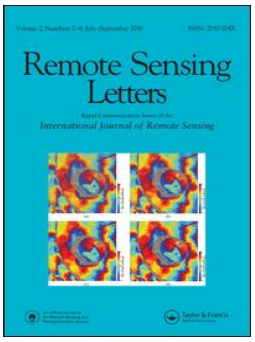

Taylor & Francis
Taylor & Francis Group

Remote Sensing Letters

ISSN: 2150-704X (Print) 2150-7058 (Online) Journal homepage: http://www.tandfonline.com/loi/trsl20

Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery

Chun Liu, Xin Huang, Dawei Wen, Huijun Chen & Jianya Gong

To cite this article: Chun Liu, Xin Huang, Dawei Wen, Huijun Chen & Jianya Gong (2017) Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery, Remote Sensing Letters, 8:9, 907-916, DOI: 10.1080/2150704X.2017.1335904

To link to this article: http://dx.doi.org/10.1080/2150704X.2017.1335904

Published online: 08 Jun 2017.

REMOTE SENSING LETTERS, 2017
VOL. 8, NO. 9, 907–916
https://doi.org/10.1080/2150704X.2017.1335904

Taylor & Francis
Taylor & Francis Group

Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery

Chun Liu[a], Xin Huang[a,b], Dawei Wen[a], Huijun Chen[b] and Jianya Gong[b]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [b]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

ABSTRACT
The ZiYuan-3 (ZY-3) satellite, equipped with three-line array scanners, is China's first civilian stereo mapping satellite. In this paper, the first assessment of the stereoscopic capacities of the ZY-3 imagery for building height estimation in heterogeneous urban areas is performed. Digital surface models (DSMs) are generated and optimized by different stereo pairs of a ZY-3 triplet for the city of Wuhan, China. The normalized DSMs (nDSMs) are computed as the height of the off-terrain objects by using morphological top-hat by reconstruction of the DSMs. We adopt random forest classification and object-based building segmentation to detect the buildings from the orthographic image, and estimate the building height by assigning the maximum nDSM value within the building regions. The actual heights of 400 buildings are used as reference data. Comparisons between the building heights extracted from the different stereo pairs are presented, indicating a higher accuracy by the nadir-forward stereo pair. The performance of the ZY-3 triplet is also compared with that of a WorldView-2 (WV-2) stereo pair, better result is achieved by ZY-3 nadir-forward image pair.

ARTICLE HISTORY
Received 22 January 2017
Accepted 21 May 2017

1. Introduction

The ZiYuan-3 (ZY-3) satellite, launched on 9th January, 2012, is China's first civilian high-resolution stereo mapping satellite (Tang et al. 2015). Designed for continuous and stable large-scale stereoscopic observation, the platform is equipped with a three-line array scanner with a favourable base-to-height (B/H) ratio of 0.85–0.95, and has a swath width of 50 km. The ground sample distance (GSD) is 3.5 m for the oblique panchromatic cameras viewing in the forward (22°) and backward directions (–22°), 2.1 m for the nadir panchromatic camera, and 5.8 m for the infrared multispectral scanner (IRMSS), respectively.

Building height is one of the most important information related to urban studies and applications, such as urban planning, dynamic monitoring, population estimation, damage assessment etc. Although some studies have investigated building height

CONTACT Xin Huang ✉ xhuang@whu.edu.cn School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.

retrieval and 3D reconstruction in urban areas, it is still a very difficult problem for large cities with dense and heterogeneous characteristics. Photogrammetric measurement of two- or multiple-view stereo images from high resolution satellites such as IKONOS, WorldView and Cartosat series provides us a practical way of estimating urban building height through constructing digital surface models (DSMs). However, building height retrieval from a DSM relies on the quality of the DSM, which can be hampered by image dissimilarities, occlusions, insufficient texture, clouds etc. The three-line array mode of the ZY-3 satellite allows for stereo mapping along-track with fixed viewing angles in a short time interval, which minimizes the radiometric variations of the multi-view images. Meanwhile, occlusions can be reduced by multi-image matching or fusion method. Lots of researchers have studied ZY-3 images for geometric accuracy verification (Tang et al. 2015) and DSM evaluation (Fratarcangeli et al. 2016), and have been reported that the vertical root-mean-square error (RMSE) of the DSM is around 5 m in the flat or urban area without ground control points (GCPs). However, research concerning building height retrieval and accuracy assessment from ZY-3 triplet in dense urban areas has not been addressed to date.

In this study, we assess the capabilities of building height retrieval from ZY-3 stereo images in heterogeneous urban areas for the first time. Comparisons are carried out between different stereo pairs from the ZY-3 triplet and also a stereo pair from WV-2 satellite.

2. Study area and data

The study area (Figure 1) is located in Wuhan, China. Wuhan is the central city in the middle part of China, characterized by intense urban growth and a dense built-up environment. The ZY-3 triplet was captured on 12 August 2013, covering the city centre, with a total area of 173.3 km^2. Moreover, a WV-2 stereo pair with an overlapping area of approximately 148 km^2 was used to perform comparison with the ZY-3 images. The WV-2 images, with the resolution of 0.5 m, were acquired along-track by a modulating camera with a convergence angle of 40.35°. To align with ZY-3 images, image co-registration procedure was conducted to minimize the distance between the ZY-3 and WV-2 images through a number of manually selected tie points. The actual heights of 400 buildings, provided by the Wuhan Urban Planning Department, were used as reference data to assess the accuracy.

3. Methodology

The workflow of our study is presented in Figure 2. From the ZY-3 stereo images, three DSMs were firstly generated by different view combinations, i.e., nadir-forward, nadir-backward and forward-backward. An optimization strategy was carried out by fusing the three DSMs to fill the unmatched areas, and the optimized DSM was then used to generate an orthographic image of the nadir image. The normalized DSM (nDSM) was computed as the height of the off-terrain objects. To extract the buildings, multi-feature indices (vegetation, building, shadow), as well as the spectral and nDSM features, were used in random forest (RF) classification. Then image segmentation and a post-filtering process were conducted to generate building objects. Building height was assigned as

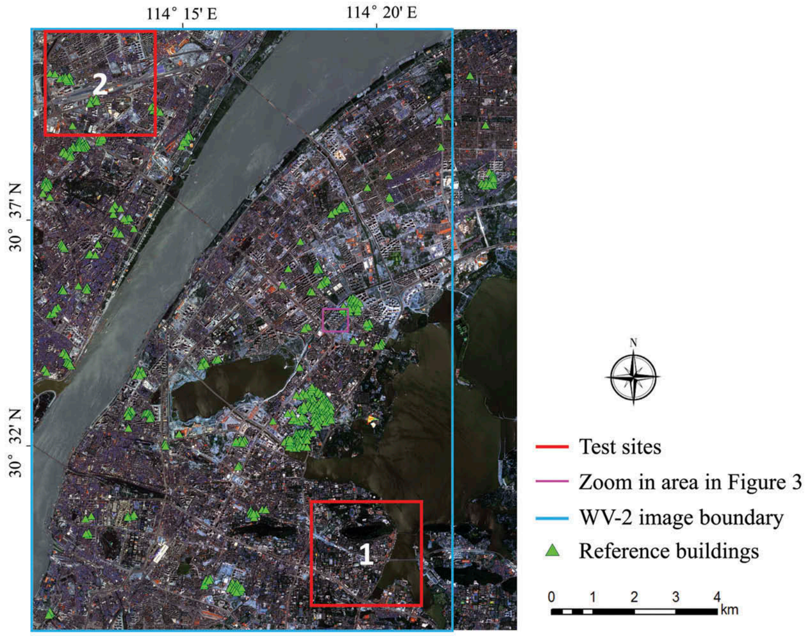

Figure 1. ZY-3 orthographic image of the study area showing the location of test site 1 and 2 (in red), the zoom in area in Figure 3 (in magenta), the boundary of the WV-2 imagery (in blue) and the reference buildings.

the maximum nDSM value within each building objects. Finally, 400 reference buildings were used to assess the accuracy of the building height estimation from ZY-3 and WV-2 stereo images.

3.1. *DSM generation*

DSM generation is the basic step for building height extraction. Existing research (Fratarcangeli et al. 2016) shows that the geolocation error of the ZY-3 triplet do not significantly affect the images' relative orientation and the DSMs generation. Since our aim is to extract the relative height of buildings from the ZY-3 triplet without any additional information, no GCP is used here.

The DSMs from ZY-3 triplet were generated by two view stereo pairs independently (i.e. nadir-forward, nadir-backward and forward-backward). Based on the relative orientation parameters of the bundle block adjustment process with the supplied rational polynomial coefficients (RPCs), the two-view images were resampled to the same resolution and rectified to an epipolar geometry. Stereo matching, the core step of the photogrammetric process, was then performed to generate a disparity map of the epipolar images. An advanced and widely used method for dense stereo matching is semi-global matching (SGM) (Hirschmüller 2008), which was used in this study to generate disparity maps of image pairs. Finally, DSMs were interpolated and resampled into a raster grid from the 3D point cloud generated by measurement of the height parallax between the homologous points.

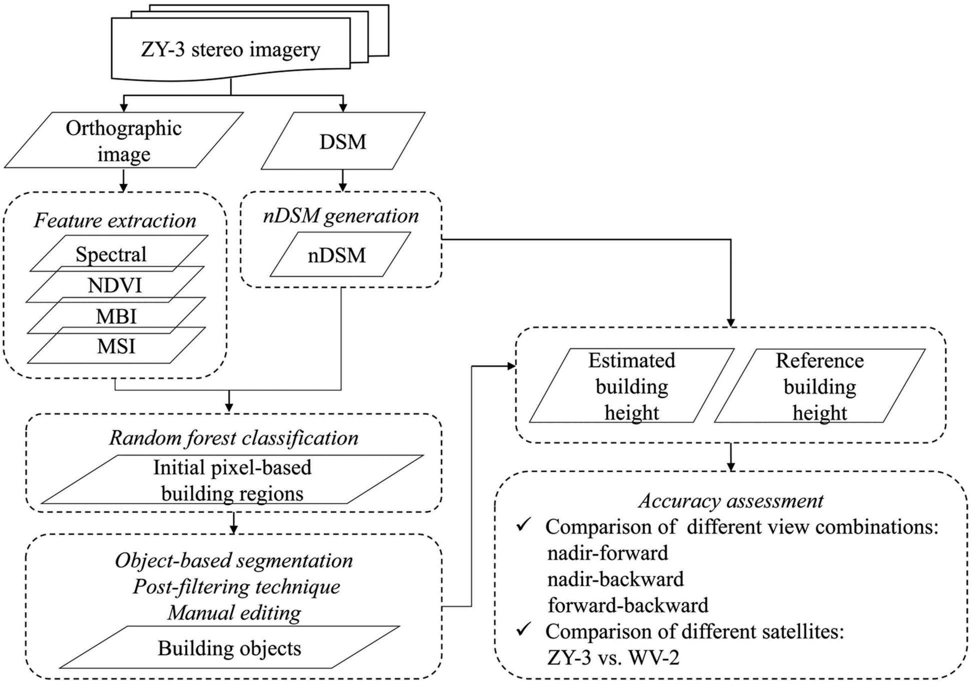

Figure 2. Workflow of the building detection and building height assessment.

3.2. *DSM fusion*

Due to the large disparity and occlusion of ground features, some areas in the images failed to generate a DSM and resulted in 'no data' or holes. Generally, such holes can be filled with a coarse DSM or an interpolation method, while the three-view directions of ZY-3 images provides a better solution. .

The fusion process was performed by taking one DSM as the base and using the other two DSMs to fill the holes of the base DSM. Specifically, in the hole area of the base DSM, we compared the other two DSMs at the same location and chose a closest value to the neighbourhood on the base DSM. Each DSM was optimized by this fusion step, and most of the holes were filled. Here we calculated DSM completeness as an index to reflect the quality of DSMs before and after fusion. The DSMs completeness is defined as the percentage of correctly matched points over the whole working area (Aguilar, Saldana, and Aguilar 2014). Water areas were masked out beforehand, since we were only concerned with the 3D reconstruction of the built-up area. Remaining small holes could be further removed by simple spline interpolation.

3.3. *nDSM generation*

Since DSM includes the height of the terrain and off-terrain objects, a normalized DSM (nDSM) needs to be generated to represent the relative height above ground. Theoretically, nDSM can be extracted by subtracting digital terrain model (DTM) from DSM. But to examine the general situation that an accurate DTM is not available, the nDSM was computed by using morphological top-hat by reconstruction of the DSM (Qin and Fang 2014).

2D grey level based top-hat reconstruction is a reconstruction filter which is able to calculate the difference of brightness between structures and their neighborhoods within the region of a structural element (SE). It is able to detect and preserve the shape of building structures in DSMs since they are brighter than the ground structures. Specifically, a mask image *f* (the DSM) is reconstructed as $R_{f,I}$ from the marker image *I*, where *I* is derived from an erosion operation ε from *f* by a SE *e*. The top-hat reconstruction is computed by subtracting $R_{f,I}$ from the mask *f*, then the peaks of *f* overlaying on *I* is extracted. The top-hat reconstruction can be written as:

$$THR(f,e) = f - R_{f,\varepsilon(f,e)} \tag{1}$$

A 'disk-shaped' SE is used with the radius estimated as the largest building radius in the scene (80 m in our experiment). Since the top-hat reconstruction cannot detect buildings that have connected to trees with similar height, a normalized difference vegetation index (NDVI) is used before the top-hat reconstruction to eliminate vegetation. Therefore, the morphological erosion is computed as:

$$\varepsilon(f,e)(i,j) = \begin{cases} \min\{f(a,b)|e(a-i,b-j) = 1, (i,j) \in D_f, (a-i,b-j) \in D_e\}, if\ NDVI(i,j) < t \\ f(i,j), otherwise. \end{cases} \tag{2}$$

whereD_f, D_e are the domain of definition of *f*, *e*, respectively, and *t* is the NDVI threshold, which was 0.23 in our case.

3.4. *Building detection*

A pixel-based building detection procedure is carried out with random forest (RF) classification method (Breiman 2001). Based on our previous research, a few basic urban classes, e.g., vegetation, buildings and shadows, can be effectively represented using a set of primitive indices. The nDSM, containing height information above ground, is also helpful to separate buildings from the terrain (e.g. road and soil). Therefore, besides the multispectral data of ZY-3 images, the NDVI, morphological building/shadow indices (Huang and Zhang 2012) and the nadir-forward nDSM are integrated to the feature stack for the RF classification.

To obtain the building objects, initial segmentation of the pan-sharpening orthographic image is conducted using the multiresolution segmentation algorithm (Benz et al. 2004). Then based on the segmented regions, the building detection result and the nDSM, building objects are extracted by using a post-filtering technique. Specifically, the regions are reserved if satisfying the following two conditions simultaneously: a) the region contains the building detection result with an area proportion higher than 10%; b) the maximum nDSM value within the region is higher than 3 m. We also slightly performed manual editing to ensure the accuracy of the building objects.

4. Results and discussion

4.1. *Generation of DSMs*

Qualitative analysis and comparison between different DSMs was carried out by visual check and DSM completeness. A zoom in area (marked in Figure 1) and its DSMs are

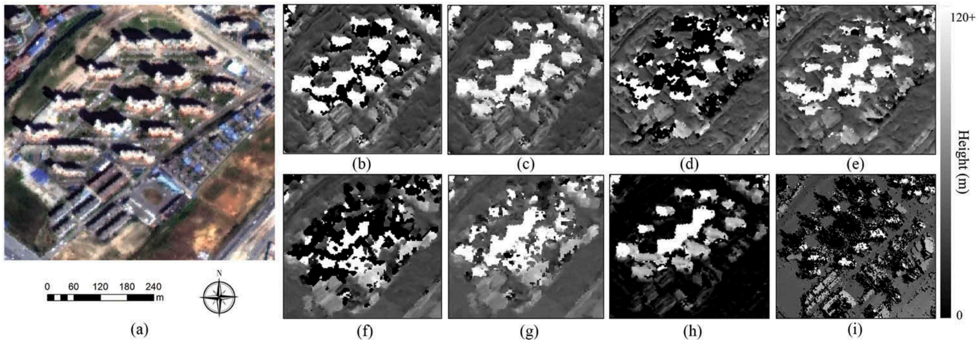

Figure 3. Zoom in view of (a) ZY-3 orthographic image, DSMs of the same area before and after fusion of the ZY-3 nadir-forward (b,c), nadir-backward (d,e) and forward-backward (f,g) stereo pairs. (h) nDSM calculated from the nadir-forward DSM and (i) DSM generated by the WV-2 stereo images.

Table 1. DSM completeness before and after fusion.

Stereo pair	Completeness before fusion (%)	Completeness after fusion (%)
ZY-3 nadir-forward	98.29	99.73
ZY-3 nadir-backward	96.92	99.74
ZY-3 forward-backward	94.29	99.72
WV-2 (no fusion)	80.18	

illustrated in Figure 3(a–g). This area is composed of high-rise buildings and a few low-medium-rise buildings. From visual inspection, the DSMs generated by nadir-forward image pair can better outline ground features than the DSMs by nadir-backward and forward-backward stereo pairs. High-rise buildings in the DSMs are usually accompanied by holes due to occlusions, while the DSM fusion method had good performance to deal with the holes and in the meantime keep the shape of features. The DSM completeness before and after fusion is illustrated in Table 1. The B/H ratio, defined as the separation of the stereo pair divided by the height of the sensor, plays a significant role in the quality of DSM generation. A larger B/H ratio is beneficial for building strong stereo geometry, while in urban areas, a smaller B/H ratio is preferred to increase similarity between the stereo pairs and thus improving the matching result. The forward-backward image pairs yielded more holes in dense urban areas due to the occlusions and large disparity. After DSM fusion, most of the holes were filled and reached the completeness higher than 99.7%. Due to the difference of the multi-angle reflectance, the nadir-forward stereo pair obtained DSM with better quality than the nadir-backward pair.

4.2. *Building detection and accuracy assessment*

From the pan-sharpened orthographic image, two test sites (marked in Figure 1) in the study area were chosen and the ground reference were collected by careful visual inspection (Figure 4). In total, 63,121 and 70,920 pixels were chosen for buildings and non-buildings, respectively. Among which, 5% random samples were used for training. The accuracies of the test sites were validated at every stage of the building detection procedure and the overall accuracy (OA) and kappa coefficient were summarized in

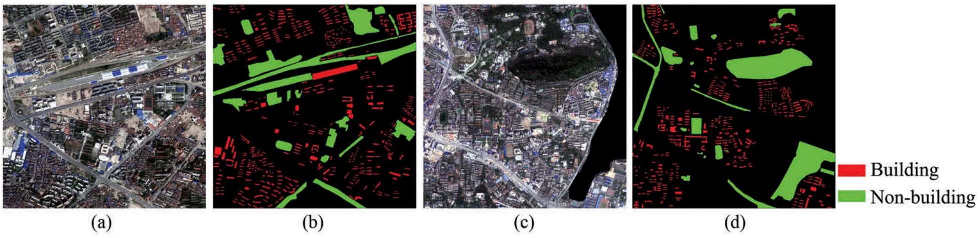

Figure 4. (a,b) Test sites 1 and the corresponding ground reference, (c,d) test sites 2 and the corresponding ground reference.

Table 2. Building detection accuracy.

	RF classification		Object-based building extraction		After manual editing	
	OA (%)	Kappa coefficient	OA (%)	Kappa coefficient	OA (%)	Kappa coefficient
Test site 1	92.23	0.85	94.59	0.86	96.47	0.87
Test site 2	92.11	0.85	93.21	0.85	95.60	0.86

Table 2. The results show that the final OA and kappa coefficient were higher than 95% and 0.86. Building height was assigned as the maximum nDSM value within each building objects. An overview of the extracted buildings and the estimated building height for the whole study area is shown in Figure 5(a) and a zoom in view of an area with buildings of different height [Figure 5(b)] were illustrated.

4.3. *Accuracy assessment of the building height estimation*

To quantitatively analyse the accuracy of the estimated building height, 400 buildings within different height levels were used as reference building heights. Accuracy

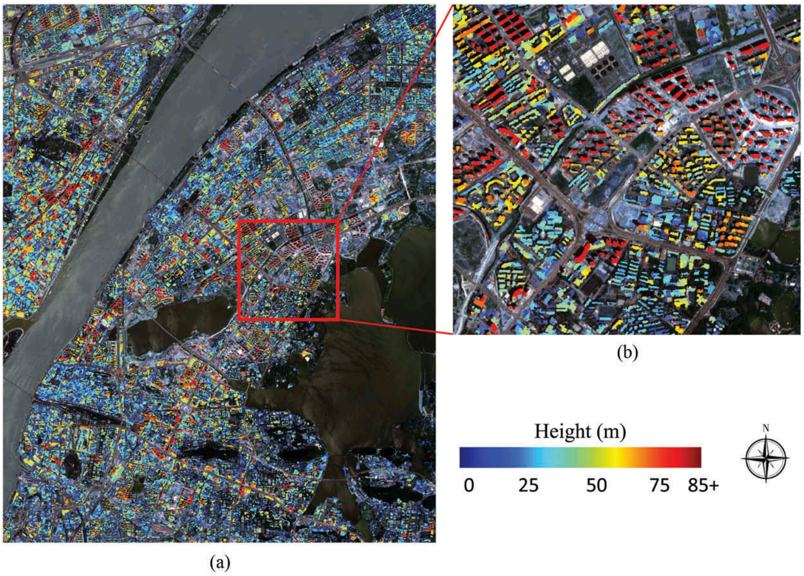

Figure 5. (a) Buildings with height information overlaid on the study area, (b) a zoom in view of an area with buildings of different height.

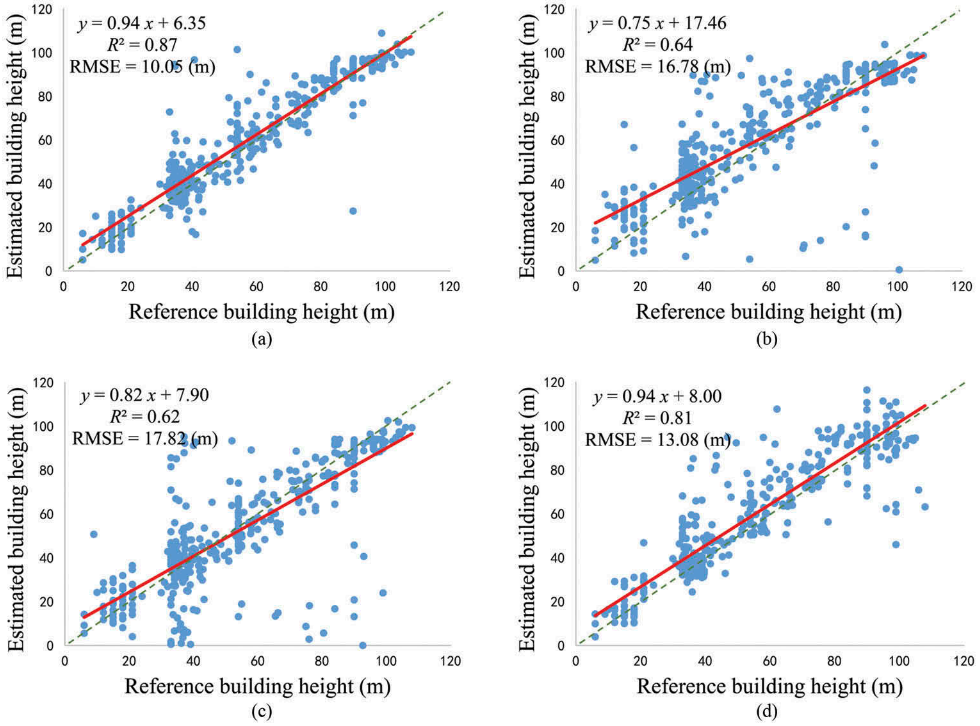

Figure 6. Comparisons of the estimated building height with the reference building height by the ZY-3 nadir-forward (a), nadir-backward (b), forward-backward (c) stereo pairs and the WV-2 stereo pair (d).

assessment was carried out, in terms of RMSE, by comparing the estimated building height with the reference height data. Linear regressions were also performed to model their relationship. The RMSEs of 10.05 m, 16.78 m and 17.82 m were obtained from the ZY-3 nadir-forward, nadir-backward and forward-backward stereo pairs, respectively, and the coefficients of determination (R^2) were 0.87, 0.64 and 0.62 for all the reference buildings [Figure 6(a-c)]. The nadir-forward stereo pair had the highest accuracy and the maximum RMSE was obtained by the forward-backward stereo pair. Most of the estimated building height were floating near the actual height, while some outliers existed and significantly reduced the accuracy. An increasing number of outliers was found from Figure 6(a–c).

We further analysed the accuracy by manually setting three categories according to the actual building height: buildings below 30 m, between 30 and 90 m, and higher than 90 m (Table 3). The minimum RMSE of 4.64 m was obtained by ZY-3 nadir-forward stereo pair for buildings above 90 m. The RMSEs of buildings from 30 m to 90 m were all higher than 10 m, indicating the unstable estimation for the medium-rise buildings.

4.4. *Comparison with WV-2*

A stereo pair of WV-2 images were used to compare with the ZY-3 triplet. The DSM generated by the WV-2 image pair is shown in Figure 3(i), where it can be observed that, for low- and medium-rise buildings, the building boundaries are better outlined than in

Table 3. RMSEs (m) of building height estimation in different height ranges.

	Below 30 m	From 30 to 90 m	Above 90 m	All
ZY-3 nadir-forward	5.73	11.19	4.64	10.05
ZY-3 nadir-backward	14.51	16.88	18.21	16.78
ZY-3 forward-backward	9.78	18.57	19.36	17.82
WV-2	7.71	12.78	18.10	13.08

the ZY-3 derived DSMs [Figure 3(b-d)], due to the high resolution of the WV-2 images. However, the area of holes next to high-rise buildings are bigger than the ZY-3 DSMs because of the large convergence angle of the WV-2 stereo pair leads to increased number of unmatched points. The DSM completeness for the WV-2 is 80.18%, much lower than the three DSMs extracted by the ZY-3 triplet.

380 reference buildings located in the overlapping area of the ZY-3 and WV-2 images, were used for the comparison. The overall RMSE was 13.08 m and the R^2 is 0.81 for the WV-2 stereo pair [Figure 6(d)], indicating a lower accuracy than the ZY-3 nadir-forward stereo pair but higher accuracy than the ZY-3 nadir-backward and forward-backward pairs. Concerning the accuracy at different height level (Table 3), the ZY-3 nadir-forward image pair had better performance especially when the buildings were higher than 90 m. Please note that, the comparison between the ZY-3 and WV-2 is only an experimental testing, and does not signify a conclusion that the accuracy of ZY-3 satellite is higher than WV-2 satellite for building height retrieval. However, 3D building reconstruction benefits from appropriate observation angles and short revisiting time, which is difficult to implement from WV-2 satellite. Comparatively, the three-line array mode of ZY-3 satellite provides a convenient and feasible way for building height estimation of large urban areas.

5. Conclusions

In this study, a validation of building height estimation from the ZY-3 triplet in a heterogeneous urban area was conducted. DSMs from the nadir-forward, nadir-backward and forward-backward image pairs were generated and optimized by a fusion method. The nDSMs were calculated to represent the height of off-terrain objects. After extracting building objects, building height were assigned as the maximum nDSM value within the objects. Compared to the reference building height, most of the estimated building height were floating near the actual height and a few outliers exist. The building height generated by the ZY-3 nadir-forward stereo pair achieved the best results, qualitatively and quantitatively, with the RMSE under 5 m for buildings higher than 90 m. The estimation for the buildings between 30 m and 90 m were not very stable with the RMSEs higher than 10 m. Furthermore, a stereo pair from WV-2 was used for comparison, resulted a lower accuracy than the ZY-3 nadir-forward image pair. Our paper conduct the first validation of the building height estimation from ZY-3 triplet in a dense urban area, indicating the potential capability of ZY-3 stereo images for large area building height retrieval. Improvement for building height extraction method from ZY-3 triplets will be further investigated according to different urban scenes in our future work.

Acknowledgement

We would like to thank the anonymous reviewers for their useful comments and suggestions.

Funding

This work was supported by the China National Science Fund for Excellent Young Scholars [grant number 41522110], by the National Key Research and Development Plan [grant number 2016YFB0501403] and by the Foundation for the Author of National Excellent Doctoral Dissertation, China [grant number 201348];

References

Aguilar, M. A., M. del Mar Saldana, and F. J. Aguilar. 2014. "Generation and Quality Assessment of Stereo-Extracted DSM from Geoeye-1 and Worldview-2 Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 52 (2): 1259–1271. doi:10.1109/TGRS.2013.2249521.

Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. "Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information." *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3): 239–258. doi:10.1016/j.isprsjprs.2003.10.002.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Fratarcangeli, F., G. Murchio, M. Di Rita, A. Nascetti, and P. Capaldo. 2016. "Digital Surface Models from Ziyuan-3 Triplet: Performance Evaluation and Accuracy Assessment." *International Journal of Remote Sensing* 37 (15): 3505–3531. doi:10.1080/01431161.2016.1192308.

Hirschmüller, H. 2008. "Stereo Processing by Semi-Global Matching and Mutual Information." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2): 328–341. doi:10.1109/TPAMI.2007.1166.

Huang, X., and L. Zhang. 2012. "Morphological Building/Shadow Index for Building Extraction from High-Resolution Imagery over Urban Areas." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (1): 161–172. doi:10.1109/jstars.2011.2168195.

Qin, R., and W. Fang. 2014. "A Hierarchical Building Detection Method for Very High Resolution Remotely Sensed Images Combined with DSM Using Graph Cut Optimization." *Photogrammetric Engineering and Remote Sensing* 80 (9): 873–883. doi:10.14358/pers.80.9.873.

Tang, X., P. Zhou, G. Zhang, X. Wang, Y. Jiang, L. Guo, and S. Liu. 2015. "Verification of ZY-3 Satellite Imagery Geometric Accuracy without Ground Control Points." *IEEE Geoscience and Remote Sensing Letters* 12 (10): 2100–2104. doi:10.1109/lgrs.2015.2450251.