WIDE-BASELINE STEREO MATCHING USING ASIFT AND POC

Jumpei Ishii, Shuji Sakai, Koichi Ito and Takafumi Aoki

Graduate School of Information Sciences, Tohoku University,
Sendai-shi 980-8579, Japan.
E-mail: ishii@aoki.ecei.tohoku.ac.jp

ABSTRACT

This paper proposes an accurate, dense and robust wide-baseline stereo correspondence matching method combining ASIFT (Affine-SIFT) and POC (Phase-Only Correlation). ASIFT-based matching is robust against perspective deformation of the stereo images, while the corresponding points are sparse. POC-based matching can find dense correspondence, while the corresponding points are not reliable in the case of the wide-baseline stereo. The complementary use of ASIFT and POC makes it possible to find accurate and dense stereo correspondence regardless of the length of camera baseline. Through a set of experiments, we demonstrate that the proposed method exhibits efficient performance compared with the conventional methods. We also apply the proposed method to 3D reconstruction from multi-view images.

Index Terms— 3D reconstruction, stereo correspondence, wide-baseline stereo, ASIFT, Phase-Only Correlation

1. INTRODUCTION

3D reconstruction using stereo vision is a technique for reconstructing a 3D model from multiple images taken by a camera or cameras. The reconstruction accuracy of stereo vision depends on both the accuracy of stereo correspondence between images and the baseline length between cameras. In order to reconstruct an accurate 3D model from images, it is important to obtain the accurate correspondence between images taken by the wide-baseline stereo camera.

In general, Scale-Invariant Feature Transform (SIFT) [1] has been used as a stereo correspondence matching method for the wide-baseline stereo. SIFT is one of the feature-based matching methods and describes a local feature characteristic for each detected feature point which is invariant to changes in rotation, scaling and illumination. Recently, Affine-SIFT (ASIFT) has been proposed, which improves robustness of SIFT against the affine transformation between images [2]. ASIFT simulates all image views obtained by varying orientation parameters of the camera axis and applies SIFT for each image pair to obtain the correspondence between the images. The use of ASIFT makes it possible to obtain the correspondence between images having large geometric deformation such as images taken by the wide-baseline stereo camera. However, the number of corresponding points is limited to the number of detected feature points, which is not suitable to reconstruct a complete 3D model. It may be hard to observe the fine 3D structure of the object, since the number of the corresponding points is few.

On the other hand, we have proposed an accurate stereo correspondence matching method using Phase-Only Correlation (POC) [3]. POC is one of the image matching methods and uses the phase components in 2D Discrete Fourier Transforms (DFTs) of given images. The important properties of POC used for image matching are that it is robust against illumination changes and noise, and it can estimate the sub-pixel translational displacement by fitting the analytical peak model of the POC function. POC is suitable to obtain the dense correspondence between the narrow-baseline stereo images, while POC may not obtain the accurate correspondence between the wide-baseline stereo images due to its large perspective deformation of the stereo images.

Bradley, et al. have proposed a dense wide-baseline stereo correspondence matching method which reduces perspective distortion by scaling the matching window [4]. This method obtains the dense correspondence by using NCC (Normalized Cross-Correlation) with changing disparities and scale factors of the matching window. The drawback of this method is its high computational cost. So, the visual-hull has to be used to limit the search range of disparities. This method is not robust against the large image deformation such as skew, since the image transformation of the local region is approximated only by scaling. Tola, et al. have proposed a local feature descriptor for dense correspondence called DAISY [5], which dramatically improves the computational efficiency of SIFT. DAISY describes the feature characteristics for all the pixels and finds the dense correspondence between images using DAISY descriptors. However, the correspondence accuracy of DAISY is pixel-level. So, the wide-baseline stereo must be required to reconstruct an accurate 3D model from images.

Addressing the above problems, this paper proposes an accurate, dense and robust wide-baseline stereo matching method combining ASIFT and POC, which is robust against the large image deformation. The proposed method corrects the local deformation between images using ASIFT-based matching and obtains the accurate and dense correspondence between images using POC-based matching. Through a set of experiments, we demonstrate that the proposed method exhibits efficient performance compared with the conventional methods. Also, we show 3D reconstruction from multi-view images using the proposed method to demonstrate its efficient performance.

2. STEREO MATCHING USING ASIFT AND POC

This section describes an accurate and dense stereo matching method using ASIFT and POC. In the proposed method, we assume that the local structure of the object can be approximated by a plane. According to this assumption, the local image deformation can be corrected by using the affine transformation. For each local region,

978-1-4673-2533-2/12/$26.00 ©2012 IEEE

the parameters of the affine transformation are estimated from the 3 corresponding point pairs obtained by ASIFT which is highly robust against the image transformation. Focusing on each local region after correcting the local image deformation, only the minute translations remain. So, we can use the POC-based correspondence matching to find accurate and dense correspondence between local regions. As a result, the use of the proposed method makes it possible to achieve accurate correspondence matching between wide-baseline stereo images and accurate 3D reconstruction from them. The followings are detailed procedure of the proposed method.

Step 1: Use ASIFT to obtain the sparse correspondence between the stereo images as shown in Fig. 1 (a).

Step 2: Generate triangle regions using by Delaunay triangulation of a set of corresponding points obtained in Step 1 as shown in Fig. 1 (b).

Step 3: For each triangle region, estimate parameters of the affine transformation from 3 corresponding points which are corners of a triangle, and correct the local image deformation of the triangle region using the affine transformation as shown in Fig. 1 (c).

Step 4: For each triangle region, find dense corresponding points using POC-based correspondence matching as shown in Fig. 1 (d).

According to the above procedure, the dense corresponding points between the stereo images can be obtained as shown in Fig. 1 (e), although the view angle between the stereo images used in Fig. 1 is relatively large. Furthermore, we introduce the following 3 techniques to improve the performance of the proposed method.

(A) Sampling

For the regions with rich texture, a lot of feature points can be detected by using ASIFT. If all the detected corresponding points are used in Delaunay triangulation, it may take much time to generate triangle regions. Addressing this problem, the detected corresponding points are sampled using Poisson disk sampling [6]. Poisson disk sampling bounds the number of neighbors within a fixed radius. The use of Poisson disk sampling makes it possible to reduce the concentration of the detected corresponding points in the local region with rich texture, and generate triangle regions with uniform size. Note that the corresponding points are sampled in ascending order of the distance from the epipolar line to preferentially select the high-accuracy corresponding points.

(B) Outlier removal

In Step 3, if the outliers are included in the corresponding points, we cannot estimate the accurate parameters of the affine transformation. Thus, it is difficult to find the accurate correspondence between the triangle regions which deformation is corrected using the wrong affine transformation. Addressing this problem, we employ the outlier removal methods using (i) the epipolar constraint and (ii) the depth information. In the case of (i), we remove the corresponding point as an outlier, if the distance from the epipolar line to the corresponding point is over the threshold. In the case of (ii), we remove outliers based on the depth of neighboring points having common side of the triangle. We remove the corresponding point as an outlier, if the difference between the depth calculated from the corresponding point and the median depth calculated from neighboring points is over the threshold.

(C) Iterative processing

The corresponding points obtained by ASIFT may concentrate in local regions with rich texture, since ASIFT can find the corresponding points only from the detected feature points. In the region with poor texture, the number of corresponding points is few,

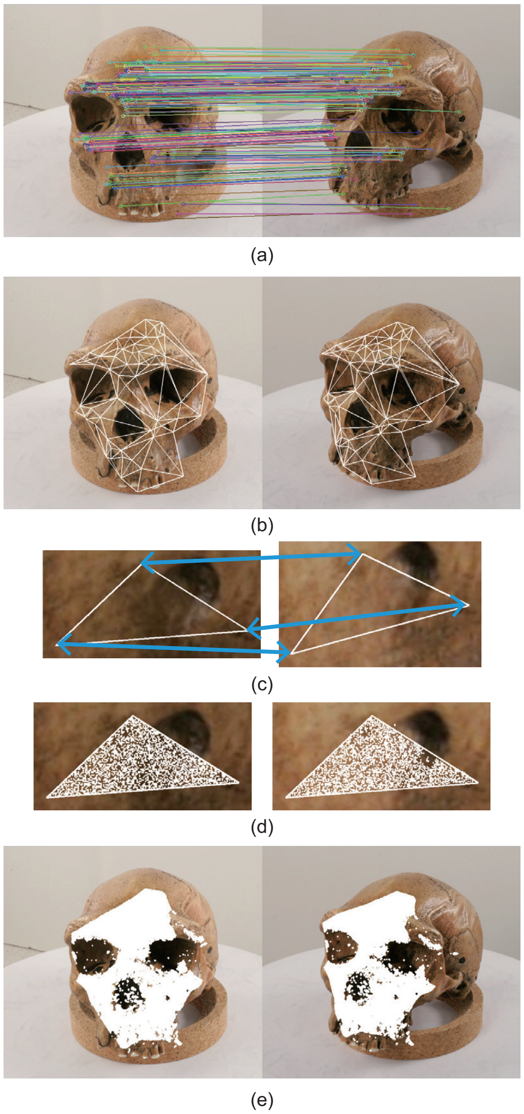


Fig. 1. Example of the proposed method: (a) corresponding points obtained by ASIFT, (b) Delaunay triangulation, (c) deformation correction using the affine transformation, (d) the correspondence matching using POC and (e) result of the correspondence matching using the proposed method.

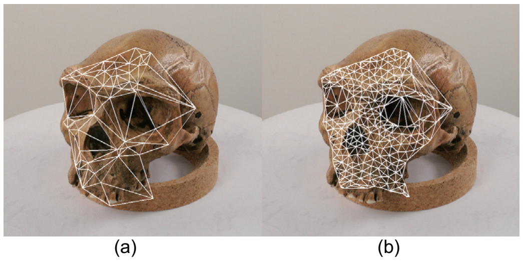
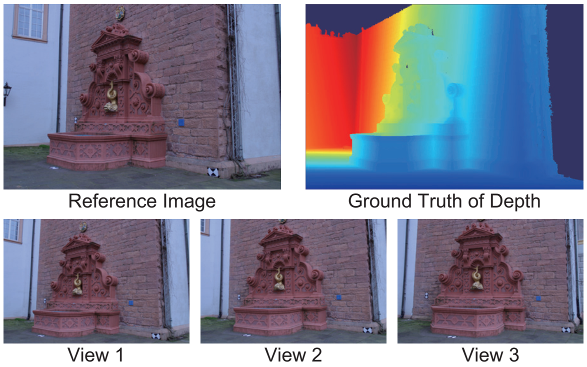


Fig. 2. Delaunay triangulation: (a) before iterative processing and (b) after iterative processing.


Reference Image

Ground Truth of Depth

View 1

View 2

View 3

Fig. 3. Fountain data set: reference image, ground truth of the depth and images taken from neighboring views.

and hence the size of the generated triangle region becomes large as shown in Fig. 2 (a). So, the image deformation of such a region cannot be approximated by the affine transformation. Addressing this problem, we iteratively perform Step 1∼5 as mentioned in the above to generate small triangle regions with uniform size. Once Step 1∼5 are performed, we can obtain more dense corresponding points than ASIFT. Using the corresponding points after performing Step 1∼5, we can generate small triangle regions with uniform size as shown in Fig. 2 (b). Note that the corresponding point whose peak value of the POC function is less than the threshold is removed as an outlier.

3. EXPERIMENTS AND DISCUSSION

In this section, we evaluate the performance of the proposed method. Fig. 3 shows a data set "Fountain" [7] used in this experiment. The image size is 768 × 512 pixels, which is 25% scaling from the original one. The reference image is the image taken from the right-most view, and is matched to the images taken from the neighboring 3 views. In this experiment, we compare the 3 correspondence matching methods: DAISY [5], POC [3] and the proposed method. The DAISY-based correspondence matching finds the corresponding points having the minimum Euclidean distance between feature vectors when searching on the epipolar line. In the proposed method, the number of iterations is 2, and the threshold values of the outlier removal using the epipolar constraint are 4 pixels and 1 pixel in the first and second times, respectively. Also, the threshold for outlier removal using the peak value of the POC function is 0.4 in this paper. We estimate the depth maps using the 3 methods and compare them with the ground truth. In this experiment, the performance of the 3 methods is evaluated by the inlier rate and the average error rate of inliers, where the inlier is defined as a corresponding point whose error rate is smaller than the threshold. The error rate is defined by

$$e = \frac{|d_e(\boldsymbol{u}) - d_g(\boldsymbol{u})|}{d_g(\boldsymbol{u})} \times 100 \ [\%],$$

where $d_e(\boldsymbol{u})$ is the estimated depth on the pixel $\boldsymbol{u}$ and $d_g(\boldsymbol{u})$ is the true depth on the pixel $\boldsymbol{u}$.

Table 1 shows the inlier rate and the average of error rates, where the threshold for inliers is 1 pixel, and Fig. 4 shows depth maps and error maps. Fig. 5 shows the inlier rates when changing the threshold for inliers. As a result, the accuracy of DAISY is low in the narrow-baseline stereo and the accuracy of POC is low in the wide-baseline stereo. Compared with DAISY and POC, the accuracy of the proposed method is high both in the narrow- and wide-baseline stereo. Also, in the case of the same baseline, the accuracy of the proposed method is higher than that of DAISY. The reason is that DAISY is the stereo correspondence matching method with pixel accuracy, while the proposed method uses the POC-based correspondence matching with sub-pixel accuracy. In the case of the wide-baseline stereo and the high-threshold of error rate, the inliner rate of the proposed method is lower than that of DAISY. As shown in Fig. 4, we observe that the estimated depth map using the proposed method is relatively small. The proposed method can estimate the depth only from the region included in the corresponding points obtained by ASIFT. So, if the corresponding points obtained by ASIFT are located on whole the image, the inlier rate of the proposed method would be significantly improved.

Table 1. Inlier rates (upper) and average error rates of inliers (lower).

View	1	2	3
DAISY	54.750%	35.164%	22.408%
	0.3428%	0.1961%	0.1240%
POC	56.450%	32.023%	1.927%
	0.2702%	0.1679%	0.1535%
Proposed	55.044%	38.635%	19.962%
	0.2271%	0.1360%	0.0937%

4. 3D RECONSTRUCTION FROM MULTI-VIEW IMAGES

We apply the proposed method to 3D reconstruction from multi-view images. The proposed method can reconstruct the accurate 3D model from multi-view images regardless of the baseline length. This advantage of the proposed method is suitable for 3D reconstruction from multi-view images. We use "nskulla" data set [8], and select 16 images from the data set. Figs. 6 (a) and (b) show examples of the image used in this experiment. We find the corresponding points between every pair of adjacent views using the proposed method and reconstruct the 3D model from them. For each reconstructed 3D point, the normal vector is estimated from the 3D point of interest and their 30 neighbors. Finally, the mesh model is generated using Poisson Surface Reconstruction [9] based on the estimated normal vectors. Figs. 6 (c) and (d) show the reconstruction results. As is observed in this result, the whole structure of the object can be reconstructed.

5. CONCLUSION

This paper has proposed the wide-baseline stereo matching method using ASIFT and POC. The complementary use of the feature-based and area-based matching methods makes it possible to achieve accurate and dense stereo correspondence regardless of the length of camera baseline. Through a set of experiments, we have demostrated that the proposed method exhibits efficient performance compared with the conventional methods. We have also applied the proposed method to 3D reconstruction from multi-view images to demonstrate the effectiveness of the proposed method.

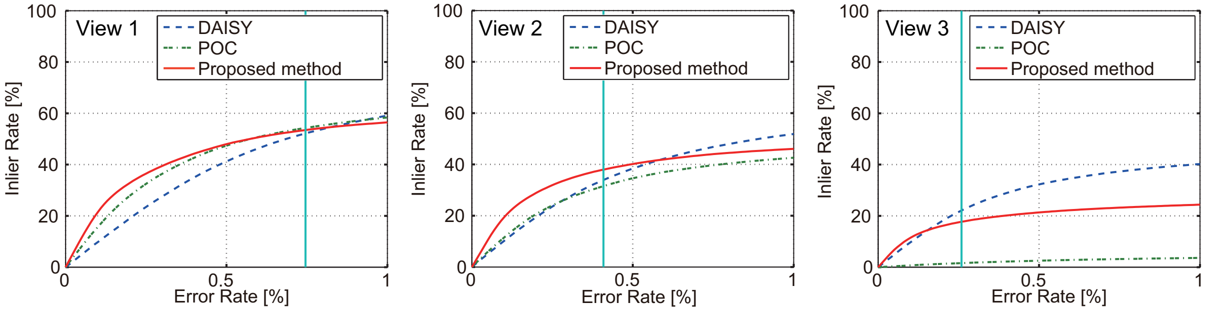


Fig. 5. Inlier rates for View 1, View 2 and View 3 when changing the threshold for inliers (the vertical line indicates the one-pixel error).

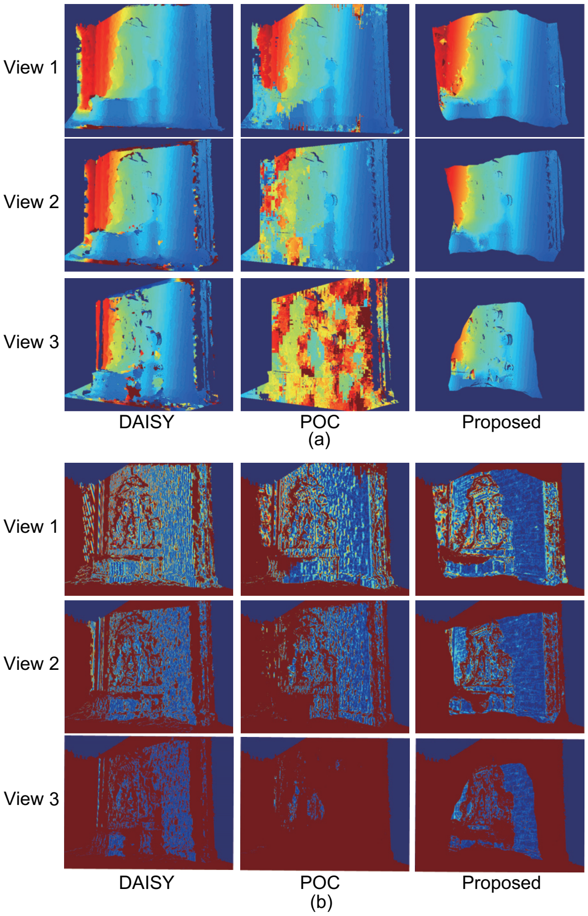


Fig. 4. (a) Depth maps and (b) error maps (The colormap ranges from blue (minimum value) to red (maximum value).).

6. REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] J.M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

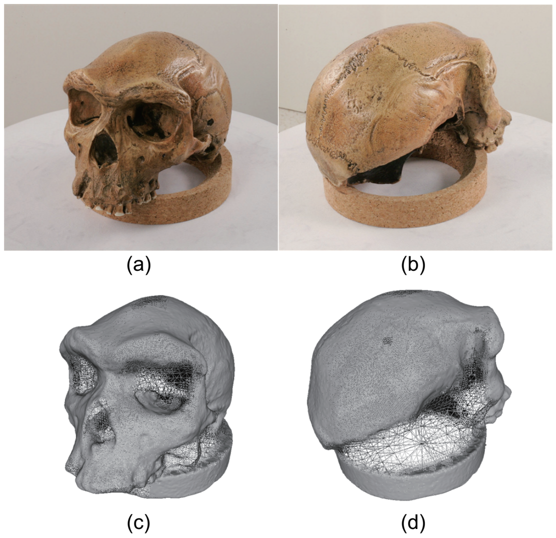


Fig. 6. 3D reconstruction results using the proposed method: (a), (b) examples of input image and (c), (d) the reconstruction results.

[3] K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi, "A sub-pixel correspondence search for computer vision applications," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1913–1923, 2004.

[4] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," *Proc. Int'l Conf. CVPR*, pp. 1–8, 2008.

[5] E.Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *Proc. IEEE Trans. PAMI*, vol. 32, no. 5, pp. 815–830, 2010.

[6] R.L. Cook, "Stochastic sampling in computer graphics," *ACM Trans. Graphics*, vol. 5, no. 1, pp. 51–72, 1986.

[7] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," *Proc. Int'l Conf. CVPR*, pp. 1–8, 2008.

[8] Y. Furukawa and J. Ponce, "3D photography dataset," `http://www.cs.washington.edu/homes/furukawa/research/mview/`.

[9] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," *Proc. Symp. Geometry Processing*, pp. 61–70, 2006.