the illuminance value E_k is calculated from the 3D model. This method has the advantage that there is no need to shoot the HDR image in order to acquire the light environment. In addition, the reproduced appearance is only used for feature-matching by natural features, so it is not necessary to reproduce the exact light environment at the time of obtaining the input camera image. Using the albedo value S_k in the database and the illuminance value E_k, appearance is reproduced from Eq. (1).

3.2.2 Feature Matching

Feature-matching between the input camera image and the reproduced appearance is performed by natural features. It is necessary to obtain a sufficient number of correct correspondences for accurate camera parameters estimation in feature-matching. So removal of incorrect correspondences is attempted based on the RANSAC (RANdom SAmple Consensus) method [8] using epipolar constraint.

Because calibration between reproduced appearance and 3D model has been taken, 2D-3D correspondences of the input camera image and 3D model are obtained from the results of feature-matching of the input image camera and the reproduced appearance.

3.2.3 Camera Parameter Estimation

The camera parameters are calculated from n 2D-3D correspondences. It is desirable that the number of correspondences n is 6 or more. Internal camera parameters $\boldsymbol{P}$ are assumed to be known by

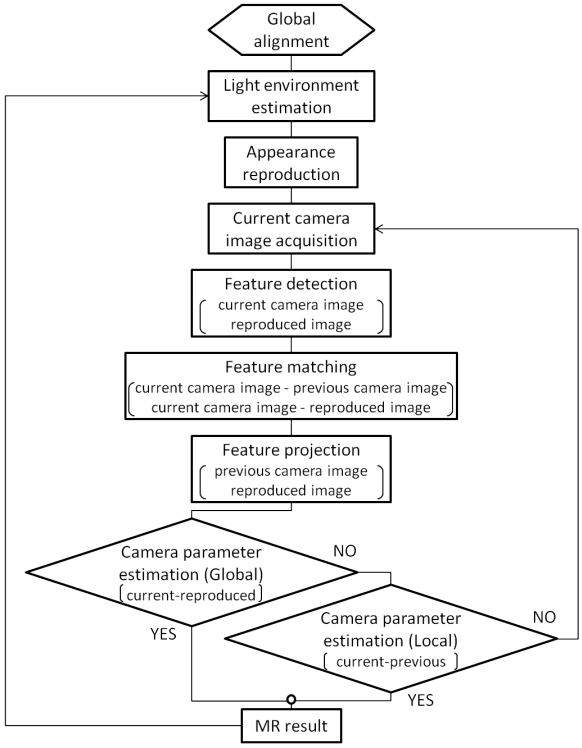


Figure 3: Flowchart of local alignment

the method of Zhang [31]. External camera parameters are configured by the translational component $\boldsymbol{T}$ for camera position and the rotational component $\boldsymbol{R}$ for camera pose. In order to determine the external camera parameters $\boldsymbol{R}, \boldsymbol{T}$, linear estimation is performed following the method of Weng et al [28]. Then, external camera parameters $\boldsymbol{R}, \boldsymbol{T}$ are calculated by minimizing the sum of squares of the reprojection error$\boldsymbol{F}_i$ using the linear estimation result as the initial value.

$$\boldsymbol{F}_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} - \boldsymbol{P} \begin{bmatrix} \boldsymbol{R}^\top & -\boldsymbol{R}^\top \boldsymbol{T} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{4}$$

$$\{\boldsymbol{R}, \boldsymbol{T}\} = \arg\min_{\boldsymbol{R},\boldsymbol{T}} \sum_{i}^{n} \|\boldsymbol{F}_i\|^2 \tag{5}$$

(u_i, v_i) are the 2D coordinates in the input camera image, and (X_i, Y_i, Z_i) are the corresponding 3D coordinates in the 3D model. The method of Levenberg-Marquardt is used for nonlinear optimization calculation.

3.2.4 Model Transformation

By converting the 3D model coordinate system using external camera parameters $\boldsymbol{R}, \boldsymbol{T}$, it is possible to adapt the coordinate systems of the real world and the virtual world. By these processes, it is possible to realize absolute alignment between the real objects and the virtual contents.

3.3 Local Alignment

After global alignment, it is necessary to keep track of the camera position and pose by local alignment. Fig. 3 shows the flow of local alignment.

3.3.1 Camera Tracking

By mapping natural features of the previous camera frame to a dense 3D model, a 3D natural features map is generated in real time. Because the calibration of previous camera frame and 3D model has been completed posteriorly, 3D coordinates can be obtained easily at the feature points of the camera image. In addition, due to natural features from the current appearance, feature-matching between the current camera frame and previous camera frame is easy, so the camera position and pose can be obtained from Eq. (5).

By repeating the camera parameters estimation and the feature projection, robust local camera position and pose tracking is possible. However, since the error of the camera parameters estimation in each frame accumulates, a gap is formed gradually. In order to eliminate this error, it is necessary to perform global realignment.

3.3.2 Global Realignment

In order to eliminate the error, appearance is reproduced in every frame as well as the global alignment, and the absolute alignment is tried by feature-matching using reproduced appearance. Because it is necessary to reproduce appearance according to the movement of the camera, appearance reproduction must be done in real time. However, because there is high computational cost in reproducing appearance by radiosity, real-time processing is difficult. Accordingly, appearance is reproduced in a short time by simple shading based on the reflection model of Lambert. In addition, light distribution is necessary in order to perform the shading, and light environment must be estimated for each frame because it always changes due to the movement of clouds and sun in an outdoor environment. There are various methods, but the proposed method estimates the light environment directly using the sky zenithal image acquired by a fisheye camera, an omnidirectional camera, or some such device [10]. Estimated light environments are also used to achieve photometric registration in MR.

The 3D model with albedo is rendered from the estimated camera parameters and light environment to reproduce appearance. If a sufficient number of accurate correspondences is acquired between the current camera frame and the reproduced appearance, global realignment is performed, and it is possible to eliminate the errors. In addition, using the fact that the calibration between the previous camera frame and reproduced appearance is performed roughly, the removal of incorrect correspondences is carried out by the geometric relationship between the current camera frame and previous camera frame in feature-matching. In this way, even if correct correspondences are considerably less with respect to incorrect, only the correct correspondences are extracted well.

4. EXPERIMENT

In order to verify the effectiveness of the proposed alignment method, experiments of the alignment between the real object and the 3D model were conducted under different light environments. In this experiment, the building was used as shown in Fig. 1 and the PC spec was CPU: Intel Core2Duo T8300 2.40GHz, RAM: 4GB, GPU: nVIDIA GeForce8800M GTX 512MB.

4.1 Database Creation

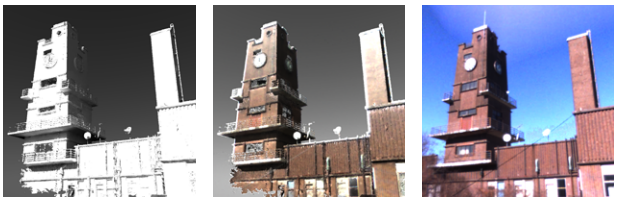
The omnidirectional HDR color image and the 3D model were necessary to build the database. Ladybug3 of Point Gray Research[2] was used to acquire the omnidirectional HDR color image. HDS3000 Leica Geosystems[3] was used as the laser range sensor to obtain the

[2]http://www.ptgrey.com/products/ladybug3/

[3]http://hds.leica-geosystems.com/en/5574.htm

(a) database image (b) illuminance image (c) albedo image

Figure 4: Albedo estimation result (11:30AM, 16th of December, sunny)



(a) illuminance image (b) reproduced image (c) input image

Figure 6: Appearance reproduction result by radiosity (11:30AM, 8th of December, sunny)

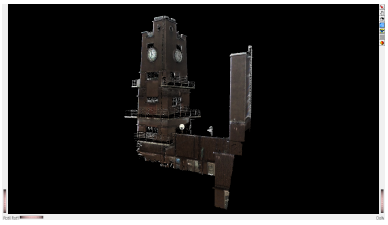


Figure 5: 3D model with albedo

3D model, and the integration of multiple distance data was carried out [16]. The 3D model has one million points. The camera parameters of the HDR image were estimated by giving 2D-3D correspondences between the HDR image and the 3D model manually.

Fig. 4 shows the result of albedo estimation. Although the HDR image had a clear difference in intensity between the front and side, the albedo image did not have that much difference. In addition, the 3D model with albedo was created by texture mapping using estimated albedo (Fig. 5). In the manner as described above, the database of omnidirectional HDR image and albedo image with 3D coordinate value was constructed.

4.2 Global Alignment

4.2.1 Appearance Reproduction

The input image was taken at a different time and date, and under different weather conditions. The illuminance image was generated using Radiance from the date, time, weather, and GPS information at the input image acquisition, and appearance was reproduced by the multiplication of the albedo image and illuminance image.

Fig. 6 shows the results of appearance reproduction by radiosity. Since the shading of the reproduced appearance is similar to that of the input image, the appearance was reproduced with correct shade and shadow. However, the process of appearance reproduction by radiosity took several minutes.

4.2.2 Feature Matching

The comparison of feature-matching between the previous method using a simple input image and the proposed method using reproduced appearance was performed. The resolution of all images is 600×600 pixels. SIFT [13], SURF [4], and CSIFT (Colored SIFT) [1] were used as the method of natural feature extraction. CSIFT is an illumination invariant feature. The number of correspondences and computational time were evaluated using images with the different light environment.

Table 1, 2 and Fig. 7, 8 show the results of feature-matching after the removal of incorrect correspondences based on the epipolar constraint by RANSAC. First, in Table 1 and Fig. 7, since the appearance of the input image and database image was not very different, correct correspondences were obtained sufficiently to estimate the camera parameters in both the proposed method and previous method. On the other hand, in Table 2 and Fig. 8, since the appearance of the input image with deep shade was very different from that of database image, correct correspondences were not obtained in the previous method. Because there were too many incorrect correspondences as opposed to correct correspondences, the removal of incorrect correspondences by RANSAC did not work well. However, since the appearance of the input image and reproduced image became close, many correct correspondences could be obtained using the proposed method, so the proposed method was shown to be effective. In addition, correct correspondences were not obtained by CSIFT in either case.

The reason for failure by CSIFT as shown in Fig. 7(c), 8(c) was thought to be a feature of a building image that had little texture where the boundary almost disappeared by removing the light environment information. In other words, it can be said that in this case, the light environment information is very important to obtain correspondences, so the proposed method to reproduce appearance was considered to be reasonable. Correspondences could be obtained from around the edge with deep shading as shown in the Fig. 8(a), 8(b). Therefore, the proposed method uses the information for 3D geometric structure as well as information about the texture of the object. Considering the number of correspondences and computational time, the proposed method using SURF was fast and sufficient.

4.2.3 Camera Parameter Estimation

The accuracy of the camera parameters estimation was evaluated by the success rate of alignment and average of reprojection error represented by Eq. (5). The resolution of camera image is 640×480 pixels. If the average of reprojection error was less than 3 pixels, alignment was regarded as a success.

As the result of 100 trials, the average of reprojection error was 120 pixels, and alignment success rate was 4% in the previous method. On the other hand, the average of reprojection error was 2.5 pixels, and alignment success rate was 89% in the proposed method, so the effectiveness of the proposed method was confirmed. Fig. 9 shows the result of the correct alignment by the proposed method. The 3D model was superimposed to the same position as the real object, so the coordinate systems of the real world and the virtual world were calibrated.

4.3 Local Alignment

Because the local camera position and pose tracking must be performed in real time, it was necessary to perform feature-matching in real time. So CUDASURF[4] implemented by GPGPU was used.

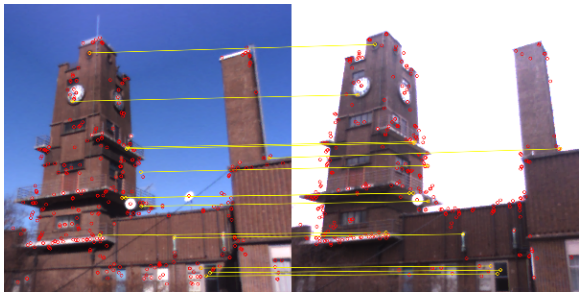
[4]http://www.d2.mpi-inf.mpg.de/surf

Table 1: Comparison of feature matching (input image: 09:30AM, 8th of December, sunny. database image: 11:30AM, 16th of December, cloudy.)

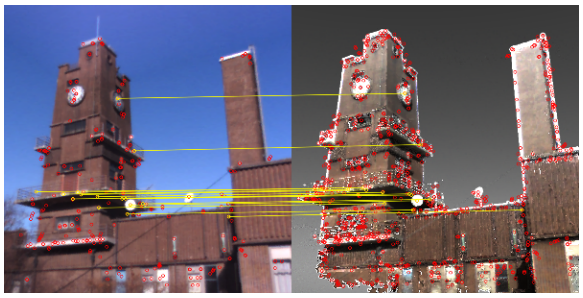
	SIFT +previous	SIFT +proposed	SURF +previous	SURF +proposed	Colored SIFT
correspondences	31	24	14	15	0
calculation time[sec]	1.5	2.0	0.7	1.0	1.8

Table 2: Comparison of feature matching (input image: 11:30AM, 8th of December, sunny. database image: 11:30AM, 16th of December, cloudy.)

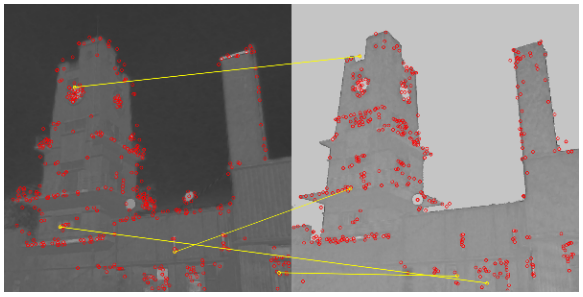
	SIFT +previous	SIFT +proposed	SURF +previous	SURF +proposed	Colored SIFT
correspondences	0	42	0	63	0
calculation time[sec]	1.8	2.4	1.0	1.2	2.4



(a) SURF matching between input image and database image



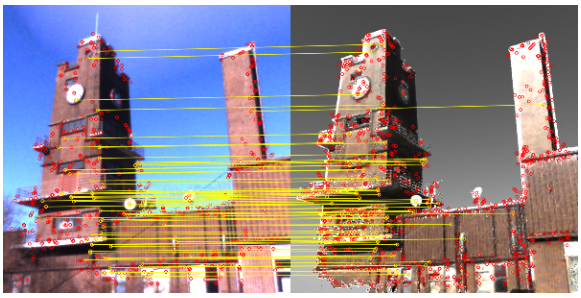
(b) SURF matching between input image and reproduced appearance



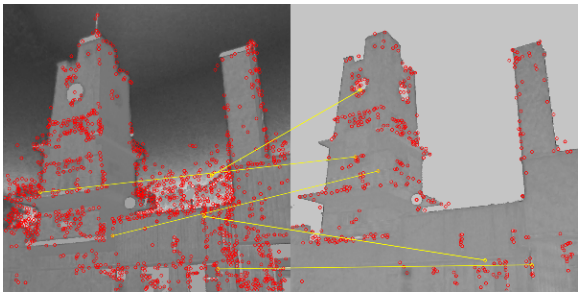
(c) Colored SIFT matching between input image and database image

Figure 7: Feature matching result (input image: 09:30AM, 8th of December, sunny. database image: 11:30AM, 16th of December, cloudy.)

(a) SURF matching between input image and database image

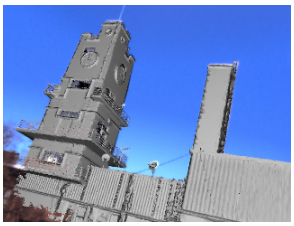


(b) SURF matching between input image and reproduced appearance

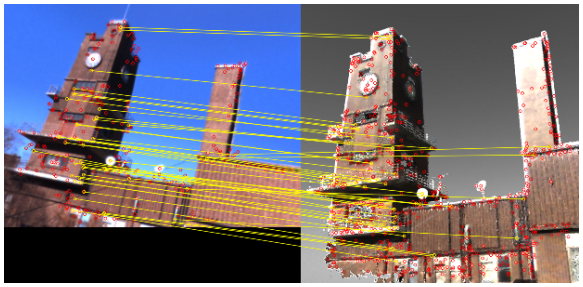


(c) Colored SIFT matching between input image and database image

Figure 8: Feature matching result (input image: 11:30AM, 8th of December, sunny. database image: 11:30AM, 16th of December, cloudy.)



(a) MR Result



(b) SURF matching between camera image and reproduced appearance by radiosity

Figure 9: Global alignment result

Table 3: Processing times of each step

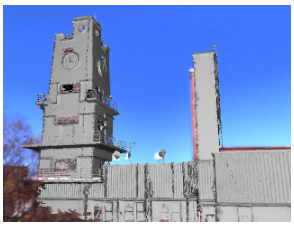
each step	time[msec]
light environment estimation	10
appearance reproduction by simple shading	180
CUDASURF detection	50
feature-matching + RANSAC	60
feature projection	10
camera parameter estimation	10
total	320

Fig. 10 shows the results of the local alignment. Enough correspondences between the current camera image and previous camera image were obtained, and the tracking by the proposed method was shown to work well (Fig. 10(b)). However, the errors of the camera parameters estimation accumulated as time increased, so deviation would arise little by little (Fig. 10(a)). If feature-matching using reproduced appearance by simple shading was successful, global realignment was performed as shown in the Fig. 11, so it is possible to prevent the accumulation of errors. Table 3 shows the processing times of each step. The local alignment ran at about 3fps, so it cannot be said to be running in real time. The appearance reproduction, feature detection, and feature matching must be more efficient.

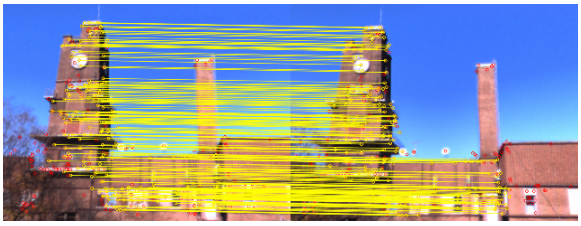
5. CONCLUSION

In this paper, a robust alignment method for light changes in an outdoor environment is proposed.

By using the 3D model of the surrounding environment, the albedo of real objects is estimated, and the appearance in current light environment is reproduced by radiosity from the albedo, so robust feature-matching for light changes is achieved. This makes it possible to achieve robust global alignment even in an outdoor environment. As an experiment result, if the light environment was very different, the alignment success rate of the previous method was



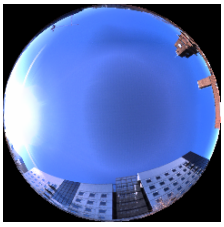
(a) MR Result



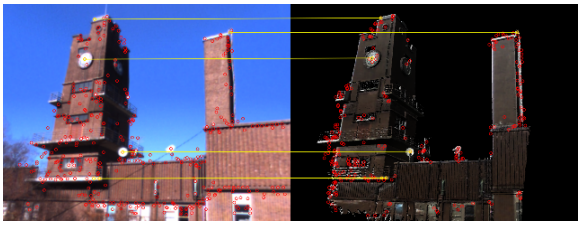
(b) SURF matching between current camera image and previous camera image

Figure 10: Local alignment result

(a) MR Result



(b) Light environment



(c) SURF matching between camera image and reproduced appearance by simple shading

Figure 11: Global realignment result

4%. On the other hand, that of the proposed method was improved by up to 89%.

In addition, by projecting natural features of the camera image to a 3D model in each frame, robust local alignment is achieved. Moreover, by generating reproduced appearance by simple shading as well as global alignment, global realignment is performed and the accumulation of errors is eliminated during local camera position and pose tracking.

Several issues need to be dealt with in the future. It is necessary to further evaluate the proposed method in various light environments. Also, the processing of each step must be speeded up more. In addition, in the appearance reproduction by simple shading, in order to estimate the light distribution, the sky zenithal image was acquired directly. However, because some areas of sky image, such as the sun, are saturated, estimated light distribution is not exact.

By solving this problem, the quality of the reproduced appearance will increase, and then the accuracy of the global realignment will increase also.

6. ACKNOWLEDGMENTS

This research received support from Digital Museum Project of Ministry of Education, Culture, Sports, Science and Technology, and Green ICT Innovation Promotion of Ministry of Public Management, Home Affairs, Posts and Telecommunications.

7. REFERENCES

[1] A. E. Abdel-Hakim and A. A. Farag. Csift: A sift descriptor with color invariant characteristics. *Computer Vision and Pattern Recognition*, 2:1978–1983, 2006.

[2] R. Azuma. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.

[3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications*, 21(6):34–47, 2001.

[4] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *European Conference on Computer Vision*, pages 404–417, 2006.

[5] G. Bleser and D. Stricker. Advanced tracking through efficient image processing and visual-inertial sensor fusion. *Virtual Reality Conference*, pages 137–144, 2008.

[6] P. Debevec. Image-based lighting. *Comput. Graph. Appl.*, 22(2):26–34, 2002.

[7] S. DiVerdi, J. Wither, and T. Hollerei. Envisor: Online environment map construction for mixed reality. *Virtual Reality Conference*, pages 19–26, 2008.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[9] K. Ikeuchi, A. Nakazawa, K. Hasegawa, and T. Ohishi. The great buddha project: Modeling cultural heritage for vr systems through observation. *International Symposium on Mixed and Augmented Reality*, page 7, 2003.

[10] T. Kakuta, T. Oishi, and K. Ikeuchi. Fast shading and shadowing of virtual objects using shadowing planes in mixed reality. *The Institute of Image Information and Television Engineers*, 62(9):1466–1473, 2008.

[11] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. *International Symposium on Mixed and Augmented Reality*, pages 1–10, 2007.

[12] V. Lepetit, L. Vacchetti, D. Thalmann, P. Fua, and C. Vrlab. Fully automated and stable registration for augmented reality applications. *International Symposium on Mixed and Augmented Reality*, pages 93–102, 2003.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[14] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *DARPA Image Understanding Workshop*, pages 121–130, 1981.

[15] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison. Dtam: Dense tracking and mapping in real-time. *International Conference on Computer Vision*, 1(6):2320–2327, 2011.

[16] T. Oishi, R. Sagawa, A. Nakazawa, R. Kurazume, and K. Ikeuchi. Parallel Alignment of a Large Number of Range Images. *3-D Digital Imaging and Modeling*, pages 195–202, 2003.

[17] S. Okura, R. Kawakami, and K. Ikeuchi. Simple surface reflectance estimation of diffuse outdoor object using spherical images. *Asian Conference on Computer Vision, Workshop on Multi-dimensional and Multi-view Image Processing*, 2007.

[18] G. Papagiannakis, S. Schertenleib, B. O'Kennedy, M. Arevalo-Poizat, N. Magnenat-Thalmann, A. Stoddart, and D. Thalmann. Mixing virtual and real scenes in the site of ancient pompeii. *Computer Animation and Virtual Worlds*, 16(1):11–24, 2005.

[19] G. Reitmayr and T. W. Drummond. Going out: Robust tracking for outdoor augmented reality. *International Symposium on Mixed and Augmented Reality*, pages 109–118, 2006.

[20] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. *European Conference on Computer Vision*, pages 430–443, 2006.

[21] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof. Global pose estimation using multi-sensor fusion for outdoor augmented reality. *International Symposium on Mixed and Augmented Reality*, pages 153–162, 2009.

[22] M. Schumann, S. Achilles, and S. Müller. Analysis by synthesis techniques for markerless tracking. *Virtuelle und Erweiterte Realität, 6. Workshop der GI Fachgruppe VR/AR*, 2009.

[23] J. Shi and C. Tomasi. Good features to track. *Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[24] G. Simon. Tracking-by-synthesis using point features and pyramidal blurring. *International Symposium on Mixed and Augmented Reality*, pages 85–92, 2011.

[25] T. Taketomi, T. Sato, and N. Yokoya. Real-time geometric registration using feature landmark database for augmented reality applications. *SPIE Electronic Imaging*, 7238, 2009.

[26] V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, D. Stricker, T. Gleue, P. Daehne, and L. Almeida. Archeoguide: An augmented reality guide for archaeological sites. *Computer Graphics and Applications*, 22(5):52–60, 2002.

[27] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. *Virtual Reality Conference*, pages 211–218, 2010.

[28] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence*, 14(10):965–980, 1992.

[29] J. A. Worthey and M. H. Brill. Heuristic analysis of von kries color constancy. *J. Opt. Soc. Am. A 3*, pages 1708–1712, 1986.

[30] H. Wuest, F. Wientapper, and D. Stricker. Adaptable model-based tracking using analysis-by-synthesis techniques. *Computer analysis of images and patterns*, pages 20–27, 2007.

[31] Z. Zhang et al. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.