LETTER *Special Section on Smart Multimedia & Communication Systems*

Acceleration of Automatic Building Extraction via Color-Clustering Analysis

Masakazu IWAI[†], ***Nonmember***, **Takuya FUTAGAMI**[††], **Noboru HAYASAKA**[†a)], ***and*** **Takao ONOYE**[††], ***Members***

SUMMARY In this paper, we improve upon the automatic building extraction method, which uses a variational inference Gaussian mixture model for performing color clustering, by accelerating its computational speed. The improved method decreases the computational time using an image with reduced resolution upon applying color clustering. According to our experiment, in which we used 106 scenery images, the improved method could extract buildings at a rate 86.54% faster than that of the conventional methods. Furthermore, the improved method significantly increased the extraction accuracy by 1.8% or more by preventing over-clustering using the reduced image, which also had a reduced number of the colors.
key words: *scenery image, building extraction, GrabCut, acceleration*

1. Introduction

In recent years, the growing usage of smartphones has enabled us to easily measure the current position of a user via global navigation satellite systems (GNSSs). However, in urban areas with many buildings, the performance of GNSS decreases because of the reflection of the satellite signal.

Therefore, to improve the measurement performance of the smartphone in urban areas, position measurement based on the recognition of buildings from scenery images, whose main subjects are buildings, is investigated. To increase the building-recognition accuracy, the regions containing the buildings are extracted from the entire image as a pre-processing step [1]. Therefore, fast and accurate building extraction from scenery images must be performed on the smartphone.

Although deep-learning-based segmentation may extract buildings accurately, it requires a sufficient amount of annotated training data. To decrease the introduction cost, we focus on methods that do not require annotated training data. In [2], a building extraction was proposed by using an adaptive thresholding method (hereafter called thresholding-based method). However, in [3], building extraction using a variational inference Gaussian mixture model (VBGMM) for a color clustering (hereafter called clustering-based method) was proposed. In [3], it was reported that the clustering-based method increased the extraction accuracy by more than 6% compared with both the thresholding-based method and deep-learning-based segmentation trained using 367 scenery images.

The building extraction has not been implemented on the smartphone in the literature. Futhemore, the computational time of the clustering-based method was not evaluated in the aforementioned studies. To assess the applicability of the building extraction in terms of the computational time, we first implement the building extraction on a PC, which is faster than the smartphone.

We observed that the clustering-based method can be improved by decreasing its computational time. Therefore, in this study, we improve the clustering-based method in terms of its computational time, without compromising with the extraction accuracy.

The target computational time of the building extraction is set to less than 10 s because the accurate position can be measured during that time in [4]. Although the position measurement is not implemented in this paper, the building extraction, which is employed for the pre-processing of the measurement based on scenery images, should be conducted for less than 10 s at least to achieve a fast and accurate camera-based positioning system.

2. Related Work

In this section, we review the thresholding-based method [2] and clustering-based method [3]. The thresholding-based method uses GrowCut [5] initialized by seeds based on a binary image via adaptive thresholding. In [2], it was reported that compared with the method developed in [1], it increased the extraction accuracy by more than 11%. However, the clustering-based method, whose flowchart is presented in Fig. 1, uses GrabCut [6] initialized by seeds based on an analysis of the color cluster generated by VBGMM. In [3], it was reported that compared with the thresholding-based method, it increased the extraction accuracy by more than 12%, via an experiment using 106 scenery images, as depicted in Fig. 2. The structures of the both the methods are similar, as both of them have two processes, namely, initial seed generation and segmentation (GrowCut or GrabCut).

Our experiment, whose details are presented in Sect. 4, showed that the clustering-based method cannot decrease the computational time compared with the thresholding-based method. This is caused by the color clustering (see Fig. 1) which accounts for more than 98% of the total computational time. In order to develop the fast and accurate

Manuscript received January 31, 2020.
Manuscript revised April 25, 2020.
[†]The authors are with the Department of Engineering Informatics, Osaka Electro-Communication University, Neyagawa-shi, 572-8530 Japan.
[††]The authors are with the Graduate School of Information Science and Technology, Osaka University, Suita-shi, 565-0871 Japan.
a) E-mail: hayasaka@osakac.ac.jp
DOI: 10.1587/transfun.2020SML0004

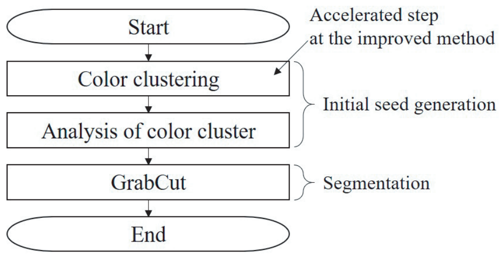


Fig. 1 Flowchart of clustering-based and improved methods.

Fig. 2 Examples of scenery images.

building extraction, the color clustering should be accelerated without sacrificing the extraction accuracy.

3. Improved Method

This section provides an overview of the improved method, whose flowchart is illustrated in Fig. 1. Compared with the clustering-based method, the improved method more effectively lowers the computational time of color clustering.

3.1 Color Clustering

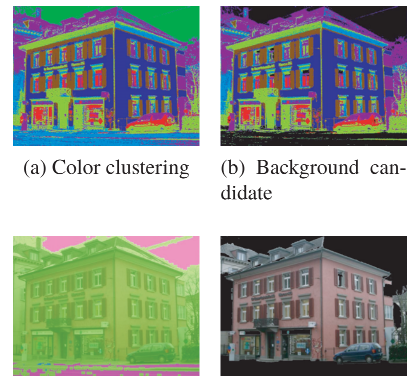
Color clustering is applied to the RGB color vector of the input scenery image using VBGMM. To accelerate color clustering, the centers of the cluster, i.e., centroids, are obtained from the image with reduced resolution at a specific rate [hereafter called reduction rate (r)]. Subsequently, each pixel of the original input image is assigned to the nearest cluster. Figure 3(a) depicts the result of applying the color clustering to the left image in Fig. 2, wherein different colors represent different clusters.

3.2 Analysis of Color Cluster

The improved method assumes that the building usually remains at the center of the scenery image, while the background, at the top and bottom. To select the background candidate, the color clusters are analyzed and those that usually appear at the top or bottom of an image are identified [3].

The analysis employs N rectangles R_n, which move from the upper region to the image center, with constant width shifts (Fig. 4). The y-coordinate of the top of rectangle R_n is denoted as y_n. Moreover, the analysis employed the content ratio ($f_n^{c_i}$), which is defined as the ratio of the number of pixels in the cluster c_i and the total number of pixels in the rectangle R_n, as in the following equation.

$$f_n^{c_i} = \frac{\sum_{x=0}^{W-1} \sum_{y=y_n}^{y+H-1} \delta(c_i, p_{x,y})}{W \cdot H}, \quad (1)$$



(a) Color clustering (b) Background candidate

(c) Initial seeds (d) Extracted building

Fig. 3 Process of the improved method.

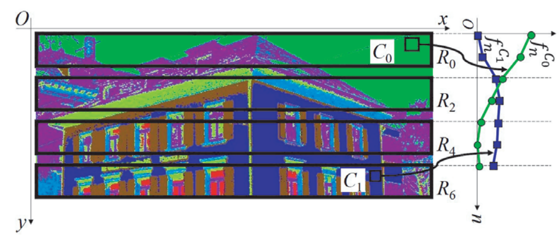


Fig. 4 Analysis of color cluster.

wherein δ and $p_{x,y}$ denote the Kronecker delta and the cluster index at pixel position (x, y), respectively. Moreover, W and H denote the width and height of rectangle R_n, respectively. The content ratio of the cluster including the background at the top or bottom, similar to the cluster in green in Fig. 4, tends to decrease with an increase in n, as the building is at the image center. On the basis of this tendency, the clusters are analyzed to identify those include backgrounds.

The principal procedure employed in the color cluster analysis is explained below. The background candidate is selected from the upper half of the scenery image. For each color cluster c_i, the pixels belonging to the cluster located at $y \leq y_{n'}$ are designated as the background candidates, if the content ratio $f_n^{c_i}$ does not increase for $0 \leq n \leq n'$. The background candidate is also selected from the lower half of the image, by applying the same step to the vertically-flipped image. The black regions in Fig. 3(b) depict the background candidates obtained by analyzing Fig. 3(a).

The initial seeds for initializing GrabCut are generated using the background candidates. The regions of the background candidates are eroded iteratively, for a specific number of times, to prevent misdeterminating the building as background by GrabCut. The eroded regions are specified as the initial seed of the background and the remaining regions as that of the building. The pink and green pixels in Fig. 3(c) depict the initial seeds of the background and building, respectively.

3.3 GrabCut

GrabCut [6] is based on the graph theory. Each pixel and color distribution of the building and background regions are designated as graph nodes. The local and global similarities are designated as the edge weights. The local and global similarities are based on the color distance between two adjacent pixels and on the probability that the pixels belong to either the building or the foreground region, respectively. GrabCut iteratively revises both the regions, by increasing the local and global similarities of each region.

The clustering-based and the thresholding-based methods employ GrabCut [6] and GrowCut [5], respectively. The improved method employs GrabCut for the image segmentation, as GrabCut operates 5 to 29 times faster than GrowCut, without compromising on the extraction accuracy [7].

The principal procedure employed in GrabCut is explained here. The initial seeds obtained from the previous subsection are used for the initialization of GrabCut. GrabCut revises both the regions. Figure 3(d) depicts the extracted building by applying GrabCut initialized by Fig. 3(c).

4. Experiment

4.1 Experimental Conditions

The performance of the improved method was evaluated in terms of computational time and extraction accuracy using 106 scenery images included in ZuBuD [8]. Three parameters, namely, the positive predictive value (PPV), negative predictive value (NPV), and accuracy (ACC), were used as the metrics of extraction accuracy. Notably, PPV and NPV measure the correct ratio of determining building and background, respectively. In addition, ACC, which measures the correct ratio of determining both building and background, is a comprehensive extraction-accuracy metric. Both the thresholding-based method and clustering-based method are used for conventional methods. The initial seed generation achieved using all the methods, respectively, was coded using Python libraries, namely, Opencv, Numpy, and Scipy. The segmentation (GrowCut and GrabCut) was coded using C++. The computational time was measured using a PC with the following specifications: Intel Core i5-4460 CPU and 16-GB RAM.

4.2 Experimental Results

Table 1 summarizes the experimental results, where "TB" and "CB" denote the thresholding-based method [2] and clustering-based method [3], respectively. In addition, "RR" denotes the reduction rate r, as discussed in Sect. 3.2; "ISG" and "SEG" denote the computational time of the initial seed generation and segmentation, respectively, as discussed in Sect. 2

As shown in Table 1, the highest ACC was achieved using the improved method at $r = 20$. However, the total computational time cannot always decrease because of the significant computational time of segmentation at $r > 20$. Therefore, we determined that the better parameter was $r =$ 20.

Table 1 Performance summary of each method.

(a) Extraction accuracy

Method	TB	CB	Improved method			
RR	–	–	5	10	20	30
PPV [%]	74.0	84.9	85.6	86.1	86.1	87.3
NPV [%]	54.0	77.8	78.6	79.2	81.2	76.9
ACC [%]	70.5	83.1	83.8	84.3	84.9	84.4

(b) Computational time

Method	TB	CB	Improved method			
RR	–	–	5	10	20	30
ISG [s]	70.06	2.66	2.38	1.03	0.72	0.63
SEG [s]	3.46	24.55	3.17	3.00	2.94	3.11
Total [s]	73.52	27.21	5.55	4.03	3.66	3.74

The clustering-based method could not demonstrate its superior computational time compared with the thresholding-based method. However, compared with the conventional methods, the improved method was found to achieve fast and accurate building extraction.

Compared with the conventional methods, the improved method decreased the computational time by 86.54% or more and increased ACC by 1.8% or more. The paired t-test and Wilcox signed-rank test (at 5% significance level) demonstrated a significant difference in ACC. Moreover, the experimental results demonstrated that only the improved method completed within the target computational time of 10 s. The computational time of the improved method might be further decreased by optimizing the implementation. The applicability of the improved method was confirmed on the PC in terms of the computational time. Therefore, for the future development, the building extraction should be implemented on a smartphone for faster and accurate camera-based measurements.

5. Discussion

5.1 Effectiveness of the Improved Method

This subsection presents a comparison of the clustering-based and improved methods and discusses the effectiveness of the improved method in terms of the extraction accuracy.

Figure 5 compares the images obtained from the processes upon application of the clustering-based method and improved method to the right image in Fig. 2. In addition, Fig. 5(a), (b), and (c) corresponds to Fig. 3(a), (b), (d), respectively. While the improved method can accurately perform building extraction, the clustering-based method suffered from a high rate of misdetermination of the background as a building at the top of the image. Because this tendency occured widely, the underlying reason is discussed here as follows.

Because the sky was unnecessarily divided into two clusters (hereafter called over-clustering) when using the clustering-based method, as depicted in the upper image in Fig. 5(a), the content ratio $f_n^{c_i}$ of the brown cluster increased with the increase of n. Hence, the brown cluster was not

(a) Color clustering (b) Background candidate (c) Extracted building

Fig. 5 Building-extraction performance (upper: clustering-based method, lower: improved method).

(a) Input (b) Extracted building

Fig. 6 Example of misdetermination.

specified as the background candidate, as depicted in the upper image in Fig. 5(b). Therefore, the clustering-based method misdetermined the sky as the building.

However, the improved method solved the over-clustering problem, as the sky was included in a single cluster (green cluster), as depicted in the lower image in Fig. 5(a). This is because the number of the color clusters provided by VBGMM decreased upon using the reduced image whose number of colors also reduced accordingly. Because the content ratio of the green cluster did not increase until rectangle R_n reached the border between the sky and building, as assumed in Sect. 3.2, almost all the part of the sky was included in the background candidate, as depicted in the lower image in Fig. 5(b). Therefore, the improved method can correctly determine the sky as the background.

This enhancement in the extraction accuracy upon solving the over-cluster problem using the improved method can be widely confirmed via other examples.

Table 2 presents the accuracy of the background candidate identification, similar to Table 1(a). The improved method appeared to have increased the initial seed ACC by 3.7% at $r = 20$ with increasing both PPV and NPV. This finding provided further evidence that the improved method exhibited enhanced extraction accuracy than the clustering-based method at Table 1(a).

5.2 Future Work

This subsection discusses potential and actual limitations of the improved method to enhance its extraction accuracy.

As the improved method was predisposed to decrease the number of the color clusters in Sect. 5.1, it might assign similarly-colored background and building to the same cluster. However, the misassignment was not observed in our experiments.

Table 2 Accuracy of background candidate.

Method	CB	Improved method			
RR	–	5	10	20	30
PPV [%]	82.8	83.9	84.6	86.1	85.1
NPV [%]	64.2	64.3	63.2	69.0	67.0
ACC [%]	78.0	78.7	78.4	81.7	80.8

The clustering-based and improved methods misdetermined the building as the background as shown in Fig. 6(b) because Fig. 6(a) has a building region at the bottom of the image, which is beyond the assumption of the clustering based and improved methods. This misdetermination should be solved by improving the initial seed generation.

6. Conclusion

To achieve faster building extraction, an improved method, which uses a reduced image to obtain the center of the clusters, was proposed. The computational time of the improved method was 86.54% shorter than that of the clustering-based method, which is the basis for the improved method. The improved method completed in 3.66 s, which was shorter than the target computational time. Furthermore, it significantly increased the extraction accuracy by 1.8%. This enhancement was provided by the fact that the improved method increased the accuracy of the background candidate by 3.7%. Our analysis shows that improvement in the extraction accuracy is attributed to the fact that the method tends to prevent over-clustering.

References

[1] J. Åhlen and S. Seipei, "Knowledge based single building extraction and recognition," Proc. Recent Advances in Computer Engineering, Communications and Information Technology, pp.29–35, Dec. 2014.
[2] D. Ueno, H. Yoshida, and Y. Iiguni, "Automated GrowCut with multilevel seed strength value for building image," Trans. Institute of Systems, Control and Information Engineers, vol.29, no.6, pp.266–274, June 2016 (in Japanese).
[3] T. Futagami and N. Hayasaka, "Automatic building extraction based on analysis of color cluster," Trans. Institute of Systems, Control and Information Engineers, vol.32, no.9, pp.349–356, April 2019 (in Japanese).
[4] Q. Wang, Q. Zhi, P. Liu, and G. Ma, "Realization of embedded RTK system and performance evaluation," Proc. China Satellite Navigation Conference, pp.639–648, May 2018.
[5] V. Vezhnevets and V. Konouchine: "GrowCut: Interactive multi-label N-D image segmentation by cellular automata," Proc. Graphicon, pp.150–156, June 2005.
[6] C. Rother, V. Kolmogorov and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol.23, no.3, pp.309–314, Aug. 2004.
[7] X. Li, K. Liu, and Y. Dong, "Superpixel-based foreground extraction with fast adaptive trimaps," IEEE Trans. Cybern., vol.48, no.9, pp.2609–2619, Sept. 2017.
[8] H. Shao, T. Svoboda, and L. Van Gool, "Zubud-Zurich building database for image based recognition," ETH Zurich Technical Report 260, April 2003.