An Efficient Image Matching Method for Multi-View Stereo

Shuji Sakai[1], Koichi Ito[1], Takafumi Aoki[1], Tomohito Masuda[2],
and Hiroki Unten[2]

[1] Graduate School of Information Sciences, Tohoku University,
Sendai, Miyagi, 980–8579, Japan
sakai@aoki.ecei.tohoku.ac.jp
[2] Toppan Printing Co., Ltd., Bunkyo-ku, Tokyo, 112–8531, Japan

Abstract. Most existing Multi-View Stereo (MVS) algorithms employ the image matching method using Normalized Cross-Correlation (NCC) to estimate the depth of an object. The accuracy of the estimated depth depends on the step size of the depth in NCC-based window matching. The step size of the depth must be small for accurate 3D reconstruction, while the small step significantly increases computational cost. To improve the accuracy of depth estimation and reduce the computational cost, this paper proposes an efficient image matching method for MVS. The proposed method is based on Phase-Only Correlation (POC), which is a high-accuracy image matching technique using the phase components in Fourier transforms. The advantages of using POC are (i) the correlation function is obtained only by one window matching and (ii) the accurate sub-pixel displacement between two matching windows can be estimated by fitting the analytical correlation peak model of the POC function. Thus, using POC-based window matching for MVS makes it possible to estimate depth accurately from the correlation function obtained only by one window matching. Through a set of experiments using the public MVS datasets, we demonstrate that the proposed method performs better in terms of accuracy and computational cost than the conventional method.

1 Introduction

In recent years, the topic of Multi-View Stereo (MVS) has attracted much attention in the field of computer vision [1–10]. MVS aims to reconstruct a complete 3D model from a set of images taken from different viewpoints. The major MVS algorithm consists of two steps: (i) estimating the 3D points on the basis of a photo-consistency measure and visibility model using a local image matching method and (ii) reconstructing a 3D model from estimated 3D point clouds. The accuracy, robustness and computational cost of MVS algorithms depend on the performance of the image matching method, which is the most important factor in MVS algorithms.

Most MVS algorithms employ Normalized Cross-Correlation (NCC)-based image matching to estimate 3D points [1, 5, 6, 8–10]. Goesele et al. [5] have applied NCC-based image matching to the plane-sweeping approach to estimate a reliable depth map by cumulating the correlation values calculated from multiple image pairs with changing the depth. Campbell et al. [8] estimated a depth map more accurately than Goesele et al. [5] by using the matching results obtained from neighboring pixels to reduce outliers. Bradley et al. [9] and Furukawa et al. [10] achieved robust image matching by transforming the matching window in accordance with not only the depth but also the normal of the 3D points.

In the MVS algorithms mentioned in the above, an NCC value between matching windows is used as the reliability of a 3D point. The optimal 3D point is estimated by iteratively computing NCC values between matching windows with changing the parameter of 3D point, i.e., depth or normal. For example, the plane-sweeping approach such as that of Goesele et al. [5] computes NCC values between matching windows with discretely changing the depth and selects the depth that has the highest NCC value as the optimal one. To estimate the accurate depth, a sufficiently small step of the depth must be employed, which significantly increases computational cost. If the step of the depth is small, the translational displacement of a 3D point is a sub-pixel on the multi-view images. Most existing methods assume that the sub-pixel resolution of a matching window is represented by linear interpolation. This assumption, however, is not always true.

In this paper, we propose an efficient image matching method for MVS using Phase-Only Correlation (POC) (or simply "phase correlarion"). POC is a kind of correlation function calculated only from the phase components in Fourier transform. The translational displacement and similarity between two images can be estimated from the position and height of the correlation peak of the POC function, respectively. Kuglin et al. [11] proposed a fundamental image matching technique using POC, and Takita et al. [12] proposed a sub-pixel image registration technique using POC. The major advantages of using POC-based instead of NCC-based image matching are the following two points: (i) the correlation function is obtained only by one window matching and (ii) the accurate sub-pixel translational displacement between two windows can be estimated by fitting the analytical correlation peak model of the POC function. By applying POC-based image matching to depth estimation, the peak position of the POC function indicates the displacement between the assumed and true depth. Hence, we can directly estimate the true depth from the results of only one POC-based window matching. By introducing POC-based image matching to the plane-sweeping approach, we need little window matching to estimate the true depth from multi-view images. In addition, the accuracy of depth estimation can be improved by integrating the POC functions calculated from multiple stereo image pairs. Thus, using POC-based window matching for MVS makes it possible to estimate depth accurately from the correlation function obtained only by one window matching. Through a set of experiments using the public multi-view stereo datasets [13], we demonstrate that the proposed method performs better in terms of the accuracy and the computational cost than the method proposed by Goesele et al. [5].

2 Phase-Only Correlation

This section describes the fundamentals of POC-based image matching. Most existing POC-based image matching methods are for 2D images. The image matching between stereo images can be reduced to a 1D image matching through stereo rectification. In this paper, we employ 1D POC function to estimate the depth from multi-view images.

POC is an image matching technique using the phase components in Discrete Fourier Transforms (DFTs) of given images. Consider two N-length 1D image signals $f(n)$ and $g(n)$, where the index range is $-M, \cdots, M$ ($M > 0$) and hence $N = 2M + 1$. Let $F(k)$ and $G(k)$ denote the 1D DFTs of the two signals. $F(k)$ and $G(k)$ are given by

$$F(k) = \sum_{n=-M}^{M} f(n) W_N^{kn} = A_F(k) e^{j\theta_F(k)}, \tag{1}$$

$$G(k) = \sum_{n=-M}^{M} g(n) W_N^{kn} = A_G(k) e^{j\theta_G(k)}, \tag{2}$$

where $k = -M, \cdots, M$, $W_N = e^{-j\frac{2\pi}{N}}$, $A_F(k)$ and $A_G(k)$ are amplitude, and $\theta_F(k)$ and $\theta_G(k)$ are phase. The normalized cross-power spectrum $R(k)$ is given by

$$R(k) = \frac{F(k)\overline{G(k)}}{\left|F(k)\overline{G(k)}\right|} = e^{j(\theta_F(k) - \theta_G(k))}, \tag{3}$$

where $\overline{G(k)}$ is the complex conjugate of $G(k)$, and $\theta_F(k) - \theta_G(k)$ denotes the phase difference. The POC function $r(n)$ is defined by Inverse DFT (IDFT) of $R(k)$ and is given by

$$r(n) = \frac{1}{N} \sum_{k=-M}^{M} R(k) W_N^{-kn}. \tag{4}$$

Shibahara et al. [14] derived the analytical peak model of 1D POC function. Let us assume that $f(n)$ and $g(n)$ are minutely displaced with each other. The analytical peak model of 1D POC function can be defined by

$$r(n) \simeq \frac{\alpha}{N} \frac{\sin\left(\pi(n + \delta)\right)}{\sin\left(\frac{\pi}{N}(n + \delta)\right)}, \tag{5}$$

where δ is a sub-pixel peak position and α is a peak value. The peak position $n = \delta$ indicates the translational displacement between the two 1D image signals

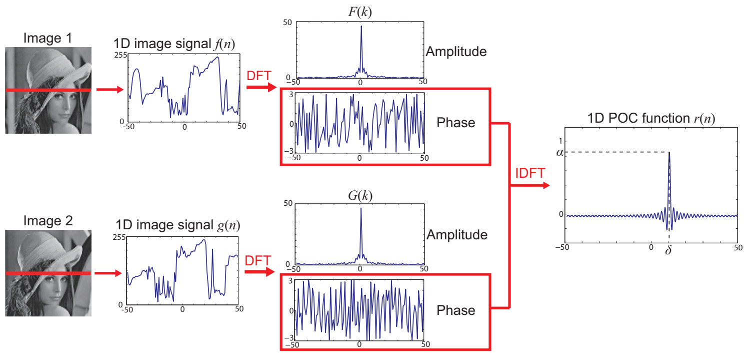

Fig. 1. Example of 1D POC-based image matching.

and the peak value α indicates the similarity between the two 1D image signals. The translational displacement with sub-pixel accuracy can be estimated by fitting the model of Eq. (5) to the calculated data array around the correlation peak, where α and δ are fitting parameters. In addition, we employ the following techniques to improve the accuracy of 1D image matching: (i) windowing to reduce boundary effects, (ii) spectral weighting for reducing aliasing and noise effects, and (iii) averaging 1D POC functions to improve peak-to-noise ratio [12, 14]. Fig. 1 shows an example of 1D POC-based image matching.

3 POC-Based Image Matching for Multi-View Stereo

In this section, we describe a POC-based image matching method for MVS. The existing algorithms using NCC-based image matching need to do many NCC computations with changing the assumed depth to estimate the accurate depth of a 3D point. On the other hand, the proposed method estimates the accurate depth only with one window matching by approximating the depth change on a 3D point by the translational displacement on the stereo image and estimating the translational displacement using POC. The proposed method also enhances the estimation accuracy by integrating the POC functions calculated from multiple stereo image pairs.

The POC functions calculated from stereo images with different view-points indicate the different peak positions due to the difference in camera positions. To integrate the POC functions, the proposed method normalizes the disparity of each stereo image and integrates the POC functions on the same coordinate system. So far, Okutomi et al. [15] have proposed the disparity normalization technique to integrate correlation functions calculated from stereo images with different viewpoints. This technique, however, assumes that all cameras are located on the same line. This assumption is not suitable in a practical situation. The disparity normalization technique used in the proposed method, which is a generalized version of the technique proposed by Okutomi et al. [15], can integrate the correlation functions calculated from stereo images with different viewpoints even if the cameras are not located on the same line.

Let $\mathbf{V} = \{V_0, \cdots, V_{H-1}\}$ be the multi-view images with known camera parameters. We consider a reference view $V_R \in \mathbf{V}$ and neighboring views $\mathbf{C} = \{C_0, \cdots, C_{K-1}\} \subset \mathbf{V} - \{V_R\}$ as input images, where H and K are the number of the multi-view images and the number of the neighboring views, respectively. The proposed method generates K pairs of a rectified stereo image and estimates the depth of each point in V_R from the peak position of the correlation function obtained by integrating the POC functions with normalized disparity. We use a stereo rectification method employed in the Camera Calibration Toolbox for Matlab [16].

Next, we describe the key techniques of the proposed method: (i) normalizing the disparity and (ii) integrating the POC functions. Then, we describe the proposed depth estimation method using POC-based image matching.

3.1 Normalization of Disparity

We consider that the camera coordinate of the reference view V_R corresponds to the world coordinate. Let $V_{R,i}^{\mathrm{rect}}$-C_i^{rect} be the rectified stereo image pair, where $V_{R,i}^{\mathrm{rect}}$ is the rectified image of V_R so as to correspond to the view angle of C_i. The relationship among the 3D point $\mathbf{M} = [X, Y, Z]^T$ in the camera coordinate of V_R, the rectified stereo image $V_{R,i}^{\mathrm{rect}}$-C_i^{rect} ($C_i \in \mathbf{C}$) with disparity d_i, and the rectified stereo image $V_{R,j}^{\mathrm{rect}}$-C_j^{rect} ($C_j \in \mathbf{C} - \{C_i\}$) with disparity d_j is defined by

$$\mathbf{M} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/d_i \\ (v_i - v_{0i})B_i/d_i \\ \beta_i B_i/d_i \end{bmatrix} = \mathbf{R}_j \begin{bmatrix} (u_j - u_{0j})B_j/d_j \\ (v_j - v_{0j})B_j/d_j \\ \beta_j B_j/d_j \end{bmatrix}, \tag{6}$$

where (u_l, v_l) is the corresponding point of $\mathbf{M}$ in $V_{R,l}^{\mathrm{rect}}$, (u_{0l}, v_{0l}) is the optical center of $V_{R,l}^{\mathrm{rect}}$, β_l is focal length and B_l is baseline length between $V_{R,l}^{\mathrm{rect}}$-C_l^{rect} $(l = i, j)$. $\mathbf{R}_l$ denotes a rotation matrix from the reference view V_R to the rectified reference view $V_{R,l}^{\mathrm{rect}}$ used in stereo rectification for $V_{R,l}^{\mathrm{rect}}$-C_l^{rect}, and is given by

$$\mathbf{R}_l = \begin{bmatrix} R_{l11} & R_{l12} & R_{l13} \\ R_{l21} & R_{l22} & R_{l23} \\ R_{l31} & R_{l32} & R_{l33} \end{bmatrix}. \tag{7}$$

From Eq. (6), we derive the relationship between d_i and d_j as follows

$$d_i = \frac{R_{i31}(u_i - u_{0i}) + R_{i32}(v_i - v_{0i}) + R_{i33}\beta_i}{R_{j31}(u_j - u_{0j}) + R_{j32}(v_j - v_{0j}) + R_{j33}\beta_j} \frac{B_i}{B_j} d_j. \tag{8}$$

From Eq. (8), the relationship between d_i and d_j is represented by the scaling factor that depends on the camera parameters and the coordinates of the corresponding points in V_R^{rect}. We define the normalized disparity d to take into

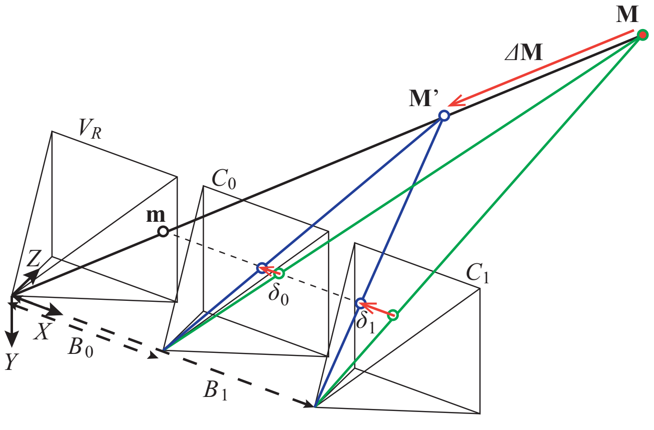

Fig. 2. Geometric relationship between the location of 3D point and the disparity on the images.

account the scale factor for each disparity. If we consider the rectified stereo image pair $V^{\mathrm{rect}}_{R,i}$-C^{rect}_i $(i = 0, \cdots, K-1)$, the relationship between d_i in each rectified stereo pair and the normalized disparity d can be written as

$$d_i = s_i d, \tag{9}$$

where s_i denotes the scale factor for the disparity d_i and is given by

$$s_i = \frac{(R_{i31}(u_i - u_{0i}) + R_{i32}(v_i - v_{0i}) + R_{i33}\beta_i)B_i}{\frac{1}{K}\sum_{l=0}^{K-1}(R_{l31}(u_l - u_{0l}) + R_{l32}(v_l - v_{0l}) + R_{l33}\beta_l)B_l}. \tag{10}$$

In this case, the 3D point $\mathbf{M}$ can be defined by

$$\mathbf{M} = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/(s_i d) \\ (v_i - v_{0i})B_i/(s_i d) \\ \beta_i B_i/(s_i d) \end{bmatrix}. \tag{11}$$

3.2 Integration of POC Function

We consider the 3D point $\mathbf{M}$ and its minutely displaced 3D point $\mathbf{M}' = \mathbf{M} + \mathit{\Delta}\mathbf{M}$, where $\mathit{\Delta}\mathbf{M} = [\mathit{\Delta}X, \mathit{\Delta}Y, \mathit{\Delta}Z]^T$ denotes the minute displacement, as shown in Fig. 2. Let d and d' be the normalized disparities of $\mathbf{M}$ and $\mathbf{M}'$, respectively. Assuming that $\mathbf{M}$ is the true 3D point, the relationship between d and d' is given by

$$d' = d + \delta, \tag{12}$$

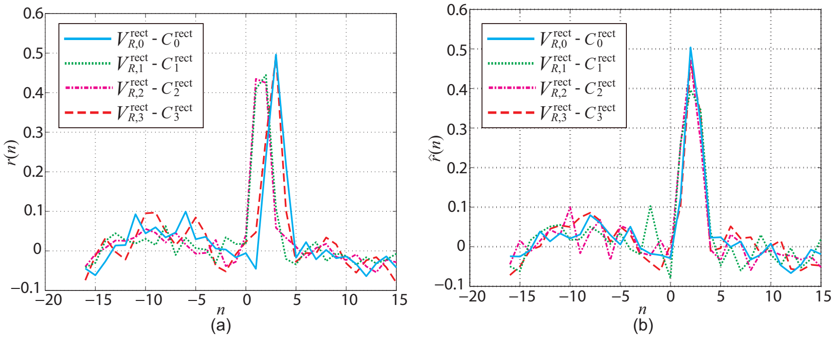

Fig. 3. Integration of the POC functions calculated from stereo image pairs with different viewpoints: (a) POC functions before disparity normalization and (b) POC functions after disparity normalization.

where δ denotes the error between the normalized disparities d and d'. For the rectified stereo image pair $V_{R,i}^{\mathrm{rect}}$-C_i^{rect} ($i \in \{0, \cdots, K-1\}$), the relationship between the 3D point $\mathbf{M}'$ and the normalized disparity d is

$$\mathbf{M}' = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/(s_i(d+\delta)) \\ (v_i - v_{0i})B_i/(s_i(d+\delta)) \\ \beta_i B_i/(s_i(d+\delta)) \end{bmatrix} . \tag{13}$$

Let f_i and g_i be the matching windows extracted from $V_{R,i}^{\mathrm{rect}}$ and C_i^{rect} centered on the corresponding point of $\mathbf{M}'$, respectively. Approximating the local image transformation by translational displacement, the translational displacement between f_i and g_i is $\delta_i = s_i\delta$. The displacement δ_i can be estimated from the correlation peak position of the POC function r_i between f_i and g_i as mentioned in Sect. 2. The different rectified stereo image pairs, however, have different translational displacements. For example, δ_i in $V_{R,i}^{\mathrm{rect}}$-C_i^{rect} and δ_j in $V_{R,j}^{\mathrm{rect}}$-C_j^{rect} ($j \in \{0, \cdots, K-1\} - \{i\}$) are not always equal. In other words, the POC functions r_i and r_j have different correlation peak positions.

Addressing this problem, we convert the coordinate system of the POC functions r_i and r_j into the same coordinate system by scaling the matching windows in accordance with each normalized disparity. Let w be the unified size of the matching window. The size of the matching windows of f_i and g_i is defined by $s_i w$. Scaling the image signals f_i and g_i by $1/s_i$, the size of the matching windows is normalized to w, where we denote $\hat{f}_i$ and $\hat{g}_i$ as the scaled version of the matching windows f_i and g_i, respectively. Hence, the correlation peak of the POC function $\hat{r}_i$ between $\hat{f}_i$ and $\hat{g}_i$ is located at δ. Similarly, for the rectified stereo image pair $V_{R,j}^{\mathrm{rect}}$-C_j^{rect}, the correlation peak of the POC function $\hat{r}_j$ between $\hat{f}_j$ and $\hat{g}_j$ is located at the same position δ, although the size of the matching window, i.e., $s_j w$, is different from that for $V_{R,i}^{\mathrm{rect}}$-C_i^{rect}, i.e., $s_i w$.

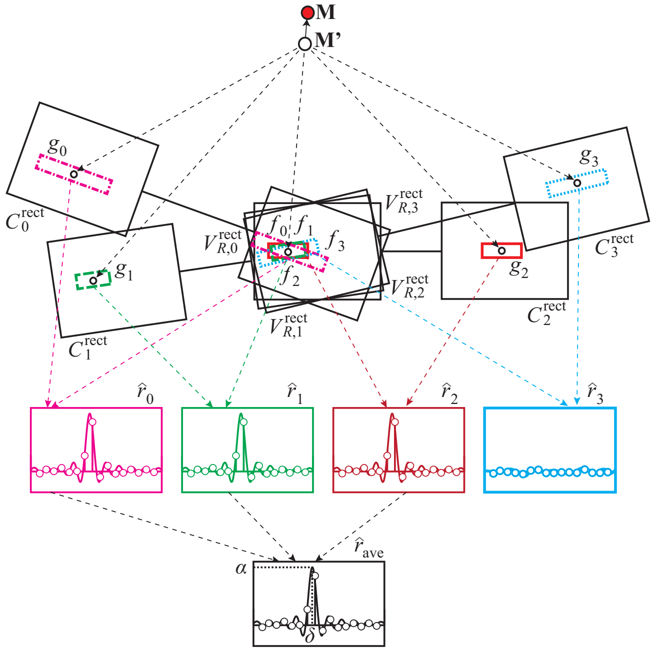

Fig. 4. Depth estimation using POC-based image matching.

Fig. 3 (a) shows the POC functions before disparity normalization. In this case, the translational displacement δ_i between matching windows is different for each view-point. Thus, the positions of the correlation peaks are also different. On the other hand, Fig. 3 (b) shows the POC functions after disparity normalization. In this case, the translational displacement δ is the same for all the viewpoints. Therefore, all the POC functions overlap at the same position.

Using disparity normalization makes it possible to integrate the POC functions calculated from rectified stereo image pairs with different viewpoints. In this paper, we employ the POC function $\hat{r}_{\text{ave}}$, which is the average of the POC functions $\hat{r}_i$ $(i = 0, \cdots, K-1)$, as the integrated POC functions.

3.3 Depth Estimation Using POC-Based Image Matching

We describe the depth estimation method using POC-based image matching with two important techniques as described above. Fig. 4 shows the flow of the proposed method. First, the initial position of the 3D point $\mathbf{M}'$ is projected onto the rectified stereo image pair $V_{R,i}^{\text{rect}}$-C_i^{rect}, and the coordinates on $V_{R,i}^{\text{rect}}$ and C_i^{rect} are denoted by $\mathbf{m}_i = [u_i, v_i]$ and $\mathbf{m}_i^C = [u_i^C, v_i^C]$, respectively, where $i = 0, \cdots, K-1$. Next, the matching windows f_i and g_i extracted from $V_{R,i}^{\text{rect}}$ centered at $\mathbf{m}_i$ with the size $s_i w \times L$ and C_i^{rect} centered at $\mathbf{m}_i^C$ with the size

$s_i w \times L$, respectively. Note that we extract L lines of the matching window to employ the technique averaging 1D POC functions to improve the peak-to-noise ratio as described in Sect. 2. Then, we apply the disparity normalization to the matching windows f_i and g_i and calculate the 1D POC function $\hat{r}_i$ between $\hat{f}_i$ and $\hat{g}_i$. The correlation peak position of the 1D POC function $\hat{r}_i$ may include a significant error if 3D point $\mathbf{M}'$ is not visible from the neighboring view $C_i \in \mathbf{C}$ or the matching window is extracted from the boundary region of an object that has multiple disparities. In this case, we observe that the correlation peak value α_i drops, since the local image transformation between the matching windows cannot be approximated by translational displacement. To improve the accuracy of depth estimation, the average POC function $\hat{r}_{\mathrm{ave}}$ is calculated from the POC functions $\hat{r}_i$ with $\alpha_i > th_{corr}$, where th_{corr} is a threshold. Finally, the correlation peak position δ with sub-pixel accuracy is estimated by fitting the analytical peak model of the POC function to $\hat{r}_{\mathrm{ave}}$. From Eq. (11), Eq. (12), and δ, the true position of the 3D point $\mathbf{M}$ is estimated by

$$\mathbf{M} = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/(s_i(d' - \delta)) \\ (v_i - v_{0i})B_i/(s_i(d' - \delta)) \\ f_i B_i/(s_i(d' - \delta)) \end{bmatrix} . \qquad (14)$$

To generate a depth map, we apply the POC-base image matching to a plane-sweeping approach, and search the depth of each pixel in V_R. Since the POC-based image matching can estimate the depth corresponding to $\pm w/4$ pixel in the neighboring-view image, we search on the ray within the bounding box with changing the depth of $\mathbf{M}'$ in stpdf of $s_i w/4$ pixel in the stereo images. We also apply the the coarse-to-fine strategy using image pyramids to the proposed method described in the above. We first esimate the approximate depth in the coarsest image layer, and then refine the depth in the subsequent image layers.

4 Experiments and Discussion

We evaluate the reconstruction accuracy and the computational cost of the conventional method and the proposed method using the public multi-view stereo image datasets [13]. In the experiments, we employ the famous method using the plane-sweeping approach proposed by Goesele et al. [5] as the conventional method.

4.1 Implementation

We describe the implementation notes for Goesele's method and the proposed methods.

Goesele's method [5]

The reconstruction accuracy and the computational cost of Goesele's method significantly depends on the step size $\varDelta Z$ of the depth. In the experiments, we employ four variations of $\varDelta Z$ such that the resolution of the disparity on the

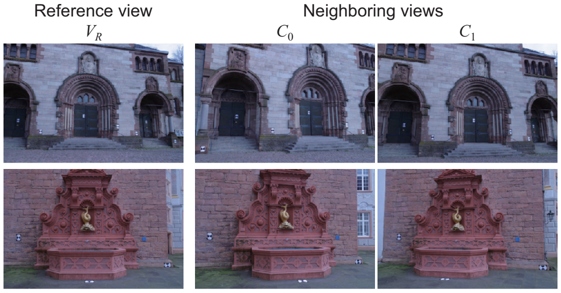

Fig. 5. Examples of reference-view image V_R and neighboring-view images $\mathbf{C}$ used in the experiments (upper: Herz-Jesu-P8, lower: Fountain-P11).

widest-baseline stereo image is $1, 1/2, 1/5$, and $1/10$ pixels. The size of NCC-based window matching is 17×17 pixels. The threshold value for averaging the NCC values calculated from stereo image pairs is 0.3.

Proposed method

The parameters for the proposed method used in the experiments are as follows. The threshold th_{corr} is 0.3, the matching window size w is 32 pixel and the number of POC functions L is 17. Note that the effective information of POC function with 32 pixels×17 lines is limited to 17 pixels×17 line, since we apply a Hanning widow with $w/2$-half width to the POC function to reduce the boundary effect as described in Sect. 2. We also employ the coarse-to-fine strategy using image pyramids. The numbers of layers are 2, 3, and 4 for 768×512, $1,536 \times 1,024$, and $3,072 \times 2,048$ pixels, respectively.

4.2 Evaluation of 3D Reconstruction Accuracy

We evaluate the 3D reconstruction accuracy using Herz-Jesu-P8 (8 images) and Fountain-P11 (11 images), which are available in [13]. The datasets Herz-Jesu-P8 and Fountain-P11 include the multi-view images with $3,072 \times 2,048$ pixels, camera parameters, bounding boxes, and the mesh model of the target object that can be used as the ground truth. For each dataset, we generate depth maps of all the view points using Goesele's method and the proposed method. We use two neighboring-view images $\mathbf{C}$ for one reference-view image V_R. Fig. 5 shows examples of V_R and $\mathbf{C}$ used in the experiments. The performance is evaluated for the three different image sizes : 768×512, $1,536 \times 1,024$, and $3,072 \times 2,048$ pixels.

We evaluate the accuracy of 3D reconstruction by the error rate e defined by

$$e = \frac{|Z_{\text{calculated}} - Z_{\text{ground truth}}|}{Z_{\text{ground truth}}} \times 100 \ [\%], \tag{15}$$

where $Z_{\text{calculated}}$ and $Z_{\text{ground truth}}$ denote the estimated depth and the true depth obtained from the ground truth, respectively. Fig. 6 shows the reconstructed 3D

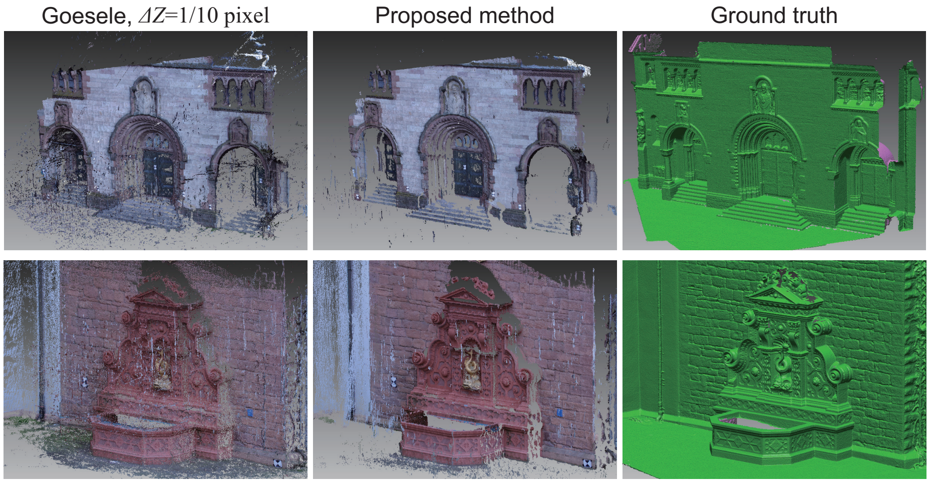

Fig. 6. Reconstruction results of $1,536 \times 1,024$-pixel images for each dataset (upper: Herz-Jesu-P8, lower: Fountain-P11).

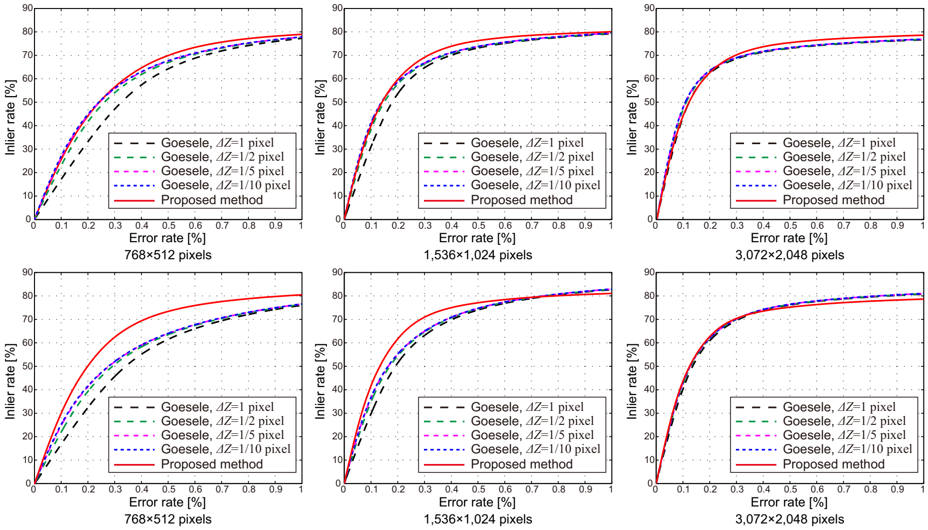

Fig. 7. Inlier rate for each dataset (upper: Herz-Jesu-P8, lower: Fountain-P11).

point clouds of Goesele's method and the proposed method for $1,536 \times 1,024$-pixel images. Fig. 7 shows the inlier rates for changing threshold of the error rates for each dataset. Fig. 8 shows the average error rates of inliers, where the inlier is defined by a 3D point whose error rate is less than 1.0%.

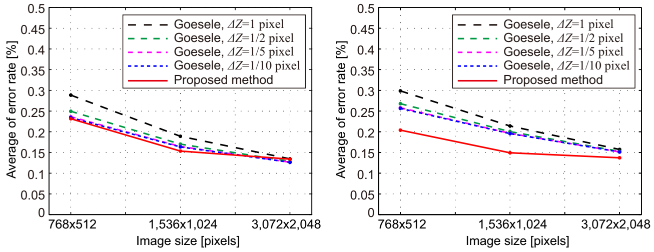

Fig. 8. Average error rates for each dataset (left: Herz-Jesu-P8, right: Fountain-P11).

For Goesele's method, the error rates of the 3D point clouds are small when the step size ΔZ is sufficiently small. For the proposed method, we observe that the reconstructed 3D points are concentrated on smaller error rates than in Goesele's method with $\Delta Z = 1/10$ pixel. We also confirm this result from the average error rates in Fig. 8. For Fountain-P11, the proposed method can estimate more accurate depth than Goesele's method for all the image sizes. In Goesele's method, to estimate the accurate depth, the sub-pixel displacement between the matching windows is represented by image interpolation. On the other hand, the proposed method employs the POC-based image matching, which can estimate the accurate sub-pixel displacement between the matching windows by fitting the analytical correlation peak model of the POC function.

As is observed in the above experiments, the proposed method exhibits higher reconstruction accuracy than Goesele's method.

4.3 Evaluation of Computational Cost

We evaluate the computational cost to estimate the depth of one point on the reference-view image for Goesele's method and the proposed method. When using the w-pixel matching window, the proposed method can estimate the displacement within $\pm w/4$ pixels for one window matching. In Goesele's method, we also estimate the displacement within $\pm w/4$ pixels using NCC-based image matching. Table 1 shows the computational cost for each method. Goesele's method with the small step size ΔZ requires high computational cost. On the other hand, the proposed method requires low computational cost that is comparable to that for Goesele's method with $\Delta Z = 1$ pixel or $\Delta Z = 1/2$ pixel. As described in Sect. 4.2, the reconstruction accuracy of the proposed method is higher than that of Goesele's method with $\Delta Z = 1/10$ pixel. Although the computational cost for Goesele's method can be reduced when ΔZ is large, the reconstruction accuracy drops significantly. Compared with Goesele's method, the proposed method exhibits efficient 3D reconstruction from multi-view images in terms of the reconstruction accuracy and the computational cost.

Table 1. Computational cost to estimate the depth of one point on the reference-view image for each method.

	Additions	Multiplications	Divisions	Square roots
Goesele, $\Delta Z = 1$ pixel	75,140	31,246	578	578
Goesele, $\Delta Z = 1/2$ pixel	150,280	62,492	1,156	1,156
Goesele, $\Delta Z = 1/5$ pixel	357,700	156,230	2,890	2,890
Goesele, $\Delta Z = 1/10$ pixel	751,400	312,460	5,780	5,780
Proposed method	40,000	34,496	2,176	1,088

5 Conclusion

This paper has proposed an efficient image matching method for Multi-View Stereo (MVS) using Phase-Only Correlation (POC). The proposed method with normalizing disparity and integrating POC functions can estimate the depth from the correlation function obtained only by one window matching. Also, the reconstruction accuracy of the proposed method is higher than that of NCC-based image matching, since POC-based image matching can estimate the accurate sub-pixel translational displacement between two windows by fitting the analytical correlation peak model of the POC function. Through a set of experiments using the public multi-view stereo datasets, we have demonstrated that the proposed method performs better in terms of accuracy and computational cost than Goesele's method. In future work, we will improve the accuracy of the proposed method to consider the normal vectors of 3D point and develop an MVS algorithm using the proposed method.

References

1. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer-Verlag New York Inc. (2010)
2. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-views stereo reconstruction algorithms. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2006) pp. 519–528
3. Strecha, C., Fransens, R., Gool, L.V.: Wide-baseline stereo from multiple views: A probabilistic account. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2004) pp. 552–559
4. Strecha, C., Fransens, R., Gool, L.V.: Combined depth and outlier estimation in multi-view stereo. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2006) pp. 2394–2401
5. Goesele, M., Curless, B., Seitz, S.M.: Multi-view stereo revisited. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2006) pp. 2402–2409
6. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. Proc. Int'l Conf. Computer Vision (2007) pp. 1–8
7. Strecha, C., von Hansen, W., Gool, L.V., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2008) pp. 1–8

8. Campbell, N.D.F., Vogiatzis, G., Hernandez, C., Cipolla, R.: Using multiple hypotheses to improve depth-maps for multi-view stereo. Proc. European Conf. Computer Vision (2008) pp. 766–779
9. Bradley, D., Boubekeur, T., Heidrich, W.: Accurate multi-view reconstruction using robust binocular stereo and surface meshing. Proc. Int'l Conf. Computer Vision and Pattern Recognition (2008) pp. 1–8
10. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Analysis and Machine Intelligence **Vol. 32** (2010) pp. 1362–1376
11. Kuglin, C.D., Hines, D.C.: The phase correlation image alignment method. Proc. Int'l Conf. Cybernetics and Society (1975) pp. 163–165
12. Takita, K., Aoki, T., Sasaki, Y., Higuchi, T., Kobayashi, K.: High-accuracy subpixel image registration based on phase-only correlation. IEICE Trans. Fundamentals **Vol. E86-A** (2003) pp. 1925–1934
13. Strecha, C.: (Multi-view evaluation) `http://cvlab.epfl.ch/data/`.
14. Shibahara, T., Aoki, T., Nakajima, H., Kobayashi, K.: A sub-pixel stereo correspondence technique based on 1D phase-only correlation. Proc. Int'l Conf. Image Processing (2007) pp. V–221–V–224
15. Okutomi, M., Kanade, T.: A multiple-baseline stereo. IEEE Trans. Pattern Analysis and Machine Intelligence **Vol. 15** (1993) pp. 353–363
16. Bouguet, J.Y.: (Camera calibration toolbox for matlab) `http://www.vision.caltech.edu/bouguetj/calib_doc/`.