ACCURATE AND FAST BUILDING DETECTION USING BINARY BAG-OF-FEATURES

Y. Hu [a, b, *], Z. Li [c], P. Li [a], Y. Ding [a], Y. Liu [c]

[a] Chongqing Geomatics Center, 401121 Chongqing, China - (huyan_dl023, 20110802080, 20160801008)@cqu.edu.cn
[b] School of Remote Sensing and Information Engineering, Wuhan University, 430079 Wuhan, China - huyan_dl023@cqu.edu.cn
[c] College of Optoelectronic Engineering, Chongqing University, 400044 Chongqing, China -
(lizhenghao, 20140813038t)@cqu.edu.cn

Commission I, ICWG II/III

KEY WORDS: Building Detection, Simple Linear Iterative Clustering, Binary Bag-Of-Features, Machine Learning, Local Feature, Descriptor, Classification

ABSTRACT:

This paper presents a non-interactive building detection approach employing binary bag-of-features (BBOF), namely, extracting building roof contours in remote sensing images automatically, rapidly and accurately. The proposed method includes two major stages, i.e., building area detection and building contours extraction. In the first stage, it contains three modules. i.e., over-segmentation, intersection point classification, building area detection. Firstly, the orthophoto is over-segmented by the Simple Linear Iterative Clustering (SLIC) superpixel segmentation method, and the intersection points is obtained. Secondly, the oriented FAST and rotated BRIEF (ORB) descriptors are generated in LAB colour space from the patches that centred on the intersection points, and the BBOF classifier is adopted to classify the intersection points into two categories. Thirdly, the area that contains of the building roof are detected through reserving the regions around the intersection points in inner parts of building roof, and eliminating the regions around the intersection points in non-building roof. At last, we can roughly generate the building area. For the second stage, it is similar to the first one while the main difference is that its classifier has three categories. Finally, we provide an evaluation between two different classifiers, including ORB+BBOF and SURF+BOF. This evaluation is conducted on orthophotos with different roof colours, texture, shape, size and orientation. The proposed approach presents several advantages in terms of scalability, suitability and simplicity with respect to the existing methods.

1. INTRODUCTION

Building detection from aerial photographs and satellite images has many applications in the fields like real-estate management, urban planning and disaster relief. In the last two decades, a large variety of methods on interactive and non-interactive building detection has been brought forward. Especially in recent years, machine learning approaches have been shown to be successful in addressing building detection with high accuracy and high robustness. From the point of view of machine learning, it is desirable to keep the user interaction at the training stage only and to fully automate the detection stage.

It is well-known that segmenting buildings in aerial images is a challenging task. This problem is generally considered when we talk about high-level image processing in order to produce numerical or symbolic information. Many techniques have been proposed in the literature. Among the techniques most frequently used, one can cite semi-automatic methods that need user interaction in order to extract desired targets or objects of interest from images. Generally, this category of methods has been introduced to overcome the problems associated with the full automatic segmentation which is usually not perfect. It consists in dividing an image into two classes: "object" and "background".

In this paper, we propose a non-interactive building detection approach employing binary bag-of-features (BBOF), namely, extracting building roof contours in remote sensing images automatically, rapidly and accurately. First the stage of building area extraction is conducted using LAB colour space and simple linear iterative clustering (SLIC). Thereafter, feature descriptors which contain of area information training samples are collected. An improved parallelepiped classification method is applied to classify the feature descriptors into building and non-building areas. Finally, the other stage of building contours detection operations is executed to obtain the accurate results.

2. BINARY BAG-OF-FEATURES

For binary bag-of-features model, common language and information retrieval are used to simplify representation. In this model, an information (such as a point in the inner building or at the age of the building) is represented as the bag (multiset) of its features, disregarding the detailed difference but keeping multiplicity. The binary bag-of-features model can be used for training a classifier.

This model describes each segmented picture using a set of features called visual vocabulary. The vocabulary is obtained by clustering local features extracted from manually segmented images, in which each resulting cluster generates a feature. An image segment is finally represented by a histogram. Each bin of this histogram corresponds to a visual word, and the associated weight represents its importance in the segment. Getting the histogram needs three steps: 1) extracting local features, 2) building visual vocabulary, 3) creating signatures.

2.1 Extracting Local Features

To extract the local features from image segments, we should detect keypoints first. Keypoints are the centres of salient patched, which is generally located around the corners or edges. In our work, the oriented FAST and rotated BRIEF (ORB) is

used to detect and describe keypoints. Our method is not specific to ORB. Even faster keypoint detector/descriptor combination may be used, although BUT remains one of the most reliable method under various image transformations. In this step, ORB keypoints are extracted from manually segmented image regions, and each keypoint is described by a vector of 128 elements summarizing local gradient information. The extracted features will be used to build the visual vocabulary.

2.2 Building Visual Vocabulary

Building the visual vocabulary means quantifying extracted local descriptors. The vocabulary is produced by clustering ORB features using the standard k-means algorithm. The size of the vocabulary is the number of clusters, and the centres of the clusters are the standard feature. Each segmented image in the database will be represented by standard features from this vocabulary.

2.3 Creating Signatures

Once the visual vocabulary is built, we index each segment by constructing its binary bag-of-features. This requires finding the weight of the visual words from the vocabulary. Each segmented image is described by a histogram, in which the k bins are the standard features and the corresponding values are the weights of the words in the image region.

3. TWO-STAGE BUILDING DETECTION

As the identified original image is very complex, it is easy to cause false detection. Non-buildings such as park trails are highly similar to real building roofs. So we use two stages to complete the final outline of the building extraction. The first stage is the extraction of the building area, and the second stage is the extraction of the building contours (Figure 1).

3.1 Building Area Detection

The purpose of the first stage is to find the almost building area from the original image. In the second stage, building contours are extracted from the in the original image and increases the accuracy of the results in the final building contour extraction.

For the first stage, the original image is reduced to a pixel resolution of 1 m before the processing is performed, subjected to a preprocessing operation. The processing of the stage with the compressed image can greatly improve the amount of information in each feature area, thus improving the accuracy of the follow-up machine learning and computing speed. The preprocessing operation is actually a border protection of the image, in order to avoid the implementation of the program when the boundary overflows.

Then, we convert the compressed picture from the BGR space to the LAB space in order to do operations from the broader colour space. In the LAB space, the compressed image is super-segmented to get the intersection. And the feature is extracted from the circular area with a radius of 31 pixels at the intersection point.

After that the oriented FAST and rotated BRIEF (ORB) feature description method is used to divide the characteristic area into non-building (S1C1) and building (S1C2). All the intersections
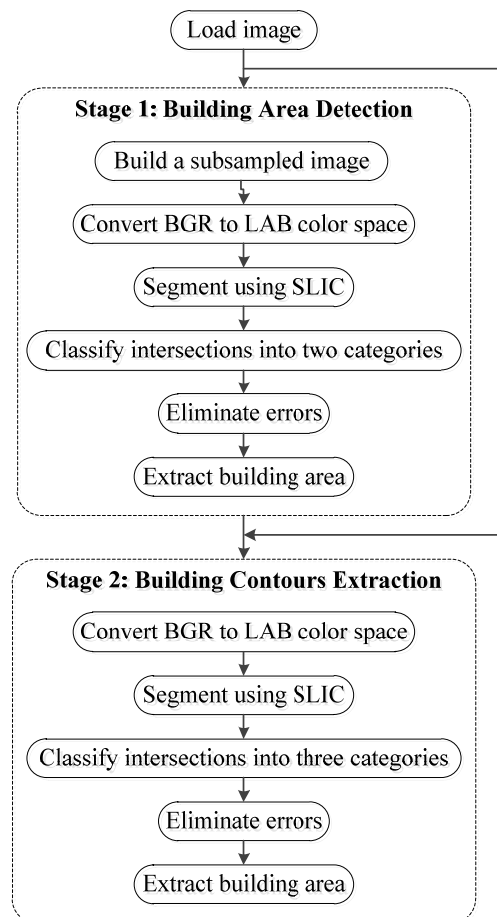

Figure 1 Proposed method for building detection

obtained after the super-segmentation is predicted automatically by the Binary Bag-Of-Features (BBOF), determined whether the point is a non-building or Building and labelled. Eventually, the approximate area of the building can be obtained according to the intensity of the point marked as the building, and the building area can be extracted by fitting rectangle. Misjudged points will be removed due to their numbers of the area pixels less than the set density threshold.

Thus, the approximate area of the building can be obtained, providing a good initial range for subsequent detection of the final contours of the building.

3.2 Building Contours Extraction

This stage is based on the fine processing of the previous stage after the coarse extraction, and the basic framework is similar to the first stage. The processed object is the extracted building area from the previous stage. The processed picture is restored from the compassed image to the original image.

Firstly, the building area image changes from BGR space to LAB space. In the LAB space, the building area image is super-segmented to get intersections. And the feature is extracted from the circular area with a radius of 31 pixels at the intersection point.

Thereafter, ORB feature description method is used to divide the characteristic area into non-building point (S1C1), the point of the building edge (S1C2) and the interior point of the building (S1C3). And then every intersection obtained after the
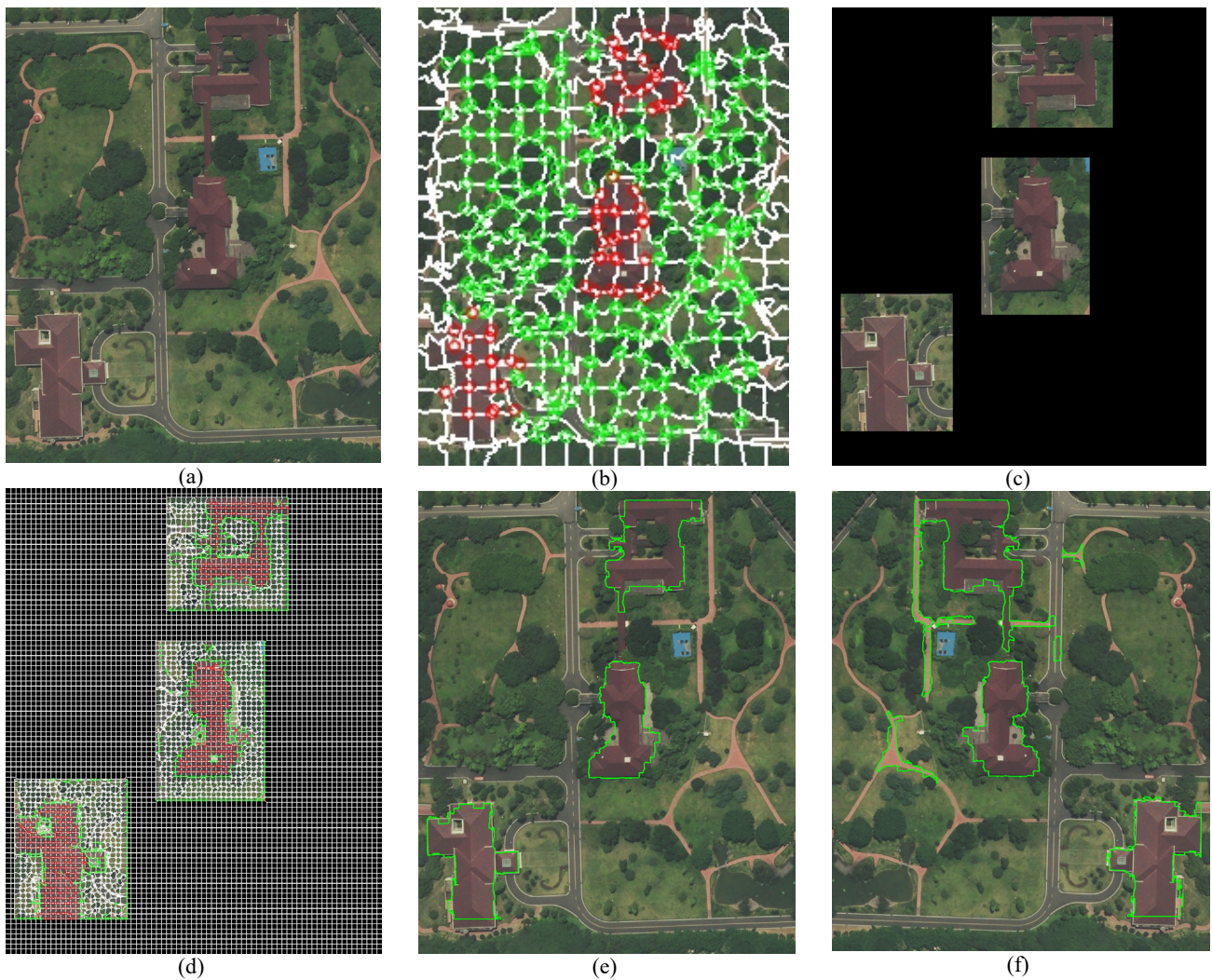
Figure 2: For image 2, (a) The original image; (b) Result of extracted intersections; (c) Result of building area detection; (d) Result of extracted intersections from detected building area; (e) Result of building contours extraction using ORB+BBOF; (f) Result of building contours extraction using SURF+BOF.

Figure 3: For image 5, (a) Result of building contours extraction using ORB+BBOF; (b) Result of building contours extraction using SURF+BOF.

Test image	Time cost (s) ORB+BBOF	Time cost (s) SURF+BOF
1	3.483272	4.034359
2	4.049904	7.918136
3	49.367466	51.019599
4	17.781792	23.535408
5	31.521437	33.286813
6	24.671056	34.391306
7	38.803565	44.248807

Table 1 Comparison of time cost

super-segmentation is predicted automatically by the Binary Bag-Of-Features (BBOF) feature bag to determine whether the point is the non-building point, the building edge point or the interior building point and labelled. Eventually, the super-segmented blocks around the non-building points are eliminated, and the edge point of the building and the interior point of the building are reserved, then the final outline of the building can be drew.

4. EXPERIMENTAL RESULTS

4.1 Image Datasets

We tested our automatic building detection method on two orthophotos whose three bands of RGB has been transferred to three bands of LAB. The test images and results of image 2 are showed in Figure 2. We provided the extracted building roofs with their contours as (e), and the results using a standard method in (f). From Figure 2, we can know the each process specifically. In Figure 3, it is the two different results of image 5, in which (a) is the result using ORB+BBOF and (b) is the result using SURF+BOF.

4.2 Results and Discussion

We illustrate the detection results of the proposed method in Figure 2 and Figure 3. Visual interpretations of the results show that our method is robust and accurate by extracting most of the building with different roof colours, texture, shape, size and orientation. In addition to the efficiency, the result in Table 1 shows the fast characteristic.

5. CONCLUSIONS

The non-interactive building extraction framework is presented for orthophotos based on energy minimization model. In our framework, the artificial building can almost be extracted and the others can be false detected hardly. Besides, with SLIC superpixel and BBOF methods, we can get foreground information as much as possible. In addition, the accuracy of intersection point classification has achieved 97.3%, which guarantees desirable subjective visual sensations of the extracted building roof contours. Experimental result also shows that our framework is robust for extract buildings with complex shapes.

ACKNOWLEDGEMENTS

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), pp. 2274-2282.

Adams, R., Bischof, L., 1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), pp. 641-647.

Ahonen, T., Hadid, A., Pietikainen, M., 2006. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037-2041.

Boykov, Y., Jolly, M., 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. *Proceedings of IEEE International Conference on Computer Vision*, pp. 105-112.

Breiman L., 2001. Random forests, *Machine Learning*, 45(1), pp. 5-32.

Bay H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer Vision & Image Understanding*, 110(3), pp. 346-359.

Felzenszwalb, P., F., Huttenlocher D., P., 2004. Efficient Graph-Based Image Segmentation. *International Journal and Computer Vision*, 59(2), pp. 167-181.

Huertas, A., Nevatia, R., 1988. Detecting buildings in aerial images. *Computer Vision, Graphics, and Image Processing*, 41(2), pp. 131-152.

Kumar, S., Hebert, M., 2003. Man-made structure detection in natural images using a causal multiscale random field. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 119-126.

Kramer, N., Braun, M., 2007. Kernelizing PLS, degrees of freedom, and efficient model selection. *Proceedings of International Conference on Machine Learning*, pp. 441-448.

Liasis, G., Stavrou, S., 2016. Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, 37(5), pp. 1127-1153.

Lee, D.S., Shan, J., Bethel, J.S., 2003. Class-guided building extraction from Ikonos imagery. *Photogrammetric Engineering and Remote Sensing*, 69(2), pp. 143-150.

Ren, X., Malik, J., 2003. Learning a classification model for segmentation. *Proceedings of IEEE International Conference on Computer Vision*, pp. 10-17.

Rother, C., Kolmogorov, V., Blake, A., 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3), pp. 309-314.

Sirmacek, Unsalan, C., 2009. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4), pp. 1156-1167.

Saito, S., Yamashita, T., Aoki, Y., 2016. Multiple object extraction from aerial imagery with convolutional neural

networks. *Journal of Imaging Science and Technology*, 60(1), pp. 0104021-0104029.

Sirmacek, B., Unsalan, C., 2008. Building detection from aerial images using invariant color features and shadow information. *Proceedings of International Symposium on Computer and Information Sciences*, pp. 1-5.

Sirmacek, B., Unsalan, C., 2009. Urban area and building detection using sift keypoints and graph theory. *Computer Vision and Image Understanding*, 47(4), pp.1156-1167.

S. Xu, T. Fang, D. Li, and S. Wang, 2010. Object classification of aerial images with bag-of-visual words. *IEEE Geoscience Remote Sensing. Letters*, 7(2), pp. 366-370.

Sirmacek B., Unsalan C., 2011. A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Transactions on Geoscience Remote Sensing*, 49(1), pp. 211-221.