An Easy-to-Use and Accurate 3D Shape Measurement System Using Two Snapshots

Mamoru Miura, Shuji Sakai, Jumpei Ishii, Koichi Ito and Takafumi Aoki
Graduate School of Information Sciences, Tohoku University
6–6–05, Aramaki Aza Aoba, Sendai-shi 980–8579, Japan
E-mail: miura@aoki.ecei.tohoku.ac.jp

***Abstract*—3D measurement using a moving camera is a technique to measure the 3D shape of an object from a set of images taken from different viewpoints. One of the well-known 3D measurement methods is Structure from Motion (SfM) using feature-based correspondence matching. However, only a limited number of 3D points are measured by this method and are not sufficient to measure the fine 3D shape of the object. Addressing this problem, this paper proposes a novel 3D measurement algorithm combining SfM using feature-based matching to estimate camera parameters and area-based correspondence matching to obtain dense correspondence. Using the proposed algorithm, this paper also proposes an easy-to-use and accurate 3D shape measurement system from two views captured with a moving consumer digital camera. Through a set of experiments, we demonstrate that the proposed system can measure the 3D shape of the object in about 20 seconds with the measurement accuracy comparable with that of the 3D laser scanner.**

***Index Terms*—3D measurement, Structure from Motion, feature-based matching, area-based matching**

I. Introduction

Recently, 3D measurement has attracted much attention in various fields of industry, medicine, etc. Existing practical 3D measurement systems can be broadly classified into active systems with laser scanning or structured light projection and passive systems with binocular or multiple cameras. The active systems require the specific measurement equipment depending on the target objects and may be liable to be expensive. Also, the application of active systems are limited, since the measurement equipment for active systems are relatively large. On the other hand, the passive systems need to select the stereo cameras depending on the target and to calibrate cameras in advance. Hence, the technical knowledge is required for users. As mentioned above, it is hard for the users without the technical knowledge to introduce the existing 3D measurement system for practical use in their daily life.

Structure from Motion (SfM) with a moving monocular camera is known as a simple approach to measure the 3D shape of an object [1], [2]. Using this approach, the user can measure 3D points of the object from multiple images captured with a monocular camera without camera calibration. So far, a variety of SfM algorithms have been proposed and the measurement accuracy of the state-of-the-art algorithms is acceptable in practical use even if the images are captured with a moving monocular camera [3]. However, only a limited number of 3D points are measured and are not sufficient to measure the fine 3D shape of the object, since SfM employs a feature-based correspondence matching such as Scale Invariant Feature Transform (SIFT) [4]. Although Multi-View Stereo (MVS) algorithms with SfM can be used to measure dense 3D points of the whole object [1], [5], most of existing MVS algorithms have the drawback of high computational cost [6].

Addressing the above problems, this paper proposes an easy-to-use and accurate 3D measurement system with a consumer digital camera for the users without the technical knowledge. The proposed system employs a 3D measurement algorithm combining SfM using feature-based matching to estimate camera parameters and area-based corresponding matching to obtain dense correspondence. The use of the proposed system makes it possible to measure accurate and dense 3D shape of the object from two views captured with a moving consumer digital camera. Through a set of experiments, we demonstrate that the proposed system can measure the 3D shape of the object with an accuracy of less than 1 mm compared with the measurement result by the laser 3D scanner. We also demonstrate that all the processes of the proposed system is finished in about 20 seconds with a mid-range laptop computer.

II. 3D Measurement System Using a Moving Camera

The proposed system consists of a consumer digital camera and a computer. The users obtain the 3D shape of an object only by capturing two images of the object using a digital camera, where most of existing consumer digital cameras can be used in the proposed system. The input images to the proposed system have to be in focus and be without any halation. The images captured with existing consumer digital cameras satisfy the above conditions, since the aperture, shutter speed and focal length of these cameras are automatically configured by their automatic focus and exposure functions. Hence, the users can capture suitable images for 3D measurement without any technical knowledge. In addition, the captured images are automatically transferred by using the digital camera with Wi-Fi access or the memory cards with wireless access such as Eye-Fi [7] so as to improve the convenience of the system.

The procedure of the proposed system consists of 3 steps: (i) camera parameter estimation, (ii) stereo rectification and (iii) 3D shape measurement as shown in Fig. 1. We describe the details for each step as follows.

(i) Camera parameter estimation

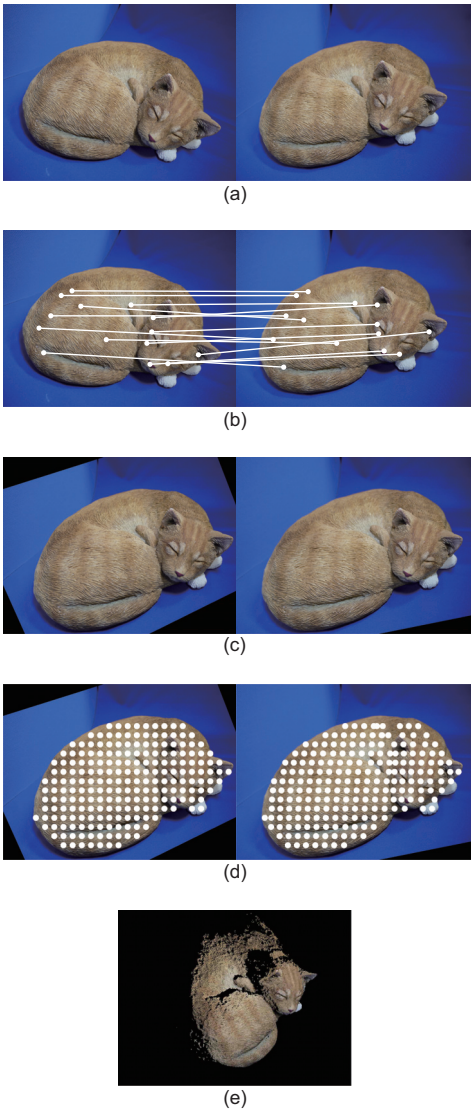

Fig. 1. Processing flow of the proposed system: (a) input stereo image pair, (b) result of feature-based correspondence matching, (c) rectified stereo image pair, (d) result of area-based correspondence matching and (e) 3D measurement result.

In this step, the camera parameters such as the intrinsic and extrinsic parameters are estimated.

First, the user captures two images of the target object with a digital camera as shown in Fig. 1 (a). Let a stereo image pair be I_1 and I_2.

Next, the intrinsic parameters of each camera are calculated. The intrinsic parameter matrix $\boldsymbol{A}$ of a camera is defined by

$$\boldsymbol{A} = \begin{pmatrix} f\frac{w}{D} & 0 & \frac{w}{2} \\ 0 & f\frac{w}{D} & \frac{h}{2} \\ 0 & 0 & 1 \end{pmatrix}, \quad (1)$$

where f is the focal length of the camera, w and h are the width and height of the image, respectively, and D is the width of the image sensor. The focal length f and the image resolution (w, h) are obtained from Exif (Exchangeable image file format) information of the image [8]. The image sensor width D depends on the camera and is given in the specification. According to Eq. (1), the intrinsic parameter matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are calculated from I_1 and I_2, respectively.

Then, the corresponding point pairs between the images I_1 and I_2 are obtained by a feature-based corresponding matching method as shown in Fig. 1 (b). We employ a feature-based corresponding matching method, since the stereo images include various geometric transformation such as scaling, rotation and nonlinear transformation due to a camera movement and a change of focal length. Using the corresponding point pairs and the intrinsic parameters $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, the extrinsic parameters $\boldsymbol{R}_{1\to2}$ and $\boldsymbol{t}_{1\to2}$ are calculated by 5-point algorithm [9] with RANSAC (RANdom SAmple Consensus) [10].

Finally, the intrinsic and extrinsic parameters are optimized by using bundle adjustment [11], since the accuracy of these parameters has an impact on the succeeding steps.

(ii) Stereo rectification

In this step, the stereo image pair I_1 and I_2 is transformed into the rectified stereo image pair I_1' and I_2' by stereo rectification as shown in Fig. 1 (c) in order to employ an area-based correspondence matching method. To obtain accurate and dense 3D points, we employ an area-based correspondence matching method in the step (iii). However, it is hard for an area-based method to obtain the correspondence between the stereo image pair having large perspective distortion. Hence, we reduce the perspective distortion between the stereo image pair by stereo rectification. Stereo rectification is to transform an image pair as if the image pair is captured with a parallel stereo camera [1], that is, the scaling in vertical direction and rotation between the image pair are reduced and the geometric deformation between the image pair is also limited to horizontal direction. Note that the correspondence search between a stereo image pair is reduced to 1D search by stereo rectification. The rectified stereo image pair I_1' and I_2' is obtained by transforming I_1 and I_2 with the homography matrix calculated from the camera parameters obtained in the step (i). Thus, the use of the stereo rectification makes it possible to reduce the perspective distortion between an stereo image pair in order to measure accurate and dense 3D points of an object.

(iii) 3D shape measurement

In this step, the dense correspondence point pairs are obtained by using an area-based correspondence matching method as shown in Fig. 1 (d). Unlike the feature-based correspondence matching method, the area-based correspondence matching method can obtain the point on the input image corresponding to the reference point placed on the arbitrary position in the reference image. Hence, when many reference points are placed on the reference image, the dense corresponding points can be obtained so as to measure the fine 3D structure of the object. The corresponding points with subpixel accuracy can be also obtained by using the model fitting technique [12], [13]. Finally, a set of 3D points are calculated from the camera parameters obtained in the step (i) and the corresponding point pairs as shown in Fig. 1 (e).

Fig. 2. Performance evaluation using a fixed stereo camera: (a) stereo images (1,280×960 pixels) and (b) ground truth 3D model measured with KONICA MINOLTA VIVID.

III. Experiments and Discussion

This section describes experiments for evaluating performance of the proposed system. We perform two experiments: (i) performance evaluation using a fixed stereo camera and (ii) performance evaluation using consumer digital cameras.

A. Performance Evaluation Using Fixed Stereo Camera

We evaluate the accuracy of the camera parameter estimation methods such as SfM using feature-based matching and the camera calibration using checkerboard patterns [14]. In general, it is difficult to directly compare the accuracy of camera parameter estimation methods, since the true values of camera parameters are unknown. Therefore, in this paper, we evaluate the accuracy of 3D measurement using a stereo image pair taken with a fixed stereo camera to compare the accuracy of camera parameter estimation methods. We also compare the measurement accuracy of 3D shape in terms of area-based correspondence matching methods.

The stereo camera is composed of two monocular cameras (Point Grey FL2G-12S2M-C), and Fig. 2 (a) shows captured images with the stereo camera. The measurement accuracy is evaluated by comparing the measurement results and the ground truth 3D mesh model measured by the laser scanner (KONICA MINOLTA VIVID) as shown in Fig. 2 (b). Using the Iterative Closest Point (ICP) algorithm [15], we align the measured 3D points and the ground truth 3D mesh model of the object. The outlier rate and RMS (Root Mean Square) error are calculated between the aligned data. In this paper, the outlier is defined by a point whose fitting error is greater than 1 pixel in the stereo images. Also the RMS errors are calculated for 3D points without outliers.

According to application, the proposed system can select (i) the feature-based correspondence matching method to estimate camera parameters and (ii) the area-based correspondence matching method to obtain the dense corresponding point pairs. As for the feature-based correspondence matching method, we compare the following 4 methods: SIFT [4], Speeded-Up Robust Features (SURF) [16], Binary Robust Invariant Scalable Keypoints (BRISK) [17] and Affine-SIFT (ASIFT) [18]. For comparison, we use the camera parameters estimated by Zhang's camera calibration method using images of a planar checkerboard [14]. As for the area-based correspondence matching method, we compare the following 4 methods: Sum of Absolute Differences (SAD) [12], Sum of Squared Differences (SSD) [12], Normalized Cross-Correlation (NCC) [12] and Phase-Only Correlation (POC) [13]. All the methods employ a coarse-to-fine strategy using image pyramids with local block matching [13]. The number of layers for coarse-to-fine search is 4. For SAD, SSD and NCC, the size of the matching window is 16 pixels × 15 lines. For POC, the size of the matching window is 32 pixels × 15 lines. The size of the matching window for POC is equivalent to that for the other methods, since the Hanning window is applied to the matching window to reduce the effect of discontinuity at signal border in DFT [13].

TABLE I
Outlier rates [%]

	SAD	SSD	NCC	POC
CALIB	42.6	26.3	1.7	1.0
BRISK	53.8	44.2	5.0	2.6
SURF	50.3	35.0	1.5	0.9
SIFT	52.5	38.3	1.5	0.9
ASIFT	57.1	44.2	3.7	2.7

TABLE II
RMS errors [mm]

	SAD	SSD	NCC	POC
CALIB	1.39	1.30	0.97	0.90
BRISK	1.19	1.17	1.03	0.95
SURF	1.23	1.23	0.69	0.67
SIFT	1.21	1.18	0.67	0.65
ASIFT	1.19	1.16	1.01	0.97

Table I shows outlier rates in the 3D measurement, where CALIB denotes the Zhang's camera calibration method. Table II shows RMS errors in the 3D measurement. As for the area-based methods, POC exhibits lower outlier rate and RMS error than others regardless of the feature-based methods. As for the feature-based methods, SIFT with POC exhibits the lowest outlier rate and RMS error. As a result, the proposed system combining SIFT and POC can measure the 3D shape of the object with the accuracy that is comparable with the laser scanner.

Tables III and IV show the computation time of camera parameter estimation and dense correspondence matching, respectively. The computation time is measured on Intel Core 2 Duo E6850 (3.00 GHz). The proposed system combining SIFT and POC, which exhibits the lowest RMS error in Table II, can perform 3D shape measurement in about 20 seconds.

B. Performance Evaluation Using Digital Cameras

We use 2 consumer digital cameras: a digital single-lens reflex camera (Panasonic LUMIX DMC-GF1) and a mobile-phone camera (Apple iPhone 4S) to evaluate performance of the proposed system. In this experiment, we employ the proposed system combining SIFT and POC, which exhibits the lowest RMS error. We use a cat carving, a dog curving and an interior tile as the target object. Fig. 3 shows 3D measurement results. As a result, the proposed system can measure the accurate and dense 3D points of various objects using the consumer digital camera. Even if we use the mobile phone

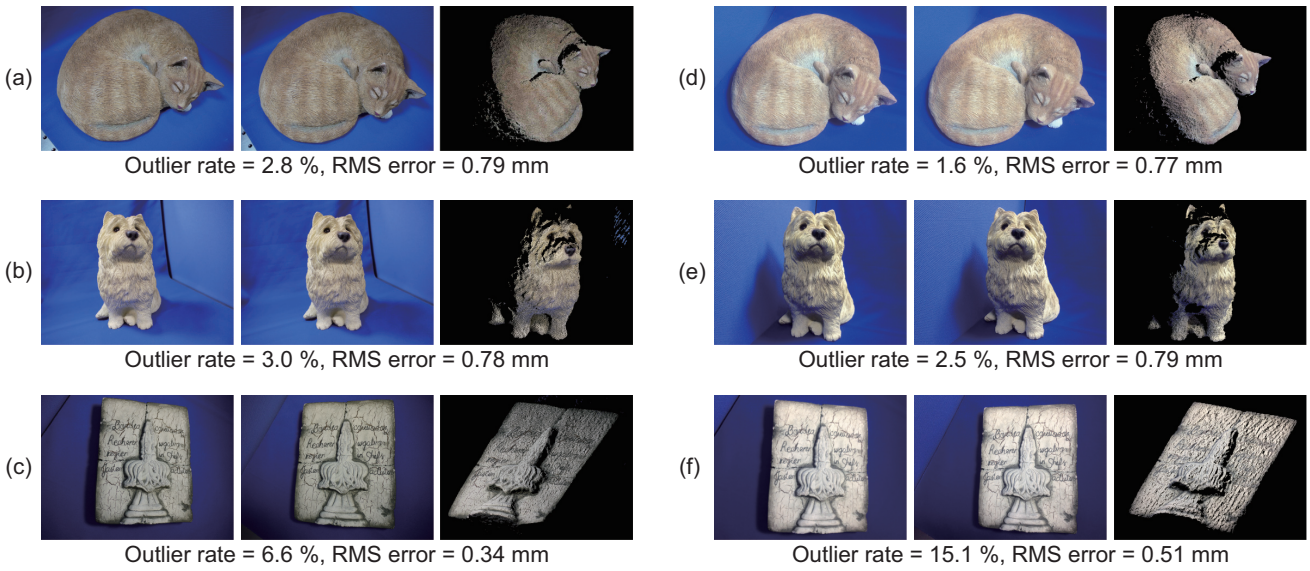

Fig. 3. Results of 3D measurement: (a), (b) and (c) with Panasonic LUMIX DMC-GF1, and (d), (e) and (f) with Apple iPhone 4S.

TABLE III
COMPUTATION TIME OF CAMERA PARAMETER ESTIMATION [S]

BRISK	SURF	SIFT	ASIFT
1.02	1.90	12.92	86.20

TABLE IV
COMPUTATION TIME OF DENSE CORRESPONDENCE MATCHING [S]

SAD	SSD	NCC	POC
1.17	1.18	3.45	5.33

camera to capture the images, the 3D measurement accuracy is below 1 mm compared with the measurement result by the laser scanner.

IV. CONCLUSION

This paper has proposed an easy-to-use and accurate 3D measurement system using a consumer digital camera. The use of the proposed system makes it possible to measure the accurate 3D shape of the object only by capturing two images without any technical knowledge. Through a set of experiments, we have demonstrated that the proposed system can measure the 3D shape of the object in about 20 seconds with the measurement accuracy comparable with that of the 3D laser scanner. The proposed system can be extended to deal with multi-view images by combining the 3D measurement results obtained from every stereo image pair on the uniform coordinate system.

REFERENCES

[1] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.
[2] R. Hartley, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2008.
[3] S. Agarwal, N. Snavely, I. Simon, S. Seitz, and R. Szeliski, "Building Rome in a day," *Proc. Int'l Conf. Computer Vision*, pp. 72–79, 2009.
[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[5] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Towards internet-scale multi-view stereo," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 1434–1441, 2010.
[6] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-views stereo reconstruction algorithms," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 519–528, 2006.
[7] Eye-Fi Inc., "Eye-Fi memory cards," http://www.eye.fi.
[8] Japan Electronics and Information Technology Industries Association, "Exchangeable image file format for digital still cameras," http://www.jeita.or.jp/.
[9] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.
[10] 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*, vol. 24, no. 6, pp. 381–395, 1981.
[11] M. I. A. Lourakis and A. A. Argyros, "SBA: A software package for generic sparse bundle adjustment," *ACM Trans. Mathematical Software*, vol. 36, no. 1, pp. 1–30, 2009.
[12] M. Shimizu and M. Okutomi, "Precise sub-pixel estimation on area-based matching," *Proc. Int'l Conf. Computer Vision*, vol. 1, pp. 90–97, 2001.
[13] T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "A sub-pixel stereo correspondence technique based on 1d phase-only correlation," *Proc. Int'l Conf. Image Processing*, vol. 5, pp. V–221–V–224, 2007.
[14] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *Proc. Int'l Conf. Computer Vision*, vol. 1, pp. 666–673, 1999.
[15] Z. Timo, J. Schmidt, and H. Niemann, "Point set registration with integrated scale estimation," *Proc. Int'l Conf. Pattern Recognition and Image Processing*, pp. 116–119, 2005.
[16] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
[17] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *Proc. Int'l. Conf. Computer Vision*, pp. 2548–2555, 2011.
[18] 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.