

An Efficient Approach for Automatic Rectangular Building Extraction From Very High Resolution Optical Satellite Imagery

Jun Wang, Xiucheng Yang, Xuebin Qin, Xin Ye, and Qiming Qin

Abstract—This letter presents a new approach for rapid automatic building extraction from very high resolution (VHR) optical satellite imagery. The proposed method conducts building extraction based on distinctive image primitives such as lines and line intersections. The optimized framework consists of three stages: First, a developed edge-preserving bilateral filter is adopted to reduce noise and enhance building edge contrast for preprocessing. Second, a state-of-the-art line segment detector called EDLines is introduced for the real-time accurate extraction of building line segments. Finally, we present a graph search-based perceptual grouping approach to hierarchically group previously detected line segments into candidate rectangular buildings. The recursive process was improved through the efficient examination of geometrical information with line linking and closed contour search, in order to obtain more reasonable omission and commission rate in building contour grouping. Extensive experiments performed on VHR optical QuickBird imageries justify the effectiveness and robustness of the proposed linear-time procedure with an overall accuracy of 80.9% and completeness of 87.3%. This method does not require user intervention and thereby has the potential to be adopted in online applications and industrial use in the near future.

Index Terms—Building extraction, geospatial object-based image analysis (GEOBIA), line segment detection, perceptual grouping, very high resolution (VHR).

I. INTRODUCTION

BUILDING extraction based on very high resolution (VHR) satellite images is one of the most challenging topics in the fields of remote sensing image understanding and computer vision. Indeed, with the advent of new-generation sensors such as IKONOS, EROS, OrbView, QuickBird, SPOT5, WorldView, GeoEye, and Pleiades, the automatic detection of geospatial objects is throughout an important research subject in geospatial object-based image analysis. Due to the complicated background interference, shooting angle, shadows, and illumination condition, it is hard to extract buildings from VHR satellite imageries. The accurate and reliable extraction of buildings remains a daunting challenge for urban-area segmen-

tation [1], [2] and classification [3], automatic map generation, military simulation, disaster management, and urban planning applications.

With the development of automatic-target-recognition techniques, various kinds of methods for building extraction from VHR optical imagery have been proposed over the past decades [4]–[6]. To the best of our knowledge, because of different materials of building roofs, it is not easy to distinguish buildings from backgrounds using spectral-based classification. In this case, regular shape and line segments of man-made buildings have been confirmed to be the most consistent and distinguishable features for recognition. Taking building boundaries as the primary clue, literature review has been conducted focused on building extraction from VHR imagery. Cote and Saeedi [7] developed a framework combining distinctive corners and variational level set formulation to extract the building outlines recently. Mayer [8] assessed building extraction approaches with well-defined criteria, taking into account both models and strategies. Meanwhile, Shufelt [9] conducted performance evaluation of four building extraction systems based on comprehensive comparative analysis and stated that new techniques were required to achieve better building extraction performances. More precisely, our investigation shows that the related previous works always encounter the main weakness arising from the low time efficiency of traditional line detection algorithms. For instance, Cui *et al.* [10] used the Hough transform (HT) [11] to extract buildings, but HT had notable drawbacks in parameter tuning and time complexity. Although Burns proposed a line segment detection algorithm using only gradient orientations [12] and it could reach relatively lower operation time, it had problems of conspicuous parameter tuning and detected numerous false positive buildings. Building extraction based on the Burns algorithm was described in [5]. After that, a *contrario* model based on a general perception principle called the Helmholtz principle was used in Desolneux *et al.*'s method [13] to improve the Burns algorithm, and it detected few false positives but was indeed an exhaustive algorithm. Therefore, the exploration of an efficient building extraction method based on rapid and accurate line detection is a possible route to avoid the common disadvantages of time-consuming process in these studies.

Based on some previously presented approaches, the intuition of this letter is to develop an efficient procedure for automatic building extraction to address the issue of time complexity in practical applications. In accordance with Marr's vision theory, a bottom-up procedure is implemented in our study. The extraction strategy starts from low-level image

Manuscript received November 15, 2013; revised March 20, 2014 and June 30, 2014; accepted August 3, 2014. Date of publication August 28, 2014; date of current version September 16, 2014. This work was supported by the National High-Tech Research and Development Program of China (2012AA121305).

The authors are with Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China (e-mail: qmqinpk@163.com).

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.2014.2347332

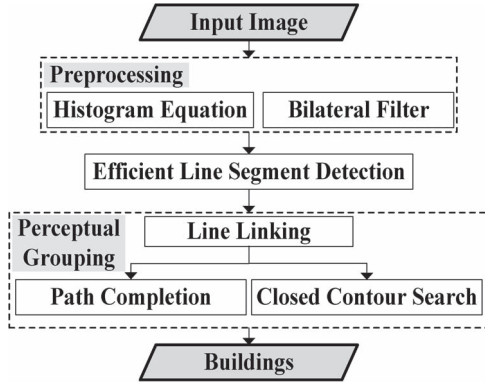


Fig. 1. Framework of the proposed procedure.

primitives (e.g., linear structures), to middle-level boundary polygon sets, and, finally, to high-level geographical semantic objects such as buildings.

II. BUILDING EXTRACTION PROCEDURE

A framework combining low-level image primitive detection with real-time line segment detection and middle-level perceptual grouping is adopted for automatic and accurate building extraction from VHR optical imagery. Fig. 1 shows the framework of the procedure.

The proposed approach consists of three stages: 1) The VHR satellite imagery is preprocessed by histogram equalization and bilateral filter [14] to enhance image contrast and decrease noises simultaneously; 2) the real-time line segment detector called EDLines proposed by Akinlar and Topal [15] is used to extract the line segments from the preprocessed image; 3) the perceptual grouping is applied to reorganize the relationship of detected line segments and to gain complete contours of buildings, which are converted to ESRI Shapefile (SHP) format for further applications.

A. Preprocessing

Considering both noise reduction and edge preserving, histogram equalization and bilateral filter are chosen to preprocess the input image. Histogram equalization is a powerful tool for enhancing global and local contrast. It is used to maintain local detail and stretch the gray levels of targets. In addition, inevitably, the Gaussian smoothing filter would weaken edges; therefore, this letter introduced a bilateral filter [14], an edge-preserving and noise-reducing smoothing filter for preprocessing, and more importantly, it is noniterative and efficient especially. Hence, compared with the traditional Gaussian filter, the bilateral filter considers the spatial distribution of image intensities to balance closeness and similarity to neighbor pixels. It can enhance contrast and preserve contours in a way tuned to human perception.

The bilateral filter operation can be represented by the following equation:

$$h(x) = k(x) \int_{\omega} f(w)c(x, w)s(f(x), f(w)) dw \quad (1)$$

where $f(x)$ and $h(x)$ represent the input image and output image, respectively. x represents the coordinates of the neighbor

center. The convolution kernels $c()$ and $s()$ are associated with the spatial and intensity domain, respectively. The calculation is computed over a nearby point ω which is a neighborhood of center pixel x . The normalization factor $k(x)$ is computed as

$$k(x) = \int_{\omega} c(x, w)s(f(x), f(w)) dw. \quad (2)$$

Here, Gaussian kernels for $c()$ and $s()$ are used to implement the closeness and similarity functions, respectively. The c kernel and s kernel can be calculated as

$$c(x, w) = e^{-\left(\frac{\|x-w\|^2}{\sigma_c^2}\right)} \quad (3)$$

$$s(f(x), f(w)) = e^{-\left(\frac{(f(x)-f(w))^2}{\sigma_s^2}\right)}. \quad (4)$$

Here, σ_c in (3) is to define the closeness of neighbor pixels, while σ_s in (4) is to describe the intensity of neighboring pixels. In this letter, σ_c and σ_s were fixed to 5 and 10, respectively.

B. Efficient Line Segment Detection

After reviewing existing line segment extracting algorithms, a state-of-the-art line segment detector called EDLines is adopted for automatic extraction of building boundary lines. EDLines is proposed by Akinlar and Topal [15]. It is an automatic algorithm without requiring any parameter tuning and runs faster than other existing line detection algorithms. EDLines made an improvement on the Burns algorithm using a false positive control criterion named the Helmholtz principle inspired by Desolneux *et al.* [13]. The criterion absorbed the advantages of previous line segment detection algorithms while avoiding most of the drawbacks; therefore, it can obtain meaningful line segments and remarkably reduced computation cost. Zhang *et al.* [16] reviewed previous line extraction methods based on repeatability and reported that EDLines was the optimal line detection algorithm at different scales, viewpoint degree, and illumination.

In the EDLines algorithm, we replace its Gaussian smoothing filter with the previously introduced bilateral filter to improve preprocessing. After that, the number of false alarms (NFA) [13] is utilized to validate the result with the Helmholtz principle. NFA is described as follows: In a $N \times N$ pixel image, suppose A to be a segment of length “ n ” with at least “ k ” points having their directions aligned with the direction of A. The NFA of A is computed as

$$NFA(n, k) = N^4 \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}. \quad (5)$$

Given that a line segment has two end points in the image and each end point can appear in any one of the $N \times N$ pixels, there exist a total of $N^2 \times N^2 = N^4$ line segments. The probability “ p ” used in the computation is the accuracy of the line direction. The validation of EDLines accepts the line segment if $NFA(n, k) \leq 1$. Otherwise, the line is rejected. The validation step of EDLines substantially controls the number of false detections.

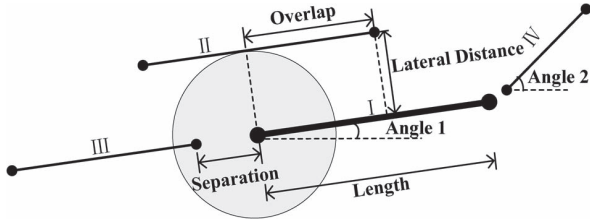


Fig. 2. Collinear line linking criteria.

C. Perceptual Grouping

Perceptual grouping is proved to be a vital tool for establishing structural buildings from line segments [4], [5], [17]. It is demonstrated to be effective in dealing with the issue of gaps and missing segments existing among the fragmented building boundaries. In order to establish a relationship between the low-level image primitives such as line segments and the higher level geospatial objects (e.g., intersections and closed boundaries), the processing step includes collinear line linking and line intersection detection. First, four improved collinear line linking criteria suggested by Izadi and Saeedi [5] are developed to perceptually group previously detected lines. The criteria of perceptual grouping for linking collinear line segments are illustrated in Fig. 2:

- 1) $\text{overlap}/\text{length} < 15\%$ for lines I and II;
- 2) lateral distance < 5 pixels (3 m) for lines I and II;
- 3) separation < 10 pixels (6 m) for lines I and III;
- 4) $|\text{Angle}1 - \text{Angle}2| \leq \pi/10$ for the angle between lines I and IV.

Based on proximity and collinearity constraints, each pair of nearby line segments is evaluated using the aforementioned criteria. Moreover, only the lines jointly satisfying these criteria are left for further grouping. All linked lines at this stage are saved for future use in the following building contour establishment.

The proposed methodology employs image primitives such as lines and line intersections, which act as potential components of building candidate hypotheses. After the line linking process, we examine their relationships using a graph search to establish a set of building rooftop hypotheses. Considering that most building roofs in VHR images appear as rectangular and rectangular shapes, a path completion for rectangular buildings is designed. The distinctive right-angle structure consisting of two perpendicular edges and a corresponding vertex is formed as a basic structure. These structures represent potential vertices of building candidate hypotheses; then, pairs of right-angle structures would produce the polygonal building outlines. The searching process will continue until a building contour loop is found. Therefore, the graph search technique serves as a bridge between low-level line segments and high-level semantic building outlines. After low-level lines are obtained, a typical demonstration of invoked searching strategy is shown in Fig. 3: A perceptual link can be formed by turning left or right to connect the complete building outlines. Over previously detected line segments, take the red and green line in Fig. 3 as an example; if separation ≤ 10 pixels and $\pi/2 - \pi/10 \leq \text{intersection angle} \leq \pi/2 + \pi/10$, the geometrically related line-corner-line groups are produced, and eventually, based on them, the boundary line intersections will be connected and formed into rectilinear building boundaries.

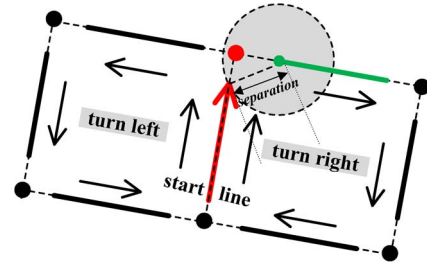


Fig. 3. Search direction and right-angle linking criteria of path completion.

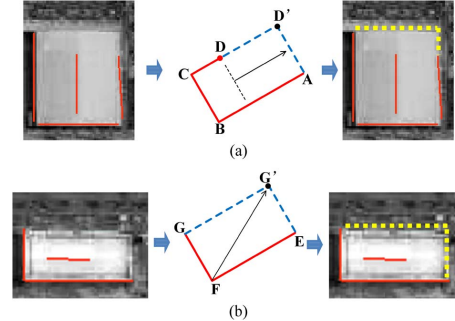


Fig. 4. Path completion of U-shape and L-shape building boundary.

However, due to image quality and background interference, some inconspicuous boundary lines (labeled in yellow in Fig. 4) failed to be detected in the previous line detection section. Although not all the boundary lines exist in the line sets, according to the prior knowledge regarding building boundary shape, we complete the unclosed U-shape [see Fig. 4(a)] contour based on the longest border AB. On the other hand, for the isolated L-shape structure [see Fig. 4(b)], the undetected vertex G' is completed using the diagonal line of the present building block. Additionally, since shadows usually appear quite dark in images, false alarm structures generated from shadow boundaries could be excluded using a gray thresholding method in the process of path completion, and the structure with an inner gray level ≤ 50 is considered as shadow. The path completion procedure has the potential to reduce omission errors to some extent compared with some previously described approaches.

Sometimes, multiple rectangles would be produced within a roof due to different materials and structures, and in order to conduct the rectangle merging, each intersecting rectangle was considered as a node. The relationship between them represents the edge. In this way, it formed an undirected graph. Then, all the connected components in the graph were computed, and the minimum bounding rectangle of each connected component was considered as the final rooftop rectangle after merging.

III. EXPERIMENTS AND DISCUSSION

To assess the performance of the proposed method, experiments were conducted on four 512×512 pixel and a 806×806 pixel pan-sharpened QuickBird satellite imageries with 0.6-m resolution. The proposed method was run on a PC with an Intel Core2 2.8-GHz CPU and 2-GB RAM, implemented in C++. As shown in Fig. 5, the images on the first row indicate the original input images, and the detection results of EDLines are shown in the second row. The third row denotes the line linking results of previously detected lines, overlaid on preprocessed

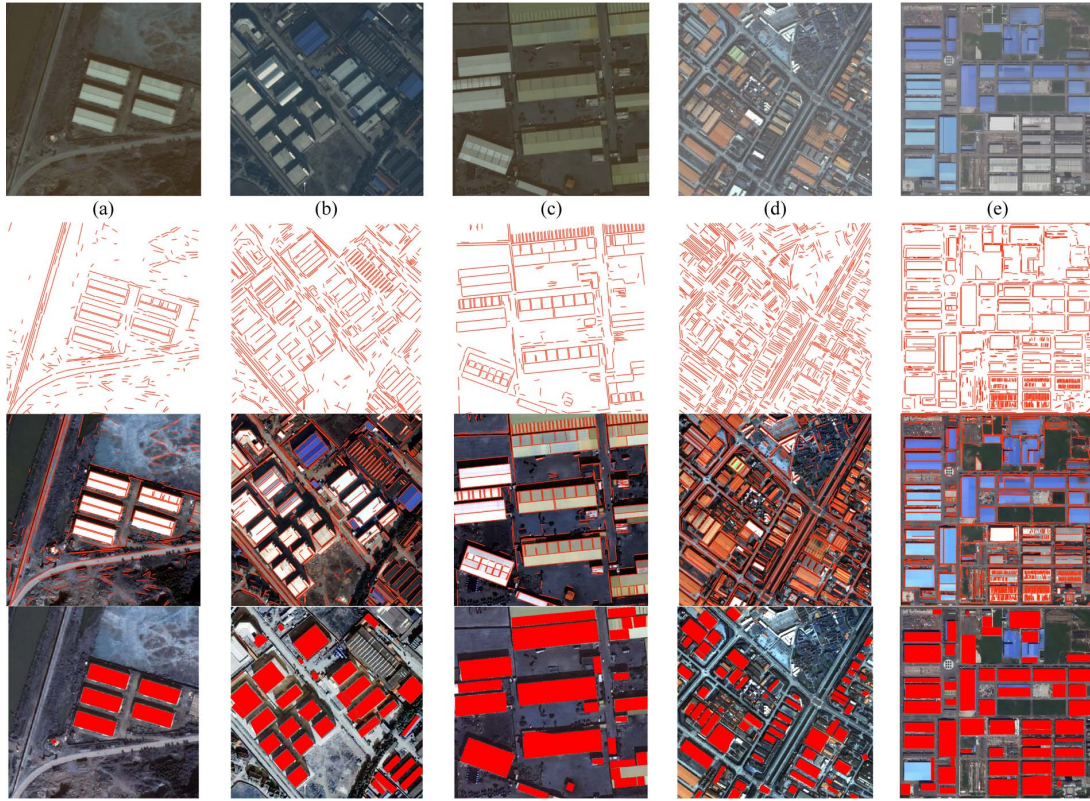


Fig. 5. (First row) Original images. (Second row) Line segment detection results. (Third row) Line linking results. (Fourth row) Building extraction results.

TABLE I
ACCURACY ASSESSMENT OF PROPOSED METHOD

| Total No. of buildings | No. of TP | No. of FP | No. of FN | detection rate (DR) | false negative rate (FNR) | Overall Accuracy |
|------------------------|-----------|-----------|-----------|---------------------|---------------------------|------------------|
| 165 | 144 | 21 | 13 | 87.27% | 8.28% | 80.90% |

images. The last row illustrates the final building extraction results. Image (e) with 806×806 pixels is zoomed to the same size of other images for demonstration. Obviously, the experimental data sets contain most of the sizes and materials of buildings to guarantee the experiment's validity and stringency.

The accuracy assessment is demonstrated in Table I. To quantify the accuracy of the building extraction results, we follow the frequently used metric which is the detection rate (DR), i.e., correctness. It measures the degree to which detected buildings indeed are real buildings. Meanwhile, the false negative rate (FNR) measures the degree of missed detections by the procedure compared to the total real buildings. Overall accuracy denotes a composite metric that takes into consideration both correctness and completeness, and these metrics are computed as follows:

$$DR = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\%$$

$$FNR = \frac{N_{FN}}{N_{TP} + N_{FN}} \times 100\% \quad (6)$$

$$Overall\ Accuracy = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \times 100\%. \quad (7)$$

Here, N_{TP} , N_{FP} , and N_{FN} represent the number of true positives (i.e., correctly extracted buildings), false positives

TABLE II
PERFORMANCE EVALUATION OF PROPOSED METHOD

| Image ID | (a) | (b) | (c) | (d) | (e) |
|-------------------------------|------|------|------|------|------|
| EDLines (s) | 0.07 | 0.08 | 0.03 | 0.08 | 0.09 |
| Number of lines | 330 | 679 | 439 | 1054 | 1256 |
| Number of lines after linking | 284 | 589 | 394 | 956 | 1123 |
| Total extraction time (s) | 13 | 15 | 11 | 12 | 16 |

(i.e., incorrectly extracted buildings), and false negatives (i.e., missing buildings), respectively.

Table II summarizes the total processing time and detection results of EDLines for all test images. The EDLines-based line segment detection only took 0.07, 0.08, 0.03, 0.08, and 0.09 s for images (a), (b), (c), (d), and (e), respectively. The nearly real-time procedure laid a solid foundation for efficient building extraction. Meanwhile, we obtained 330, 679, 439, 1054, and 1256 lines for each scene in turn, and collinear line segments were merged based on linking criteria. Finally, the whole extraction procedure cost 13, 15, 11, 12, and 16 s for images (a), (b), (c), (d), and (e), respectively. Thus, the performance evaluation shows that the proposed method offers a slight performance improvement with an average processing time of approximately 13 s for images with the number of pixels between 262 144 (512×512) and 649 636 (806×806).

In computing the aforementioned results, Table III compares the performances of some previous works with that of the proposed method. Although these listed algorithms were not run directly, their elapsed time and performance presented in relevant papers were taken and compared to the proposed

TABLE III
COMPARISON OF THE PERFORMANCE WITH PREVIOUS WORK

| Method | Noronha [4] | Izadi [5] | OK [6] | Cote [7] | Mayunga [17] | Our method |
|---------------|----------------|--------------|-------------|-------------|-----------------|---------------|
| time (s) | 225 | 1387 | 184 | 120 | 60 | 13 |
| DR | 95.4% | 95.2% | 83.6% | 91.2% | 91.0% | 87.3% |
| Image size | 800× 400 | 400× 400 | 641× 863 | 400× 400 | 600× 600 | 512× 512 |

approach under similar computing environments, considering an average input image size of around 500×500 , which is about half a million pixels with spatial resolution finer than 1 m. The algorithms in [4], [5], and [17] showed fine extraction results with higher DR, but both cost a notably longer operation time. Meanwhile, it takes, on average, around 120 s per rooftop to complete extraction in a recent study [7]. It can be observed that our method delivers an average efficiency better than others with acceptable accuracy.

Based on the visual assessment, the proposed approach can successfully pinpoint the true positive buildings. According to quantitative evaluation results in Table I, buildings can be detected within acceptable time with an overall accuracy of 80.9% and detection rate of 87.3%. As expected, the extracted buildings had a high positional accuracy and shape precision. However, due to the background interference and other factors, 21 buildings failed to be extracted because of their low contrast to the surrounding urban impervious surface. Moreover, in dense residential areas, those closely distributed small buildings might be merged and extracted as a single building patch and hard to be separated accurately, which inevitably led to some missed detections. On the other hand, the proposed method might extract and label several nonbuilding regions such as crossroads, parking lots, lawns, etc. Given this, there existed 13 missing detected buildings with a false negative rate of 8.28% for the total 165 buildings in several different images. Nevertheless, quantitative evaluations demonstrate that the output of the proposed approach can provide rapid and accurate building outlines, which proves to be robust and flexible for building extraction tasks. Moreover, the promotion of building extraction efficiency may become more apparent with the increase of image size. Generally, similar to the automatic method in [18], user intervention is not required in our process, and hence, it can be easily extended through a simple modification of the image term; a whole data set can be effectively dealt with through a batch-processing mode. Taking into account the computation cost, the described system is a relatively preferable solution for efficient building extraction.

IV. CONCLUSION

This letter concerns the problem of time complexity in practical applications. In this letter, an efficient approach is proposed for automatic rectangular building extraction from VHR satellite imagery using the state-of-the-art line segment detector coupled with developed perceptual grouping strategy. The experiment results reveal that the proposed procedure can remarkably improve building extraction efficiency. Furthermore, through the combination of collinear line linking, path completion, and perceptual contour search, we have obtained promising building extraction results with an overall accuracy

of 80.9%. Thus, the proposed approach leads to a notable reduction of computation cost with similar performance compared with traditional solutions. Moreover, user intervention is not required in our method, which makes it automatic enough for further image understanding, target detection, and scene classification tasks. In particular, it can be introduced to online applications and industrial usage in the near future.

Future works are required on the following issues. First, an integration of corner detection with line segment detection is required for more accurate extraction of arbitrarily shaped buildings on larger data sets. In addition, with the development of segmentation algorithms, modeling for the combination of boundary and partitioned region should be our future work. As a final future work, a really synthetic approach with fusion of spatial relationship, shadow, spectral, and texture features may improve the overall extraction accuracy.

REFERENCES

- [1] L. Weizman and J. Goldberger, "Urban-area segmentation using visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 388–392, Jul. 2009.
- [2] R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical texture-based segmentation of multiresolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2129–2141, Jul. 2009.
- [3] T. Novack, T. Esch, and H. Kux, "Machine learning comparison between Worldview-2 and Quickbird-2-simulated imagery regarding object-based urban land cover classification," *Remote Sens.*, vol. 3, no. 10, pp. 2263–2282, Oct. 2011.
- [4] S. Noronha and R. Nevatia, "Detection and modeling of buildings from multiple aerial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 5, pp. 501–518, May 2001.
- [5] 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.
- [6] A. O. Ok, "Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 21–40, Dec. 2013.
- [7] M. Cote and P. Saeedi, "Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 313–328, Jan. 2013.
- [8] H. Mayer, "Automatic object extraction from aerial imagery—A survey focusing on buildings," *Comput. Vis. Image Understand.*, vol. 74, no. 2, pp. 138–149, May 1999.
- [9] J. A. Shufelt, "Performance evaluation and analysis of monocular building extraction from aerial imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 311–326, Apr. 1999.
- [10] S. Y. Cui, Q. Yan, and P. Reinartz, "Complex building description and extraction based on Hough transformation and cycle detection," *Remote Sens. Lett.*, vol. 3, no. 2, pp. 151–159, Mar. 2012.
- [11] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, 1981.
- [12] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 4, pp. 425–455, Jul. 1986.
- [13] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 7–23, Oct. 2000.
- [14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
- [15] C. Akinlar and C. Topal, "EDLines: A real-time line segment detector with a false detection control," *Pattern Recognit. Lett.*, vol. 32, no. 13, pp. 1633–1642, Oct. 2011.
- [16] Y. S. Zhang, Y. Liu, and Z. R. Zou, "Comparative study of line extraction method based on repeatability," *J. Comput. Inf.*, vol. 8, no. 24, pp. 10097–10104, 2012.
- [17] S. D. Mayunga, D. J. Coleman, and Y. Zhang, "A semi automated approach for extracting buildings from QuickBird imagery applied to informal settlement mapping," *Int. J. Remote Sens.*, vol. 28, no. 10, pp. 2343–2357, May 2007.
- [18] K. Karantzas and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 133–144, Jan. 2009.