Buildings extraction from high resolution remote sensing images based on superpixels graphcut

Fenghua Huang[2,3,4], Ying Yu[2,3]

Abstract. There are typically many shortcomings in high resolution remote sensing (HRRS) images, such as large imaging oblique angles, significant noise interference and shadows from artificial surface objects. This always adds complexity to the extraction of buildings from HRRS images directly. Traditional pixel-based methods for the extraction of buildings from HRRS images are inaccurate and inefficient. In order to address these problems, this paper first uses a simple and efficient clustering technique, Simple Linear Iterative Clustering (SLIC), to pre-segment HRRS images into a set of sub-blocks (superpixels). Next, a new method for buildings extraction from HRRS images based on superpixels Graphcut (BEHISPG) is proposed. In BEHISPG, common pixels are replaced with superpixels and a Graphcut segmentation cost function is customized to further segment HRRS images and improve segmentation effectiveness based on building features extracted from the superpixels. Finally, segmented results are filtered by computing and analyzing the metrics of rectangle degree, aspect ratio and area, for the purpose of effectively extracting buildings. In order to verify the performance of BEHISPG, comparisons are made with two other detection algorithms on the double-temporal HRRS images that are relatively captured over four experimental zones. Experimental results show that BEHISPG has an average recall of 95.05% and an average accuracy of 88.83%, which proves that it is superior to the other two algorithms. Hence, the proposed algorithm is suitable for the extraction of buildings from HRRS images that have complicated backgrounds.

Key words. Superpixel segmentation, building extraction, simple linear iterative clustering, graphcut algorithm.

[2] Workshop 1 - Information Engineering College,Yango University,Fuzhou 350015, China
[3] Workshop 2 - Spatial Information Engineering Research Centre of Fujian Province,Yango University,Fuzhou 350015, China
[4] Corresponding author: Fenghua Huang; e-mail: fenghuait@sina.com

1. Introduction

In high resolution remote sensing (HRRS) images, it is more challenging to extract buildings than other surface objects (e.g., roads and nudation). This is due to the lack of directly relevant 3D data, having different spectrums for the same buildings in different remote sensing images, diverse building appearances, and complex surrounding scenes in urban regions [1,2]. Substantial works have been done worldwide on automatic detection of buildings from HRRS images, which have achieved remarkable results. Currently available methods include the region segmentation algorithms [2-3], edge extraction methods [4], corners matching algorithms [5-6], and the supplementary knowledge algorithms [7-9]. But the performance of the above methods for extracting buildings is limited in terms of accuracy, efficiency, generality and automation. In order to address these problems, a new method for buildings extraction from HRRS images based on Simple Linear Iterative Clustering (SLIC) and superpixels Graphcut (BEHISPG) is proposed to improve extraction accuracy and efficiency.

2. Principles of Graphcut Algorithm

Graphcut is an interactive graph segmentation method. The basic idea is to convert images segmentation into undirected graphs segmentation. In Graphcut for image segmentation, let a weighted undirected graph G $\{V, E, W\}$ represent an image, where V denotes a set of nodes in the graph (each node in the graph corresponds to a specific pixel in the image), E denotes a set of edges in the graph (each edge in the graph connects two neighboring nodes), and W denotes a set of weights for each edge (the weight of each edge denotes how close the relationship between two nodes of the edge). Graphcut graph always has two types of nodes, i.e., ordinary nodes and target nodes [10], and it has two more target nodes than the common graph, represented by S and T. The ordinary nodes correspond to the pixels in the image relatively. The Graphcut graph also includes two types of edges. Edges that connect ordinary nodes in two neighborhoods (corresponding to every two neighboring pixels in the image) are called n-links. The edges that connect the ordinary nodes and the target nodes are called t-links. In Graphcut graph segmentation, nodes can be classified into user-labeled foreground nodes set, background node set and unlabeled nodes set [11]. Foreground nodes denote the target object that users need to extract. The purpose of interactive image segmentation is to determine whether unlabeled nodes belong to the foreground or background based on the information of the user-labeled foreground nodes and background nodes. Fig.1 illustrates the Graphcut-based segmentation process for a 3×3 image [10].

In Graphcut-based segmentation, the maximum flow / minimum cut algorithm is used to find the minimum cut set from the constructed undirected graphs [12]., S E, for the purpose of segmenting the image by minimizing the sum of all edge weights in the cut set S. Consider that the undirected graph G has a set of n nodes, V, and that there exists a binary vector $A=(A_1, A_2, ..., A_i, ..., A_n)$, where A_i denotes the index of node i in node set V. A_i can be the index of an ordinary node or a
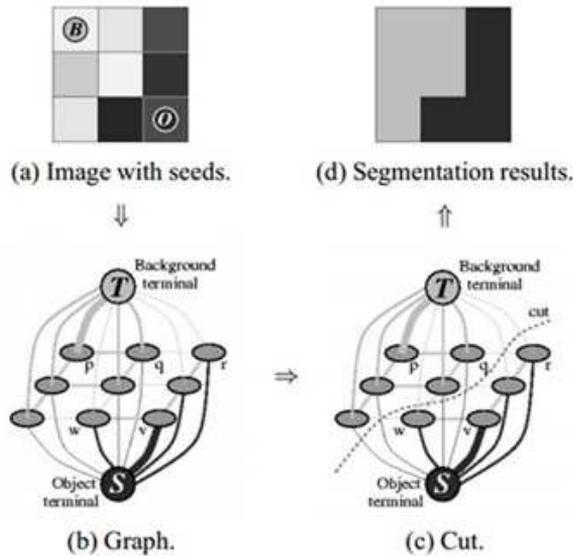
Fig. 1. Graphcut-based segmentation process for a 3×3 image

target node, S or T. The cost function that minimizes the cut is called the energy function, $E(A)$, which is a weighted combination of the regional term $R(A)$ and the boundary term $B(A)$ [12]. Among all segmentations that meet the conditions, the segmentation that achieves the global minimum of the cost function (the minimum cut) is considered the optimal segmentation. Similarity between the nodes (pixels) in image segmentation can be measured by gray scale, illumination and other features, or a combination of some features. Currently available methods that implement Graphcut include Goldberg-Tarjan, Ford-Fulkerson and their improved algorithms [10-12].

3. Buildings Extraction from HRRS Images based on Superpixels Graphcut (BEHISPG)

In BEHISPG, Simple Linear Iterative Clustering (SLIC) is first used to pre-segment the HRRS images into a set of superpixels, which are then used as a substitute for ordinary pixels. SLIC is a simple and efficient clustering technique [13], in which an adaption of k-means clustering method is adopted to generate superpixels with homogeneity and compactness. Then the Graphcut algorithm (maximum flow/minimum cut algorithm) is used to cluster the above superpixels of HRRS images with their extracted or combined features. Finally, segmentation results are filtered by jointly considering the three metrics of rectangle degree, aspect ratio, and area of the segmented objects for the purpose of extracting buildings more effectively.

average of A component in LAB color space(A_avg),average of B component in
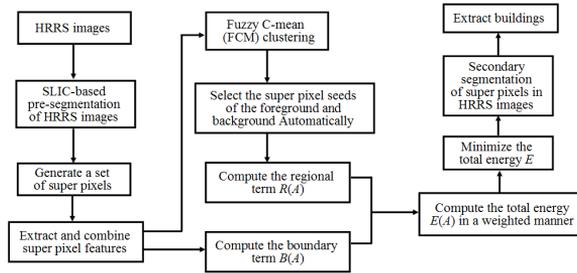
Fig. 2. Steps of BEHISPG

LAB color space (B_avg), and spatial distance between the centers of two superpixels (sps_dist). The extracted features are combined to form a vector, $T = [L\_avg\ A\_avg\ B\_avg\ sps\_dist]$. Let $A = (A_1, A_2, ..., A_i, ..., A_n)$ represents the set of superpixels generated after pre-segmentation, where n denotes the number of superpixels in the set. $A_p = 1$ means that the pth superpixel belongs to the foreground, and $A_p = 0$ means that the pth superpixel belongs to the background. In the case of segmenting HRRS images using the superpixel Graphcut strategy, the total cost function $E(A)$ [12] is defined as equation (1):

$$E(A) = \lambda R(A) + B(A) \tag{1}$$

Where $R(A)$ and $B(A)$ denote the regional and boundary terms respectively, and the weighted parameter λ denotes the relative importance of the regional and boundary terms. A high λ value means that the regional term is dominant in the total energy, and a low λ value means that the boundary term is dominant in the total energy.

3.1. Calculation of regional term R(A)

Non-supervised learning is performed with extracted feature samples using the more effective clustering algorithm, fuzzy C-means (FCM)[10], in order to automatically obtain superpixel seeds of the foreground and background (a superpixel seed is located in the center of the each cluster). By computing the Euclidean distance between non-labeled superpixels and the two superpixel seeds (all the superpixels can be divided into two types: the foreground represents building and the background represents other objects on the ground), we can compute the regional term of the cost function [11], $R(A)$, as equation (2):

$$R(A) = \sum_{i \in V} R_i(A_i) \tag{2}$$

If A_i belongs to the foreground, then $R_i(obj) = 0$ and $R_i(bkg) = 0$;
If A_i belongs to the background, then $R_i(obj) =$ and $R_i(bkg) = 0$;
If A_i does not belong to the foreground or background, then $R_i(obj)$ and $R_i(bkg)$ can be computed as equations (3) and (4)[10]:

$$R_i(obj) = D_{obj}/(D_{obj} + D_{bkg}) \quad (3)$$
$$R_i(bkg) = D_{bkg}/(D_{obj} + D_{bkg}) \quad (4)$$

3.2. Calculation of boundary term B(A)

The boundary term [11] is computed as equation (7):

$$B(A) = \sum_{(i,j) \in E} B(i,j)|A_i - A_j| \tag{3}$$

This equation implies that $B(A)$ depends primarily on the weight of the edge connecting superpixels i and j, $B(i, j)$. In this paper, $B(i, j)$ is computed as the weighted combination of lightness weight $B_i(i, j)$, color weight $B_{ab}(i, j)$ and the weight $B_d(i, j)$ which represents the spatial distance between superpixels i and j. Thus, $B(i, j)$ can be computed as equation(8):

$$B(i,j) = \sqrt[3]{B_l(i,j) \times B_{ab}(i,j) \times B_d(i,j)} + \alpha B_l(i,j) + \beta B_d(i,j) \tag{4}$$

Where α and β denote the weighted coefficients of the energy terms for lightness and inter-block spacing respectively. The weights of $B_i(i, j)$, $B_{ab}(i, j)$ and $B_d(i, j)$ can be defined as follows:

Definition of Bl(i, j)

Lightness weight, $B_l(i, j)$, is defined as equation (9):

$$B_l(i,j) = e^{-\frac{(L_i - L_j)^2}{\sigma_l^2}} \tag{5}$$

Where L_i and L_j respectively denote the average of LAB lightness components for superpixels i and j, and σ_i denotes the variance of global lightness. If the lightness of two superpixels is similar and B_l is large, there is a high probability that the two superpixels belong to the same type.

Definition of $B_{ab}(i, j)$

Color weight, $B_{ab}(i, j)$, is defined as equation (10):

$$B_{ab}(i,j) = e^{-\frac{\sqrt{((A_i - A_j)^2 + (B_i - B_j)^2)}}{\sigma_{ab}^2}} \tag{6}$$

Where A_i and A_j respectively denote the average values of component A in LAB color space for superpixels i and j, B_i and B_j respectively denote the average values of component B in LAB color space for superpixels i and j, and σ_{ab} denotes the average global variance of components A and B in LAB color space. If the color space of two superpixels is similar and B_{ab} is large, there is a high probability that the two pixels belong to the same type.

Definition of $B_d(i, j)$

The spatial distance weight between superpixels i and j, $B_d(i, j)$, is defined as

equation (11):

$$B_d(i,j) = e^{-\frac{\|C_i - C_j\|^2}{\sigma_d^2}} \tag{7}$$

Where C_i and C_j denote the central locations of the minimum enclosing rectangles for superpixels i and j respectively, and σ_d denotes the variance of global distance between the two superpixels. If the distance between two superpixels is small and B_d is large, there is a high probability that the two superpixels are overlapped.
Based on Equations (7) and (8), we can compute the boundary term $B(A)$ as equation (12):

$$B(A) = \sum_{(i,j)\in E} |A_i - A_j| (\sqrt[3]{B_l(i,j) \times B_{ab}(i,j) \times B_d(i,j)} + \alpha B_l(i,j) + \beta B_d(i,j)) \tag{8}$$

3.3. Superpixels segmentation and extraction of buildings

Once the energy function is designed, the Graphcut algorithm (maximum flow/ minimum cut method) can be used to obtain the global optimal solution of the energy function in equation (1). By using this process, each superpixel can be recognized to belong either to the foreground (buildings) or background (other surface objects), and the further segmentation based on superpixels in HRRS images can be achieved. After further segmentation, the images consist of only two types of objects: candidate buildings and background objects.

The shape of the superpixels is not considered in the further segmentation. As a result, some non-building blocks may be incorrectly included into superpixel combination. Hence, it is necessary to further filter candidate buildings obtained in further segmentation so as to finalize the buildings extraction. Buildings always exhibit regular shapes, and their areas, rectangle degrees and aspect ratios also often fall within special ranges.

3.4. Evaluation of Extracted Buildings

The accuracy and efficiency of building extraction is measured using the metrics of recall (R), accuracy (P), F-Score (F) and time consumption (T). Let N_1, N_2 and N_3 denote the amount of the buildings correctly extracted, the total amount of buildings extracted and the actual amount of buildings in the experimental zones respectively. Hence, metrics P, R, and F can be computed as equations (9), (10) and (11):

$$P = N_1/N_2 \times 100\% \tag{9}$$

$$R = N_1/N_3 \times 100\% \tag{10}$$

$$F = \frac{2 \times P \times R}{P + R} \times 100\% \tag{11}$$

F is the harmonic mean of P and R, which reflects the overall performance of

the algorithm since the two metrics are considered comprehensively. F alleviates possible conflict between P and R to a certain extent. A high value of F means that the performance of the algorithm is excellent . Time consumption, T, is another important performance metric that implies the efficiency of building extraction.

## 4. Experimental Results and Analysis

4.1. Experimental basic data

The experimental data of this paper consists of two WorldView2 images of Shenzhen city (placecountry-regionChina) collected in November 2012 (time-phase 1) and August 2013 (time-phase 2). Each image has three bands (red, green and blue) and is captured at a resolution of. Due to differences in imaging time and angle, the image captured at time-phase 1 had many shadows, and the shadows from some different buildings overlapped each other. But, the image captured at time-phase 2 contained a lot of building side walls, which added difficulty to the extraction of buildings. In this work, four typical experimental zones are selected from the two-temporal images above for buildings extraction. Four pairs of representative sub-images (A1-B1, A2-B2, A3-B3, and A4-B4) are selected from image A (captured at time-phase 1) and image B (captured at time-phase 2) as test images, all of which are. For A1-B1 and A2-B2, the images contain a lot of buildings, nudation, vegetation, regular building shapes (approximately rectangular) and large spacing between buildings. Shadows, building side walls and the bright cement ground constitute the major sources of noise interference.

4.2. Analysis of experimental results

BEHISPG has three main parameters: weighted coefficient of regional term in Equation (1), λ, weighted coefficients of lightness and inter-block distance energy in Equation (8), α and β. With A1-B1, A2-B2, A3-B3 and A4-B4 as test images, the above parameters first are optimized by Cross validation method, and the optimal parameter settings are $\lambda 0.45$, $\alpha=0.54$ and $\beta=0.41$. Then, based on the optimal parameter settings, BEHISPG is used to extract buildings from the four pairs of test images (i.e., A1-B1, A2-B2, A3-B3, and A4-B4). Results are given in Fig.3~6. Correctly recognized buildings are marked with red closed polylines, incorrectly recognized buildings are marked with blue closed polylines, and omitted buildings are marked with green closed polylines. Note that there may exist more than one building in a red closed polylines.

As shown in Fig.3~6, each of the four pairs of test images refer to HRRS images of the same region are collected at different time phases. Due to different camera angles, the same building in a pair of HRRS images may be deviated, rotated and distorted; there is also a difference in their sizes, brightness, and colors. But most buildings in the four pairs of test images are correctly extracted. Table 1 shows the values of performance metrics of BEHISPG when extracting buildings in the four pairs of test images. In Table 1, N_1, N_2 and N_3 respectively denote the amount of

A1(Time-Phase1)   Buildings extraction result for A1

B1(Time-Phase2)   Buildings extraction result for B1
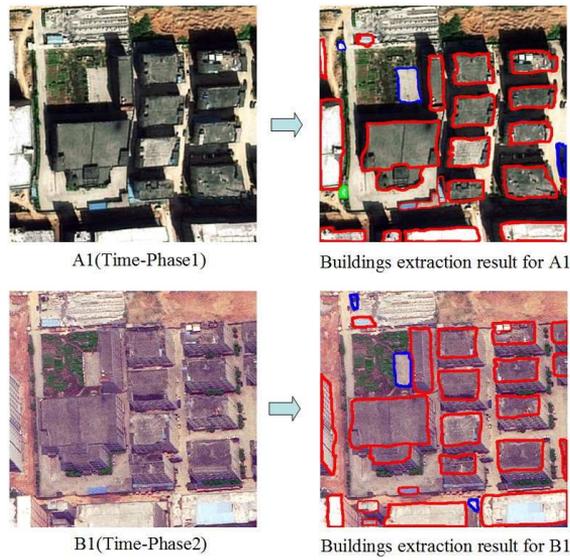
Fig. 3. Extracted buildings in images A1 and B1

A2(Time-Phase1)   Buildings extraction result for A2

B2(Time-Phase2)   Buildings extraction result for B2
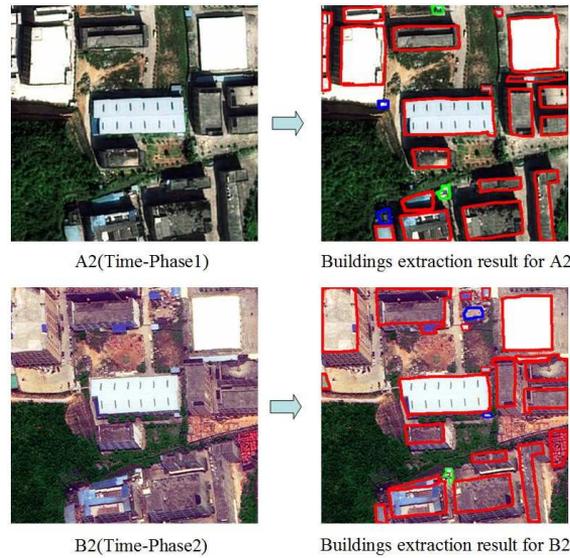
Fig. 4. Extracted buildings in images A2 and B2

buildings correctly extracted, the total amount of buildings extracted and the actual amount of buildings in the experimental zones.

Table 1 shows that the average accuracy(Avg_C), average recall(Avg_R), average F-score(Avg_F) and average time consumption (Avg_T) for the four pairs of test images are up to 88.83%, 95.05%, 0.9178 and 137.94 seconds. This demonstrated that BEHISPG can effectively extract buildings from HRRS images captured
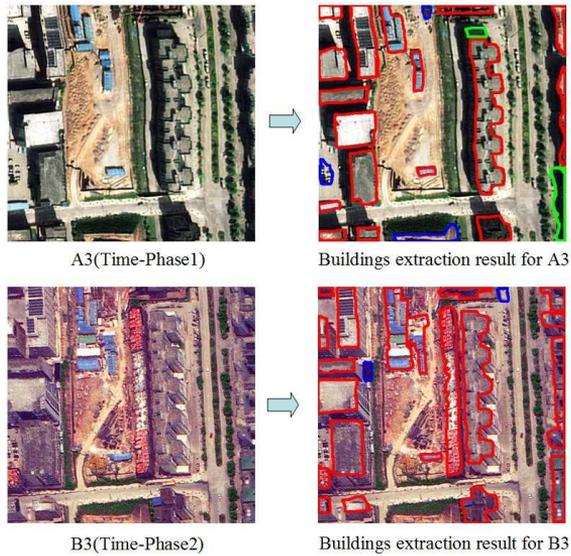
Fig. 5. Extracted buildings in images A3 and B3
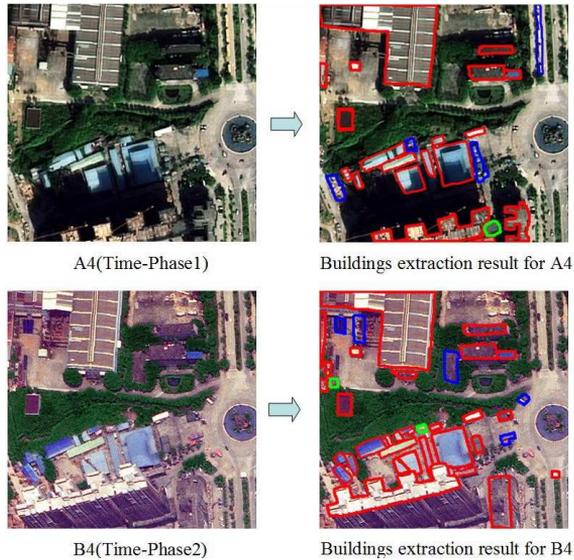
Fig. 6. Extracted buildings in images A4 and B4

at varying angles even if the buildings have distinct sizes and colors, and are interfered by shadows and nearby surface objects. The values of Avg_R for A2,A3 and B4 is low and obviously less than the average level. This is because in the above images there are several small buildings that are segmented into incomplete partitions by the image edges. Since the conditions of area, aspect ratio and rectangle degree are not met, these buildings are omitted in the buildings extraction process.

Table 1. Performance metrics of BEHISPG for extracting buildings in test images

Test images	N_1	N_2	N_3	P (%)	R (%)	F (%)	T (s)
A1 (phase 1)	22	25	23	88.00	95.65	0.9167	134.3
B1 (phase 2)	22	25	22	88.00	100.00	0.9362	133.5
A2 (phase 1)	19	21	21	90.48	90.48	0.9048	127.6
B2 (phase 2)	22	24	23	91.67	95.65	0.9362	134.1
A3 (phase 1)	26	29	29	89.66	89.66	0.8966	137.1
B3 (phase 2)	32	34	32	94.12	100.00	0.9697	154.3
A4 (phase 1)	19	23	20	82.61	95.00	0.8837	118.4
B4 (phase 2)	31	36	33	86.11	93.94	0.8986	164.2
Average				(Avg_C) 88.83	(Avg_R) 95.05	(Avg_F) 0.9178	(Avg_T) 137.94

4.3. Comparative analysis of experimental results

In order to verify the performance of different algorithms, BEHISPG is compared with two other representative algorithms, OBUBLE [1] and NCUTBE [2]. OBUBLE uses object-oriented methods to extract buildings based on total neighborhood variation. NCUTBE is similar to BEHISPG, but unlike BEHISPG, which pre-segments images using the marker watershed algorithm, NCUTBE does not perform pre-segmentation, but directly map each pixel in the HRRS images to a node in the graph. Some experiments are carried out to compare the performance of BEHISPG, OBUBLE and NCUTBE with the same four test image pairs, and the results are shown in Table 2.

Table 2 shows that, although BEHISPG's average time consumption is slightly larger than OBUBLE and obviously less than NCUTBE, it outperforms the other two algorithms in terms of average recall, average accuracy and average F-score. NCUTBE's accuracy is similar to BEHISPG, but its recall is inferior to the latter. This is mainly due to the fact that in NCUTBE the calculation of edge weights only takes pixels position, gray scale and edge features into account for the Ncuts-based segmentation, while neither RGB nor LAB color space is considered in NCUTBE. Thus, some buildings whose colors are very distinct will be omitted. Furthermore, BEHISPG uses SLIC to pre-segment images so as to generate much fewer superpixels than ordinary pixels in the original images, which decreases the number of nodes in graph segmentation. Although BEHISPG introduces the additional step of pre-segmentation and considers more features when computing the edge weights of undirected graphs, its average time consumption is still shorter than NCUTBE. This implies that BEHISPG is more efficient than NCUTBE. OBUBLE is vastly inferior to the other two algorithms in terms of average recall, average accuracy and average F-score. If the buildings have large inclination and obvious interference from shadows and other surface objects, OBUBLE is prone to misdetections and omissions in building extraction. This is due to the necessity of OBUBLE to follow image seg-

mentation with shape analysis-based pre-extraction of buildings, multi-directional morphology-based road filtering and prior model-based post-extraction of buildings. These steps are ineffective for HRRS images with complex backgrounds and are prone to suffer from error accumulation, which will decrease the recall and accuracy for building extraction. Hence, in HRRS images, especially for buildings that have large inclination and obvious interference from shadows and other surface objects, BEHISPG has better performance than NCUTBE and OBUBLE in general.

Table 2. Performance Comparison of BEHISPG, OBUBLE and NCUTBE

Extraction algo- rithms	Avg_C (%)	Avg_R (%)	Avg_F	Avg_T (s)
BEHISPG	88.83	95.05	0.9178	137.94
NCUTBE	88.00	91.67	0.8980	163.5
OBUBLE	86.96	83.33	0.8511	127.6

5. Analysis and Discussion

The proposed BEHISPG algorithm outperforms other two algorithms (NCUTBE and OBUBLE), but it still has some limitations. First, BEHISPG is inaccurate when extracting buildings that have irregular shapes or whose rooftops are broken. In addition, some surface objects (such as vehicles and containers) whose shapes are similar to buildings are difficult to be eliminated, which also can cause some false detection and omission. Second, the influence of shadows and side walls on image matching is alleviated by removing noise through bilateral filtering, but the influence is still non-negligible and no complete solution is currently available. Finally, BEHISPG method can more efficiently extract the building clusters in which each building roof is intersected with others or embedded in another one, but it is temporarily unable to accurately divide the closely connected building clusters into individual buildings. In the future, we will jointly consider these problems and propose a more accurate, efficient and automatic method for extraction of buildings from HRRS images.

6. Conclusion

In this paper, a superpixel Graphcut-based method is used to extract buildings from HRRS images with significant background interference by jointly considering SLIC segmentation, superpixels and graph segmentation theories. First, SLIC is used to pre-segment HRRS images into a set of superpixels. Next, a new Graphcut method based on superpixels (maximum flow/minimum cut algorithm) and an appropriate segmentation cost function are employed for further segmentation. Finally, in order to achieve effective extraction of buildings, segmentation results are filtered by computing the segmentation object's rectangle degree, aspect ratio and area. Extraction accuracy and efficiency are improved substantially, as demonstrated by an

average recall of 95.05% and an average accuracy of 88.83%, although shadows and side walls have a non-negligible influence on extraction performance. Buildings that have irregular shapes or whose rooftops are broken are prone to be omitted, and surface objects that have similar shapes to buildings are typically detected wrongly. These problems need to be addressed in future work.

References

[1] SHI. W. Z,MAO. Z. Y: *Research on Building Change Detection from High Resolution Remotely Sensed Imagery Based on Graph-cut Segmentation.* Journal of Geo-Information Science *18* (2016), 423–432.

[2] REN. K. Y,SUN. H. X,JIA. Q. X: *Building Recognition from Aerial Images Combining Segmentation and Shadow.*IEEE conference on Intelligent Computing and Intelligent System(2009),578–582.

[3] CHENG. L,GONG. J. Y: *Building Boundary Extraction Using Very High Resolution Images and LiDAR.* Acta Geodaetica et Cartographica Sinica *37* (2008),391–393.

[4] SIRMACEK. B,UNSALAN. C: *Urban Area Detection Using Local Feature Points and Spatial Voting, IEEE Geoscience and Remote Sensing Letter.* IEEE Geoscience and Remote Sensing Letter *7* (2010), 146–150.

[5] SIRMACEK. B,UNSALAN. C: *Urban-Area and Building Detection Using SIFT Keypoints and Graph Theory.* IEEE Transaction on Geoscience and Remote Sensing*47* (2009),1156–1167.

[6] ZHANG. Y: *Optimisation of Building Detection in Satellite Images by combining Multispectral Classification and Texture Filtering.*ISPRS Journal of Photogrammetry & Remtoe Sensing *54* (1999),50–60.

[7] ZHOU. J. Q, LI. Z. J: *Spatial Relation-Aided Method for Object-Oriented Extraction of Buildings from High Resolution Image.*Journal of Applied Sciences *30* (2012),511–516.

[8] JIN. X, DAVIS. C: *Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural,Contextual, and Spectral Information.* EURASIP Journal on Applied Signal Processing *14* (2005), 2196–2206.

[9] TIAN. Y, WANG. C: *Interactive Image Segmentation Method Based on Fuzzy C-Means and Graph Cuts.*Journal of Engineering Graphics *31*, (2010),123–127.

[10] SU. J. L,WANG. Z. H: *An Image Segmentation Method Based on Graph Cut and Super Pixels in Nature Scene.*Journal of PlaceNameplaceSoochow PlaceTypeUniversity (Natural Science Edition) *28*, (2012),27–33.

[11] SONG. Z. G,ZHAN. Y. W: *An Image Segmentation Method Based on Graph Cut and Super Pixels in Nature Scene.* Computer Systems & Applications *21* (2012),206–209.

[12] ACHANTA. R, SHAJI. A, SMITH. K, LUCCHI. A, FUA. P,SSSTRUNK. S: *SLIC superpixels compared to state-of-the-art superpixel methods.* IEEE Transactions on Pattern Analysis & Machine Intelligence*34* (2012),2274–2282.