This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Fusion of WorldView-2 Stereo and Multitemporal TerraSAR-X Images for Building Height Extraction in Urban Areas

Yong Xu, Peifeng Ma, Edward Ng, and Hui Lin, *Senior Member, IEEE*

***Abstract*—We investigated the joint use of the high-resolution WorldView-2 optical satellite images and the multitemporal TerraSAR-X synthetic aperture radar (SAR) satellite images to extract building height information in high-density urban areas. The main idea of the proposed fusion approach is to take full advantage of both data sets in building height retrieval. The proposed approach includes two main stages. First, initial building height estimates are extracted from WorldView-2 stereo images and multitemporal SAR images. These initial results are then combined using a novel object-based fusion approach, in which the heights of points for the same building footprint are retrieved and integrated. Experiments on the Mong Kok area of Hong Kong showed that the proposed approach using both data sets outperforms the use of either stereo images or SAR images alone. According to the results of the proposed approach, the average absolute height retrieval error is 6.53 m, which is much lower than using stereo and SAR images (9.08 and 12.24 m, respectively). The proposed fusion approach is suitable for building height retrieval in urban areas where single satellite data have limitations.**

***Index Terms*—Building height, object-based fusion, synthetic aperture radar (SAR), WorldView-2.**

I. Introduction

BUILDING height retrieval from satellite data has become a hot topic in the remote sensing community. According to the available satellite data, there are three categories of method for building height retrieval, namely, stereo photogrammetry technology with pairs of optical images (stereo images), synthetic aperture radar (SAR) technology, and light detection and ranging data (LiDAR) technology.

Stereo images have been widely used for building height retrieval since they became available [1]. This method usually has four main stages, namely, data preprocessing, stereoscopic matching, digital surface model (DSM) generation, and above-ground building height extraction. Studies using stereo images for building height extraction have mainly focused on parts of the aforementioned stages. Some studies have emphasized on the development of advanced stereoscopic matching methods to improve the building height retrieval accuracy [2], whereas others have investigated the use of multiple stereo images to identify common targets for height retrieval [3], [4]. In addition, some studies have made use of ancillary data, such as approximate digital elevation modeling or prior building footprints [5]. In addition to stereo images, monocular satellite image can be used to retrieve the height of buildings in areas where the shadows of the buildings can be measured [6]. However, the accuracy of this method is affected by the quality of the measured shadows and the surrounding environments, which limits its practical use.

In addition to stereo images, SAR images have been widely used for building height retrieval in recent decades, and various SAR techniques have been explored. According to a data source on the usage of SAR, there are several different processing strategies, such as interferometry with inference SAR images [7]–[10], radargrammetry with stereo SAR images [11], monoscopic SAR images [12]–[14], and multiaspect SAR images [15]. However, due to the intrinsic characteristics of SAR images (e.g., layover effect) and the effect of the mutual inference of the surrounding environment, studies using SAR data to derive building height are still limited to simple scenes of isolated buildings [5], [10], [12], [13], [16].

The advent of LiDAR has opened up a new phase of building detection research. LiDAR provides huge point clouds that significantly improve the accuracy of building detection [17]. It is not only applicable to the retrieval of building height but can also be used to reconstruct accurate 3-D building models [18]. Although the results obtained using LiDAR data are promising, it is still expensive, and restrictions on flight plans in some countries limit its application in large urban areas.

In summary, automatically extracting building heights from satellite data still has its challenges. The main limitations using different kinds of data can be summarized as follows. First, stereo images tend to underestimate the height of tall buildings, and taller buildings have larger prediction errors [19]. Second, SAR interferometry can only provide limited results (noisy and incomplete data), particularly for high-density urban areas where the mutual inference of surrounding buildings is significant [10], [16]. Third, LiDAR data are still expensive, and flight plan restrictions limit its application to large urban areas [17]. These difficulties restrict the practical use of single satellite data for building height retrieval in large urban areas, particularly for high-density areas with complex scenarios.

Manuscript received January 21, 2015; revised March 24, 2015; accepted April 22, 2015. This work was supported in part by the Hong Kong Research Grants Council under Grant CUHK 14408214, in part by the National Key Basic Research Program of China under Grant 2015CB954103, and in part by the Innovation and Technology Support Program of HKSAR under Grant ITS/075/13. *(Corresponding author: Edward Ng.)*

Y. Xu is with the Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong, China.

P. Ma and H. Lin are with the Institute of Space and Earth Information Science and Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China.

E. Ng is with the School of Architecture, the Institute of Environment, Energy and Sustainability (IEES), and the Institute of Future Cities (IOFC), The Chinese University of Hong Kong, Hong Kong, China (e-mail: edwardng@cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2015.2427738

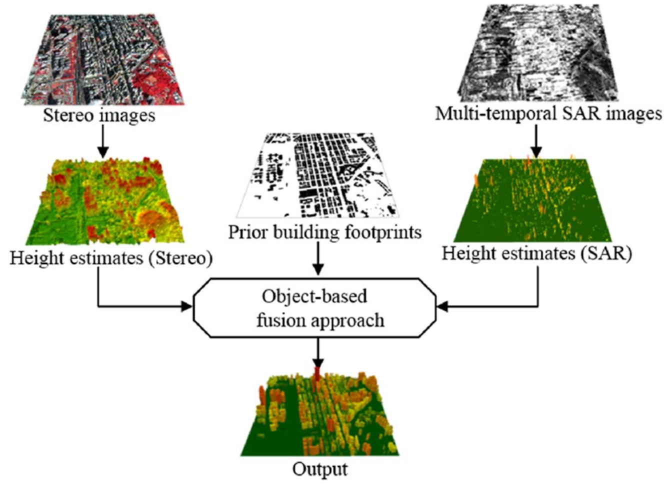

Fig. 1. Procedure of the proposed approach for building height retrieval.

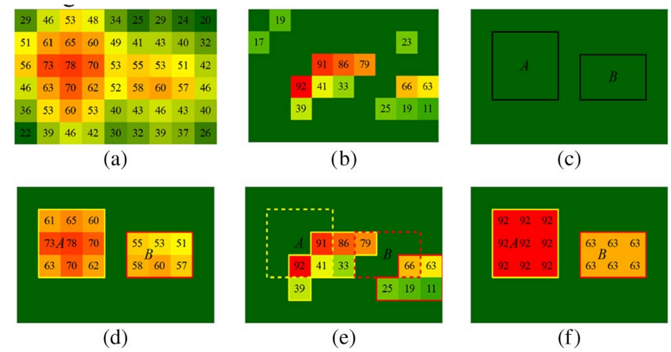

Fig. 2. Illustration of the object-based fusion approach, in which two building height estimates from both stereo and SAR images are combined to generate a fused result. (a) Initial height estimates obtained from stereo images (unit: meters). (b) Initial height estimates obtained from SAR images (unit: meters). (c) Prior building footprints. (d) Extracted height estimates (stereo). (e) Extracted height estimates (SAR). (f) Fused result.

To achieve a more promising result, scholars have investigated the integration of different data sources for building height retrieval, such us SAR and optical image fusion [16], [20] and LiDAR and optical image fusion [17]. However, most of the fusion approaches take advantage of optical images for building footprint extraction rather than building height retrieval. For example, in [16], Sportouche *et al.* proposed a method for detecting and reconstructing 3-D buildings in urban scenes using both high-resolution optical and SAR images. Nevertheless, in their studies, optical images were limited to generate 2-D building footprints instead of heights of buildings. Therefore, we investigated the joint use of stereo optical images and SAR images for building height retrieval in urban areas, where single data have limitations.

The remainder of this paper is organized as follows. In Section II, a novel object-based fusion approach for building height retrieval is introduced. The results and the performance analysis are given in Section III. Conclusions with a discussion of the proposed approach are presented in Section IV.

II. Methodology

This section presents the fusion-based approach for retrieving the heights of buildings in urban areas using both stereo and SAR images, under the assumption that the building footprints are known. As shown in Fig. 1, the approach is implemented using stereo images, multitemporal SAR images, and prior building footprints as the inputs, and it includes two main stages. First, initial building height estimates are extracted from stereo images and multitemporal SAR images. The initial results are then combined according to an object-based fusion approach, in which points with heights belonging to the same building footprint are retrieved and combined to generate a fused result. Details of the approach are given as follows.

A. Initial Building Height Estimates

In this stage, two initial building height estimates were extracted from stereo and SAR images. For stereo images, an empirical model was used to extract the building height estimate using the built-in rational polynomial coefficient (RPC). In this approach, a group of rational functions was used to model the relationship between the coordinates of image points and the coordinates of corresponding object points on the ground. After the same image points of stereo images were extracted as the input of the aforementioned rational functions, the heights of corresponding object points could be retrieved.

The TomoSAR technique was used to derive the vertical heights of scatters from multitemporal SAR images. The tomography technique extended the conventional interferometry SAR technique in processing the data in two range–azimuth dimensions to three range–azimuth–height dimensions. By making use of multitemporal SAR images with redundant observations, targets at different elevations in the same range–azimuth unit can be distinguished. Hence, the TomoSAR technique performs well in distinguishing different height targets in one SAR signal and is particularly suitable for retrieving the heights of buildings in urban areas. Details of the TomoSAR technique for building height extraction are given in [8], in which scatter points with heights are extracted from SAR images.

As an example, two initial building height estimates obtained from stereo images and multitemporal SAR images are given in Fig. 2(a) and (b) (with height unit as meters). Compared with the height estimates by stereo images, height estimates obtained by SAR images are more likely to have missing data due to SAR image noise and distortions.

B. Object-Based Fusion Approach

In this stage, both of the initial building height estimates were combined to generate a fused result for the same building footprint. The literature review and our experimental studies proved that the height result with stereo images has no significant geometrical error; thus, the height points within building footprints can be directly assigned to the right buildings. However, the results from SAR images tended to have a larger positioning error. As shown in Fig. 2(b), one SAR point has a height of 79 m. Considering that a set of close points with similar heights (or near arithmetic progression) likely belong to the same building, this point is likely to belong to building A as it is more similar with the point with a height of 86 m of building A than the point with a height of 66 m of building B. However, it is incorrectly located on building B in this study. Therefore, spatial distortion of SAR results poses the challenge of allocating the heights of points to the right buildings. Moreover, some height points belonging to other targets (such as viaducts) may wrongly locate in one building area due to the geometrical error, which also causes prediction error. To get a robust result, an object-based fusion approach was developed, consisting of three steps.

First, connected components are extracted from the initial SAR height results using the connected components analysis method [21]. Only the components that have more than a certain number of points (e.g., 4) are considered as candidate objects for buildings. As shown in Fig. 2(e), there were two components extracted as candidate objects.

Second, possible candidate objects are matched with prior building footprints. If object i is matched with building k, they should meet the following two criteria: 1) building k is the closest building to object i; and 2) the spatial distance between object i and building k is within a certain distance. If one object has more than one building within the same close distance, it should be divided into several small objects, and the matched buildings are found again. As shown in Fig. 2(e), the matched objects (with solid line) and buildings (with dashed line) are marked with the same color: yellow for building A and red for building B.

Third, the initial height results from both stereo and SAR images are fused based on prior building footprints and the matched objects. For the same building footprint, we can achieve two sets of height models: one comes from the stereo result within the building footprint, as shown in Fig. 2(d), and the other one comes from the SAR result within the matched object, as shown in Fig. 2(e). To avoid height bias arising from positioning error, we simply use the highest point as the building height. The fusion operation is based on the two highest values of both results. For example, the heights of buildings A and B are 78 and 60 m based on the stereo images and 92 and 66 m based on the SAR images. If their difference is small, the average value is the fused result. If there is a large difference between the two results, the larger value is used as the predicted height. Therefore, if we specify a threshold of 10 m, the fused result is 92 m for building A and 63 m for building B in this study. The final fused result is given in Fig. 2(f).

Some buildings still have missing data as the matched objects from the SAR result are incomplete and do not cover all of the buildings. For these buildings, height points within the building footprints are extracted directly from both the stereo and SAR images, and the fusion process is the same as in Step 3. In particular, if no SAR height points are achieved, then the height result from the stereo images is assigned directly.

The object-based fusion approach has two benefits in building height retrieval. First, some SAR height points not located in the actual buildings, due to geometrical distortions, can still be retrieved and applied to the building height prediction, which helps to provide more accurate results. Second, SAR height points belonging to the right buildings but wrongly located on neighboring buildings can be removed using the proposed approach, which may cause a large prediction error for the neighboring buildings. As shown in Fig. 2(e), there is a point with a height of 79 m in the footprint of building B, but it should belong to building A. With the proposed approach, the incorrect assignment of this point to building B can be avoided, and this point has been rightly labeled as building A in this study.

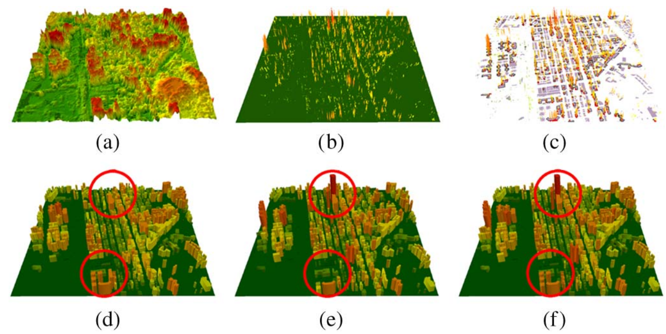

Fig. 3. Building height retrieval using different approaches. (a) Initial height result using stereo images (unit: meters). (b) Initial height result using SAR images (unit: meters). (c) Prior building footprints, in which matched buildings are marked with black color, whereas matched SAR objects are marked with red color. (d) Result using the stereo images. (e) Result using the SAR images. (f) Result using the fusion approach.

III. Results and Performance Analysis

A. Study Area and Data Collection

The study area was located at 22° N, 113° E in the Mong Kok district of Hong Kong, a high-density urban area with an average building height around 45 m and where the heights of some buildings reach several hundred meters. Using this study area, we collected the actual building data (including 2-D building footprints and building height information) from 2010, the 14 TerraSAR-X images from 2011, and a pair of stereo images on 2014. The actual building height data from 2010 were used as validation data. The actual building height information reflects the surface height in this study. Therefore, DSMs from both the stereo and SAR images are used as the heights of buildings, rather than subtracting the ground surface elevation from the estimated DSMs as in other studies [5].

B. Experimental Results

We used both SAR images and stereo images separately to obtain the initial height models. Fig. 3(a) shows the generated DSM using the stereo images, whereas Fig. 3(b) shows some points with heights using the SAR images. With the proposed object-based fusion approach, the scatter points of the SAR result are first organized as objects. Next, all of the matched SAR objects are retrieved according to prior building footprints. As shown in Fig. 3(c), the matched SAR objects and buildings are highlighted with red and black colors. Finally, the initial height results from both stereo and SAR images are fused based on prior building footprints and the matched objects. The fused result was shown in Fig. 3(f). In comparison, the results generated with the stereo images and the SAR images are also provided [see Fig. 3(d) and (e)].

C. Performance Analysis

The average absolute difference (AAD) was used to assess the quality of the result by comparing the predicted result with the actual result. The lower the AAD value, the better the quality of the predicted result. As the last row of Table I shows, the overall accuracy using the stereo images, the SAR images, and both data sets is 9.08, 12.24, and 6.53 m, respectively,

TABLE I
AAD IN PREDICTED RESULTS FROM DIFFERENT SATELLITE DATA COMPARED WITH THE ACTUAL BUILDING HEIGHTS

	Stereo images	SAR images	Proposed
AAD (H>=100)	86.67	1.88	1.88
AAD (50<H<100)	14.25	15.98	10.76
AAD (H<=50)	6.27	11.57	6.09
Overall AAD	**9.08**	**12.24**	**6.53**

indicating that the proposed approach using both data sets can achieve more accurate results than using only one single data set. Visual inspection of the fused images in Fig. 3 also confirms the better performance of the proposed approach using both data sets. As shown in Fig. 3(d), the results using the stereo images tend to underestimate the tall buildings, whereas the results using the SAR images have the problem of missing data [see Fig. 3(e)]. All of the aforementioned restrictions have been improved in the fused result using both the SAR and stereo images, and most of the underestimated tall buildings and the missing buildings of the SAR result are well constructed using the proposed fusion approach.

To better evaluate the performance of the different satellite data in building height retrieval, all results are grouped into three categories according to the actual building heights: heights below 50 m, within 50 to 100 m, and above 100 m. The statistical results are given in Table I, from which three findings can be summarized. 1) Stereo images provide promising result when the heights are below 50 m, as the AAD index is 6.27 m, indicating that these images are suitable for height retrieval in low-density urban areas where most of the buildings are below tens of meters. 2) SAR images achieve extremely good results for tall buildings with heights above 100 m, as the AAD index is 1.88 m compared with the actual data. This result reflects that SAR images are good at predicting the heights of tall buildings and are suitable for building height retrieval in areas where different types of buildings are sparsely distributed. 3) Using both data sets can achieve better prediction results in all height ranges, particularly for the range between 50 and 100 m, where both SAR and stereo images have certain limitations. The AAD index for the fused result using both data sets is 10.76 m in the range of 50–100 m, whereas the AAD indexes for the results with the SAR and stereo images are 14.25 and 15.98 m, respectively. Therefore, the fused results using both the SAR and stereo images are most suitable for building height retrieval tasks in a complex urban scenario where a single data source has limitations.

Scatterplots were produced for all of the results, with the x-coordinates being the actual building height and the y-coordinates the predicted value. Fig. 4(a)–(c) reflects the scatterplots of three different results generated from the SAR images, the stereo images, and both images together, respectively. We can see that the result from the fused approach is closer to the 1–1 line than the results from using the stereo images or the SAR images alone. From the scatterplots, it is apparent that the tall buildings are underestimated by the stereo images, and the SAR result has a large prediction error in buildings with heights below 100 m. Nevertheless, all of the results fall close to the 1–1 line, indicating that each data set achieved an acceptable

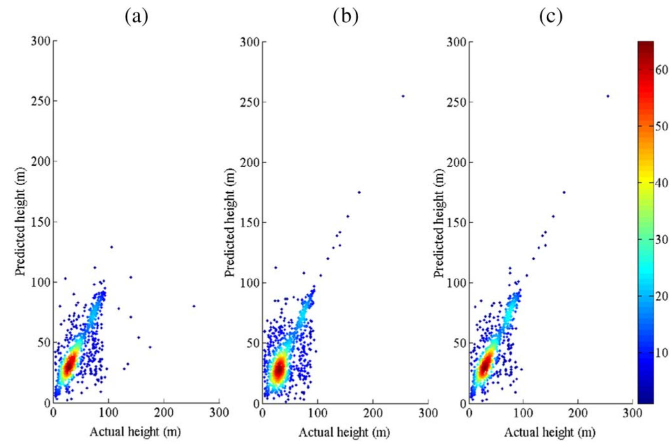

Fig. 4. Scatterplots of the predicted results compared with the actual building height, where the colors indicate the density of the points. (a) Result using the stereo images. (b) Result using the SAR images. (c) Fused result with the proposed approach using both the stereo and SAR images.

result. The linear fitting gains are 0.743, 0.845, and 0.898 for Fig. 4(a)–(c), respectively.

IV. Conclusion and Discussion

In this paper, both stereo and SAR images were jointly used to retrieve the heights of buildings in a high-density urban area. The results show that the fused approach outperforms the use of either stereo images or SAR images alone. The AAD using the proposed approach is 6.53 m, much lower than those using stereo images or SAR images alone, where the average absolute errors are 9.08 and 12.24 m, respectively. The benefit of the proposed fused approach is that the advantages of both stereo and SAR images are well combined, as SAR images are good at predicting the height of tall buildings, whereas stereo images can provide the complementary height information of buildings for which SAR is missing.

The proposed approach is a novel object-based fused approach, in which the fusion process is conducted based on the extracted objects of the SAR results and matched buildings. This approach can help to find more accurate height points for building height retrieval, and at the same time, the predicted results wrongly located on neighboring buildings can be removed, which improves the final fused result. Compared with the results using single data sets, the fused results using SAR and stereo images are more suitable for building height retrieval tasks in complex urban scenarios. By classifying the results into three categories, we also found that stereo and SAR images are still suitable for retrieving the heights of buildings for some urban scenarios. In particular, stereo images are suitable for height retrieval in low-density urban areas where most of the buildings are below tens of meters, whereas SAR images are suitable for areas where different types of buildings are sparsely distributed.

Although the proposed fusion approach has been validated as effective for building height retrieval, one limitation of the approach is that prior building footprints are required to find matched objects for height retrieval. Therefore, further research should integrate both building footprints and building height extraction when using the proposed fusion approach.

References

[1] T. E. Avery and G. L. Berlin, *Fundamentals of Remote Sensing and Airphoto Interpretation*. New York, NY, USA: Macmillan, 1992.

[2] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[3] L. Zhang and A. Gruen, "Multi-image matching for DSM generation from IKONOS imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 60, no. 3, pp. 195–211, Feb. 2006.

[4] G. A. Licciardi *et al.*, "Retrieval of the height of buildings from WorldView-2 multi-angular imagery using attribute filters and geometric invariant moments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 71–79, Feb. 2012.

[5] C. Q. Zeng, J. F. Wang, W. F. Zhan, P. J. Shi, and A. Gambles, "An elevation difference model for building height extraction from stereo-image-derived DSMs," *Int. J. Remote Sens.*, vol. 35, no. 22, pp. 7614–7630, Jul. 2014.

[6] M. Izadi and P. Saeedi, "Three-dimensional polygonal building model estimation from single satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2254–2272, Jun. 2012.

[7] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[8] P. F. Ma, H. Lin, H. X. Lan, and F. L. Chen, "On the performance of reweighted L1 minimization for tomographic SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 895–899, Apr. 2015.

[9] S. Guillaso, L. F. Ferro-Famil, A. Reigber, and E. Pottier, "Building characterization using L-band polarimetric interferometric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 2386–2395, Jul. 2005.

[10] E. Colin-Koeniguer and N. Trouve, "Performance of building height estimation using high-resolution PolInSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5870–5879, Sep. 2014.

[11] E. Simonetto, H. Oriot, and R. Garello, "Rectangular building extraction from stereoscopic airborne radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2386–2395, Oct. 2005.

[12] R. Guida, A. Iodice, and D. Riccio, "Height retrieval of isolated buildings from single high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2967–2979, Jul. 2010.

[13] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, "Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1487–1504, Mar. 2010.

[14] S. Auer, S. Hinz, and R. Bamer, "Ray-tracing simulation techniques for understanding high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1445–1456, Mar. 2010.

[15] F. Xu and Y. Q. Jin, "Automatic reconstruction of building objects from multiaspect meter-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2336–2353, Jul. 2007.

[16] H. Sportouche, F. Tupin, and L. Denise, "Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3932–3946, Oct. 2011.

[17] G. Q. Zhou and X. Zhou, "Seamless fusion of LiDAR and aerial imagery for building extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7393–7407, Nov. 2014.

[18] A. Sampath and J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1554–1567, Mar. 2010.

[19] S. Eckert and T. Hollands, "Comparison of automatic DSM generation modules by processing IKONOS stereo data of an urban area," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 2, pp. 162–167, Jun. 2010.

[20] J. D. Wegner, J. R. Ziehn, and U. Soergel, "Combing high-resolution optical and InSAR features for height estimation of buildings with flat roofs," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5840–5854, Sep. 2014.

[21] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB [M]*. Delhi, India: Pearson Education, 2004.