

POINT CLOUD AND DIGITAL SURFACE MODEL GENERATION FROM HIGH RESOLUTION MULTIPLE VIEW STEREO SATELLITE IMAGERY

K. Gong, D. Fritsch

Institute for Photogrammetry, University Stuttgart, 70174 Stuttgart, Germany - (Ke.gong, dieter.fritsch@ifp.uni-stuttgart.de)

Commission II, WG II/2

KEY WORDS: Multiple View Stereo, Satellite Imagery, Rational Polynomial Coefficients, Semi-Global Matching

ABSTRACT:

Nowadays, multiple-view stereo satellite imagery has become a valuable data source for digital surface model generation and 3D reconstruction. In 2016, a well-organized multiple view stereo publicly benchmark for commercial satellite imagery has been released by the John Hopkins University Applied Physics Laboratory, USA. This benchmark motivates us to explore the method that can generate accurate digital surface models from a large number of high resolution satellite images. In this paper, we propose a pipeline for processing the benchmark data to digital surface models. As a pre-procedure, we filter all the possible image pairs according to the incidence angle and capture date. With the selected image pairs, the relative bias-compensated model is applied for relative orientation. After the epipolar image pairs' generation, dense image matching and triangulation, the 3D point clouds and DSMs are acquired. The DSMs are aligned to a quasi-ground plane by the relative bias-compensated model. We apply the median filter to generate the fused point cloud and DSM. By comparing with the reference LiDAR DSM, the accuracy, the completeness and the robustness are evaluated. The results show, that the point cloud reconstructs the surface with small structures and the fused DSM generated by our pipeline is accurate and robust.

1. INTRODUCTION

High resolution satellite (HRS) sensors have made tremendous development over the last decade. More HRS satellites, like Sentinel-2, WorldView-3, Pléiades and so on, are launched by private companies or space agencies. These Earth observation satellites can cover most areas of our planet. Moreover, along with the launch of WorldView-3 satellite, the ground sample distance (GSD) of HRS panchromatic imagery has formally stepped into 30cm level. The better GSD largely enriches the surface features contained in the HRS optical imagery. The HRS satellites fly over a certain site frequently, which provides a large number of image collections and makes the multiple view stereo (MVS) satellite imagery acquisition feasible. Although the terrain surface might change over seasons, the MVS satellite images captured on different dates is a convenient data source of large scale surface reconstruction. Because of the large coverage and sub-meter GSD, MVS high resolution satellite images are valuable for global 3D mapping, environment monitoring, urban planning, terrain change detection and so on.

Recently, a publicly MVS benchmark for commercial satellite imagery has been released by the John Hopkins University Applied Physics Laboratory (JHUAPL), USA (Bosch et al., 2016). The benchmark contains fifty WorldView-3 panchromatic and multispectral images. The benchmark imagery covers a 100 square kilometres area close to San Fernando, Argentina. The GSD of the dataset is approximately 30cm for nadir images. All those images are captured from November 2014 to January 2016. Most images are collected on different dates. LiDAR point cloud and digital surface model (DSM) at 30cm GSD are also provided by the benchmark as the ground truth. This well-organized MVS high resolution satellite benchmark motivates us to learn and test the methods of the point cloud and DSM generation from MVS satellite data.

In this paper, we propose a pipeline based on the pair-wise multi-view method to process the MVS high resolution satellite

imagery. The point clouds and DSMs, which are generated from different preselected pairs, are fused to obtain the final DSM product. Related work is presented in section 2. Section 3 introduces the methodology of the proposed pipeline. Section 4 demonstrates the results generated from a test area of the benchmark data and some analysis. Finally, we draw some conclusions in section 5.

2. RELATED WORK

As known, the MVS imagery reconstruction methods can be classified into two categories. The first category solves the multi-view triangulation problem for all the images simultaneously (Furukawa and Hernandez, 2015). The second category uses binocular stereos. It processes the stereo pairs separately and then fuses the output point clouds and DSMs to a final result (Haala, 2013). The true multi-view method is more rigorous but more complicated. Some researchers have investigated and compared both 3D reconstruction methods on satellite MVS images (Ozcanli et al., 2015). In their implementation, the pair-wise multi-view reconstruction method generates better results than the other method. Because of the efficiency and stable performance of the semi-global matching (SGM) algorithm (Hirschmüller, 2008), most researchers implement the binocular stereo method for satellite MVS datasets (d'Angelo and Kuschik, 2012; Kuschik, 2013; Qin, 2017; Facciolo et al., 2017). In this paper, we also apply the binocular stereo method for MVS satellite imagery reconstruction.

As we mentioned before, the MVS satellite images are collected on different dates. The difference of satellite's geometric configuration, atmosphere condition, illumination situation can cause negative effects on the image matching. The surface modification and seasonal change also affect the final DSM fusion. Therefore, a critical preparation procedure for DSM generation of the MVS satellite imagery is the image selection. P. d'Angelo et al (2014) suggest that image pairs with intersection angle between 15 to 25 degrees can achieve good results. In order to learn the factors of image selection, G.

Facciolo et al (2017) computed the DSMs of all possible pairs and evaluated the accuracy. According to their observation, the temporal proximity, maximum incidence angle and the intersection angle are three main factors that affect more on the dense image matching side. They suggest the selected images should be acquired on nearby dates with incidence angles less than 40 degrees, and the intersection angle should be between 5 and 45 degrees. Qin (2017) also agreed that the intersection angle play a big role in the quality of the generated DSMs. He chose the image pairs with intersection angles from 10 to 30 degrees. Although the MVS satellite imagery may contains hundreds or even thousands possible stereo pairs, only some well selected pairs will finally be used to generate the fused DSM.

In the standard MVS reconstruction work flow, structure from motion (SFM) or camera model orientation is the essential first step. Unlike frame cameras, the HRS sensors usually provide 80 Rational Polynomial Coefficients (RPCs) along with the images to the users, instead of the exterior and interior parameters. The RPCs build pure mathematic relations between the image coordinates and the object coordinates as a ratio of two polynomials, although they have no physical meanings. Many researches have verified that the RPC model can replace the physical model and maintain the accuracy (Grodecki and Dial, 2001; Hanley and Fraser, 2001; Fraser, et al., 2002). Grodecki and Dial (2003) proposed the bias-compensated RPCs bundle block adjustment. Many researchers applied this RPC based bundle block adjustment for the orientation (d'Angelo and Kuschik, 2012; Ozcanlil et al., 2015; Gong and Fritsch, 2016). The bundle block adjustment needs sufficient tie points and ground control points (GCP). Because the GCPs are not always available, relative orientation is a better solution for the MVS satellite images. Franchis et al. (2014) point out that the RPC model's error is presented as the shift between corresponding point and epipolar line in image domain. This shift is named as the relative pointing error and it is measured as a simple translation if the image size is small. Their relative orientation method removes the relative pointing error of the small image by the median of the translations. Qin (2017) applied pair-wise bias-compensation using tie points and conducted least squares minimization for the registration of the generated DSM and the reference DSM. The parameters of the registration are reused to calculate a translation in image space for the RPCs refinement. Gong and Fritsch (2017) proposed the relative bias-compensated model without GCPs. Several tie points are detected first. The virtual ground control information are generated with tie points and uncorrected RPCs. The RPCs are refined by additional affine models calculated according to the quasi ground control points and tie points. We applied the relative bias-compensated model in this paper.

Differing from the traditional frame images, HRS images are hard to generate the epipolar geometry because of the changing perspective center and the attitude. As Kim (2000) explained in his work, the epipolar curves of satellite pushbroom sensor are more like hyperbola curves than straight lines, and the epipolar pairs only exist locally. The projection-trajectory method supported by the RPCs is commonly used to find the corresponding epipolar pairs (Wang, et al., 2010). With the satellite epipolar pairs, the semi-global matching (SGM) like algorithm is the most popular solution for the dense image matching problem. Many researches have proved that SGM can provide good quality dense point clouds and DSMs from satellite data (d'Angelo and Reinartz, 2011; Wohlfeil, et al., 2012; d'Angelo and Kuschik, 2012; Gong and Fritsch, 2016).

For the binocular stereo reconstruction method, the DSM fusion is the last but critical procedure. Usually, the DSMs of each satellite images pair are put into a regular spaced and discretized grid in a UTM coordinate system. Kuschik (2013) selected the simple median filter to get the final height of every cell. Qin (2017) proposed an adaptive depth fusion method that considers the spatial consistency. Instead of assigning the median value of the height of an individual cell, they define a window centered at this cell. All the cells within the window are candidates for the height values filtering. Facciolo et al (2017) propose a clustering-based method. The height of each cell is estimated by the k-medians clustering with increasing number of clusters until the clusters are close enough to the predefined precision. We applied the classical median filter in our pipeline.

3. METHODOLOGY OF THE PROPOSED PIPELINE

The proposed pipeline matches the selected satellite image pairs separately and then fuses all the generated DSMs to the final result. Generally, our pipeline is divided to four steps: image selection, relative orientation and image rectification, dense image matching and Triangulation and DSM fusion. The pipeline is implemented by our self-programed C++ codes and the workflow is presented in Figure 1.

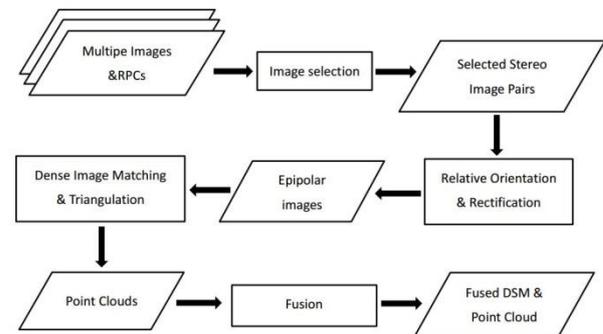


Figure 1. Workflow of the point cloud and DSM generation from the MVS satellite imagery

3.1 Image Selection

As we have learned from former research work (d'Angelo et al., 2014; Facciolo et al., 2017; Qin, 2017), the image capture date, the incidence angle and the intersection angle of the image pairs' views are the three main factors that could affect the quality of the generated DSM.

We start our investigation from the satellite images' capture date. First, all the images are sorted in chronological order. As we have already learned, the closer the image capture dates are, the better results we can obtain. Actually, according to our observation, there are two exceptions:

1. The interval between the capture dates are near, but the images are collected in different seasons. For instance, there are two images captured on October 07, 2015 and October 22, 2015 in the benchmark dataset. The capture dates are relatively close, but the dense image matching has a lot of failure matches. Because there are apparently season changes from the first image to the second image. The images and generated DSM is shown in Figure 2.
2. The interval between the capture dates are far, but the images are captured in the same season. For instance, two images captured on November 14, 2014 and December 18,

2015, has acceptable performance on dense image matching. If the terrain surface does not change a lot in the close years, the image captured in the same season will be matched well. The images and generated DSM is shown in Figure 3.

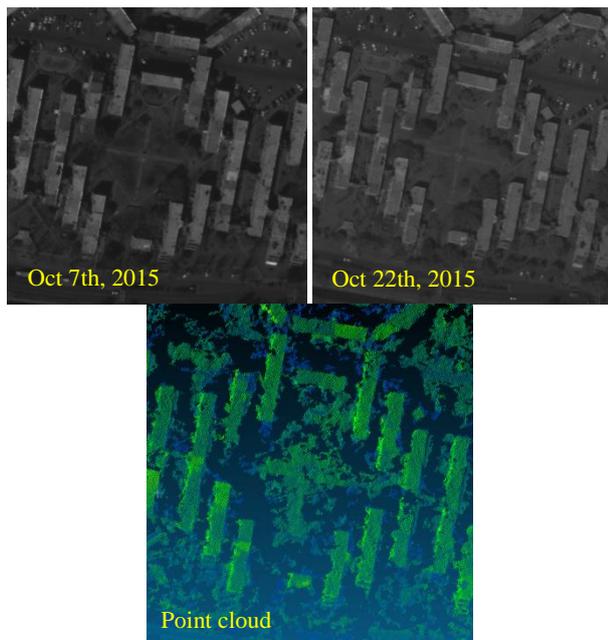


Figure 2. Images and point cloud from different season but same year

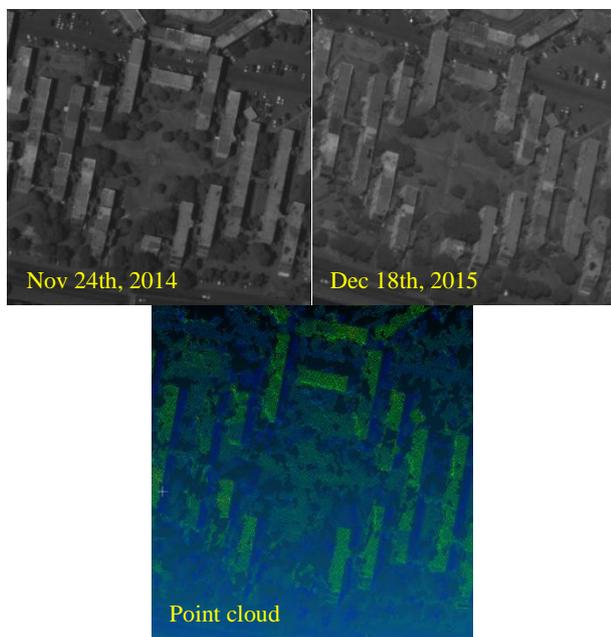


Figure 3. Images and point cloud from same season but different year

Next, we check the incidence angle of the satellite images. The spatial resolution of the image is lower when the incidence angle is larger. We found that the images, which have an incidence angle larger than 35 degree, have bad performance in our dense image matching procedure.

At last, we test the influence caused by the intersection angles of different image views. According to our experiments, the image pairs which have intersection angles larger than 40 degrees provide little help to our final results. Follow some former researches (d'Angelo et al., 2014; Facciolo et al., 2017), the image pairs are less useful if the intersection angle is less than 5 degrees. In our experiments, we also select the image pairs have the intersection angles larger than 5 degrees as our input images.

According to our experiments, we apply the image selection strategy like this: First, we eliminate the images that have an incidence angle larger than 35 degrees. The rest images are divided to two seasons – winter and summer group. Inside the groups, we ignore the year of data collection and order the images by month. The satellite images that are captured in the same month are marked to compose our selected input image pairs. At last, we kept the image pairs have the intersection angle between 5 and 40 degrees.

3.2 Relative Orientation and Image Rectification

The bias-compensated RPCs bundle block adjustment can easily solve the orientation problem of satellite stereo images (Grodecki and Dial, 2003; Fraser and Hanley, 2005). The biggest limitation of this method is the requirement of GCPs. Unfortunately, the ground control information is not always available for satellite datasets. We apply a relative bias-compensated model for the relative orientation of MVS satellite images.

First we select some tie points in all the input images. A pair of same-date stereo imagery is selected to generate the virtual ground surface with the unrefined RPCs. By using the image coordinates of the tie points, we trace the ray from the image point to the virtual ground surface and find the corresponding object coordinates. Then, these generated object points are used as the virtual ground control points to do the bias-compensated bundle block adjustment for all the input images. The virtual ground surface has a 3D translation to the true ground. It is used as a reference for our relative orientation, so the adjustment removes the relative but not absolute bias for different images. All the DSMs generated from different stereo images are aligned to the virtual ground surface. So no point cloud or DSM registration is needed before the DSM fusion. As well as the absolute bias-compensated bundle block adjustment, the relative model also estimates an additional affine model to compensate the relative bias for each image, and at least 4 to 6 ground control points are needed (Fraser and Hanley, 2005). In our previous research (Gong and Fritsch, 2017), the accuracy of the relative bias-compensated model with virtual ground control information can reach sub-pixel level.

We generate the epipolar images for each input stereo pair. The projection-trajectory method is used to search the corresponding epipolar curves. Our work applied the modified piecewise epipolar resampling strategy to generate the epipolar images. The epipolar curves are approximated as multiple segments. We resample the epipolar images along those epipolar segments from bottom to top and align the epipolar segment pair of stereo image to the same row. The details are referred to (Gong and Fritsch, 2017).

3.3 Dense Image Matching and Triangulation

To get very dense point clouds, our pipeline applies a modified Semi-Global Matching (SGM) method tSGM (Rothemel, et al.,

2012). The tSGM method is implemented in the C++ library libTsgm, which is also the core algorithm of the DIM software SURE. The usage of the library is authorized by nFrames GmbH. The tSGM algorithm selects the 9×7 Census cost instead of the Mutual Information. The Census cost is insensitive to parametrization and provides robust results (Zabih and Woodfill, 1994). The tSGM algorithm implements a hierarchical coarse-to-fine method to limit the disparity search ranges. When the matching of the higher resolution pyramid is started, the results of lower resolution pyramids are introduced as the priors to determine the disparity search ranges (Rothermel, et al., 2012). The tSGM algorithm reduces the computing time and optimizes the memory efficiency.

The tSGM method generates a disparity map for each stereo pair. The corresponding pixels on the stereo images are derived according to the disparity map. The dense point clouds are generated by forward intersection with the corresponding pixels.

3.4 DSM Fusion

As explained in section 3.2, no further registration is needed for the point clouds, because they are already aligned on the virtual ground truth surface. The point clouds are projected into a regular spaced and discretized grid in UTM coordinate system. In our implementation, a simple median filter is applied for the DSM fusion. The median height of each cell is selected as the final height of the fused DSM.

4. EXPERIMENTS

We choose a subset region from the MVS satellite benchmark for testing. The details of the test region and the reference data are presented in section 4.1. The result of the proposed pipeline and the analysis are demonstrated in 4.2.

4.1 Test Site and Evaluation Method

The MVS satellite imagery benchmark provides fifty WorldView-3 panchromatic images. The spatial resolution is about 30cm. These images are collected from November 2014 to January 2016. The distribution of the data collection date covers every month. The subset images extracted from the MVS satellite images are as big as 3000×3000 pixels. The test site is close to Fernando, Argentina and the area is ca. 1 km^2 . It contains different terrain types like fields, residential buildings and vegetation.



Figure 4. The reference LiDAR DSM

The benchmark dataset also provides LiDAR data as the ground truth. The LiDAR data is collected in June, 2016. The nominal point spacing is 20cm. The reference DSM at 30cm GSD is generated from the LiDAR data. The reference DSM of the test site is shown in Figure 4.

4.2 Results and Analysis

Following the pipeline described in section 3, we select the input image pairs first. The images having an incidence angle larger than 40 degrees are excluded. We sort the rest images into summer and winter group. Figure 5 presents the same area with tall trees in the point clouds generated from summer and winter images and the reference LiDAR DSM. According to Figure 5, we learn that the plants in the reference DSM are flourishing. The DSMs derived from the images captured in winter will have big height differences in vegetation areas. So we only use the imagery in summer group as our input images. We ignore the year of the data collection and only concern about the month. All the images captured in the same month are selected to compose stereo pairs. Among these image pairs, that have intersection view angles less than 5 degrees or larger than 40 degrees, are eliminated. Finally, we select 53 stereo pairs from all the possible pairs as our input data.

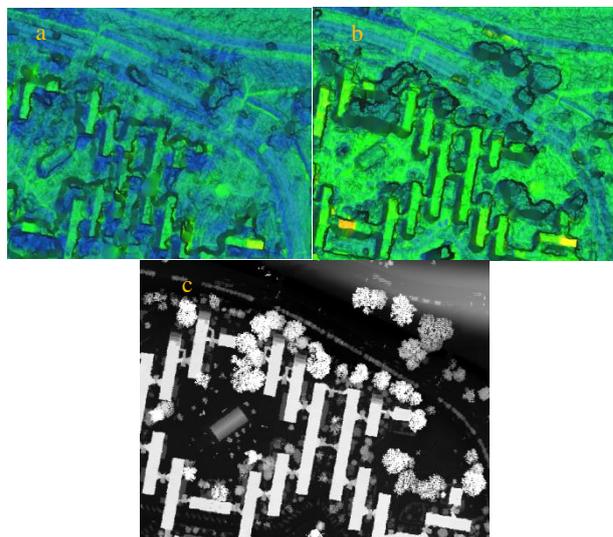


Figure 5. Area with tall trees in: a. Point cloud of winter images b. Point cloud of summer images c. Reference DSM

In order to conduct the relative orientation, we detect 23 tie points in all the selected images. Nine tie points are chosen as the virtual ground control points and the rests are regarded as the check points. The virtual ground control points are distributing evenly in the image space. Two satellite images captured on 18th December, 2015 are selected to produce the virtual ground surface. We trace the ray of the virtual ground control points from one image and find the intersection on the virtual ground surface. In this way, we get the object coordinates of the virtual ground points. Then the relative bias-compensated model is applied for all the selected images.

Epipolar stereo images are generated by our modified piecewise epipolar resampling strategy. By applying the tSGM algorithm, we have the disparity maps of every input stereo pair. The point clouds are derived from the disparity maps, and they are all aligned to the virtual ground surface. Then, we compute the root mean square errors (RMSE) of the height difference between the point cloud and the reference DSM. To be noticed, the point clouds need to be aligned to the reference DSM before the

comparison of the heights. We apply the registration tool provided by the MVS satellite benchmark. The registration tool is implemented by a coarse-to-fine method. In different spatial resolution level, it removes the translational shift by searching through several possible translations and finding the position which can minimize the median error between the point cloud and the ground truth (Bosch et al., 2016). In this way, we minimize the effect of the shift between the point cloud and the ground truth caused by our relative orientation procedure. According to the RMSE of the height difference, all the input image pairs are ranked.

	Completeness (%)	Median error (m)	RMSE (m)
1 image pair	66.26	0.662	2.826
2 image pairs	65.79	0.640	2.628
5 image pairs	71.47	0.491	2.622
10 image pairs	69.67	0.582	2.589
20 image pairs	70.02	0.599	2.690
50 image pairs	67.87	0.635	2.716

Table 1. Evaluation results of the fused point clouds generated from different number of stereo image pairs

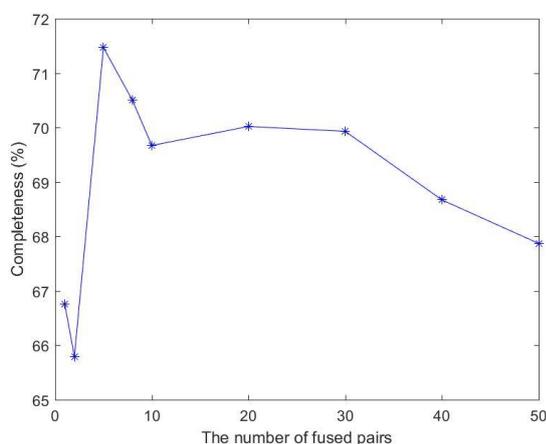


Figure 6. Relation between the completeness and the number of fused stereo image pairs

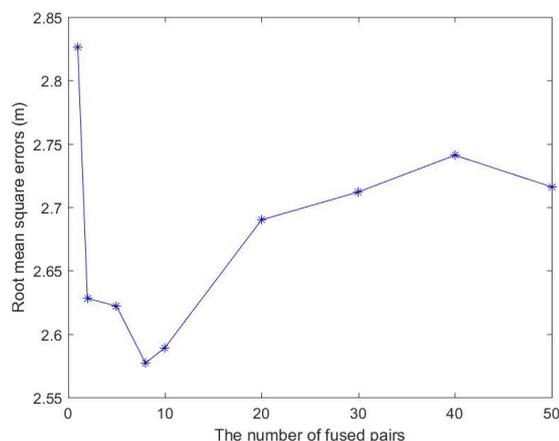


Figure 7. Relation between the completeness and the number of fused stereo image pairs

To investigate the relation between the quality of the final DSM and the image pairs applied for DSM fusion, we generate the fused point clouds from the best one, the best two, the best five, the best ten, the best twenty and the best fifty image pairs according to the image pair rank. Point clouds are fused via a simple median filter. In order to undertake the quantitative analysis of the fused point cloud, we compute the accuracy and the completeness of the height difference to the reference DSMs. The RMSE and the median errors of all the points' height differences are evaluated for the accuracy computation. The completeness is the percentage of the points which have less than 1m height difference to the ground truth. The evaluation result is demonstrated in Table 1. We generate more fused point clouds from different number of stereo image pairs and depict Figure 6 and Figure 7. The relation between the completeness and the number of used image pairs is shown in Figure 6. Figure 7 present the relation between the RMSE and the number of the image pairs.

According to Table 1, we find that the point cloud fused from five image pairs has the highest completeness and the lowest median error. The point cloud fused from the best ten stereo image pair has the lowest RMSE. Observing the relations presented in Figure 6, the completeness is increasing when more stereo image pairs are added to our fusion procedure. The completeness reaches the peak when the numbers used stereo image pairs is five. After that, the completeness becomes worse and worse. We can see from Figure 7, the RMSE is decreasing when the number of fused image pairs is increasing at first. It reaches the lowest error when the number is between five and ten image pairs. Similar to the completeness, the RMSE is decreasing if more image pairs are applied. Therefore, the optimal number of the satellite image pairs applied for point cloud fusion is about five to ten. The accuracy and completeness is decreasing when more stereo image pairs are involved into the fusion procedure, because more errors are introduced by some low quality image pairs. In our work, we select the best eight stereo image pairs to generate the final fused point cloud. We convert the fused point cloud into a discretized and regular spaced grid in the UTM coordinate system to obtain the fused DSM. The fused DSM of the test site, which are generated from the best eight image pairs, is exhibited in Figure 8.



Figure 8. The fused DSM

In order to verify the quality of our fused DSM, we accomplish a comparison of the height differences between the fused DSM and the reference LiDAR DSM. The distribution of the height difference is depicted as histogram presented in Figure 9. The accuracy and completeness of the results are computed. Moreover, we compute the normalized median deviation (NMAD), 68% and 95% quantiles of the absolute errors to evaluate the robustness of the fused DSM. The statistical evaluation result is displayed in Table 2.

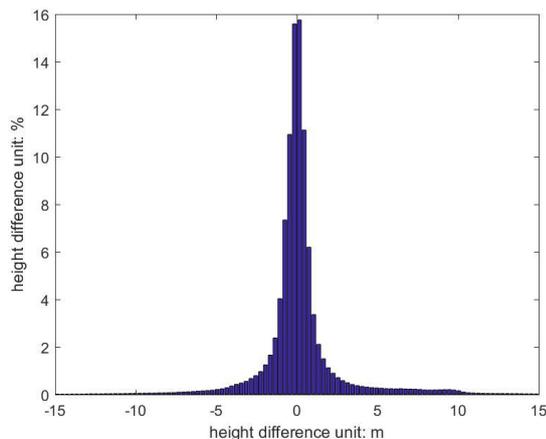


Figure 9. The distribution of the height difference

Completeness (%)	Median error (m)	RMS E (m)	NMAD (m)	Aq68 (m)	Aq95 (m)
70.55	0.520	2.572	0.807	0.943	5.908

Table 2. Evaluation result of the fused DSM

According to Figure 9, most points have small height elevation errors to the ground truth. The 95% quantile of the distribution is about 5.9m, so the points have height differences larger than 6 should be the outliers. As Table 2 shows, over 70% completeness and half meter median error is quite good compared to the results shown in Table 1. The RMSE of the final fused DSM is the lowest among all the fuse images options exhibited in Table 1 and Table 2. Although 2.57m RMSE is still not good enough, considering about the WorldView-3 image's sub-meter GSD.

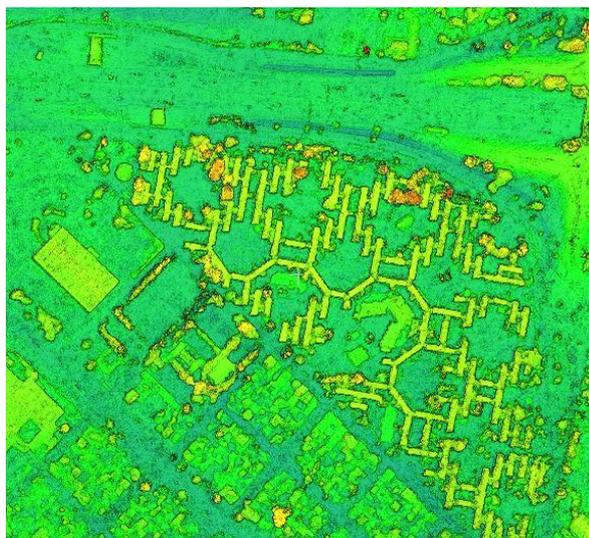


Figure 10. The 3D model of the fused point cloud

The fused point cloud is visualized by the open source software CloudCompare. For better visualization, we generate the 3D mesh of the fused point cloud in CloudCompare, which is displayed in Figure 10. Additionally, several sub-areas are extracted from the 3D model and the reference DSM to analyze the reconstructed details of the fused point cloud. The sub-areas of the fused point cloud and the reference DSM are shown in Figure 11.

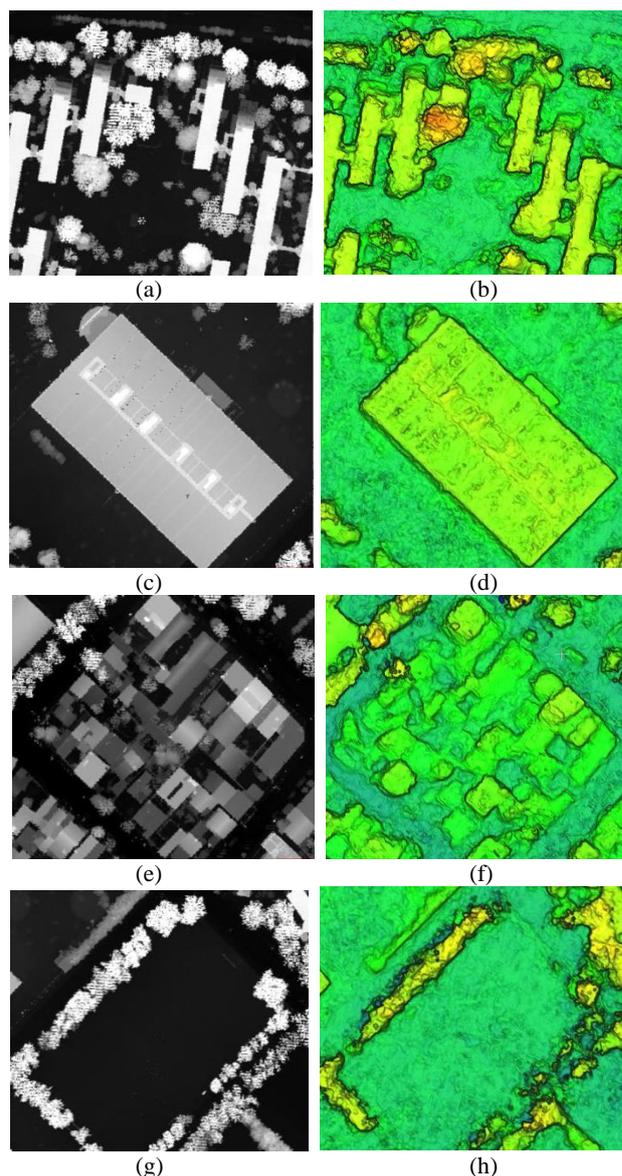


Figure 11. Detail comparison between the reference DSM and the generated 3D model

Figure 11 (a), (c), (e), (g) are the sub-areas extracted from the reference LiDAR DSM, and Figure (b), (d), (f), (h) are the corresponding areas on the 3D model generated from the fused point cloud. Figure 11(a) and (b) display an area which contains buildings and trees. We can observe that most buildings and trees are well reconstructed and easy to distinguish. But the reconstructed buildings and trees will get mixed and cause some errors, if the trees are too close to the buildings. In Figure (c) and (d), we can find an isolated large building in the extracted area. The reconstructed building's edge is sharp in Figure 11 (d). Comparing the generated 3D model to the reference DSM, the building's small structures on the roof of the building are also reconstructed. Figure (e) and (f) presents the residential areas.

In this area, the buildings are low and intensive. The 3D reconstruction performs worst in this kind of areas. Because the buildings are too close, the shadows of the buildings are often cast on the building nearby. It is hard to reconstruct the residential area as separate buildings. The reconstructed buildings are connected to each other and introduce errors of the height. The last extracted area is a football field surrounded by trees, which is displayed in Figure (g) and (h). Generally, the vegetation is reconstructed well in our point cloud. Notice the left corner of the area, there are two rows of trees. The trees are planted close, and the shadows or the trees nearby will cause the mismatch during the dense image matching. Although we can distinguish the two rows from the 3D model, some of the trees are missing if we compare it to the reference DSM. Generally, the fused points reconstruct the terrain surface well, but it has bad performances when the objects on the ground are too close to each other.

5. CONCLUSION AND OUTLOOK

We propose a pipeline for point cloud and DSM generation from the MVS satellite images in this paper. The methods of the pipeline are implemented by self-programmed C++ codes. The experiments are carried on a test site provided by JHUAPL's MVS satellite image benchmark. Considering the season change and the incidence angles of the view, we propose a strategy of image selection. Those images having large incidence angles, large or too small intersection angles and the images are captured in different seasons are eliminated. We apply the relative bias-compensation model for the relative orientation, which align the point clouds to a virtual ground surface. No further point cloud is needed in the fusion step. Following the pipeline, all the point clouds are generated. The influence of the number of the involved stereo image pairs is investigated. In our experiment, the results show that the optimal number of the fused image pairs is from 5 to 10 (for example, 8 pairs in our test). More additional image pairs introduce errors because some image pairs are of low quality. The fused point cloud is generated by applying the median filter on the selected images' point clouds. The fused point cloud reconstructs the terrain surface well, for example some small structures can be restored. But it has bad performance when the objects are close to each other and easy to be affected by the shadows. The fused point cloud is converted into grids in a UTM system and derives the DSM. The completeness of the DSM is over 70%. According to our result, the accuracy of the DSM is generally good. The median error is half meter and the RMSE is 2.5 meter. The RMSE is large compare to the input WorldView-3's GSD. The errors can be introduced by the intensive residential areas. The NMAD of the DSM is 0.8m and the 68% quantile of the height difference distribution is 0.94m, so the result is robust and most points have small difference to the ground truth.

There are still some aspects needed to improve our work. First, the images that collected in winter are abandoned in our pipeline. Actually, the buildings in the winter images stay the same as the summer images. Only vegetation has tremendous differences because of the season change. In the future, we will bring the winter images to our pipeline and verify the performance. Second, we apply the simple and classical median filter for point cloud fusion. Because the MVS satellite images are collected in different dates, there could be better solution than taking the median height values of the cell for the fusion. Except the test site we applied in this paper, more MVS test sites contain different terrain types, are provided by the benchmark. We will test our methods on these MVS stereo images in the future.

ACKNOWLEDGEMENTS

The authors would like to thank John Hopkins University Applied Physics Lab for providing the well-organized MVS satellite imagery benchmark. The authors also would like to thank Dr. Mathias Rothermel for his kindly advices about MVS reconstruction. The first author is funded by the Chinese Scholarship Association, which is gratefully acknowledged.

REFERENCES

- Bosch, M., Kurtz, Z., Hagstrom, S., Brown, M., 2016. A multiple view stereo benchmark for satellite imagery. In Applied Imagery Pattern Recognition Workshop (AIPR), 2016 IEEE, pp. 1-9.
- d'Angelo, P., Kusch, G., 2012. Dense multi-view stereo from satellite imagery. In Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, pp. 6944-6947.
- d'Angelo, P., Reinartz, P., 2011. Semiglobal matching results on the ISPRS stereo matching benchmark. Report on ISPRS Workshop "High-Resolution Earth Imaging for Geospatial Information", Hannover, Stuttgart.
- d'Angelo, P., C. Rossi, C. Minet, M. Eineder, M. Flory and I. Niemyer, 2014. High Resolution 3D Earth Observation Data Analysis for Safeguards Activities. In: Symposium on International Safeguards, pp. 1-8.
- de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J. M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), 49.
- Facciolo, G., De Franchis, C., Meinhardt-Llopis, E., 2017. Automatic 3D Reconstruction from Multi-Date Satellite Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 57-66.
- Fraser, C., Hanley, H., Yamakawa, T., 2002. Three-dimensional geopositioning accuracy of Ikonos imagery. *The Photogrammetric Record*, 17(99), pp. 465-479.
- Fraser, C., Hanley, H., 2005. Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery. *Photogrammetric Engineering & Remote Sensing*, 71(8), pp. 909-915.
- Gong, K., Fritsch, D., 2016. A Detailed Study about Digital Surface Model Generation Using High Resolution Satellite Stereo Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 69.
- Gong, K., Fritsch, D., 2017. Relative Orientation and Modified Piecewise Epipolar Resampling for High Resolution Satellite Images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 579.
- Grodecki, J., Dial, G., 2001. IKONOS geometric accuracy, Proceedings of Joint International Workshop on High Resolution Mapping from Space, Hannover, Germany, pp. 77-86 (CD-ROM).

Haala, N., 2013. The Landscape of Dense Image Matching Algorithms. IN: *Photogrammetric Week'13*, Ed. D. Fritsch, Wichmann/VDE Verlag, Berlin/Offenbach, pp. 271-284.

Hanley, H., Fraser, C., 2001. Geopositioning Accuracy of IKONOS Imagery: Indication from Two Dimensional Transformations. *Photogrammetric Record*, 17(98), pp. 317-329.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 30(2), pp. 328-341.

Kim, T., 2000. A study on the epipolarity of linear pushbroom images. *Photogrammetric Engineering and Remote Sensing*, 62(8), pp. 961-966.

Kuschik, G., 2013. Large scale urban reconstruction from remote sensing imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5, W1.

Ozcanli, O. C., Dong, Y., Mundy, J. L., Webb, H., Hammoud, R., Tom, V., 2015. A comparison of stereo and multiview 3-D reconstruction using cross-sensor satellite imagery. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference, pp. 17-25.

Qin, R., 2017. Automated 3D recovery from very high resolution multi-view satellite images. ASPRS 2017 Annual Conference.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. Sure: Photogrammetric surface reconstruction from imagery. In Proceedings LC3D Workshop, Berlin, pp. 1-9.

Wang, Y., 1999. Automated triangulation of linear scanner imagery. In Joint Workshop of ISPRS WG I/1, I/3 and IV/4 on Sensors and Mapping from Space, pp. 27-30.

Wohlfeil, J., Hirschmüller, H., Piltz, B., Börner, A., Suppa, M., 2012. Fully automated generation of accurate digital surface models with sub-meter resolution from satellite imagery [J]. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci.* 34-B3: pp. 75-80.

Zabih, R., Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. In: J.-O. Eklundh (ed.), *Computer Vision ECCV '94, Lecture Notes in Computer Science*, Vol. 801, Springer Berlin Heidelberg, pp. 151–158.