# Accurate and Dense Wide-Baseline Stereo Matching Using SW-POC

Shuji Sakai, Koichi Ito, Takafumi Aoki
Graduate School of Information Sciences,
Tohoku University, Sendai, 980–8579, Japan
Email: sakai@aoki.ecei.tohoku.ac.jp

Hiroki Unten
Toppan Printing Co., Ltd.
Bunkyo-ku, Tokyo, 112–8531, Japan

***Abstract*—This paper proposes an accurate and dense wide-baseline stereo matching method using Scaled Window Phase-Only Correlation (SW-POC). The wide-baseline setting of the stereo camera can improve the accuracy of the 3D reconstruction compared with the short-baseline setting. However, it is difficult to find accurate and dense correspondence from wide-baseline stereo images due to its large perspective distortion. Addressing this problem, we employ the SW-POC, which is a correspondence matching method using 1D POC with the concept of Scale Window Matching (SWM). The use of SW-POC makes it possible to find the accurate and dense correspondence from a wide-baseline stereo image pair with low computational cost. We also apply the proposed method to 3D reconstruction using a moving and uncalibrated consumer digital camera.**

***Index Terms*—3D reconstruction, stereo correspondence, wide-baseline stereo, Scaled Window Phase-Only Correlation.**

## I. Introduction

3D reconstruction using stereo vision is a technique to reconstruct the surface shape of the objects from multiple view images taken by a camera [1]. The accuracy of 3D reconstruction depends on (i) the accuracy of correspondence matching between images and (ii) the length of the stereo camera baseline. For the purpose of high-accuracy 3D reconstruction using stereo vision, it is important to find accurate correspondence between the stereo image pair taken by the camera with wide-baseline setting.

The feature-based correspondence matching such as Scale-Invariant Feature Transform (SIFT) [2] has been employed to obtain the correspondence from wide-baseline stereo image pairs, since the feature-based approaches are robust against the geometric distortion due to wide-baseline setting. In this case, only a limited number of corresponding points is obtained. The sparse corresponding points can be used to estimate the camera parameters, while these are not sufficient to reconstruct the fine 3D structure of the objects. The area-based matching has been employed to obtain the dense correspondence between the narrow-baseline stereo image pairs [3]. In the case of wide-baseline setting, the area-based approaches cannot obtain the accurate correspondence due to its large perspective distortion of the stereo image pairs.

Bradley et al., have proposed a dense wide-baseline stereo 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. In practical situation where the visual-hull cannot be used, it is difficult to apply the Bradley's method. Furthermore, Bradley's method requires more computational cost to obtain correspondence with sub-pixel accuracy.

Addressing the above problems, we propose a novel accurate and dense stereo correspondence matching method using Scaled Window Phase-Only Correlation (SW-POC). The proposed method employs the 1D POC-based correspondence matching [5] with the concept of Scaled Window Matching (SWM) [4] to find the accurate correspondence from a wide-baseline stereo image pair with low computational cost. A set of experiments demonstrates that the proposed method exhibits accurate and robust correspondence matching in both short- and wide-baseline stereo pairs. We also show the 3D reconstruction results of objects using the proposed method, where the stereo image pairs are taken by a moving and uncalibrated consumer digital camera.

## II. Correspondence Matching Using SW-POC

In this section, we briefly introduce 1D Phase-Only Correlation (POC) [5] and describe the correspondence matching technique using SW-POC.

### A. 1D Phase-Only Correlation: POC

POC is an image matching technique using the phase information obtained from DFT (Discrete Fourier Transform) of images. In the case of a rectified stereo image pair, the disparity can be limited to horizontal direction. The use of 1D POC makes it possible to achieve high-accuracy correspondence matching with low computational cost.

Let $f(n)$ and $g(n)$ be the 1D image signals, where $-M \leq n \leq M$ and the signal length is $N = 2M + 1$. Then, the normalized cross-power spectrum $R(k)$ is defined as

$$R(k) = \frac{F(k)\overline{G(k)}}{|F(k)\overline{G(k)}|} = e^{j(\theta_F(k)-\theta_G(k))}, \qquad (1)$$

where $F(k)$ and $G(k)$ are the 1D DFTs of $f(n)$ and $g(n)$, $\overline{G(k)}$ denotes the complex conjugate of $G(k)$, and $-M \leq$

$k \leq M$. The 1D POC function $r(n)$ between $f(n)$ and $g(n)$ is the 1D Inverse DFT (1D IDFT) of $R(k)$ and is given by

$$r(n) = \frac{1}{N} \sum_{k=-M}^{M} R(k) W_N^{-kn}, \tag{2}$$

where $W_N = e^{-j\frac{2\pi}{N}}$. Assume that $f(n)$ and $g(n)$ are minutely displaced with each other by $\delta$, we can derive the analytical peak model of the 1D POC function between $f(n)$ and $g(n)$ as follows

$$r(n) \simeq \quad \frac{\alpha}{N} \frac{\sin(\pi(n+\delta))}{\sin(\frac{\pi}{N}(n+\delta))}. \tag{3}$$

The above equation represents the shape of the peak for the 1D POC function between the 1D image signals that are minutely displaced with each other. This equation gives a distinct sharp peak. When $\delta = 0$, Eq. (3) becomes the Kronecker delta function. We can show that the peak value $\alpha$ decreases (without changing the function shape itself), when small noise components are added to the images. Hence we assume $\alpha \leq 1$ in practice. The peak position $n = -\delta$ of the 1D POC function reflects the displacement between the two 1D image signals. Thus, we can compute the displacement $\delta$ between signals $f(n)$ and $g(n)$ by estimating the true peak position of the 1D POC function $r(n)$. We have also proposed the important techniques for improving the accuracy of 1D image matching for sub-pixel correspondence matching: (i) function fitting for high-accuracy estimation of peak position, (ii) windowing to reduce boundary effects, (iii) spectral weighting for reducing aliasing and noise effects, (iv) averaging 1D POC functions to improve peak-to-noise ratio and (v) coarse-to-fine strategy for robust correspondence search [5].

### *B. Scaled Window-POC*

SW-POC is an image matching technique to handle the perspective distortion of wide-baseline stereo image pairs by scaling the size of matching window depending on the shape of the object. After rectifying a stereo image pair, the reference and corresponding points have the same vertical coordinates. Therefore, we only have to consider perspective distortion in the horizontal direction. Assuming that the local distortion can be approximated by horizontal scaling, the perspective distortion can be reduced by scaling the size of the matching window as shown in Fig. 1. And then, the accurate displacement between scaled matching windows can be estimated by using 1D POC. Combining the image matching technique using SW-POC and the coarse-to-fine strategy using image pyramids, we can find accurate correspondence form a wide-baseline stereo image pair as well as a short-baseline stereo image pair.

### *C. Scale Factor Estimation for SW-POC*

This subsection describes how to determine the scale factor for SW-POC. The scale factor between the matching windows depends on the surface structure, i.e., the surface normal $\mathbf{n}$, and the distance from the cameras to the object surface. Focusing on the 3D point $\mathbf{M} = (X, Y, Z)$, the scale factor $s$ is defined by

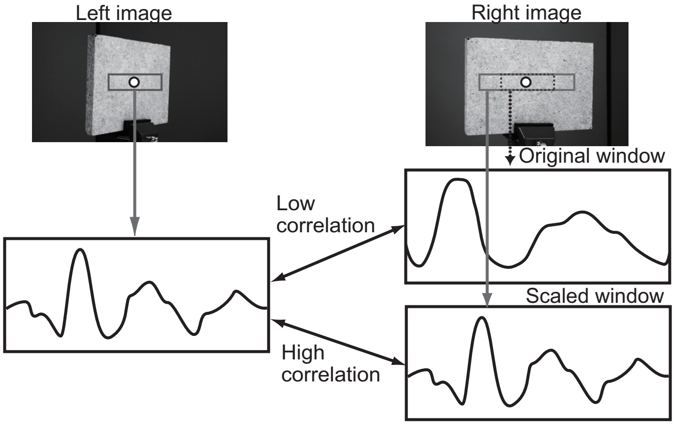


Fig. 1. Overview of SW-POC.

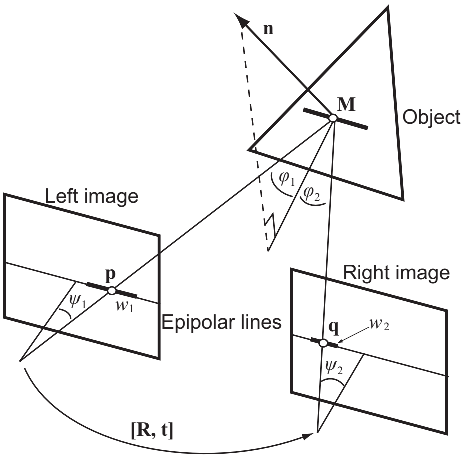


Fig. 2. Geometric relationship among left and right images and object.

$$s = \frac{\cos\psi_1}{\cos\psi_2} \frac{\cos\phi_2}{\cos\phi_1}, \tag{4}$$

where $\psi_i$ is the incident angle of the viewing ray on the respective image planes, and $\phi_i$ is the angle between the viewing ray and the projection of the surface normal $\mathbf{n}$ into the epipolar plane [4]. $\psi_i$ is obtained from the intrinsic and extrinsic parameters of the camera $i$ and the 3D point $\mathbf{M}$, while $\phi_i$ is obtained from the extrinsic parameters of the camera $i$, the 3D point $\mathbf{M}$ and the surface normal $\mathbf{n}$ on the point. The geometric relationship is illustrated in Fig. 2.

In general, the surface normal $\mathbf{n}$ is unknown information for 3D reconstruction using stereo vision, so we cannot estimate the scale factor $s$ using Eq. (4). Addressing this problem, we estimate the scale factor $s$ using a peak value $\alpha$ of the 1D POC function. The window matching of SW-POC is performed with changing the scale factor $s$. We select the scale factor $s$ having the largest peak value of 1D POC function. It results in the increase of the computational cost, since this approach needs the iterative window matching. In order to reduce the

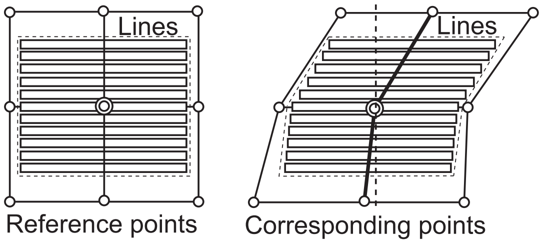


Fig. 3. Matching window to reduce skew caused by the 3D shape of objects.

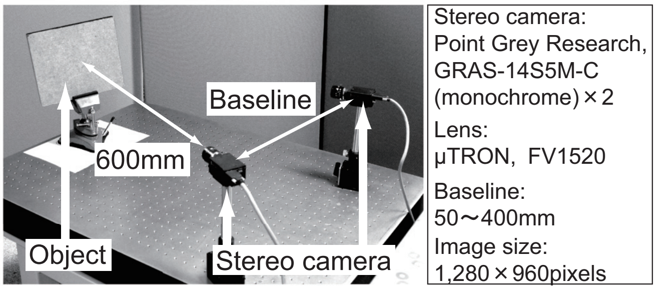
computational cost, the scale factor $s$ is estimated only in the coarsest layer.

### D. Reducing Computational Cost Using Coarse-to-Fine Strategy

The method discussed the above needs the iterative window matching with changing the scale factor and the initial disparity in the coarsest layer to estimate the accurate scale factor $s$ for SW-POC. The computational cost of the above method is still higher than that of the conventional method using 1D POC. In order to achieve further reduction of the computational cost, we adopt the coarse-to-fine strategy for 3D reconstruction of the objects. First, we obtain the sparse correspondence between images using SW-POC with various scale factors and initial disparities, and reconstruct a coarse 3D shape of the objects. Next, we calculate the scale factors and the initial disparities for dense reference points from the coarse shape. Then, we obtain the dense correspondence between images using SW-POC with the scale factors and the initial disparities calculated as above, and reconstruct a fine 3D shape.

### E. Averaging POC Functions to Consider 3D Shape

In practical situation, one pair of 1D image signals is not sufficient to find the accurate correspondence due to poor image quality such as noise, blur, etc. resulted in degraded Peak-to-Noise Ratio (PNR) of 1D POC function. We can improve PNR by averaging a set of 1D POC functions evaluated at distinct positions around the reference and corresponding points [5]. As described in the above, SW-POC assumes that the perspective distortion between 1D image signals can be approximated by displacement and horizontal scaling. As for a single line which is parallel to the horizontal axis, this assumption is proper. However, when skew occurs between the matching windows, the local distortion cannot be approximated by displacement and horizontal scaling. Although we empirically confirm that the method described in Sect. II-D is robust against a certain level of skew, the corresponding error is increased by large skew. Therefore, we obtain the correspondence using the method described in Sec. II-D with rectangular matching windows, and then update the position of corresponding points using SW-POC with matching windows where each line is translated according to the 3D shape of the objects to reduce skew as shown in Fig. 3. When adjacent corresponding points include outliers, the accuracy of the correspondence matching may be decreased due to the above updating process. Thus, we update only the corresponding points whose peak value of 1D POC function after updating is greater than that before updating.



Stereo camera:
Point Grey Research,
GRAS-14S5M-C
(monochrome) × 2
Lens:
μTRON, FV1520
Baseline:
50〜400mm
Image size:
1,280 × 960pixels

Fig. 4. Stereo vision system used in experiments.

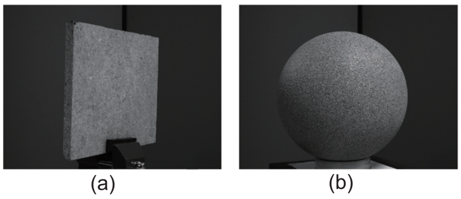


Fig. 5. Examples of left camera images: (a) plane, (b) sphere.

## III. Experiment and Discussion

In this section, we evaluate the accuracy and the computational cost of the proposed method compared with the 1D POC-based correspondence matching method [5] and Bradley's method [4].

The parameters for each method are empirically optimized. For 1D POC and SW-POC, the length of 1D image signal is $N = 32$ pixels and the number of 1D image signals to be averaged is $L = 15$, the number of the image pyramid layers is 3. For Bradley's method, the size of the matching window is $16 \times 16$ pixels. Note that the size of the matching window for 1D POC and SW-POC is equal to that for Bradley's method, since the 1D Hanning window is applied to 1D image signals for 1D POC and SW-POC to reduce the effect of discontinuity at signal border in 1D DFT [5]. The search range is $\pm 40$ pixels at the coarsest layer for SW-POC and $\pm 160$ pixels for Bradley's method, where these search ranges are equivalent. The scale factors of the scaled window matching are $1/2$, $1/\sqrt{2}$, $1$, $\sqrt{2}$ and $2$ both for SW-POC and Bradley's method.

Fig. 4 shows the stereo vision system used in our experiments. We reconstruct a solid plane and a solid sphere as reference objects, where the distance between the camera and the reference object is around 600 mm. We capture the stereo images of reference objects with changing the baseline length from 50 mm to 400 mm, where we fix the left camera and

TABLE I
OUTLIER RATES [%] IN 3D RECONSTRUCTION OF A PLANE AND A SPHERE.

| Baseline [mm] | | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|---|
| Plane | 1DPOC | 1.69 | 11.94 | 34.04 | 93.67 | 92.75 |
| | SW-POC | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Bradley | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Sphere | 1DPOC | 0.00 | 0.98 | 98.89 | 98.76 | 99.05 |
| | SW-POC | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 |
| | Bradley | 0.00 | 0.00 | 0.18 | 6.04 | 18.63 |

TABLE II
RMS ERRORS [MM] IN 3D RECONSTRUCTION OF A PLANE AND A SPHERE.

| Baseline [mm] | | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|---|
| Plane | 1DPOC | 0.3224 | 0.3298 | 0.4720 | 0.5276 | 0.4153 |
| | SW-POC | 0.2010 | 0.1601 | 0.1320 | 0.1130 | 0.1117 |
| | Bradley | 0.4513 | 0.3864 | 0.1889 | 0.1271 | 0.1385 |
| Sphere | 1DPOC | 0.2240 | 0.1319 | 0.6104 | 0.4634 | 0.3893 |
| | SW-POC | 0.2203 | 0.1119 | 0.0688 | 0.0564 | 0.0588 |
| | Bradley | 0.3479 | 0.1998 | 0.1455 | 0.1404 | 0.1211 |

TABLE III
COMPUTATIONAL COST FOR FINDING A SINGLE CORRESPONDING POINT.

| | Additions | Multiplications | Divisions | Square roots |
|---|---|---|---|---|
| 1D POC | 70,784 | 61,056 | 3,840 | 1,920 |
| SW-POC | 200,555 | 172,992 | 10,880 | 5,444 |
| Bradley | 3,109,300 | 1,302,200 | 25,500 | 25,500 |

move the right camera so that the left camera image is not changed. Fig. 5 shows examples of the left camera images used in the experiments. We rectify the stereo images for each baseline setting, and find the correspondence for the reference points on the object of left images. In the experiments, we place the reference points in a grid with a spacing of 10 pixels. The proposed method adopts the coarse-to-fine strategy described in Sect. II-D. For the proposed method, we place the sparse reference points in a grid with a spacing of 30 pixels and the dense reference points in a grid with a spacing of 10 pixels.

### A. Accuracy of 3D Measurement

We evaluate the 3D measurement accuracy by fitting errors of plane and sphere models. Table I shows outlier rates for 1D POC, SW-POC and Bradley's method, where the outlier is defined by a point whose fitting error is greater than 1 pixel. Fig. 6 shows examples of the 3D points reconstructed by each method from the stereo image pairs whose baseline length is 50 mm and 400 mm. The outlier rates for 1D POC are increased with increasing the length of baseline, since 1D POC assumes that the local distortion between stereo image pairs is only displacement. On the other hand, the outlier rates for SW-POC and Bradley's method are not increased with increasing the length of baseline. Furthermore, in the case of a sphere, the outlier rates for Bradley's method are gradually increased, since the local distortion can not be approximated by horizontal scaling when skew occurs between the matching windows, while the outlier rates for SW-POC is significantly small, since the proposed method consider the skew between the matching windows as described in Sect. II-E.

Table II shows the RMS (Root Mean Square) of fitting errors in 1D POC, SW-POC and Bradley's method, where RMS errors are calculated for 3D points without outliers. RMS errors in SW-POC and Bradley's method are reduced with the wider-baseline setting, since the use of the wide-baseline settings makes it possible to suppress the influence of correspondence errors on the reconstruction results. From Table II and Fig. 6, the RMS errors in SW-POC are less than those for Bradley's method, especially for a plane with short-baseline settings and a sphere with wide-baseline settings. This is because the sub-pixel accuracy of SW-POC is higher than that of Bradley's method and the proposed method consider the local distortion such as skew. As is observed in the above experiments, the proposed method can find accurate correspondence from both short- and wide-baseline stereo image pairs.

### B. Computational Cost

We evaluate the amount of computation required for finding a single corresponding point. Note that the computational cost of SW-POC is defined by the mean cost for sparse and dense reference points, since the proposed method employ the coarse-to-fine strategy as described in Sect. II-D.

Table III shows the number of additions, multiplications, divisions and square roots for each method. The computational cost of SW-POC is smaller than that of the Bradley's method. Although the proposed method employs the iterative window matching to estimate the scale factor and the initial disparity, the computational cost of SW-POC is not significantly large compared with that of 1D POC. This is because the proposed method suppresses the increase in computational cost by estimating the scale factor and the initial disparity only in the coarsest layer and by using the coarse-to-fine strategy as described in Sect. II-D.

## IV. APPLICATION

In this section, we present simple, accurate and dense 3D reconstruction using a moving and uncalibrated consumer digital camera as an application of the proposed method.

In the case of 3D reconstruction using a consumer digital camera, a stereo image pair is obtained by capturing two images from different viewpoints, where the baseline length between two images depends on the camera motion. So, a correspondence matching technique robust against the baseline length is required for the 3D reconstruction from a moving camera. Thus, we employ the correspondence matching method using SW-POC proposed in this paper. Since SW-POC assumes the use of rectified stereo image pairs, the camera parameters must be estimated. In the case of a moving camera, we need to estimate the camera parameters from the captured images, since it is difficult to calibrate a camera in advance. Addressing this problem, we employ the Structure from Motion (SfM) using SIFT [2], [6], [7]. Also, the 3D points reconstructed by the proposed method are accurate and dense, so we can generate the mesh model from the

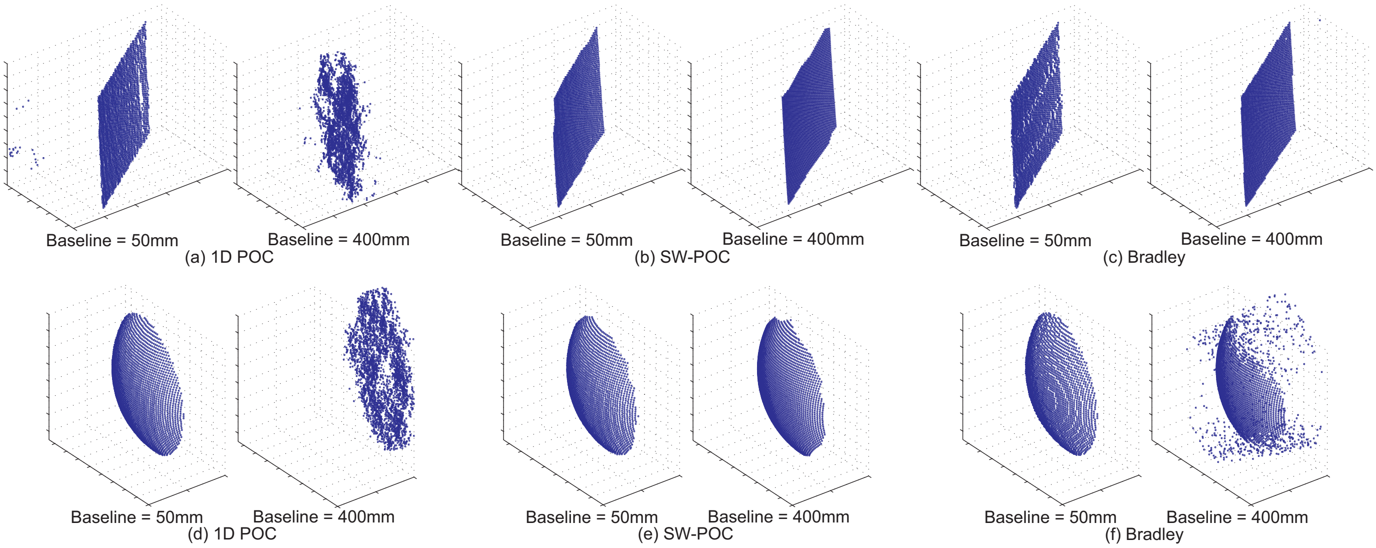


Fig. 6. Reconstruction results: (a)–(c) reconstructed 3D points of a plane, and (d)–(f) reconstructed 3D points of a sphere.

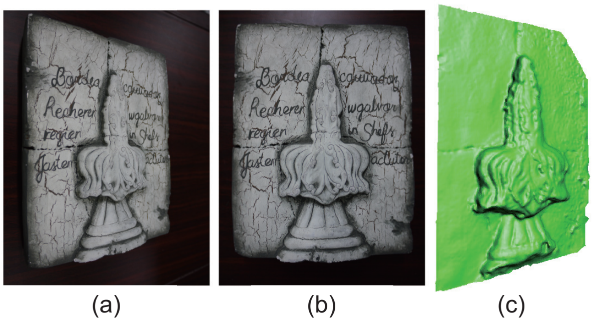


Fig. 7. Reconstruction result of the tile: (a) left image, (b) right image, and (c) reconstructed mesh model

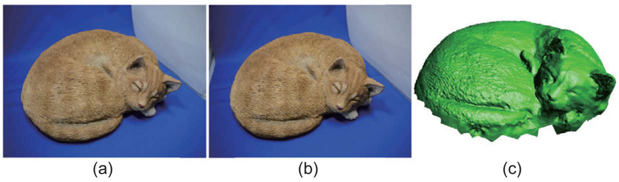


Fig. 8. Reconstruction result of the cat: (a) left image, (b) right image, and (c) reconstructed mesh model

reconstructed 3D points using Poisson Surface Reconstruction [8], where a surface normal of each point is calculated from neighboring points.

For example, we reconstruct 3D shape of a interior tile and a cat carving from two views captured by a consumer digital camera, where the size of images is 2,000 × 1,500 pixels. Fig. 7 and Fig. 8 show stereo images and reconstructed mesh models of the tile and the cat, respectively. Note that the baseline length in Fig. 7 (a) and (b) is wide, while that in Fig. 8 (a) and (b) is short. As a result, for objects having complicated structure such as the tile and the cat, the proposed method can reconstruct accurate 3D mesh models regardless of the difference in baseline setting.

## V. Conclusion

This paper has proposed an accurate and dense stereo correspondence matching method using SW-POC, and demonstrated the high-accuracy 3D reconstruction using SW-POC with lower computational cost than Bradley's method. We have also demonstrated 3D reconstruction from a moving camera using the proposed method. The use of the proposed method makes it possible to reconstruct accurate and dense 3D structure of objects with very simple operation such as two shots of a consumer digital camera.

## References

[1] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer-Verlag New York Inc., 2010.

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

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'l J. Computer Vision*, vol. 47, no. 1–3, pp. 7–42, Apr. 2002.

[4] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[5] 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*, pp. V–221–V–224, 2007.

[6] M. Brown and D. G. Lowe, "Unsupervised 3D object recognition and reconstruction in unordered datasets," *Proc. 5th Int'l Conf. 3-D Digital Imaging and Modeling*, pp. 56–63, 2005.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge University Press, 2004.

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