

OBJECT-BASED SEMI-GLOBAL MATCHING FOR MULTI-VIEW HIGH RESOLUTION SATELLITE IMAGES

Tee-Ann Teo¹ and I-Ling Kuo²

¹Associate Professor, Dept. of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan 30010,
Email: tateo@mail.nctu.edu.tw

²Master Student, Dept. of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan 30010,
Email: lilian810319@gmail.com

KEY WORDS: Object-based Semi-Global Matching (OSGM), Multi-view satellite images, WorldView-2

ABSTRACT: Multi-view WorldView-2 images are capable for producing high spatial resolution surface model. Semi-Global Matching (SGM) is considered to be a powerful stereo matching algorithm to dense points from stereo images. However, the limitation is that traditional SGM is usually applied for frame images and it is a pairwise image matching algorithm. In order to deal with these limitations, an image-based SGM is extended to Object-based Semi-Global Matching (OSGM) for multi-view push-broom satellite images. This study develops an OSGM for multi-view WorldView-2 images in 3-D point clouds generation. The idea of OSGM is to divide the object space into dense voxels. Then, each voxel is back-projected to different images to obtain the gray values. The cost function of OSGM is calculated in the object space rather than the epipolar image space. Therefore, the cost function considers all images simultaneously. The proposed scheme considers all multi-view images to improve the reliability of image matching. The test area is located in Taipei city, Taiwan. Three images are acquired by a WorldView-2 satellite. First, the impact of OSGM parameters is examined. Then, the impact of different similarity indices is analyzed. Finally, the point clouds derived from OSGM are compared with lidar data in accuracy analysis.

1. INTRODUCTION

To compare aerial and satellite images, the standard scene of satellite image provides larger coverage than the aerial image. The development of optical satellite sensors focuses on spatial, spectral and temporal resolutions. For example, very high resolution satellite WorldView-2 provides both 0.5m panchromatic image and 2.0m multi-spectral (i.e. coastal, blue, green, yellow, red, red edge, nir1, and nir2) image. Moreover, most of the optical satellites are capable to acquire in-track multi-view images by body rotation. Therefore, the multi-view high resolution satellite images are capable for producing high spatial resolution surface model for large area.

Several studies (Haala et al., 2010, Laberl, et al. 2010) indicated that the multi-view aerial images are capable for producing high quality surface model by dense matching algorithm. Dense matching (i.e. SGM) is a kind of full-size matching and determines the conjugate points for every pixel in image space. Therefore, multi-view images are adopted to improve the correctness in dense matching. In order to determine reliable results from multi-view images, the multi-view image matching methods includes image-based (Wenzel et al., 2013) approach and object-based (Paparoditis et al., 2000) approach. Wenzel et al., (2013) performed multiple pair-wise images matching from multi-view images. Then, the conjugate points from all the image-pair were combined in space intersection. Paparoditis et al., (2000) generated a large number of voxels in object space, then, each voxel was back-projected to all images to determine the correlation between all images. The idea of image-based method is to divide the multi-view images into many stereo images while the object-based method matches all the images at the same time.

The image matching technique can be classified into three categories: local method (Scharstein et al., 2002), global method (Pierrot-Deseilligny and Paparoditis, 2006) and semi-global matching (SGM) (Hirschmuller, 2008). The local method considered the similarity in a certain size of window; the low texture or low contrast area usually caused ambiguity. The idea of global method is to determine the optimal solution by minimization of global cost function. This method usually requires highly computational resource and computational time for large coverage. To improve the computational performance of global matching, the semi-global matching was proposed to aggregate the cost functions in 8 directions only. Besides, Rothermel et al., (2012) proposed a tSGM method by image pyramid to reduce the computational time. SGM is considered to be a powerful stereo matching algorithm to dense points from stereo images. However, the limitation is that traditional SGM is usually applied for frame images and it is a pair-wise image matching algorithm in image space. In order to deal with these limitations, an image-based SGM is extended to Object-based Semi-Global Matching (OSGM) for multi-view push broom satellite images.

As the epipolar geometric for push-broom linear array (e.g. WorldView-2) is different from projective frame camera, pseudo epipolar image is usually adopted for push-broom scanner. The proposed method overcomes the problem of epipolar generation for image-SGM for WorldView-2 image (Kuo and Teo, 2015). This study develops an OSGM for WorldView-2 images in 3-D point clouds generation. The idea of OSGM is to divide the object space into dense voxels. Then, each voxel is back-projected to different images to obtain the gray values. The cost function of OSGM is calculated in the object space rather than the epipolar image space. Therefore, the cost function considers all images simultaneously.

2. METHODOLOGY

The proposed method develops an object-based SGM for push-broom scanner. There are four major steps in OSGM: (1) voxelization in object space, (2) back-projection, (3) determination of cost function, and (4) cost aggregation. The following section introduces the details of OSGM.

2.1 Voxelization in Object Space

First, we use the metadata of satellite image to define the boundaries (i.e., E_{min} , E_{max} , N_{min} , N_{max} , H_{min} , and H_{max}) in object space. The grid size (i.e., S_{gsd}) of voxels is the ground sampling resolution of image. For example, the grid size of WorldView-2 panchromatic is defined as 0.5m. Then, we use these predefined parameters to divide the object space into many voxels. For example, the voxels 1, 2 and 3 in Figure 1.

2.2 Image Back-Projection

The second step is to assign the attribute of voxel by image back-projection. The geometrical sensor model in our study is rational function model (RFM). We use rational polynomial coefficients (RPCs) from image metadata and ground control points (GCPs) from large scale topographic map to build up the precise transformation (Teo, 2011) from object space (E, N, H) to satellite image space (S, L). Figure 1 illustrates the image back-projection from voxels to image space. Every image has its own voxels. The attribute of voxel is interpolated from the corresponding image pixel values. After the voxel-by-voxel back-projection and resampling, all the voxels are filled in image pixel values in object space.

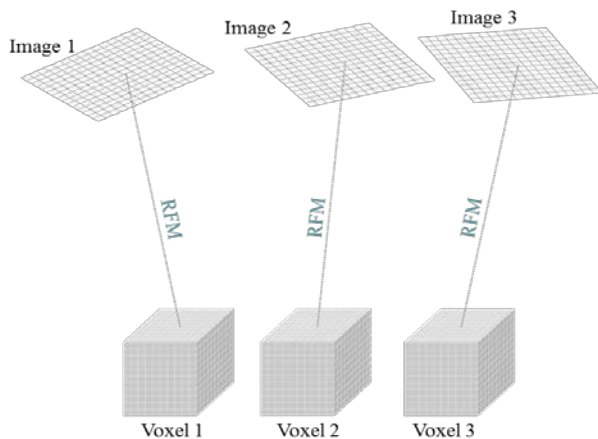


Figure 1. Image back-projection

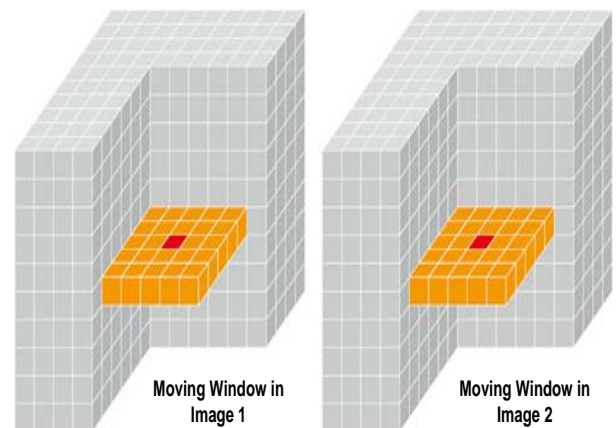


Figure 2. Moving voxels for similarity index

2.3 Determination of Cost Function

The third step is to calculate the similarity indices between voxel from different images. The similarity indices in our study include: (1) absolute difference (AD), (2) census (Humenberger et al., 2009), and (3) normalized cross correlation (NCC) (Richards, and Jia, 2013). The AD simply calculates the absolute difference between two windows; Census uses the differences between two windows to calculate the percentage of same value; NCC calculates the correlation between two windows. The similarity indices are calculated between two voxels at the same

location (see Figure 2) in a predefined window size. The moving window is not only moving on E and N directions, but also moving on H direction. To compare the image content of moving window in original image space and object space, the advantage of using object space is to avoid non-uniform pixel size and different orientation in similarity measurement. The similarity can be AD, Census or NCC. Equation 1 shows the similarity indices ρ between image i and image j . The cost function $C(E,N,H)$ in equation 2 is the average for similarity indices from different combinations. For example, the combinations of 3 images are images 1 and 2, images 1 and 3, images 2 and 3. So, the number of combination n is 3.

$$\rho_{i,j}(E,N,H) = \text{similarity}(\text{Voxel}_{imgi}, \text{Voxel}_{imgj}), i \neq j \quad (1)$$

$$C(E,N,H) = \frac{-\sum \rho_{i,j}(E,N,H)}{n} \quad (2)$$

2.4 Cost Aggregation

The last step is to determine the optimal solution by energy function. The idea of OSGM is to determine the minimum energy by aggregate the cost function from different directions. This method uses aggregations in 8 directions to approximate the global minimum energy. The main difference between image-based SGM and OSGM is that OSGM find the minimum energy in object space (see Figure 3) rather than the epipolar image space. Besides, image-based SGM uses disparity to compute object coordinates while OSGM obtains the object coordinates directly. The energy function of OSGM is shown as equation 3. The $E(Z(X,Y))$ is an energy function aggregated from 8 directions. The cost aggregation function in direction r is $L_r(p, Z(X,Y))$ (see equation 4). P_1 and P_2 are constant penalties. If the distance between two voxels is smaller than ΔZ , then, P_1 is used, otherwise, P_2 is applied.

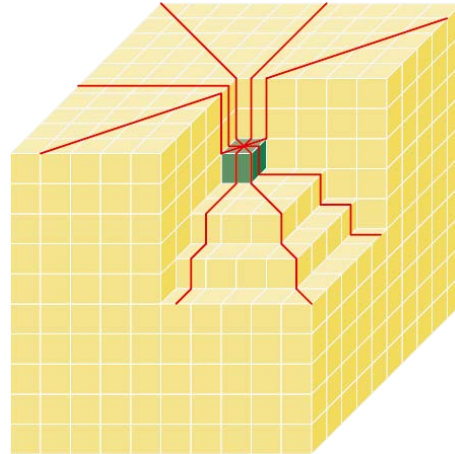


Figure 3. Cost aggregation from 8 directions

$$E(Z(X,Y)) = \sum_p C(p, Z_p(X,Y)) + \sum_{q=N_p} P_1 T[|Z_p(X,Y) - Z_q(X,Y)| = \Delta Z] + \sum_{q=N_p} P_2 T[|Z_p(X,Y) - Z_q(X,Y)| > \Delta Z] \quad (3)$$

$$L_r(p, Z(X,Y)) = C(p, Z_p(X,Y)) + \min\{L(p-r, Z(X,Y) - \Delta Z), L(p-r, Z(X,Y)) + P_1, L(p-r, Z(X,Y) + \Delta Z) + P_1, \min(L(p-r, Z(X,Y) \pm i * \Delta Z) + P_2)\}, \quad (4)$$

$i > 1$

Where,

$C(p, Z(X,Y))$ is a cost (similarity index) of an object point p at $Z(X,Y)$;

$L_r(p, Z(X,Y))$ is cost aggregation of an object point p in direction r ;

P_1 and P_2 are constant penalties;

r is direction of cost aggregation from 1 to 8;
 ΔZ is grid size in z direction.

3. EXPERIMENTAL RESULTS

3.1 Test Data

The multi-view WorldView-2 images (Figure 4) are located in Taipei City and the test size is about 600m by 400m. The production level is Level 1B and the ground sampling distance is 0.5m. The image acquisition time was Feb. 5, 2010. The mean satellite elevation angles for these three images are 81.4, 67.6, and 79.4 degrees. Besides, the mean satellite azimuths for these three images are 106.1, 174.1 and 147.9 degrees. All the view directions is within 90~180 degrees and there is no opposite direction. We combine elevation angles and azimuth angles to calculate the convergence angle between the images. The convergence angles for images 1 and 2, images 1 and 3, images 2 and 3 are 20.59, 7.07 and 13.58 degrees, respectively. The maximum base-to-height ratio is about 0.35. Since the base-to-height ratio is smaller than 0.35, these multi-view images are challenge data set for surface matching. In order to do the verification, we use 1m DSM from airborne lidar (Figures 4d and 4e) as reference data. This lidar data was also acquired in year 2010. The result of OSGM is shown as Figure 4f. The spacing of DSM from OSGM is 0.5m.

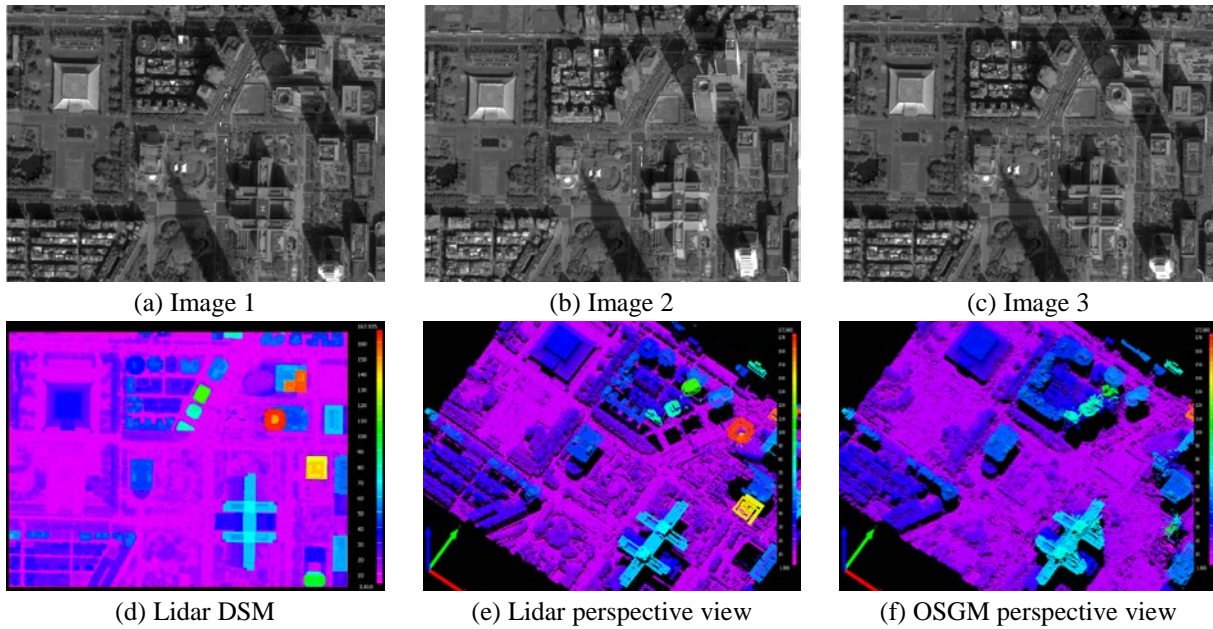


Figure 4. Test images and Lidar DSM

Table 1 shows the accuracy of geometric correction using RFM. We use more than 30 control point to ensure the geometrical consistence between the images. These three images use the same check points to evaluate the accuracy of image back projection. The root-mean-square-errors (RMSE) for independent check points are less than 1 pixel. This result indicated that the image back-projection reached subpixel accuracy and it is suitable for multiple image matching.

The input data for this study is multi-view WorldView-2 images. The output result is dense point in object space from OSGM. The evaluations were carried out in three aspects. First, this study analyzed the impact of different parameters, such as window size, P1 and P2 penalties. Therefore, the importance of parameters can be analyzed. Second, the performance of different similarity indices for cost function was discussed. Third, this study compares the accuracy of OSGM and traditional local matching approach.

Table 1. Accuracies of image back projection

	Control Points			Check Points		
	Number of Point	RMSE_Sample (pixel)	RMSE_Line (pixel)	Number of Point	RMSE_Sample (pixel)	RMSE_Line (pixel)
Image 1	34	0.61	0.48	22	0.14	0.16
Image 2	44	0.47	0.62	25	0.21	0.10
Image 3	51	0.58	0.47	25	0.17	0.26

3.2 Parameter Selection

The major parameters for OSGM are window size, P1 and P2. Table 2 summaries the description and the test values in this study. For window size, we use 5 x 5 to 11 x 11. It is equivalent to 2.5m x 2.5m to 5.5m x 5.5m. The larger window size will produce smoothness surface and the smaller window size will produce rough surface. For P1 and P2, these two penalties are smoothness parameters. The impact of P1 is for the neighboring (i.e. Z-1 and Z+1) voxels only. Hence, the impact of P2 is more significant than P1. We use lidar DSM as reference data. The evaluation index is the standard deviation of delta Z between lidar and OSGM DSM. The similarity index in used is NCC.

Table 2. Parameter selection

Parameters	Description	Values
Window size	Window size for similarity measurement	5x5, 7x7, 9x9, 11x11
P1	Smoothness parameter for neighboring voxels. In this study, the similarity indices are within -1 to 1, therefore, this parameter is less than 1.	0.1, 0.3, 0.5
P2	Smoothness parameter for non-neighboring voxels. As the larger distance between voxels is more penalized than neighboring voxels, P2 is usually larger than P1.	0.3,0.6,1.2, 1.8,2.4, 3.0

Figure 5a compares the standard error for different P1 and different window sizes. In this case study, the larger window size (i.e. 11 x 11) has lower accuracy. Therefore, we use smaller window size (i.e. 5 x 5) in this study. To compare the P1 from 0.1 to 0.5, the larger P1 has lower accuracy. As the similarity index in this study is NCC, the numerical values are ranged from -1 to 1. The P1=0.5 is a significant value when compare to NCC. Therefore, we use smaller P1 (i.e. 0.1) in this study. The test values for P2 are ranged from 0.3 to 3.0. The experimental result (Figure 5b) clearly presented that the value of P2 is inverse proportion to accuracy in all window sizes. Therefore, we use larger P2 (i.e. 3.0) in this study.

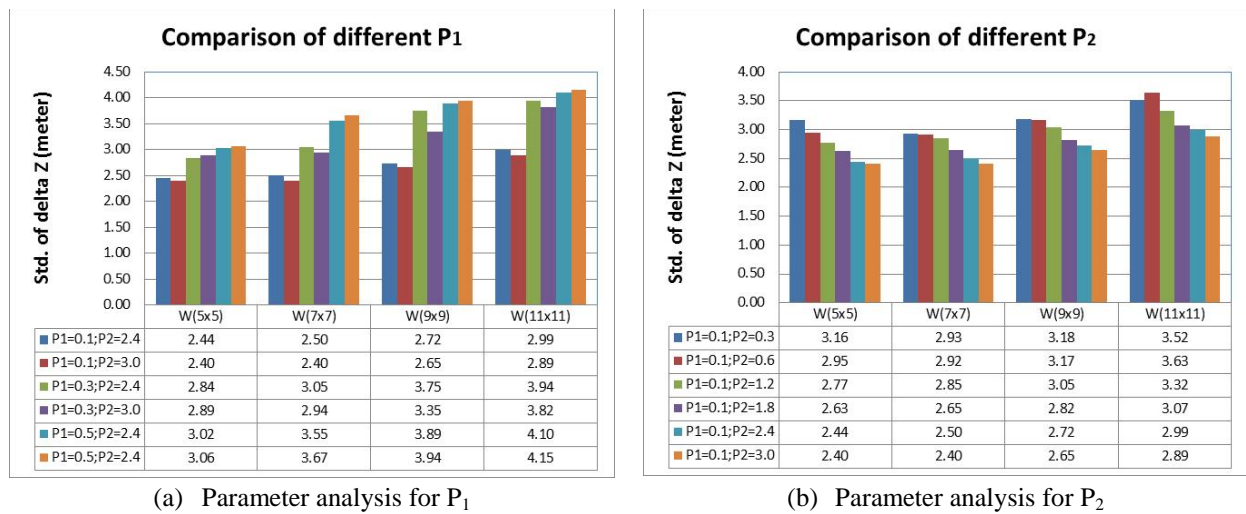
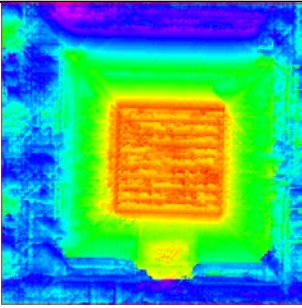
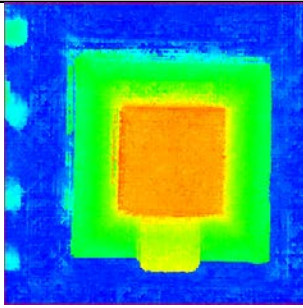
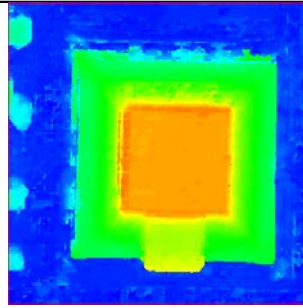
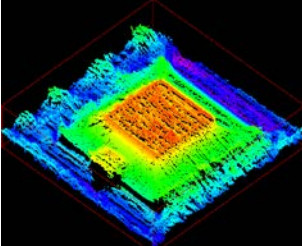
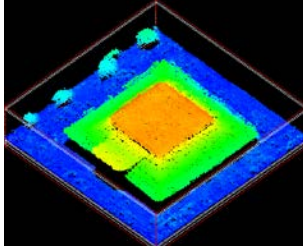
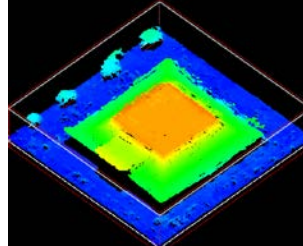
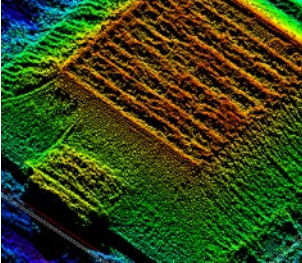
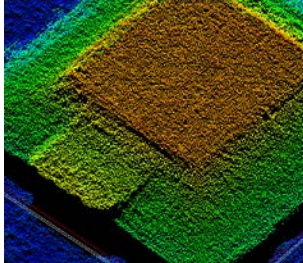
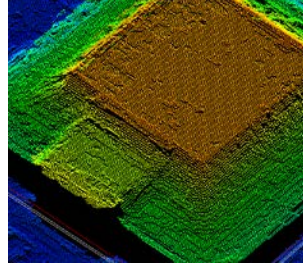


Figure 5. Results of parameter selection.

3.3 Comparison of Different Cost Functions

Once the parameters have been determined, we use the parameters from previous section. In order to understand the effect of different cost functions in OSGM, this study used standard deviation of delta Z between lidar and OSGM DSM in evaluation. We compare three similarity indices: AD, Census, and NCC. All these method use the same parameters for OSGM. Table 3 compares the results from these three indices. AD is faster than the others. However, this method has higher noise than the others. For Census, the numerical result is similar to NCC, but it has larger local variance than NCC in visual inspection. Overall, NCC has better accuracy than the other methods. Therefore, we suggest using NCC in OSGM.

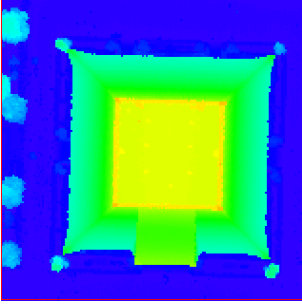
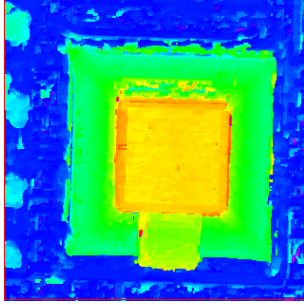
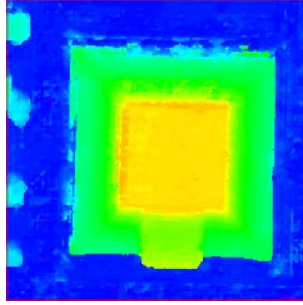
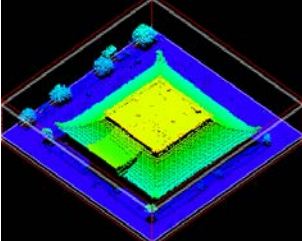
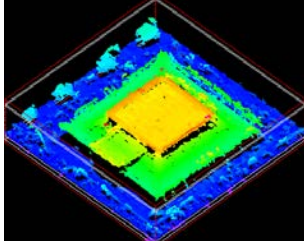
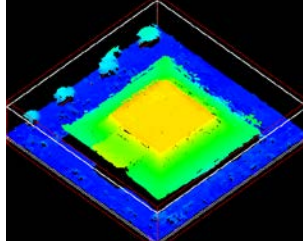
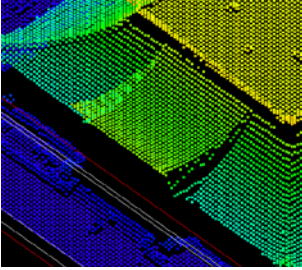
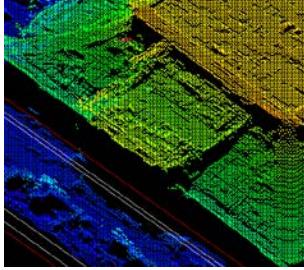
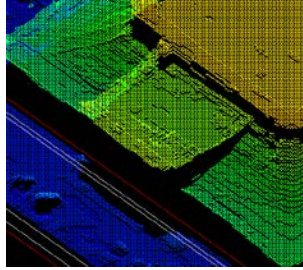
Table 3. Comparison of different cost functions

	AD	Census	NCC
Top-view			
Perspective-view			
Zoom-in			
Accuracy: Mean \pm Std. (meter)	3.66 \pm 4.22	2.17 \pm 2.54	2.22 \pm 2.40

3.4 Comparison of OSGM and Local Matching

As mentioned in section 1, image matching includes local matching, global matching and semi-global matching. This section compares the traditional local matching approach and proposed OSGM. The traditional approach only considers the highest similarity index as final object coordinates while OSGM use the cost aggregation to determine the semi-global optimal object coordinates. Table 4 compares lidar, results of local approach and OSGM. The local approach has higher random noise than OSGM. Besides, the standard error of local approach is also larger than OSGM, the difference reached 0.91m. The experimental result shows that the proposed OSGM is more reliable than traditional local approach.

Table 4. Comparison of OSGM and local matching

	Lidar	Local Matching (NCC)	OSGM (NCC)
Top-view			
Perspective-view			
Zoom-in			
Accuracy: Mean \pm Std. (meter)		2.60 \pm 3.31	2.22 \pm 2.40

4. CONCLUSIONS AND FUTURE WORKS

This study proposed an object-based semi-global matching (OSGM) for multi-view high resolution satellite images. The impacts of parameters were analyzed, and different cost functions were compared in this study. Moreover, the reliability of traditional local approach and proposed OSGM were also discussed. The major contributions of this study are to (1) establish an object-based semi-global matching scheme for push-broom linear scanner; (2) analyze the impact for different parameters and different methods. The proposed scheme integrated multi-view image to determine the optimal object coordinate by energy function, subsequently improving the accuracy of dense matching. The conclusions are summarized as follows:

- (1) Comparing the result of OSGM and lidar DSM, the standard deviation of delta z is about 2.5m. The accuracy is referring to the base-to-height ratio (i.e. 0.35) of test images.
- (2) By analyzing different parameters for OSGM, the suggested window size is about 2.5m by 2.5m. The suggested constant penalties for P1 and P2 are 0.1 and 3.0 in this case study.
- (3) Comparing the traditional local matching and OSGM approach, the overall improvement is about 0.9m. Moreover, the result of local matching has higher noise than the proposed OSGM.

The limitations of this study are summarized as follows; future works will focus on these limitations:

- (1) The proposed method is not able to extract vertical façade. This method only determines an object height in vertical direction.
- (2) The multispectral high resolution images were not contributed to image matching scheme. Additional spectral features were adopted to improve the results of image matching.

ACKNOWLEDGEMENTS

The authors would like to thank DigitalGlobe for providing the test data sets.

REFERENCES

- 1 Haala, N., Hastedt, H., Wolff, K., Ressel, C., and Baltrusch, S., 2010. Digital photogrammetric camera evaluation – generation of digital elevation models, *Photogrammetrie - Fernerkundung – Geoinformation (PFG)*, 2:99-115.
- 2 Hirschmuller, H. 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.
- 3 Humenberger, M., Engelke, T., and Kubinger, W. 2010. A census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 77-84.
- 4 Kuo, I.L. and Teo, T.A., 2015. Dense Matching for Worldview-3 Multispectral Stereo Images, *Proceedings of the 36th Asian Conference on Remote Sensing*, Oct. 19-23, Quezon City, Metro Manila, Philippines, USB Flash drive.
- 5 Leberl, F., Irschara, A., Pock, T., Gruber, M., Scholz, S., and Wiechert, A., 2010. Point clouds: lidar versus 3D vision. *Photogrammetric Engineering and Remote Sensing*, 76, 1123-1134.
- 6 Pierrot-Deseilligny, M., and Paparoditis, N. 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(1/W41), Available at http://www.isprs.org/proceedings/xxxvi/1-w41/makaleler/Pierrot_multiresolution_matching.pdf
- 7 Richards, J.A., and Jia, X., 2013. *Remote Sensing Digital Image Analysis: An Introduction*, 5th edition, Springer, 494 pages.
- 8 Rothermel, M., Wenzel, K., Fritsch, D., and Haala, N. 2012. SURE: Photogrammetric surface reconstruction from imagery. In *Proceedings LC3D Workshop, Berlin (Vol. 8)*, Available at http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf
- 9 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.
- 10 Teo, T.A., 2011, Bias Compensation in a Rigorous Sensor Model and Rational Function Model, *Photogrammetric Engineering and Remote Sensing*, 77(12), 1211-1220.
- 11 Wenzel, K., Rothermel, M., Fritsch, D., and Haala, N. 2013. Image acquisition and model selection for multi-view stereo. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 251-258.