EVALUATION OF SELECTED COST AGGREGATION METHODS ON HIGH RESOLUTION SATELLITE STEREO IMAGES

N. Tatar [1]*, H. Arefi [1]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran – (n.tatar, hossein.arefi)@ut.ac.ir

Commission III, WG III/1

ABSTRACT:

Dense stereo processing requires a critical step that called cost aggregation or cost optimization. Most of the cost aggregation methods are evaluated on close range stereo images from Middlebury or KITTI datasets. While the effect of cost aggregation on high resolution satellite stereo processing has not yet been sufficiently evaluated. In this paper, three typical cost aggregation methods together with another approach which is a combination of these methods are evaluated on high resolution satellite stereo images and then are compared with LiDAR ground truth. These methods including Semi-Global Matching (SGM), Guided Filtering (GF), iterative GF (IGF), and SGM followed by GF (SGM-GF) with Census and Zero Normalized Cross Correlation (ZNCC) cost functions. Although the Census cost function has a good performance on the border of the objects and low blurring effects, the results of both cost functions, i.e. Census and ZNCC, have same treatment on all stereo methods. Also, in order to make an impartial assessment, for all stereo methods, we do not perform any disparity map refinement. The bad-pixel criteria with an absolute difference height error greater than 2 meters for SGM, GF, IGF, and SGM-GF methods is 36.7%, 34.8%, 33.8%, and 28.6% respectively. Also, the Normalized Median Absolute Difference (NMAD) error for SGM, GF, IGF, and SGM-GF is 1.29, 1.15, 1.06, and 0.94 meters, respectively. Overall, the experimental results on WV III stereo images demonstrate that the SGM method has lower accuracy and SGM-GF method is accurate than other methods.

1. INTRODUCTION

Dense stereo matching is one of the most significant steps for generating 3D point cloud from satellite stereo images. Automatic stereo matching procedure for high resolution aerial and satellite stereo images containing the shadow effects, geometric and radiometric changes increase the complication of stereo matching (Zhang and Gruen, 2006). Although numerous stereo matching algorithms have been developed in digital photogrammetry and computer vision, however, these difficulties have not still been solved completely (Hamzah and Ibrahim, 2016; Remondino et al., 2014; Tippetts et al., 2016).

The common stereo matching methods comprise four main steps: matching cost computation, cost aggregation (or optimization), disparity computation and disparity map refinement (Scharstein and Szeliski, 2002). In the matching cost computation step, a cost function is employed to measure the similarity of matching between two pixels. In general, matching costs contain an ambiguity due to the radiometric and geometric distortions, repetitive patterns, as well as textureless areas (Boykov et al., 2001; Scharstein and Szeliski, 2002). There are exist two strategies available to solve the ambiguity of matching costs through the connectivity of the neighbouring pixels, including local cost aggregation and global cost optimization (Scharstein and Szeliski, 2002).

Usually, the local algorithms aggregate matching costs around a certain area according to the constant weight or radiometric and geometric weights. For local cost aggregation, various methods have been developed such as multiple windows, adaptive support weight, cross-based local stereo matching, and guided filtering (He et al., 2013; Hosni et al., 2013; Yoon and Kweon, 2006; Zhang et al., 2009). Often the weakness of the local matching algorithms is appearing in textureless areas, occlusion,

and radiometric changes. However, the global cost optimization algorithms are formulated as minimizing a global energy or probability function, which includes data and smoothness terms (Boykov et al., 2001). Typically, global stereo matching methods do not perform local cost aggregation and they optimize matching costs directly. Until now, many global stereo matching methods have been proposed in computer vision society, such as Graph Cuts (GC), dynamic programming, and Belief Propagation (Boykov et al., 2001; Pierrot-Deseilligny and Paparoditis, 2006; Scharstein and Szeliski, 2002; Sun et al., 2003). On the other hand, as alternative to the global methods, Semi-Global Matching (SGM) algorithm (Hirschmüller, 2008) is faster than global algorithms and also popular for top view stereo images due to the higher efficiency compared with other stereo matching methods (Alobeid, 2011; Rothermel et al., 2012). However, the SGM or other global stereo matching methods usually require large amounts of temporal memory and computational resources.

Finally, the disparity map is computed using winner takes all (WTA) by selecting the disparity of lowest optimized matching costs or aggregated costs. Although, cost aggregation or optimization reduce the ambiguity of matching costs, in some cases such as large textureless areas, the border of occlusions and shadow areas the disparity map should be refined (Hamzah and Ibrahim, 2016; Tatar et al., 2017).

Cost aggregation methods for stereo matching has been found to be talented in earlier photogrammetric methods but has not yet been sufficiently evaluated and still holds potential for more investigations on satellite stereo images. In this paper, three well known cost aggregation methods are evaluated on high resolution satellite stereo images, including SGM, Guided Filtering (GF), and iterative GF (IGF). In addition, an approach

with combination of SGM and GF methods are proposed to be evaluated with other methods. In the following section, the cost aggregation methods are described briefly and then the result of experiments is presented. The paper will end with the discussion and conclusions.

2. COST AGGREGATION METHODS

In this section, we compare four cost aggregation methods including SGM, GF, IGF, and SGM followed by GF for high resolution satellite stereo processing. However, the cost aggregation methods are distinguished from the methods that use these cost aggregations. These methods are popular in computer vision due to their accuracy, reliability or fastening for runtime. However, the performance of these stereo methods has not been evaluated on high resolution satellite stereo images.

Before presenting the evaluation procedure, it is important to note that the epipolar images are inputs for stereo methods. The epipolar images are resampled along the epipolar geometry to omit the vertical disparity; so that each row of the left image corresponds to the same row at the right image (i.e. parallax y is equal to zero). Unlike the images with perspective geometry, the epipolar geometry of satellite stereo images could not be considered as a straight line. Recently, for linear consideration of epipolar geometry in satellite stereo images, an image tiling strategy has been proposed to produce epipolar images from high resolution satellite stereo image (de Franchis et al., 2014; Tatar and Arefi, 2019). In this paper, the high resolution satellite stereo images are also rectified based on these approaches.

Matching cost computation is a critical step in dense stereo matching. However, we select the Census and Zero Normalized Cross Correlation (ZNCC) with 9×9 kernel size due to their good results on satellite stereo images (Zhu et al., 2011). On the other hand, in order to avoid the unfair comparison there not exist a disparity map refinement step. Also, to estimate the disparity in sub-pixel level, after obtaining minimum aggregated cost, a quadratic curve is fitted to the neighbouring of lowest aggregated costs, and the position of the minimum is obtained as final disparity. The disparity map also requires uniqueness constraints and therefore similar to the original SGM, a left-right check consistency is applied after computing left and right disparity maps. However, the details of each method is described in the following sub sections.

2.1 Semi-Global Matching

Generally, the capability of computed matching costs to find correspondences has an ambiguity due to equal matching cost for neighbouring pixels. So, the aggregation of matching costs from 8 or 16 paths has been proposed in SGM Method (Hirschmüller, 2008). To this end, matching costs are aggregated based on matching costs and their disparities. Equation 1 is used to aggregate the matching costs in each path and then using equation 2 the costs are aggregated in all paths.

$$L_r(p,d) = C(p,d) +$$

$$+ \min \begin{cases} L_r(p-r,d) \\ L_r(p-r,d\pm1) + p_1 - \min_k L_r(p-r,k) \\ \min_i L_r(p-r,i) + p_2 \end{cases} \quad (1)$$

Where; p, d, $L_r(p, d)$, C(p, d), r, and k are location of the interest pixel, disparity value, cost path toward the actual path, pixel-wise matching cost, actual path, and pixels in each path, respectively. moreover, P1 and P2 are two

penalties for penalizing disparity changes between neighboring pixels of one pixel and more than one pixel, respectively.

$$S_{SGM}(p,d) = \sum_{r=1}^{8or16} L_r(p,d) \quad (2)$$

2.2 Guided Filtering (GF)

One of the major cost aggregation methods is guided image filtering (He et al., 2013; Hosni et al., 2013). This method uses image content to preserve the edges and increases the efficiency, however it is considered as a local stereo method. The grayscale (left or right epipolar) image is selected as the guidance for filtering. The kernel of this filter is obtained as follows:

$$W_{p,q} = \frac{1}{N^2} \sum_{q \in W} \left(1 + \frac{(I_p - \bar{I})(I_q - \bar{I})}{\sigma^2 + \varepsilon} \right) \quad (3)$$

where, I_p and I_q represent the image intensity of interest pixel and neighbouring pixel respectively. Also $\bar{I}$, σ, and N are mean, standard deviation, and number of intensities in the support window respectively. The ε is a smoothness parameter (He et al., 2013). The matching costs are aggregated based on equation 5 as follows:

$$S_{GF}(p,d) = \sum_{q \in W} W_{p,q} C(q,d) \quad (4)$$

2.3 Iterative GF (IGF)

It is important to note that the GF cannot completely reduce or remove noises from matching costs. Hence, an iterative GF (IGF) has been proposed to compensate the noises from aggregated costs (Hamzah et al., 2017). This method aggregates matching costs or aggregated costs in an iterative solution. The aggregated costs in k^{th} iteration are obtained as following:

$$S_{iGF}^k(p,d) = \begin{cases} S_{GF}(p,d) & if & k=1 \\ \sum_{q \in W} W_{p,q} S_{iGF}^{k-1}(q,d) & if & k>1 \end{cases} \quad (5)$$

2.4 SGM followed by GF (SGM-GF)

Simply applying GF or IGF cannot truncates the low entropy pixels from matching costs or aggregated costs. However, the SGM is adaptive to the noises due to its nature for selecting minimum costs. The SGM method also couldn't find the correct matching in large low entropy areas. Hence, some post processing on disparity map is performed to solve this problem in the previous methods (Hirschmüller, 2008). On the other hand, performing SGM in an iterative framework requires a lot of time which is not efficient pleasant. In this paper, after solving the matching costs ambiguities by SGM method, the GF is also applied (SGM-GF). This procedure helps to find more precise corresponding pixels and preserve the edges that is very important in photogrammetric mapping. The GF could be implemented in a fast way on GPU and it helps to save runtimes. This procedure is mathematically expressed as follows:

$$S_{SGM-GF}(p,d) = \sum_{q \in W} W_{p,q} S_{SGM}(q,d) \quad (6)$$

3. EXPERIMENTS AND ANALYSIS

This section presents the datasets and the result of the assessment of the above mentioned stereo methods on satellite stereo images. All stereo methods and tests have been coded by

MATLAB and C++, on Z500 Lenovo notebook with five-core 2.6 GHz processor. In particular, the computed matching costs by ZNCC are normalized between 0 to 100. Hence, the same penalties could be used for both Census and ZNCC in SGM method. The small and large penalties for SGM method are selected as 12 and 30, respectively. The parameter of GF is fixed due to their recommendations on original papers (He et al., 2013; Hosni et al., 2013). However, the effect of support windows is analysed.

3.1 Evaluation procedure and dataset

High resolution satellite stereo images from WV III are used in the experiments. The stereo images are selected from a public benchmark dataset for multi-view satellite images (Bosch et al., 2016). The Ground Sampling Distance (GSD) of images is 0.3 meter and they are captured over an urban area in San-Fernando city in Argentina. Most of the images were taken at the different dates, however we select the images with the different looking angles which could be considered as stereo images (cf. Figure 1). The first image captured on 1 September 2015 and the second one captured on 3 October 2015. Also, airborne LiDAR ground truth with GSD of 0.2 m has been prepared for evaluation of stereo methods.

After stereo matching, the corresponding points are projected into UTM Zone 21 South in order to compare the results with LiDAR Digital Surface Model (DSM). Therefore, due to the existence of rational polynomial coefficients (RPCs) for satellite stereo images, the corresponding points are intersected by RPCs (Grodecki and Dial, 2003) and the point cloud is generated accordingly.

The generated point cloud by RPCs has a small 3D shift from the ground truth due to the error of the independent RPCs for each satellite image (Afsharnia et al., 2017). To compensate this error and subsequently perform an accurate assessment, the generated point cloud is aligned with ground truth. In this paper a simple 3D alignment strategy is employed similar to previous methods (Facciolo et al., 2017). The 3D alignment strategy uses the ZNCC cost function to measure the similarity between DSM and the generated point cloud. To estimate a robust 3D shift, the outliers are removed using a dual rank filter (Zhang et al., 2003).

The results are evaluated by measuring the Euclidian distance between generated point cloud and ground truth DSM. As described in section 2, the disparity map is generated without the disparity map refinement step, therefore there may be several outliers in final point cloud. Hence, the Normalized Median Absolute Difference (NMAD) (Höhle and Höhle, 2009) is selected as a robust metric. The NMAD is computed as following:

$$NMAD = 1.4826 \, median\left(\left|\Delta Z - median(\Delta Z)\right|\right) \qquad (7)$$

Another criterion for evaluation of results is the percentage of bad-pixels with an absolute difference height error greater than 2 meters. This criterion is computed using following equation:

$$Bad-pixels = \frac{\sum T\left[\left|\Delta Z\right| > \delta_h\right]}{N} \qquad (9)$$

where, N is the number of all corresponding points, δ_h is a threshold that we select it 2 meters, and T is a logical operator which is 1 if its argument is true and 0 otherwise.

3.2 Experimental results

In the pre-processing step, the stereo images are resampled along the epipolar geometry. Due to the small size of the selected stereo image tiles, the epipolar images are generated using a fundamental matrix (Tatar and Arefi, 2019). The evaluation of epipolar resampling show that the epipolar images are generated with sub-pixel accuracy. The generated epipolar images are shown in Figure 1.
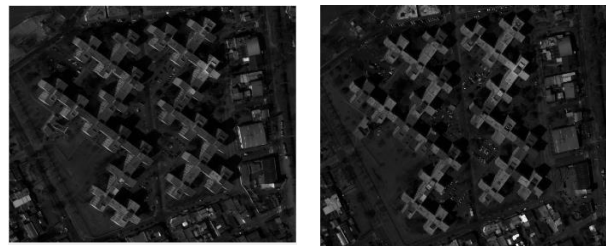

Figure 1: The epipolar image from WV III images. The left image captured on 1 September 2015 and the right image captured on 3 October 2015.

Before comparing the cost aggregation, we should select the optimum support window size for GF method. For this purpose, the point cloud is generated and subsequently, the NMAD and Bad-pixels are computed for all support windows with the range of 11×11 to 31×31. The result of this procedure for both ZNCC and Census is illustrated in Figure 2. Both NMAD and Bad-pixels criteria show that the optimum support window for *GF* method is 31×31. Therefore, we select this window size for all comparisons in the following. Figure 3 shows the result of GF method by Census cost function with different support windows.
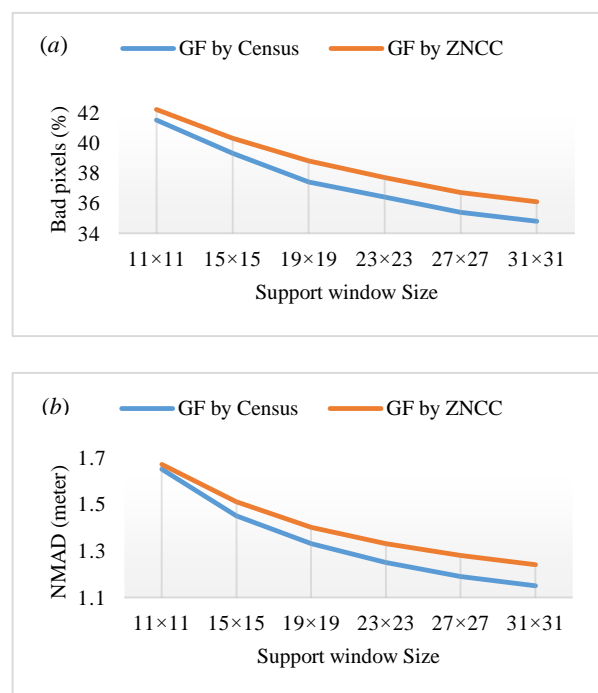

Figure 2: Evaluation of *GF* method by selecting different support windows. The vertical axis of the figure (a) denotes the percentage of Bad-pixels and the figure (b) represents the metric of NMAD.
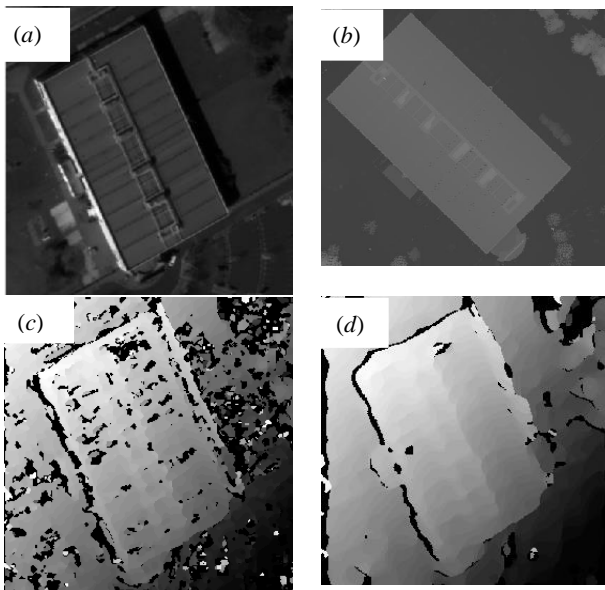
Figure 3: The effect on different support windows on the generated disparity maps by GF method with Census cost function. a) left epipolar image; b) LiDAR ground truth; c) and d) disparity maps by GF method with 13×13 and 31×31 support windows respectively.
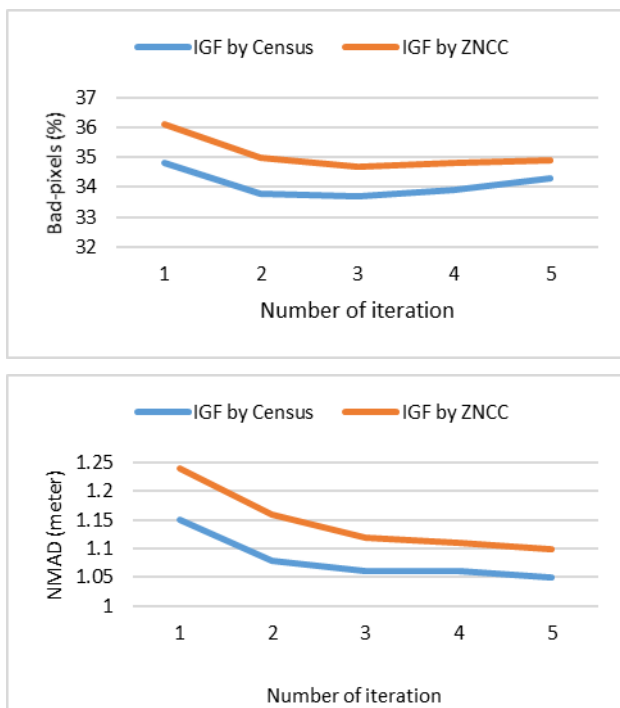

Figure 4. Evaluation of IGF method with different iteration

Another parameter which should be tuned is the number of iterations in IGF method. We evaluated the IGF method in 5 iterations as shown in Figure 4. The NMAD and Bad-pixels are computed according to the ground truth and this evaluation shows that the IGF does not improve the GF significantly. However, we selected the result of the IGF method with 3 iterations for further comparison.

The reference data and the generated disparity maps by compared stereo methods are shown in Figure 5. These results show that the large textureless areas in satellite stereo images

cannot be matched successfully using the SGM method. However, the result of SGM-GF is more appropriate. Also, Table 1 presents the evaluation for these results. Both NMAD and Bad-pixels criteria demonstrate that the SGM-GF method with census cost function provides good performance in the case of satellite stereo processing.

Methods	Census cost function		ZNCC cost function	
	NMAD (meter)	Bad-pixels (%)	NMAD (meter)	Bad-pixels (%)
GF	1.15	34.8	1.24	36.1
iGF	1.06	33.8	1.12	34.7
SGM	1.29	36.7	1.31	38.6
SGM-GF	0.94	28.4	1.03	31.3

Table 1: statistical evaluation of cost aggregation methods on satellite stereo images

4. DISCUSSION

The spectral similarity between neighbouring objects is a challenging issue in high resolution satellite stereo processing. In this paper, three selected cost aggregation methods and a combination of these methods have been evaluated on satellite stereo images in order to find the best cost aggregation method in this task. In the experiments, the NMAD and Bad-pixels were used to evaluate the results due to available LiDAR ground truth.

It is important to note that the high resolution satellite stereo processing depends on many factors such as radiometric resolution, study area, variation of object sizes and minimum size of smaller objects. Among the different study areas, the urban areas have a big challenge, therefore a stereo pair from the high resolution WV III images over an urban area has been selected to evaluate the cost aggregation methods. The study area includes different buildings with different shapes and sizes.

Although, the Census cost function has a good performance on the border of objects and low blurring effects; however, the results of both cost functions have same behavior on all cost aggregation methods.

Figure 2. presents the effect of different support windows on GF method. Due to the higher accuracy for larger windows, the main reason is the lower noise percentage in large areas. Also, blurring the border of objects for large windows demonstrates that the spectral information plays an important role in GF method. This is mainly because the GF method has a constant geometric weight for all elements. Hence, focusing on geometric weight such as dissimilarity between disparities for GF method could be developed in the future.

The effectiveness of more than one iteration has been examined in IGF method. Although increasing the number of iteration in IGF method reduces the noises, it also increases the blurring effects around the border of the objects. Generally, the IGF method is similar to GF method when GF method uses the large support windows.

Comparing the stereo methods, the SGM-GF method provide good matching results. The SGM method has low accuracy which might be because of the lack of parameter optimization in its original paper. The satellite images were taken on different dates, and consequently there are some radiometric changes in stereo images which might be as another reason for low accuracy of SGM method.

One of the problems of compared stereo methods on satellite stereo images is that the blurring effect occurs around the border of the objects which may be solved by some post processing or disparity map refinement techniques. Therefore,

investigation of disparity map refinement on satellite stereo matching and proposing new methods could be considered in the future. Due to the urban study area, the selected methods do not examine in a mountainous or forest areas. So, the evaluation

of stereo methods for forest inventory or point cloud generation in mountainous areas can be evaluated in the future.
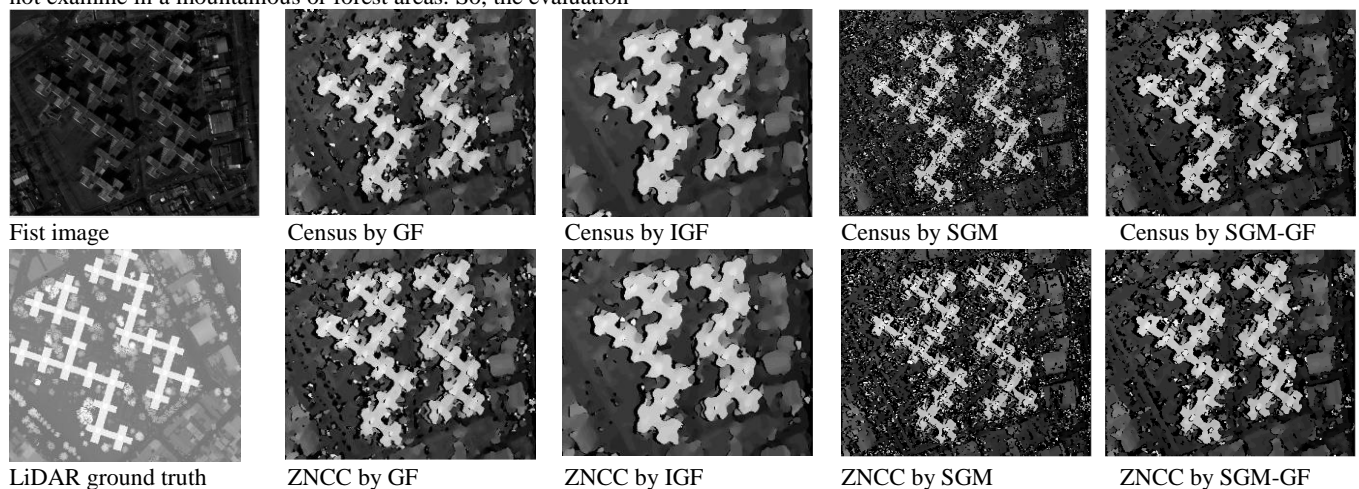

| Fist image | Census by GF | Census by IGF | Census by SGM | Census by SGM-GF |

| LiDAR ground truth | ZNCC by GF | ZNCC by IGF | ZNCC by SGM | ZNCC by SGM-GF |

Figure 5: The comparison of the generated disparity maps from the selected cost aggregation methods. First and second raw include the result for Census and ZNCC cost functions, respectively.

5. CONCLUSION

In this paper, we have evaluated three typical cost aggregation methods on urban satellite stereo images with airborne LiDAR ground truth. Also, a combination of these methods has been proposed due to their strongest properties. All stereo methods use of Census and ZNCC as matching cost function. The cost aggregation methods including GF, IGF, SGM, and SGM-GF have the same behaviour for both cost functions. The compared methods do not perform any disparity map refinement and therefore, it causes the evaluation of stereo methods to be an impartial assessment.

The results of IGF and SGM-GF methods have taken on best performance due to their parameter optimization in this paper. Among the compared stereo methods, the SGM-GF method have good performance and the original SGM method has lower accuracy. The result of experiments also demonstrates that the combination of SGM and GF methods improves the result of the original SGM algorithm in case of high resolution satellite stereo processing.

REFERENCES

Afsharnia, H., Arefi, H., Sharifi, M.A., 2017. Optimal weight design approach for the geometrically-constrained matching of satellite stereo images. Remote Sensing 9, 965.

Alobeid, A., 2011. Assessment of matching algorithms for urban DSM generation from very high resolution satellite stereo images. Fachrichtung Geodäsie und Geoinformatik der Leibniz-Univ., Honnover, Germany.

Bosch, M., Kurtz, Z., Hagstrom, S., Brown, M., 2016. A multiple view stereo benchmark for satellite imagery, Applied Imagery Pattern Recognition Workshop (AIPR), 2016 IEEE. IEEE, pp. 1-9.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23, 1222-1239.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. On stereo-rectification of pushbroom images, Image Processing (ICIP), IEEE International Conference on. IEEE, Paris, France, pp. 5447-5451.

Facciolo, G., De Franchis, C., Meinhardt-Llopis, E., 2017. Automatic 3D Reconstruction from Multi-Date Satellite Images, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 57-66.

Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogrammetric Engineering & Remote Sensing 69, 59-68.

Hamzah, R.A., Ibrahim, H., 2016. Literature survey on stereo vision disparity map algorithms. Journal of Sensors 2016.

Hamzah, R.A., Ibrahim, H., Hassan, A.H.A., 2017. Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. Journal of Visual Communication and Image Representation 42, 145-160.

He, K., Sun, J., Tang, X., 2013. Guided image filtering. IEEE transactions on pattern analysis and machine intelligence 35, 1397-1409.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, 328-341.

Höhle, J., Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS Journal of Photogrammetry and Remote Sensing 64, 398-406.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C., Gelautz, M., 2013. Fast cost-volume filtering for visual correspondence and

beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 504-511.

Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36.

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

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

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision 47, 7-42.

Sun, J., Zheng, N.-N., Shum, H.-Y., 2003. Stereo matching using belief propagation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25, 787-800.

Tatar, N., Arefi, H., 2019. Stereo rectification of pushbroom satellite images by robustly estimating the fundamental matrix. International Journal of Remote Sensing 40, 8879-8898.

Tatar, N., Saadatseresht, M., Arefi, H., 2017. The Effect of Shadow Area on Sgm Algorithm and Disparity Map Refinement from High Resolution Satellite Stereo Images. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 281-286.

Tippetts, B., Lee, D.J., Lillywhite, K., Archibald, J., 2016. Review of stereo vision algorithms and their suitability for resource-limited systems. Journal of Real-Time Image Processing 11, 5-25.

Yoon, K.-J., Kweon, I.S., 2006. Adaptive support-weight approach for correspondence search. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 650-656.

Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., Zhang, C., 2003. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. Geoscience and Remote Sensing, IEEE Transactions on 41, 872-882.

Zhang, K., Lu, J., Lafruit, G., 2009. Cross-based local stereo matching using orthogonal integral images. IEEE transactions on circuits and systems for video technology 19, 1073-1079.

Zhang, L., Gruen, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. ISPRS Journal of Photogrammetry and Remote Sensing 60, 195-211.

Zhu, K., d'Angelo, P., Butenuth, M., 2011. A performance study on different stereo matching costs using airborne image sequences and satellite images, Photogrammetric Image Analysis. Springer, pp. 159-170.