Neurocomputing 168 (2015) 808–822

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A novel fast image segmentation algorithm for large topographic maps

Qiguang Miao [a], Pengfei Xu [b,*], Tiange Liu [a], Jianfeng Song [a], Xiaojiang Chen [b]

[a] School of Computer Science and Technology, Xidian University, Xi'an 710071, China
[b] School of Information Science and Technology, Northwest University, Xi'an 710127, China

ARTICLE INFO

Article history:
Received 20 December 2014
Received in revised form
29 March 2015
Accepted 12 May 2015
Communicated by Jinhui Tang
Available online 22 May 2015

Keywords:
Image segmentation
Large topographic maps
Randomized sampling
Fuzzy classification
Multilevel image fusion

ABSTRACT

Image segmentation for topographic maps is challenging due to their low quality, high degrees of mixed and false coloring. Besides, many pixels cannot be explicitly separated from each other because of the fuzziness of their colors. Therefore the algorithms based on fuzzy theory are suitable to process such images utilizing their ability to deal with the blurring effect. However, there are still some problems of large-scale data, high time complexity and inaccurate classification. In order to overcome these problems, we propose a novel algorithm for segmenting large topographic maps based on the ideas of fuzzy theory, randomized sampling and multilevel image fusion. In this algorithm, the large topographic map is randomly sampled first. Then the optimal clustering centers are acquired by fuzzy C-means (FCM) clustering. Further, the map is segmented with the fuzzy classification method. Finally, multilevel image fusion is used to fuse the segmented images into the final segmentation maps. Randomized sampling is used to reduce the amount of data, and improve the efficiency of image segmentation. Multilevel fusion can make the final segmentation maps more accurate. The experiments show that our method has higher efficiency and accuracy than the traditional ones. It provides a reliable image segmentation method for large topographic maps.

1. Introduction

The topographic map is a special information carrier. In a map, a wide variety of geographical elements intertwines in the spatial domain. However, the biggest difference between topographic maps and nature images is that different kinds of geographical elements are expressed by different colors. Therefore, color features provide the key to extracting geographic information, which can be detected using image segmentation methods based on color information [1]. At present, many image segmentation methods for topographic maps have been proposed. These methods make use of the color information and the spatial information [2]; they choose different color spaces [3–5], and apply other theories of color separation to segment topographic maps [1,6–9], which achieved good results in the corresponding aspects, such as the segmentation accuracy of the pixels or the segmentation efficiency, etc.

By constructing the color histogram [10–14], Bernd and Serguei used various methods to find the optimal threshold to segment topographic maps [11,12]. These methods have the advantages of reduced computation and no prior information is required. But when the maps contain noise, it is difficult to determine the threshold of the histogram. On the other hand, image segmentation can be considered as the classification of the objects. So pattern classification techniques are used for segmenting topographic maps [15,16]. The most common methods are k mean clustering and FCM clustering [5, 17–20][], and many improved algorithms based on them have been proposed. Zheng et al. proposed an image segmentation method by combining the spatial and color information, and the 2-D histogram is used in the FCM clustering algorithm with fuzzy restriction [2]. Xie et al. presented a strategy to adjust membership on the basis of weighted FCM clustering algorithm, and the 2-D color histogram is used to determine the weighting factors [21]. Xin proposed a segmentation method based on the class center constraints for maps [22]. But the color features of topographic maps contain many uncertain factors [23], and there exist color distortion and scattering. So the color information of the maps is fuzzy. If the computer can deal with the blurring effect in the process of color reorganization, the segmentation method could perform well. Therefore, the fuzzy classification methods can solve the problems of graduated color and parti-color to some extent.

At present, for small topographic maps or areas of the maps, different image segmentation methods can achieve good results. But for large topographic maps, there are difficulties for these methods because of the large-scale data and long computation time, which may be far beyond the acceptable limit. These factors would make these methods unsuitable. In order to improve the efficiency and the segmentation results, this paper proposes an image segmentation method for large topographic maps via randomized sampling and multilevel image fusion (IS-RSMIF).

* Corresponding author.
E-mail address: pfxu@nwu.edu.cn (P. Xu).

http://dx.doi.org/10.1016/j.neucom.2015.05.043
0925-2312/

The organization of this paper is as follows: In Section 2, we analyze the traditional topographic map segmentation methods, and propose our solutions to the problems existing in the most commonly used method. The proposed method is described in Section 3. In Section 4, we compare the experimental results in terms of accuracy and efficiency among different state-of-the-art techniques. Finally, the concluding remarks are given in Section 5.

2. The existing problems of the commonly used methods

There are many methods for segmenting topographic maps, and most of these methods are based on histogram, fuzzy classification, k Mean Clustering, and classification by the characteristics. They perform well to a certain degree. However problems remain, when these methods are used to segment large topographic maps. In this section, we analyze three main problems of topographic map segmentation, which are the acquisition of the prior data, the self-adapting determination of the optimal class number and the fusion of the segmented images. On that basis, we put forward the corresponding solutions to these problems, respectively.

2.1. The acquisition of the prior data

The current segmentation methods have good performance when they are applied to segment some small maps. But when the maps are larger, in the order of millions of pixels, these methods would become unsuitable because of high computation cost, which may be far beyond the acceptable limit. The simplest solution to this problem is to extract the prior data, which may be some areas or pixels sampled from the maps. The prior data is used to obtain the clustering centers by some clustering algorithms [2,5]. Then the large maps are segmented according to the clustering centers. But in order to make the clustering centers optimal, the prior data should meet the following conditions:

- The amount of the prior data should be enough;
- The prior data should contain enough information of different colors;
- The ratios among different colors should remain unchanged.

The traditional method is a process of manual selection; some areas in the large maps are selected as the prior data. But this method is difficult for the operators to determine the positions of the areas in the large maps and the size of the prior data. Pyramid decomposition [24,25] can reduce the account of the data by $2^l \times 2^l$, where l is decomposition levels. In this method, the prior data is extracted by interlaced and separated sampling. Since there exist lots of lines distributed in the maps horizontally and vertically, a large number of pixels on these lines may be removed. This makes the prior data unreasonable.

According to the analysis above, this paper uses a sample method based on the random distribution model to obtain the prior data. A matrix R is generated randomly, and follows an uniform distribution. The size of R is the same as that of the large maps, and the value of $R(i,j)$ is in the range $[0, 1]$, where (i, j) is the coordinate of an element in R. Then a sampling probability P is set. The value of P is determined by the following equation.

$$P \approx \frac{T}{M \times N} \tag{1}$$

If $M \times N \le T$, P is set to be 1. P is in the range $(0, 1]$. We define a new matrix as the sampling matrix, if $R(i,j) \le P$, then $R'(i,j) = 1$, else $R'(i,j) = 0$. So the sampling data is directly related to the sampling probability P, which can be computed according to the relationship between the threshold T and the size of the maps, and T is used to control the amount of the prior data. If the value of T is closer to $M \times N$, the sampling data can represent the information of the original map better, but it also has higher computation cost. And its value is determined according to the maps, if the maps have higher image quality, T can be smaller; otherwise, T should be bigger. In general, the value of T can be set as 10,000. So if the value of T is chosen properly, the sampled data can be used as a good representation of the original map.

The prior data obtained by this method can guarantee that the ratios between different colors remain unchanged, and the prior data is appropriate for large topographic maps. Suppose that the color K has N_{Ak} pixels in the large map A (with the size of $M \times N$). The ratio of the color K in A is

$$P_{Ak} = \frac{N_{Ak}}{M \times N} \tag{2}$$

The selected probability of all the pixels is P. So the sampled prior data has $M \times N \times P$ pixels, and the color K sampled has $N'_k = N_{Ak} \times P$ pixel. The ratio of the color K in the sampled data is

$$P'_k = \frac{N'_k}{M \times N \times P} = \frac{N_{Ak} \times P}{M \times N \times P} = \frac{N_{Ak}}{M \times N} = P_{Ak}. \tag{3}$$

So the ratios between different colors remain the same, and the amount of the prior data can be adjusted flexibly by P.

2.2. The self-adapting determination of the optimal class number

Most topographic map segmentation methods are based on clustering or classification, and it is very important to determine the optimal class number, which directly affect the accuracy of the segmented results. According to the cartography of the Chinese maps, Zhu Wenzhong proposed that the class number should be nine [26]. Black is one of them, and all the colors overlay by which become black; another is white as the background; and the others are produced by overlapping with each other. So the number is

$$1 + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} + 1 = 9 \tag{4}$$

where $\binom{x}{y}$ represents mathematical combination, which means that there are x kinds of colors in the map, and y kinds of colors overlap each other.

However, in different countries, the cartography rules are different; the number of colors, which are used to express different geographical elements, is also different. On the other hand, in some maps, the colors overlayed by black are not always black, and the background is not white, so nine is not the optimal class number. In this section, an improved calculating method is designed to obtain the optimal class number. If there are m kinds of colors in the maps, the optimal class number $Cluster\_n$ should be defined as

$$Cluster\_n = \binom{m}{1} + \binom{m}{2} + \binom{m}{2} + \cdots + \binom{m}{m} + 1 \tag{5}$$

where $\binom{m}{1}$ is the number of the original colors, $\binom{m}{2} + \binom{m}{3} + \cdots + \binom{m}{m}$ is the number of the produced colors due to color overlaying, and "1" is the number of the background color.

2.3. The fusion of segmented images

Due to the existence of varied colors, secondary colors and graduated colors, the number of the segmented images is more than that of colors in topographic maps. So these segmented