# Accurate Outline Extraction of Individual Building from Very High Resolution Optical Images

Xuebin Qin, Shida He, Xiucheng Yang, Masood Dehghan, Qiming Qin, and Martin Jagersand

Abstract—This letter presents a novel approach for extracting accurate outlines of individual buildings from Very-High-Resolution (VHR, 0.1-0.4 m) optical images. Building outlines are defined as polygons here. Our approach operates on a set of straight line segments that are detected by a line detector. It groups a subset of detected line segments and connects them to form a closed polygon. Particularly, a new grouping cost is defined firstly. Second, a weighted undirected graph G(V,E) is constructed based on endpoints of those extracted line segments. The building outline extraction is then formulated as a problem of searching for a graph cycle with the minimal grouping cost. To solve the graph cycle searching problem, the "Bi-Directional Shortest Path (BDSP)" method is utilized. Our method is validated on a newly created dataset which contains 123 images of various building roofs with different shapes, sizes and intensities. The experimental results with average Intersectionover-Union of 90.56% and average alignment error of 6.56 pixels demonstrate that our approach is robust to different shapes of building roofs and outperforms state-of-the-art method.

*Index Terms*—Building recognition, outline extraction, perceptual grouping, graph optimization.

# I. INTRODUCTION

CCURATE building outline extraction is important to urban planning, cadastral surveying and other related applications. Building outlines extraction from very high resolution (VHR) images has become a popular yet challenging topic in the fields of photogrammetry, remote sensing, geographic information system (GIS) and computer vision.

There already exist many methods for building outlines extraction from aerial and satellite images. Image segmentation based methods [1], [2], [3], [4], [5] is a popular class of methods. Cote and Saeedi [6] combine distinctive corners detection with level-set method to fit the best possible boundaries of building rooftop. Song and Shan [7] adopt active contour models and intensity based cluster to extract building boundaries from satellite images. Yang and Wang [8] extract building contours using shape priors constrained level set method. These methods are sensitive to initialization, local minimum and noise. In addition, perceptual grouping [9] [10] and deep learning based approaches [11] have also been proposed. However, the approach proposed in [9] is not

Qiming Qin is with Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China e-mail: qmqinpku@163.com.

Xiucheng Yang is with ICube Laboratory, University of Strasbourg, Strasbourg, 67081, France e-mail:xiucheng.yang@etu.unistra.fr.

able to extract non-rectangular building outlines. Because its grouping rules are designed according to the basic attributes of rectangles: four sides with right angles. Learning based building extraction methods [11] require large number of training data and their results are usually region masks with relatively coarse contours. Little attention has been paid to obtaining accurate and detailed outlines of individual buildings.

This work focuses on extracting accurate outlines of individual buildings with different shapes. The accurate outline extraction is formulated as a salient object detection problem. There are mainly two classes of salient object detection methods: intensity based and edge based.

The intensity based methods detect objects using the saliency measure defined on the difference or contrast between foreground and background pixels. Yang et al. [12] utilize superpixels of a given image as nodes to construct a close-loop graph. The saliency maps are obtained by ranking these nodes according to their similarities to background and foreground queries based on affinity matrices. Zhu et al. [13] propose a background measure to characterize the spatial layout of image regions with respect to image boundaries. They integrate multiple low level cues and the background measures into their optimization problem. Srivatsa and Babu [14] estimate the foreground regions using objectness proposals and then other pixels/regions are weighted by their proposed saliency measure. They integrate these weights into an optimization framework to obtain the final saliency map. Zhang et al. [15] develop a salient object detection method by solving an approximate Minimum Barrier Distance (MBD) Transform, which achieves 100X speedup over the exact MBD algorithm. The outputs of these methods are usually saliency maps with coarse or uneven boundaries of the target objects.

The edge based methods usually define saliency measure based on the properties of to-be-extracted object boundaries, such as curvature, gap, length, enclosed region area, etc. Gestalt laws [16] and perceptual grouping are their theoretical basis.

Kokkinos [17] proposes a fractional-linear programming approach to find the most salient boundary. Lu *et al.* [18] create a contour saliency measure subject to completeness and tightness constraints, and optimize it using dynamic programming in polar coordinate system. However, the transformation between Cartesian coordinate system and polar coordinate system decreases the accuracy of boundaries. Wang *et al.* [9] extract rectangular building outlines by grouping straight line segments according to their distances and angles, but only simple rectangular buildings can be extracted and several parameters have to be tuned carefully. Wang *et al.* [19], [20]

Xuebin Qin is supported by China Scholarship Council (CSC) and Alberta Innovates Technology Features (AITF).

X. Qin, S. He, M. Dehghan and M. Jagersand are with the Dept. of Computing Science, University of Alberta, Edmonton, AB, T6G 2R3, Canada e-mail: {xuebin, shida3, masood1, mj7}@ualberta.ca.

develop a graph-based method for salient closed boundaries detection. This method selects and connects a subset of edge fragments sequentially to form a closed boundary with the saliency maximum. Although they can handle more irregular shapes than [9], the cost function in their optimization is not adapted for building outline extraction.

As buildings are artifacts with rich straight line features, our method is based on grouping detected straight line segments, similar to [19], [20]. The main contributions of this work are: (1) a novel grouping cost for individual building outline extraction; (2) a novel optimization framework that is able to handle cost functions tailored for grouping costs with different formats; Part of the optimization algorithm is presented in [21] for real-time tracking of objects in videos.

The rest of this letter is organized as follows. Section II describes the details of proposed grouping cost and the optimization approach. Experimental results and comparison are given in Sec. III, with the concluding remarks in Sec. IV.

## II. PROPOSED METHOD

Given a VHR image, our goal is to group a subset of detected line segments and form a polygon which describes the accurate outline of a target building. First, a novel grouping cost, which is composed of the completeness (or closure) and smoothness (or continuation) [16], is defined. Then the straight line segments are detected from the given image to construct a weighted undirected graph G(V, E). Finally, the grouping process is formulated as searching for the optimal cycle in the weighted undirected graph.

#### A. Derivation of Grouping Cost

1) Completeness: An intuitive definition of completeness requires the ratio of non-edge pixel number over the total pixel number along a closed curve to be sufficiently small (the optimum is zero) [18]. Given a closed curve  $\Gamma$ , the completeness cost  $C(\Gamma)$  is defined:

$$C(\Gamma) = \frac{\oint_{\Gamma} e(s)ds}{\oint_{\Gamma} ds} \tag{1}$$

where

$$e(s) = \begin{cases} 0 & s \text{ is an edge pixel} \\ 1 & s \text{ is not an edge pixel} \end{cases}$$
(2)

2) Smoothness: The smoothness cost  $S(\Gamma)$  of a closed curve  $\Gamma$  is given by the Elastic prior [17], [22]:

$$S(\Gamma) = \frac{\oint_{\Gamma} |\kappa(s)| ds}{\oint_{\Gamma} ds}$$
(3)

where  $\kappa(s)$  is the curvature of edge pixel. The numerator denotes the total absolute curvature and measures how far the curve is from a convex curve [23].

Our method operates on a set of straight line segments that are extracted by a line detector. As the curvature at straight line intersections is not well defined, we approximate the total absolute curvature term of (3) as:

$$\oint_{\Gamma} |\kappa(s)| ds \propto \frac{\oint_{\Gamma} ds}{\iint_{R} dA} \tag{4}$$



Fig. 1. Graph construction and optimization

where R is the region enclosed by contour  $\Gamma$ . The smoothness term can be further simplified using (3) and (4) to:

$$S(\Gamma) \propto \frac{1}{\iint_R dA}$$
 (5)

In other words, this term is reciprocal to the area of region R enclosed by  $\Gamma$ .

3) Grouping Cost: Grouping cost is typically formulated as a weighted summation of completeness and smoothness [18]:

$$G(\Gamma) = C(\Gamma) + \lambda S(\Gamma) = \frac{\oint_{\Gamma} e(s)ds}{\oint_{\Gamma} ds} + \lambda \cdot \frac{1}{\iint_{R} dA}$$
(6)

where  $\lambda$  is a weight that balances the completeness and smoothness. However, (1) and (5) are two incomparable measures with different units. This makes the tuning of  $\lambda$  a hard process for obtaining good results. In fact, the choice of optimal  $\lambda$  is shape-dependent and differs for different images. This will be further elaborated in the experimental results.

This motivate us to propose a new Grouping Cost  $G(\Gamma)$  as the multiplication of these two terms:

$$G(\Gamma) = \frac{\oint_{\Gamma} e(s)ds}{\oint_{\Gamma} ds} \cdot \frac{1}{\iint_{R} dA}.$$
(7)

In this work, the building outlines are depicted by polygons. The Grouping Cost  $G(\Gamma)$  of a polygon, comprised of n detected line segments and k gaps, is defined as:

$$G(\Gamma) = \frac{\sum_{j}^{k} e_j}{\left(\sum_{i}^{n} l_i + \sum_{j}^{k} e_j\right)} \cdot \frac{1}{A}$$
(8)

where  $e_j$  denotes the length of a gap between two sequentially connected line segments and  $\sum_{j}^{k} e_j$  is the total gap length along polygon,  $l_i$  indicates the length of a line segment and  $(\sum_{i}^{n} l_i + \sum_{j}^{k} e_j)$  is the perimeter of the polygon and A is the area of the polygon. The polygon with the smallest  $G(\Gamma)$  is taken as the final optimal outline.

### B. Graph Construction

Given a VHR image, its straight line segments are first detected by EDLines [24]. Fig. 1(a) shows the detected line segments of an VHR image. Each line segment is represented by a pair of endpoints. As detected line segments are not connected, Delaunay Triangulation (DT) [25] is utilized to fill the gaps between them (see Fig. 1(b)). Then detected line segments are superimposed on a generated Triangular Irregular Network (TIN) to construct an undirected graph G(V, E) (see Fig. 1 (c)). Each node in the graph corresponds to an endpoint. A graph edge correspond to either a TIN edge (blue line) or a detected line segment (red lines). Note that there are two types of graph edges and their weights are set differently similar to [20]. For a TIN edge, the weight is set to its geometric length which is in fact the gap length. The weight of a line segment is set to zero. This means that line segments are more likely to be a part of building outlines.

#### C. Optimal Outline Search

A building outline corresponds to a special cycle, which has the minimal grouping cost (8), in the weighted undirected graph G(V, E). Minimum Ratio Weight Cycles (MRWC) [19], [26] are most commonly used algorithms for the grouping cost optimization. However, the denominator of our grouping cost (8) is multiplication of two terms which cannot be solved with MRWC methods. This section describes our algorithm for finding the optimal cycle based on the grouping cost (8). Our search algorithm has two steps: (1) Cycle Candidates generation; (2) optimal outline retrieval.

In the first step, a set of cycle candidates are generated based on our recent method [21] called "Bi-Directional Shortest Path (BDSP)". As shown in Fig. 1 (d), a cycle is considered as a candidate if it consists of the following three components: 1) a zero-weight edge  $E_i$  (white edge); 2) a shortest path from the start node  $V_{is}$  of the edge  $E_i$  to a third node  $V_j$  (yellow path); 3) a shortest path from the end node  $V_{ie}$  of the edge  $E_i$  to the same third node  $V_j$  (green path). For each zero weighted edge, we first set its weight to infinity and then traverse all of the third nodes on the graph to generate cycle candidates by BDSP. This means that for a graph with n zero-weighted edges, there are  $n \times 2(n-1)$  cycle candidates.

In the second step, the grouping cost (8) for each cycle candidate is computed. The cycle candidate with the minimal grouping cost is taken as the final optimal building outline.

## **III. EXPERIMENTAL RESULTS**

#### A. Dataset and Error Metric

To assess the performance of our approach, a new dataset for building outline extraction on VHR aerial and satellite images of urban areas is built using [27]. The resolution of these images varies from 0.1-0.4 m. The dataset<sup>1</sup> contains 123 images of buildings with different shapes. The accurate *Ground Truth* (GT) is manually labeled for all the images.

To quantify building outline extraction results, two accuracy measures are used: a region-based and an edge-based measure.



Fig. 2. Results based on grouping cost (6): the size of top row image is  $116 \times 108$ , the size of bottom row image is  $501 \times 439$ 

The region-based measure is *Intersection-over-Union* (*IoU*) which is defined as the relative region coincidence [19]:

$$IoU = \frac{|R \cap R_{GT}|}{|R \cup R_{GT}|} \times 100\%$$
(9)

where  $R_{GT}$  and R are the regions enclosed by the ground truth and extracted building outline, respectively. |R| indicates the area of R.

The edge-based measure used is the average Alignment Error  $(E_{aveAl})$  which is defined as

$$E_{aveAl} = \frac{Dist_{GT} \otimes I_{\Gamma}}{|I_{\Gamma}|} \tag{10}$$

where  $Dist_{GT}$  is the distance transform map of the ground truth region boundary.  $I_{\Gamma}$  is the extracted binarized outline map and  $|I_{\Gamma}|$  is the perimeter of the outline  $\Gamma$ .

# B. Evaluation and Comparison

Both summation-based grouping cost (6) and multiplicationbased grouping cost (8) can be optimized using our "BDSP" algorithm. However, in (6), the choice of the weight  $\lambda$  has to be tuned carefully for different images, especially when these buildings have big differences in their shapes and sizes. This is further illustrated in Fig. 2, where the results of the algorithm for summation-based cost is shown for different values of  $\lambda$ . It can be seen that it is not possible to achieve a good accuracy on both images using the same  $\lambda$ . Our multiplication-based grouping cost (8), on the other hand, resolves this problem. It works well on both of these two buildings without parameters tuning, see the first column in Fig. 2.

Our method is compared with other five state-of-the-art methods: (i) a regional information combined ratio contour method (RRC) [20], (ii) a minimum barrier salient object detection method (MB+) [15], (iii) a saliency detection method via graph-based manifold ranking (MR) [12], (iv) an objectness measure based salient object detection method (SO) [14], and (v) a saliency optimization based method (wCtr) [13]. The RRC method is the state-of-the-art line-based grouping method. The direct outputs of our method and RRC are grouped polygons. To facilitate comparison, they are presented as binarized region maps in Fig. 3. Methods (ii)-(v) are regional intensity-based methods. Their original results are saliency maps represented by normalized gray scale images (0-255). In our experiments, these saliency maps are thresholded

<sup>&</sup>lt;sup>1</sup>https://webdocs.cs.ualberta.ca/~xuebin/building\_extraction.html



Fig. 3. Sample results of different methods: the first column is the original image, the second column is the ground truth, the third column is the result of our method, column four to column eight are results produced by RRC, MB+, MR, SO and wCtr.



Fig. 4. Region and edge based evaluations. (a), (b) and (c), (d) are region based and edge based evaluations respectively. (a) and (c) are ordered IoU and  $E_{aveAl}$  of the testing images. (b) shows the number of images where the IoU is greater than certain threshold. (d) shows the number of images where the  $E_{aveAl}$  is less than certain threshold.

TABLE I. Overall average IoU and  $E_{aveAl}$ 

Method	Ours	RRC	MB+	MR	SO	wCtr
IoU	90.57	82.67	66.67	72.23	76.36	64.32
$E_{aveAl}$	6.54	9.52	20.17	17.76	13.87	15.87

(the threshold is set to 125 because it provides almost the best overall performance of these methods according to our thresholding tests.) to obtain binarized buildings' regional segmentation. The building outlines are then obtained via edge extraction from those binarized images.

Fig. 3 shows sample results of different methods. The key challenge of building outlines extraction by using grouping based method is to resist the impacts of many detected noisy line segments which are close to the target building. RRC fails easily in complex and concave shapes, see row 1-6.

It is sensitive to noisy line segments and prone to group shorter boundaries. Our method is more robust than RRC to this kind of noisy line segments. MB+, MR, SO, wCtr are more dependent on homogeneous colors or intensities. They are good at extracting building roofs with unified intensities. Theoretically, they are more robust to complex shapes. However, VHR images contains very detailed structures of building roofs and these structures usually have different colors, which confuse the intensity based methods. Detecting building roofs with heterogeneous intensities is difficult for these intensity based methods, see results in row 2 and 4. They are also prone to take salient background as the building roofs, see row 5. It can be seen that our method is robust to different shapes and intensity variations compared to other methods.

Table I summarizes the average IoU and  $E_{aveAl}$  of all the methods over the dataset. Our method achieves 90.56% of



Fig. 5. Illustration of failure cases: Red and green outlines are ground truth and our results, respectively. (a) failure caused by complex structures inside the roof outline. (b) failure due to non-building noisy line segments. (c) failure due to missing edges. (d)-(e) failure due to large surroundings area that includes specific textures or structures; (d) is of size  $281 \times 277$  and (e) is of size  $589 \times 611$ .

the average IoU and 6.56 pixels of  $E_{aveAl}$  (more than 30% improvement in  $E_{aveAl}$ ), which outperform other methods. The overall IoU and  $E_{aveAl}$  trends of each method are shown in Fig. 4a and 4c. To further highlight the robustness of our method, the curves of successfully extracted image numbers with respect to different IoU and  $E_{aveAl}$  thresholds are shown in Fig. 4. It can be seen that our method outperforms all of the others in terms of both IoU and  $E_{aveAl}$ . Although our method outperforms state-of-the-art grouping methods in most cases, there are still extreme cases that result in failure of accurate outline extraction, some of which are shown in Fig. 5.

#### **IV. CONCLUSIONS**

This letter addresses the problem of accurate extraction of complex building outlines from VHR aerial and satellite images. A new outline grouping cost is proposed in terms of a ratio that is normalized relative to outline length and area. Then, a novel and simple framework is introduced for graph construction and outline searching. The results on our newly built dataset demonstrate that our method is robust to buildings with different intensities and shapes. Currently, our method can only extract the most salient building outline from a given image. Hence, the input image has to be roughly cropped around the target building. This prerequisite somehow limits the applications of our method. Future work will focus on extending our method to multiple building outlines extraction from large scale images by integrating oriented object detection methods.

#### REFERENCES

- M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [2] S. Osher and J. A. Sethian, "Fronts propagating with curvaturedependent speed: algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [3] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [4] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, 2005, pp. 430–436.
- [5] Y. Zheng, B. Jeon, D. Xu, Q. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy c-means algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 28, no. 2, pp. 961–973, 2015.

- [6] M. Cote and P. Saeedi, "Automatic Rooftop Extraction in Nadir Aerial Imagery of Suburban Regions Using Corners and Variational Level Set Evolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 313–328, Jan. 2013.
- [7] Y. Song and J. Shan, "Building extraction from high resolution color imagery based on edge flow driven active contour and jseg," *IAPRSIS*, vol. 37, pp. 185–190, 2008.
- [8] J. Yang and Y.-H. Wang, "Towards automatic building extraction: variational level set model using prior shape knowledge," in 2012 International Conference on Image Analysis and Signal Processing. IEEE, 2012, pp. 1–6.
- [9] J. Wang, X. Yang, X. Qin, X. Ye, and Q. Qin, "An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery," *IEEE Geosci. Remote Sensing Lett.*, vol. 12, no. 3, pp. 487–491, 2015.
- [10] X. Yang, J. Wang, X. Qin, J. Wang, X. Ye, and Q. Qin, "Fast urban aerial image matching based on rectangular building extraction," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 4, pp. 21–27, 2015.
- [11] J. Yuan, "Automatic building extraction in aerial scenes using convolutional networks," CoRR, vol. abs/1602.06564, 2016.
- [12] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, "Saliency detection via graph-based manifold ranking," in 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013, 2013, pp. 3166–3173.
- [13] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, 2014, pp. 2814–2821.
- [14] R. S. Srivatsa and R. V. Babu, "Salient object detection via objectness measure," in 2015 IEEE International Conference on Image Processing, ICIP 2015, Quebec City, QC, Canada, September 27-30, 2015, 2015, pp. 4481–4485.
- [15] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Měch, "Minimum barrier salient object detection at 80 fps," in *IEEE International Conference on Computer Vision(ICCV)*, 2015.
- [16] G. Kanizsa, Organization in Vision. New York: Praeger 1979.
- [17] I. Kokkinos, "Highly accurate boundary detection and grouping," in The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010, 2010, pp. 2520–2527.
- [18] C. Lu, S. Liu, J. Jia, and C. Tang, "Contour box: Rejecting object proposals without explicit closed contours," in 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, 2015, pp. 2021–2029.
- [19] S. Wang, T. Kubota, J. M. Siskind, and J. Wang, "Salient closed boundary extraction with ratio contour," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 546–561, 2005.
- [20] J. S. Stahl and S. Wang, "Edge grouping combining boundary and region information," *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2590– 2606, 2007.
- [21] X. Qin, S. He, C. P. Quintero, A. Singh, M. Dehghan, and M. Jägersand, "Real-time salient closed boundary tracking via line segments perceptual grouping," *CoRR*, vol. abs/1705.00360, 2017.
- [22] T. Schoenemann, S. Masnou, and D. Cremers, "The elastic ratio: Introducing curvature into ratio-based image segmentation," *IEEE Trans. Image Processing*, vol. 20, no. 9, pp. 2565–2581, 2011.
- [23] A. Brook, A. M. Bruckstein, and R. Kimmel, "On similarity-invariant fairness measures," in *Scale Space and PDE Methods in Computer Vision, 5th International Conference, Scale-Space 2005, Hofgeismar, Germany, April 7-9, 2005, Proceedings,* 2005, pp. 456–467.
- [24] C. Akinlar and C. Topal, "Edlines: A real-time line segment detector with a false detection control," *Pattern Recognition Letters*, vol. 32, no. 13, pp. 1633–1642, 2011.
- [25] F. P. Preparata and M. I. Shamos, *Computational Geometry An Introduction*, ser. Texts and Monographs in Computer Science. Springer, 1985.
- [26] I. Jermyn and H. Ishikawa, "Globally optimal regions and boundaries as minimum ratio weight cycles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1075–1088, 2001.
- [27] X. Qin, S. He, Z. V. Zhang, M. Dehghan, and M. Jägersand, "Bylabel: A boundary based semi-automatic image annotation tool," in 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, Lake Tahoe, NV, USA, March 12-15, 2018, 2018, pp. 1804–1813.