

## DEVELOPMENT OF MULTI-VIEW STEREO CONSIDERING ACCURACY OF EXTERIOR ORIENTATION ELEMENTS

T. Fuse<sup>1,\*</sup> and H. Ikezawa<sup>1</sup>

<sup>1</sup> Dept. of Civil Engineering, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan – fuse@civil.t.u-tokyo.ac.jp

### Commission II

**KEY WORDS:** Multi-View Stereo, Accuracy of Exterior Orientation, Global Structure from Motion, Patch Optimization, Bundle Adjustment, Unmanned Aerial Vehicle, Oblique Camera

### ABSTRACT:

Structure from motion (SfM) has been widely used to achieve automatic 3D reconstructions. However, as the 3D point clouds obtained via SfM are sparse, multi-view stereo (MVS) was developed to compensate for this sparseness. The accuracy of the 3D surface depends on the accuracy of the orientation elements based on the SfM. Additionally, in the case of an unmanned aerial vehicle (UAV), SfM exhibits a decrease in the accuracy of the orientation elements during complex camera movements. This paper proposes a patch-based MVS (PMVS) method considering the accuracy of the orientation elements. The proposed method involves applying the global SfM, estimating accuracy of exterior orientation (EO) elements, and introducing the accuracy of EO elements to PMVS. The PMVS approximates an object surface by using small rectangular patches, namely local tangent plane approximation. The patches are optimized by minimizing the sum of the photometric discrepancy scores. The accuracy of the EO elements is introduced to the patch optimization as weighting function. This accuracy is defined using the variances of the estimated parameters in the bundle adjustment. We also investigate the types of weighting functions. The results indicate that the proposed method is capable of considering geometric conditions during patch estimation. The proposed method was applied to the three types of image datasets, i.e., images captured using an SLR camera at ground level, images captured using a UAV equipped with a SLR camera, and images captured using an airplane equipped with an oblique camera. Through the experimental results, the improved accuracy and the effectiveness of the proposed method were confirmed.

### 1. INTRODUCTION

3D models have been widely used due to extensive availability of multiple free software. These software are mainly based on structure from motion (SfM), which is used to achieve automatic 3D reconstructions. As SfM calculates the 3D coordinates only for feature points, the results of the 3D point clouds are sparse. Therefore, dense image matching techniques have been developed to compensate for this sparseness by estimating the 3D surface corresponding to all pixels in the images. These dense image matching techniques are based on stereo matching methods. The stereo matching methods can be extended to methods that can address multiple images, known as multi-view stereo (MVS). Thus far, many MVS methods such as voxel-based approach, deformable polygon meshes, multiple depth maps, and patch-based methods (Remondino et al., 2012; Remondino et al., 2014; Seitz et al., 2006) have been developed.

A majority of MVS methods generally consist of 4 steps (Scharstein and Szeliski, 2002): (1) matching cost (similarity measure) calculations, (2) cost aggregation in the peripheral domain, (3) calculation and optimization of disparity, and (4) improvement of disparity (filtering). Based on the process of calculating and optimizing disparity, the MVS methods are categorized as local, global, or semi-global methods (Szeliski, 2011).

Global methods estimate disparities based on the formulation of the optimization problem, such as energy minimization, in the entire image rather than cost aggregations. A regularization

term can be introduced in the energy function for a stable estimation. The computation load depends on the shape complexity of the objects. Semi-global methods (Hirschmüller, 2008; Bethmann and Luhmann, 2014) were proposed to reduce computational load.

Local methods are advantageous in terms of their computation load. They estimate the disparities of pixels by minimizing matching costs, such as intensity difference, in the peripheral domain, i.e., winner-take-all optimization. One of the most popular method among these local methods is the patch-based MVS (PMVS) (Furukawa and Ponce, 2010). The PMVS approximates the object surface by using a large number of small patches with normal vectors. The patches are expanded from the results of SfM. The PMVS is robust against occlusions and mismatches, and it does not require an initial approximation.

The accuracy of the 3D surface reconstruction obtained via PMVS depends on the accuracy of the orientation elements based on the SfM. In the case of an unmanned aerial vehicle (UAV), SfM exhibits a decrease in the accuracy of orientation elements when the camera movement is complicated. Additionally, the image should be selected according to the accuracy to apply the method to enormous image sets (Fuse and Harada, 2016). This paper proposes a PMVS method considering the accuracy of the orientation elements. Specifically, the accuracy of the exterior orientation (EO) elements obtained by SfM is introduced to the patch extension in PMVS.

\* Corresponding author

## 2. ESTIMATING EXTERIOR ORIENTATION ACCURACY THROUGH BUNDLE ADJUSTMENT

The proposed method introduces the EO accuracy into the patch extension in PMVS. This section explains the procedure of estimating the EO accuracies. In this study, global SfM (Moulon et al., 2013; Martinec and Pajdla, 2007), including bundle adjustment, is applied to estimate EO accuracies. The global SfM is robust against cumulative error. In this study, interior orientation elements (intrinsic parameters) are estimated in advance using Zhang's method (Zhang, 2000).

### 2.1 Image Matching

Initially, image matching performed for the image pairs from the set of images, and the image coordinates of the feature points that are common to the image pairs are determined. Scale invariant feature transform (SIFT) (Lowe, 2004) is adopted as the image matching method. Even if SIFT is applied for feature point extraction and matching, incorrect matching still exist. The feature points matching is refined using random sample consensus (RANSAC) (Fischler and Bolles, 1981). The RANSAC algorithm employs a method of outlier removal. Finally, the remaining points are accepted as feature points.

### 2.2 Global SfM

Using the image coordinates of each feature point provided by image matching, the EO elements are estimated through relative orientations. The global SfM estimates relative rotations between all images as the essential matrix by using the five-point algorithm (Nistér, 2004) and the epipolar graph. The accuracy of rotation is fairly independent of the baseline length (Enqvist et al., 2011). The relative rotation estimation is followed by applying rotation consistency (Zach et al., 2010).

The translation is then estimated through the trifocal tensor method (Moulon et al., 2013). The trifocal tensor method minimizes the reprojection error of the 3D coordinates of feature points  $X_j$  as compared to the image coordinates  $\{(x_j^i, y_j^i)\}_{i \in \{1,2,3\}}$  in the three images:

$$\rho(t_i, X_j) = \left\| \left( x_j^i - \frac{R_i^1 X_j + t_i^1}{R_i^3 X_j + t_i^3}, y_j^i - \frac{R_i^2 X_j + t_i^2}{R_i^3 X_j + t_i^3} \right) \right\|, \quad (1)$$

$$\min_{t_i, X_j} \gamma, \quad (2)$$

s.t.  $\rho(t_i, X_j) \leq \gamma, R_i^3 X_j + t_i^3 \geq 1, t_i = (0, 0, 0)$

where  $t_i$  is the translation, and  $R_i$  is the rotation of image  $i$ . The relative orientation  $t_{ij}$  between image  $i$  and  $j$  is transformed to the world coordinates  $T_i$  (i.e., translation registration) as follows:

$$\min_{T_i, \lambda_{ij}} \gamma, \quad (3)$$

s.t.  $|T_j - R_j T_i - \lambda_{ij} t_{ij}| \leq \gamma, \lambda_{ij} \geq 1, T_i = (0, 0, 0)$

where  $\lambda_{ij}$  is the scale factor. Finally, the bundle adjustment is applied to all images.

Through the trifocal tensor and bundle adjustment, the translation vector  $\mathbf{t}$ , rotation matrix  $R$ , and 3D coordinates of feature points  $\mathbf{X}$  can be calculated. When a camera position is

translated with  $\mathbf{t}$  and rotated with  $R$  in a fixed world coordinate system, the 3D coordinates  $\mathbf{X} = (X, Y, Z)^T$  are projected to the image coordinates  $\mathbf{x} = (x, y)^T$ . Let the intrinsic parameter matrix of the camera be  $K$ . The geometric relationship between the 3D coordinates and the image coordinates are expressed by using the homogeneous coordinates  $\tilde{\mathbf{X}} = (X \ Y \ Z \ 1)^T$  and  $\tilde{\mathbf{x}} = (x \ y \ 1)^T$ .

$$\tilde{\mathbf{x}} \approx KR^T(I | -\mathbf{t})\tilde{\mathbf{X}}. \quad (4)$$

### 2.3 Estimating EO Accuracy

The EO accuracy is calculated according to the 3D coordinates of the feature points and EO elements. Prior to calculating the accuracy, the bundle adjustment with ground control points is applied again.

The accuracy of the EO elements is defined by the variances of the estimated parameters in the bundle adjustment. Let  $\mathbf{x}^*$  be the estimated parameter vector. The covariance matrix  $Cov(\mathbf{x}^*)$  can be expressed as the inverse of the Fisher information matrix  $F(\mathbf{x}^*)$ :

$$Cov(\mathbf{x}^*) = F^{-1}(\mathbf{x}^*) = \sigma^2 (J^T(\mathbf{x}^*)J(\mathbf{x}^*))^{-1}, \quad (5)$$

where  $\sigma^2$  = variance of observation error  
 $J(\mathbf{x}^*)$  = Jacobian matrix of parameters

3D models obtained via SfM possess gauge ambiguity. Due to this gauge ambiguity, the Jacobian matrix of the parameters becomes rank deficient. To avoid this rank deficiency, QR decomposition is applied, and the ground control points are fixed.

The translation accuracy is defined as the variance of translation. The representation matrix of the Rodrigues' rotation formula is adopted for calculating rotation accuracy. These accuracies are introduced to the PMVS.

## 3. PATCHE-BASED MVS CONSIDERING EXTERIOR ORIENTATION ACCURACY

### 3.1 Basic Models

The proposed method is based on PMVS (Furukawa et al., 2010). The PMVS employs a patch model, a photometric discrepancy function, patch optimization, and an image model.

The patch model approximates the surface of the object by using small rectangular patches, namely local tangent plane approximation. The patch model  $p$  includes the coordinates of patch centre  $c(p)$  and normal unit vector  $n(p)$ . The normal unit vector is in the same direction as that of the corresponding camera.

Let  $V(p)$  be the image ( $I$ ) set, which contains the patch  $p$ . The photometric discrepancy score  $g(p)$  defines the dissimilarity between the images in  $V(p)$ .

$$g(p) = \frac{1}{|V(p) \setminus R(p)|} \sum_{I \in V(p) \setminus R(p)} h(p, I, R(p)) \quad (6)$$

where  $R(p)$  = reference images  
 $h(p, I, R(p))$  = dissimilarity between  $I$  and  $R(p)$

The dissimilarity score  $h$  is computed using the normalized cross correlation (NCC), and it is defined as 1-NCC. When the value of  $h$  is higher than a threshold, the patch is regarded as the wrong image and discarded.

The patches are optimized by minimizing the sum of the photometric discrepancy scores regarding  $c(p)$  and  $n(p)$ :

$$\min_p g(p) \quad (7)$$

The conjugated gradient method is normally used for this optimization.

For efficient computation, the reprojected patch locations on the image are saved (image model), and the image  $i$  is divided into image cells  $C_i(x,y)$ . When the reprojected patch location is in the cell, it is associated with the patch set  $Q_i(x,y)$ .

### 3.2 PMVS Algorithm

The PMVS algorithm mainly consists of feature matching, patch expansion, and patch filtering. The patches are created based on the photometric discrepancy function, as mentioned in previous sections, and then they are expanded to obtain a dense 3D reconstruction.

**3.2.1 Feature matching:** Feature matching creates a sparse point cloud with patches. For feature matching, the difference-of-Gaussian filter and the Harris operator are applied, as in a previous study (Furukawa et al., 2010). The epipolar line is also used for an efficient search. Through the feature matching process,  $c(p)$  and  $n(p)$  of the patches can be determined.

**3.2.2 Patch expansion:** Patch expansion creates new patches adjacent to the previous sparse patches. For a sparse patch  $p$ , the neighbouring image cells  $C(p)$  in all images are defined according to the following condition:

$$C(p) = \{C_i(x',y') \mid p \in Q(x,y), |x-x'| + |y-y'| = 1\} \quad (8)$$

If the neighbouring image cell does not contain patches, a new patch  $p'$  is created. The direction of the normal vector  $n(p')$  is the same as that of the light ray through the corresponding image cell centre. The patch centre  $c(p')$  is determined as the intersection between the line including the normal vector  $n(p')$  and the plane of the neighbouring patch  $p$ .  $c(p')$  and  $n(p')$  are optimized by minimizing the sum of the photometric discrepancy scores, similar to that described in the previous section. During optimization,  $c(p')$  is constrained such that the projected point of  $c(p')$  on the image cell does not change. According to the new patches,  $V(p)$  and  $Q_i(x,y)$  are also updated.

**3.2.3 Patch filtering:** Filters are utilized to eliminate the incorrect patches. The first filter is based on visibility consistency. To apply visibility consistency, the neighbouring patches are defined as follows:

$$\left| (c(p) - c(p') * n(p)) \right| + \left| (c(p) - c(p') * n(p')) \right| < 2\rho_1 \quad (9)$$

When the patches  $p$  and  $p'$  are included in same image cell but do not satisfy equation (9), the two patches are defined as the inconsistency set  $U(p)$ . Additionally, if the following condition is satisfied, the patch  $p'$  is removed as an outlier.

$$|V(p)|(1-g(p)) < \sum_{p_i \in U(p)} 1-g(p_i) \quad (10)$$

The second filter is also related to visibility consistency. When the number of the image set  $|V(p)|$  is less than a threshold value, the corresponding patch is removed.

### 3.3 Incorporating EO Element Accuracies and PMVS

The accuracy of the EO element is introduced to the patch optimization of PMVS. The EO element accuracy is converted to a weighting function for the sum of the photometric discrepancy scores (Figure 1):

$$g(p) = \frac{1}{|V(p) \cap R(p)|} \sum_{I \in V(p) \cap R(p)} w_{IR(p)} * h(p, I, R(p)) \quad (11)$$

where  $w_{ij}$  is the weighting function for image  $i$  and  $j$ .

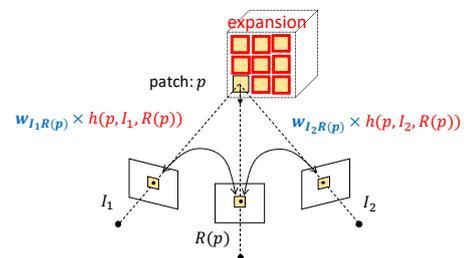


Figure 1. Weighting sum of photometric discrepancy scores for patch optimization

When the accuracy of an image is high, the weight also becomes high. This indicates that the contributions of the image with high accuracy to the goal function is higher. Let  $(\sigma_\phi^2, \sigma_\theta^2, \sigma_\kappa^2)$  and  $(\sigma_x^2, \sigma_y^2, \sigma_z^2)$  be variances of each rotation and translation, respectively. The EO accuracies of rotation  $\sigma_R^2$  and translation  $\sigma_t^2$  are defined as follows:

$$\sigma_R^2 = \sigma_\phi^2 + \sigma_\theta^2 + \sigma_\kappa^2, \quad (12)$$

$$\sigma_t^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2. \quad (13)$$

The suffix  $i, j, avg$  are added for the image  $i, j$ , and average of all images. We also investigate the types of weighting functions:

$$w_{ij}^1 = \alpha \frac{2 * \sigma_{R\_avg}^2}{\sigma_{R\_i}^2 + \sigma_{R\_j}^2} + \beta \frac{2 * \sigma_{t\_avg}^2}{\sigma_{t\_i}^2 + \sigma_{t\_j}^2}, \quad (14)$$

$$w_{ij}^2 = \alpha \frac{(1/\sigma_{R\_i}^2) + (1/\sigma_{R\_j}^2)}{2 * (1/\sigma_{R\_avg}^2)} + \beta \frac{(1/\sigma_{t\_i}^2) + (1/\sigma_{t\_j}^2)}{2 * (1/\sigma_{t\_avg}^2)}, \quad (15)$$

$$w_{ij}^3 = \alpha \frac{\sigma_{R\_avg}^2}{\sigma_{R\_i}^2} + \beta \frac{\sigma_{t\_avg}^2}{\sigma_{t\_i}^2}, \quad (16)$$

$$w_{ij}^4 = \alpha \frac{(1/\sigma_{R\_i}^2)}{(1/\sigma_{R\_avg}^2)} + \beta \frac{(1/\sigma_{t\_i}^2)}{(1/\sigma_{t\_avg}^2)}, \quad (17)$$

where  $\alpha$  and  $\beta$  are the balancing parameters, and  $\alpha + \beta = 1$ .  $w_{ij}^1$  and  $w_{ij}^2$  are simple ratio of accuracy and inverse of accuracy, respectively. Furthermore,  $w_{ij}^3$  and  $w_{ij}^4$  only consider the accuracy of image  $i$ . Image  $j$  is regarded as a reference image. The accuracy of the reference image can be considered to be constant; thus, it can be ignored. As a result, the proposed method can consider geometric conditions during patch optimization.

#### 4. EXPERIMENTAL RESULTS

The proposed method was applied to the following three types of datasets.

##### 4.1 Experiment 1: Images Taken at Ground Level

Images were captured using an SLR camera from the ground, and the laser scanner data was prepared for evaluation (Figures 2, 3). Table 1 presents the specifications of the camera. The accuracy of the laser scanner data is 5 mm. A total of 20 points were picked from the laser scanner data up to be used as ground control points (12 points) and test points (8 points).



Figure 2. Image taken at ground level



Figure 3. Laser scanner data for evaluation

Table 1. Camera specifications

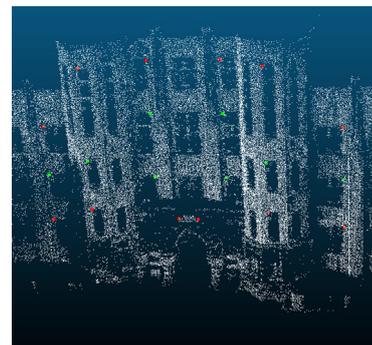
Camera	CANON Eos 60D
Number of pixels	18 M [pix]
Focal length	18 [mm]
Image size	3456 × 2304 [pix]
Ground resolution	10 [mm/pix]
Number of images	18
Overlap	95 [%]

The accuracy of the test points was verified, and the 3D points acquired by using the proposed method and the laser scanner were compared. Table 2 and Figure 4 depict the results of the proposed method.

The results showed that the accuracy was improved in a majority of the cases. Through a detailed comparison between the results of the proposed method and the laser scanner data, the higher effectiveness of the proposed method was confirmed at the flat parts and the edge parts of the building. However, a lower accuracy was observed at the patterned and indented surface.

Table 2. Accuracy evaluation of experiment 1

Weights	$(\alpha, \beta)$	Horizontal [mm]	Vertical [mm]
No weight		12.86	20.95
$w_{ij}^1$	(1.0, 0.0)	13.57	21.15
	(0.5, 0.5)	14.63	23.15
	(0.0, 1.0)	12.82	20.42
$w_{ij}^2$	(1.0, 0.0)	14.53	21.67
	(0.5, 0.5)	13.94	22.97
	(0.0, 1.0)	13.37	20.08
$w_{ij}^3$	(1.0, 0.0)	12.74	19.88
	(0.5, 0.5)	13.17	20.38
	(0.0, 1.0)	12.95	20.25
$w_{ij}^4$	(1.0, 0.0)	13.42	20.70
	(0.5, 0.5)	14.19	21.44
	(0.0, 1.0)	12.41	18.94



(a) SfM



(b) PMVS

Figure 4. Result of images taken at ground level

##### 4.2 Experiment 2: UAV Images

Images were captured from a UAV equipped with an SLR camera (Figure 5). Table 3 lists the specifications of the camera. A total of 58 points were surveyed as ground control points (13 points) and test points (45 points).



Figure 5. UAV image

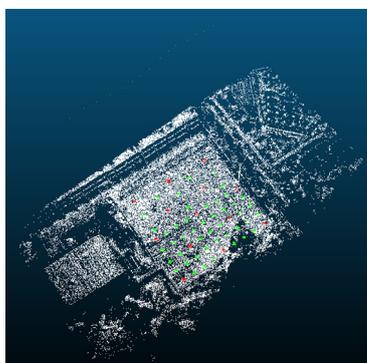
Table 3. Camera specifications of UAV

Camera	CANON Eos Kiss X7
Number of pixels	18 M [pix]
Focal length	20 [mm]
Image size	4608 x 3456 [pix]
Ground resolution	22 [mm/pix]
Number of images	15
Overlap	95.6 [%]

The results of the second experiment showed that the proposed method achieved sub-pixel accuracy (as shown in Table 4 and Figure 6). However, the results of the proposed method were slightly improved as compared to those of the original PMVS method. In this case, vertical images were used; thus, the accuracy of EO elements was approximately the same for all images.

Table 4. Accuracy evaluation of experiment 1

Weights	$(\alpha, \beta)$	Horizontal [mm]	Vertical [mm]
No weight		6.38	23.07
$w_{ij}^1$	(1.0, 0.0)	6.50	23.91
	(0.5, 0.5)	6.47	23.78
	(0.0, 1.0)	6.64	23.19
$w_{ij}^2$	(1.0, 0.0)	6.40	23.45
	(0.5, 0.5)	6.54	24.41
	(0.0, 1.0)	6.45	24.16
$w_{ij}^3$	(1.0, 0.0)	6.43	23.10
	(0.5, 0.5)	6.30	22.80
	(0.0, 1.0)	6.48	22.95
$w_{ij}^4$	(1.0, 0.0)	6.40	24.57
	(0.5, 0.5)	6.47	24.69
	(0.0, 1.0)	6.54	24.42



(a) SfM



(b) PMVS

Figure 6. Result of UAV images

In the first and second experiments, we also investigated the types of weighting functions. The weighting function  $w_{ij}^3$  that was based on the ratio of the variance of the target image and average variance was more effective than others. However, the weighting function  $w_{ij}^2$  based on the harmonic average of variance exhibited better results at the flat section of the object.

### 4.3 Experiment 3: Images Captured Using an Oblique Camera

In the third experiment, images were captured using an airplane equipped with an oblique camera (Figure 7). Table 5 lists the specifications of the camera.

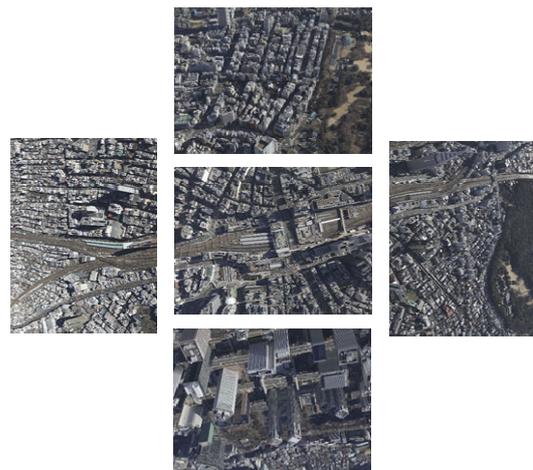


Figure 7. Image captured via oblique camera

Table 5. Oblique camera specifications

Camera	IGI Penta-DigiCam
Number of pixels	60 M [pix]
Focal length	50 (nadir), 82 (oblique) [mm]
Image size	8956 x 6708 [pix]
Ground resolution	100 [mm/pix]
Number of images	100–200
Overlap	75 [%]

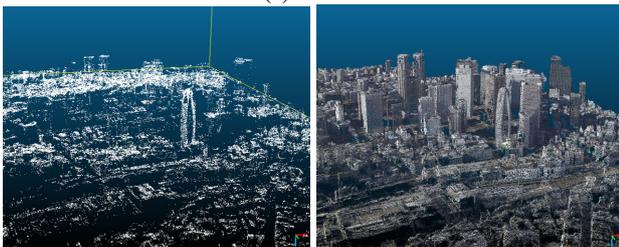
The third experiment demonstrated the applicability of the proposed method to large data. The proposed method was applied to images captured using an oblique camera. The computation was stable, and significantly denser results were obtained as compared with those obtained using SfM. Figure 8 presents the results of the proposed method.

## 5. CONCLUSIONS

In this study, a multi-view stereo considering the accuracy of exterior orientation elements was developed. The proposed method is based on global SfM and patch-based multi-view stereo (PMVS). The accuracy of the EO elements obtained by global SfM is incorporated in the patch optimization of PMVS. This accuracy is defined as the variances of the EO elements, namely the geometric conditions that can be considered in the PMVS. Through three types of experiments, the accuracy improvement of the patches and their applicability to enormous images were confirmed.



(a) PMVS



(b) SfM

(c) PMVS

Figure 8. Results of oblique camera images

Future studies should focus on the application of the method to short baseline image pairs as the PMVS computation of these pairs is unstable. The proposed method is expected to be applicable to various types of images such as shared images on the Internet. In such cases, the interior orientation elements should be introduced. To address the additional elements, the definition of the weighting function should be investigated. Furthermore, the relationships between the efficiency and accuracy should be analysed. These endeavours will lead to a further improved applicability of the PMVS.

#### ACKNOWLEDGEMENTS

This research was funded by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (B) 18H01554. The UAV images and oblique camera images used in this study were provided by Professor Hirofumi Chikatsu (Tokyo Denki University) and Asia Air Survey Co.,Ltd., respectively. We would like to express our gratitude for their cooperation.

#### REFERENCES

Bethmann, F. and Luhmann, T., 2014. Object-based Multi-Image Semi-Global Matching - Concept and first results. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5, 93-100.

Enqvist, O., Kahl, F. and Olsson, C., 2011. Non-sequential structure from motion. *IEEE International Conference on Computer Vision Workshops*.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp.381-395.

Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362-1376.

Furukawa, Y., Curless, B., Seitz, S.M. and Szeliski, R., 2010. Towards internet-scale multi-view stereo. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1434-1441.

Fuse, T. and Harada, R., 2016. Development of image selection method using graph cuts. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 641-646.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-342.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.

Martinec, D. and Pajdla, T., 2007. Robust rotation and translation estimation in multiview reconstruction. *2007 IEEE Conference on Computer Vision and Pattern Recognition*.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6), 756-770.

Moulon, P., Monasse, P. and Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. *Proceedings of the IEEE International Conference on Computer Vision*.

Remondino, F., Del Pizzo, S., Kersten, T.P. and Troisi, S., 2012. Low-cost and open-source solutions for automated image orientation - a critical overview. *Lecture Notes in Computer Science*, 7616, 40-54.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F. and Nex, F., 2014. State of the art in high density image matching. *Photogrammetric Record*, 29(146), 144-166.

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 519-528.

Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3), 7-42.

Szeliski, R., 2011. *Computer Vision - Algorithms and Applications*. Springer, Heidelberg, Germany.

Zach, C., Klopschitz, M. and Pollefeys, M., 2010. Disambiguating visual relations using loop constraints. *Proceedings of Computer Vision and Pattern Recognition*, 2.

Zhang, Z., 2010. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.