Line Separation from Topographic Maps Using Regional Color and Spatial Information

Pengfei Xu1, Qiguang Miao2, Tiange Liu3, Xiaojiang Chen1 and Dingyi Fang1
1 School of Information Science and Technology, Northwest University, China
2 School of Computer Science and Technology, Xidian University, China
3 School of Information Science and Engineering, Yanshan University, China
pxu@nwu.edu.cn, qgmiao@xidian.edu.cn, liutiange@ysu.edu.cn,
xjchen@nwu.edu.cn, dyf@nwu.edu.cn

Abstract
The lines in topographic maps are difficult to be separated from each other because of their confusing colors. To solve this problem, we propose a novel line separation method using their regional color and spatial information. Firstly, we divide the lines into lots of circular regions with a certain diameter, and consider these regions as the basic processing units. Then based on a new concept of regional color confusion, we classify all the divided circular regions into two kinds of regions by whether the color is pure or mixed. Further, for pure color regions, a fuzzy clustering algorithm with Gaussian kernel can be used to cluster them into different lines based on their color information. Meanwhile, we determine the memberships of the mixed color regions according to their spatial relations with the clustered pure color regions. The concept of regional color confusion is proposed to reduce the influences of the confusing colors to line separation, and the spatial relations are utilized to solve the problems of the membership determination of the mixed color regions. The experimental results demonstrate that our method can achieve higher accuracy compare with other two state-of-the-art methods, which provides a novel idea for line element segmentation from scanned topographic maps.

1 Introduction
Topographic map carries a wealth of information of this planet over a very long period [Chiang et al., 2014]. There are a large number of historical topographic maps, which have been created in the past several centuries, and most of those maps were scanned and stored as raster images1,2,3,4. The information contained in those digital maps including landscape ecology, land-cover changes and urbanization is very precious for many research topics, which leads to the automatic recognition of geographic elements in those topographic maps become essential. In topographic maps, a wide variety of geographical elements including contour lines, rivers, lakes, roads, and etc., are intertwined recorded [Chiang et al., 2011][Chen et al., 2006]. In order to efficiently and accurately recognize different geographic elements, element separation should be performed in the first place [Chiang et al., 2014]. However, color distortion happens due to the oxidation and pollution, during the long-term storage. Besides, false color caused by RGB channels misalignment and color aliasing induced by the scanner’s point spread function also frequently happen in the scanning process. All these facts result in the poor quality of obtained map images (As shown in Figure 1). So our main challenge is how to obtain satisfied element separation results for these poor map images. In topographic maps, line elements carry most geographic information [Liu et al., 2016a][Miao et al., 2013]. Thus, in this paper we focus on segmenting lines including all the straight lines and other curve lines.

At present, most of element separation methods for the topographic maps regard maps as a set of pixels [Khotanzad and Zink, 2003][Chang et al., 2014a], and utilize the color of pixels and very limited spatial information to classify each pixel into different layers. These pixel-based methods rarely consider the regional features of geographic elements, and thus hardly deal with the problems of color distortion, false color and color aliasing. The results of these methods usually show with “pepper and salt” noise and poor continuity of lines, as shown in Figure 2.
In 2010, Chen et al. proposed an object-oriented automatic segmentation algorithm [Chen et al., 2010], which considers line segments as the processing units for the first time. Compared with the traditional pixel-based algorithms, their results are significantly improved. However, the color information of each line segment is expressed by the average color of these lines, which would bring statistical mistakes when false color, color distortion, and even elements overlapping happened, as shown in Figure 3. In order to solve this problem, Liu et al. in 2016 [Liu et al., 2016a] defined the majority color of the lines as their main color, which is used to separate the lines. However, this kind of statistical method works well only for the long line segments with many pixels, but it is not applicable for the short lines with color distortion. Further, Liu et al. proposed another line separation algorithm based on region growing [Liu et al., 2016b], which has the idea of separating lines under the guidance of regional information. Nevertheless, this method does not make a special process for the regions with more chaotic colors (especially for node regions), which results in a number of fractures at the intersection points in the results.

According to human visual system, there are some obvious reasons for that people can quickly and accurately identify the lines representing different geographic elements[Thorpe et al., 1996][Isik et al., 2014][Chang et al., 2014b]. Firstly, human focus on not only the single pixel, but also the details of a local region. They also analyze and synthesize the interrelationships among the regions, rather than sticking to the isolated regions for identification. Moreover, the pixels and their neighboring ones affect each other, and constitute some regions with certain color and texture features, as shown in Figure 4. Therefore, how to make full use of the regional features and the relationships among these regions is the cue that motivates us to accomplish element separation, especially for line separation focused by us in this paper.

Based on all the analysis above, we propose a novel line separation method based on regional color and spatial information. This new method can be regarded as a line separation process with two paths, which are responsible for separating two types of regions respectively. In this method, we divide the lines extracted from topographic maps into small circular regions, which are classified into pure color regions and mixed color regions according to their color distribution. And then, a fuzzy clustering algorithm is used to separate pure color regions into different map layers, and the memberships of all the mixed color regions are determined based on their spatial relations with the separated pure color regions. Our method not only breaks through the limitation of considering the pixels as the basic processing units, but also overcomes the roughness of taking the whole line segments as the processing objects.

Our contributions can be summarized as follows:

1. We put forward the concept of regional color confusion, which is utilized to express the color distribution in a region. Based on this concept, the lines are classified into pure color regions and mixed color regions. Unlike the same processing strategy used for all the pixels or regions in previous existing methods.

2. We propose a new strategy for line separation from topographic maps: we put forward the mixed color regions, and we separate these special regions based on their spatial relationships between them and their neighbor pure color regions. Unlike the color information used for all the regions in these existing methods.

The remainder of this paper is organized as follows. In section 2, we review the related works. In Section 3, we analyze the color characteristics of different regions on the lines extracted from topographic maps and describe the proposed method in detail. The experimental results and analysis are given in Section 4 before concluding in Section 5.

2 Related Work

The geographic element separation from topographic maps has recently received much more attention [Chen et al., 2006][Chiang et al., 2009]. Various methods based on color histograms are applied to find an optimal threshold to segment topographic maps [Levachkine et al., 2001]. These methods have the advantages of lower computational complexity, and do not need any prior information. However, they are difficult to determine an optimal threshold, especially for the maps with poor quality. There also are several methods, which consider element separation as object classification [Mignotte, 2008], and in these methods, fuzzy classification techniques are more appropriate to solve the problem of the ambiguous color information in some topographic maps [¨Ozdemir and Akarun, 2001][¨Ozdemir and Akarun, 2002]. In 2015, Miao et al. proposed a fast image segmentation algorithm based on random probability sampling to solve the
problem that the existing methods have high computational complexity for the large maps, and achieved good performances [Miao et al., 2015].

However, the ambiguous color information of the pixels is always the bottleneck in the process of element separation. In order to solve this problem, Chen proposed an object-oriented algorithm for topographic map segmentation [Chen et al., 2010]. This algorithm takes independent lines as the processing units, which breaks through the limitation of using single pixel as the basic processing unit in the traditional pixel-based methods. Based on this idea, in 2016, Liu et al. proposed an algorithm based on the major color features of the lines [Liu et al., 2016a]. In his algorithm, he takes the main color of the lines for line separation instead of their average color. Moreover, by combining the color and spatial information of the lines, Liu proposed another contour-line color layer separation algorithm based on fuzzy clustering and region growing in 2016 [Liu et al., 2016b]. However, these methods still have difficulty in dealing with line separation in the complex topographic maps. For instance, they have no strategies for the determination of the memberships of the node regions. In this paper, we propose a novel line separation method based on regional color and spatial information to deal with some special regions better.

3 Line Separation Using Regional Color and Spatial Information

In general, there are four different types of regions on the lines extracted from topographic maps: the correct pure color regions (Figure 5 (a)), the false pure color regions contaminated by other elements (Figure 5 (b)), the mixed color regions contaminated by background (Figure 5 (c)), and those contaminated by other lines (Figure 5 (d)).

In order to separate all these four different types of regions into their own map layers, we design a new line separation method using regional color and spatial information, and the framework of the proposed method is shown in Figure 6.

3.1 The Region Division and Classification

Before dividing regions, we extract all the lines from the original topographic maps using the linear element extraction algorithm proposed by Miao [Miao et al., 2013]. Then all the extracted lines are divided into a number of consecutive circular regions by taking the pixels on their skeleton lines (which is obtained by morphological thinning algorithm [Gonzalez et al., 2014], as shown in Figure 7) as the center points and their average width as the diameter. As shown in Figure 8 (a), assuming that there are \( m \) pixels on the skeleton line \( L \), then a circular region \( R \) on \( L \) can be obtained by taking a certain pixel \( (x_j, y_j) (j = 1, 2, ..., m) \) on \( L \) as the center point and the width \( d \) of \( L \) as the diameter. Further, all pixels in this circular region are grouped together as the minimum processing unit for line separation later. Since the skeleton lines basically retain the original shapes of the lines, it is possible for us to trace these lines from \( (x_j, y_j) \) to obtain a series of circular regions along the skeleton lines, which can be seen in Figure 8 (b).

Furthermore, we put forward the concept of regional color confusion to represent the complexity of color distribution in a certain region, and it can be expressed by formula (1).

\[
E_R = - \sum_{j=1}^{c} P(h_j) \log_2 P(h_j) \tag{1}
\]

where \( E_R \) is the value of regional color confusion within the region \( R \), \( P(h_j) \) is the proportion of the pixels with the color \( h_j \) in region \( R \), and \( c \) is the total number of colors in \( R \).

Compared with the mixed color region, a pure color region has much less complexity of color distribution, therefore, \( E_R \) can be utilized to determine whether a circular region is a pure color region or a mixed color region, and which can be expressed by formula (2).

\[
\begin{align*}
E_R & \leq T & \text{R is a pure color region} \\
E_R & > T & \text{R is a mixed color region}
\end{align*} \tag{2}
\]

![Figure 7: The skeleton lines](image)

![Figure 8: Lines and their circular regions. (a)A circular region R; (b)Lots of regions.](image)
where the value of $T$ is about 4.0, which is an experimental value obtained from lots of these two types of regions.

### 3.2 Region Separation for The Pure Color Regions

The color information of the whole line is affected by all its pixels, but they have varying degrees of influence to the line’s color. Normally, the pixels far from the skeleton line have lower ability to represent the true color of this line than those near to the skeleton line, since they are easily influenced by background. Therefore, we calculate the regional color features based on the spatial relationships of each pixel from the central pixels.

It is assumed that there is a pure color region with a pixel $l_O(x_O, y_O)$ on its skeleton line as the center point, and the line width $d$ as the diameter, as shown in Figure 9. This region has $m_o$ skeleton pixels $l_G$, $G \in \{1, 2, ..., m_o\}$, and $n_o$ non-skeleton pixels $p_F$, $F \in \{1, 2, ..., n_o\}$.

For a non-skeleton pixel $p_F$, its corresponding weight $\omega_{F_l}$, which can express its influence to the line’s color, is calculated by formula (3),

$$\omega_{F_l} = \frac{1}{D_{F_l}}$$  \hspace{1cm} (3)

where $D_{F_l}$ is the minimum Euclidean distance from the non-skeleton pixel $p_F$ to the skeleton line $l_G$.

$$D_{F_l} = \min(||p_F - l_G||)$$  \hspace{1cm} (4)

Then the color features of the pure color region $C_R$ can be calculated by formula (5):

$$C_R = \left( \sum_{F=1}^{n_o} \frac{\omega_{F_l}}{\sum_{F=1}^{n_o} \omega_{F_l}} \times C_F \right) + \frac{\sum_{G=1}^{m_o} C_G}{m_o}$$  \hspace{1cm} (5)

where $C_F$ is the color information of the non-skeleton pixel $p_F$, and $C_G$ is the color information of the skeleton pixel $l_G$. It should be noted that all the color features are calculated on CIE L*a*b* color space [Liu et al., 2016b].

When the color features of all the pure color regions are obtained, a fuzzy clustering algorithm is used to separate these regions [Miao et al., 2015].

### 3.3 Region Separation for The Mixed Color Regions

Unlike region separation for the pure color regions, the mixed color regions, of which their centers connect to each other in the skeleton line, are grouped together as a region group $G_{R_m}$, as shown in Figure 10. Subsequently, these groups are classified based on their spatial relationships between them and their neighbor pure color regions which have been separated before.

Since a mixed color region may be composed of several lines, it is allowed to belong to more map layers, which can accurately show the true attributes of each mixed color region. In topographic maps, there are four cases of the spatial relations between the groups of the mixed color regions and their neighbor pure color regions, so we can determine the memberships of these groups by performing the following procedures.

1) For the isolated groups, there are no neighboring pure color regions, as shown in Figure 11 (a), then $G_{R_m} \in \Phi$, and their memberships need to determine manually. However, this situation, which shows the overlap between two lines with the same shape from the top to end, is rarely appeared in topographic maps.

2) When there is a line segment with parts of the pure color regions labeled $L_1$, and end with a group of the mixed color regions, as shown in Figure 11 (b), then $G_{R_m} \in L_1$.

3) If there is one group of the mixed color regions connected between two line segments of the pure color regions with the same label $L_1$, as shown in Figure 11 (c), then $G_{R_m} \in L_1$.

4) In a more complex case, if one group of the mixed color regions connects two or even more line segments of the pure color regions with different labels, which may be $L_1, L_2, ..., L_Q$, as shown in Figure11 (d), then $G_{R_m} \in L_1$.

Up to now, we have accomplished the separation of all the pure color regions and the mixed color regions, and the final separation results can be obtained.

### 3.4 Summary of The Proposed Method

In summary, there are mainly five formulas in the proposed method which being successively performed to achieve the segmentation. Formula (1) is used to calculate the complexity of color distribution in a certain region of the lines. Formula (2) is utilized to determine whether a circular region is a pure color region or a mixed color region. Formula (3) and...
Formula (4) is combined to calculate the weights, which can express their confidence of representing the areas’ colors. Finally, formula (5) is used to calculate the color features of the pure color region, which will be put into the subsequent clustering algorithm to separate these pure color regions. For further demonstrate the algorithm, the pseudo code is listed in the following.

Algorithm 1: Line separation using regional color and spatial information

**Input:** Input image I (an original topographic map)  
The total number C of colors in image I.

**Output:** The separated maps: $M_1, M_2, ..., M_C$

1. Apply line extraction algorithm to obtain the linear element image $I_{line}$
2. Divide lines of $I_{line}$ into n small regions $R_i, i = 1, 2, ..., n$
3. Classify $R_i$ into p pure color regions $R^pure_i, l = 1, 2, ..., p$, and q mix color regions $R^mix_i, m = 1, 2, ..., q$
4. For $R^pure_i$, apply Fuzzy clustering method to separate $R^pure_i$ into C classes $M_c, c = 1, 2, ..., C$
5. For $R^mix_i$, calculate the relationship between $R^mix_i$ and all $M_c$, and merge $R^mix_i$ into the corresponding $M_c$
6. Obtain the separated maps: $M_1, M_2, ..., M_C$

4 Experiments and Analysis

In order to verify the effectiveness and feasibility of the proposed method, three original topographic maps with different sizes are used as testing images, as shown in Figure 12. Besides, FCM algorithm using a single pixel as the basic processing unit [Miao et al., 2015] and Liu’s method [Liu et al., 2016a] are exploited as the comparison methods. Moreover, we make subjective and objective evaluations on the experimental results respectively. Since the contour lines are the most important geographical elements in geographic information systems, the brown maps of contour lines are given for Map B and Map C with large sizes, while the other separation maps are not shown because of space limitation of this paper.

From the selected experimental results which are shown in Figure 13-17, we can see that FCM algorithm has unsatisfied performances for these maps: (1) There are lots of noise in the result maps due to the complex color distribution in topographic maps, as shown in Figure 14(b), Figure 16(b) and Figure 17(b). (2) Many gaps appear on the segmented lines. The performance of Liu’s method has a great improvement. There is much less noise in the results, and the lines have better continuity as well. However, there is still some inaccurate separation in some places, especially for the lines with many nodes, which results in the gaps appear around the nodes, as shown in Figure 15(c), Figure 16(c), and Figure 17(c). Compared with these two methods, our method can achieve better performances. We use two types of regions as the basic processing units instead of pixels, which leads to less noises in the results. Besides, the spatial information is utilized to determine the memberships of the mixed color regions, and some mixed color regions may have more than one memberships, which not only can improve the separation accuracy, but also handle the nodes to get much better continuity of the result lines, as shown in Figure 13(d), Figure 16(d), and Figure 17(d).

Furthermore, we use the rate of correct separated pixels (RCP), the rate of false separated pixels (RFP) [Miao et al., 2015] and the rate of connected components (RCC) to evaluate the result maps.

Figure 11: The separation of the groups of the mixed color regions. (a) $G_{RM} \in \Phi$; (b) $G_{RM} \in L_1$; (c) $G_{RM} \in L_2$; (d) $G_{RM} \in \{L_1, L_2, ..., L_4\}$

Figure 12: The testing topographic maps. (a) Map A; (b) Map B; (c) Map C.

Figure 13: The brown results obtained from Map A with zoomed-in patches. (a) The ground truth; (b) FCM; (c) Liu’s method; (d) our method.

Figure 14: The black results obtained from Map A with zoomed-in patches. (a) The ground truth; (b) FCM; (c) Liu’s method; (d) our method.
The rate of the connected components:

\[
RCC = \frac{LN - TLN}{\max \{LN_1, LN_2, \ldots, LN_m\}}
\]  

where \(LN\) is the number of the connected components in the result maps, \(TLN\) indicates the number of the connected components in their corresponding ground truth, and \(\max \{LN_1, LN_2, \ldots, LN_m\}\) is the maximum value of \(LN\) in the result maps obtained by \(m\) kinds of methods.

In the following, we evaluate all the results with the above metrics, and the corresponding evaluations are shown in Figure 18 to Figure 20. It can be seen that FCM has lower RCP in most result maps, and has much higher RFP and much worse RCC in all the results. The results obtained by Liu’s method have obvious improvement, the lines in the result maps are more continuous, and fewer false separated pixels appear. So his method works better in RCC and RFP than FCM. In contrast to these two methods, our method has the best performances. The lowest RCC shows its separated lines has the most continuity, and its higher RCP and lower RFP show that it can separate more pixels into their own maps correctly.

5 Conclusion

There are many challenges for line separation from topographic map with low quality. This paper proposes a new
line separation method by utilizing regional color and spatial information. It divides the lines into two types of regions, which are then separated to their own result maps respectively. Especially for the mixed color regions, the spatial relations between them and the pure color regions are used to determine their memberships, which overcomes the problem of the false separation of the contaminated pixels by background or other elements, and the experimental results show that our method has a better performance.

Acknowledgments

The work was jointly supported by the National Natural Science Foundations of China under grant No. 61502387, 61702415, 61772396, 61472302; Natural Science Foundation of Shanxi Province, under grant No.2016JQ6029, 2017JQ4003; The 59th China’s Post-doctoral Science Fund No. 2016M592832. Natural Science Foundation of Hebei Province of China under grant No.F2018203096, the 61th Foundation of Shaanxi Province, under grant No.2016JQ6029, 2016M592832. Natural Science Foundation of Hebei Province of China under grant No.F2018203096, the 61th China’s Post-doctoral Science Fund No.2017M611188

References