Extracting contour lines from topographic maps based on cartography and graphics knowledge

Rui-Qing WU

School of Electronic Engineering, University of Electronic Science & Technology of China, Chengdu, 610054, P.R.China

Xi-Ran CHENG, Cun-Jian YANG

Department of Geography &Resources Science, Sichuan Normal University, Chengdu, 610068, P.R.China

Abstract: This paper addresses the problem of contour line extraction from scanned topographic maps in poor condition. A novel method is developed, using knowledge of cartography and graphics to extract contours by removing other layers which overlay the contours, and reconstructing the contour lines. The contributions of this paper are the supplementation of the use of knowledge discovery for extraction on the scanned topographic maps. Examples are presented from diverse applications to show that the developed algorithm can work effectively.

Keywords: extraction, contour line, cartography, graphics, knowledge base, symbolic

1. Introduction

There has been an increase in research on automated extraction of contour lines, and certainly, advancements have been made. Early literature about contour lines extraction focuses on the thinning method [1]. But this method is inefficient for vectorizing real-life scanned raster images. Dori and Liu present a thinningless sparse pixel vectorization (SPV) algorithm that needs post-processing [2]. J. Song et al.and Kök present a different algorithm without post-processing [3], [4]. However, as has been demonstrated, noise in the input image and line junctions are the most common obstacles to achieving satisfactory vectorization performance [5]. To provide more information and better readability, much research has been devoted to perform on color topographic maps. Most segmentation methods emphasize color space selection, as demonstrated by Spinello and Pasca [6]. Chen et al. extract the contour line segments by extraction of all of the linear features with gray-level histogram analysis [7]. However, the method uses semiautomatic extraction that requires more manual processing. It is used for color maps without blurred and/or distorted lines.

Recent developments on fully automatic feature extraction and object recognition from scanned neutral topographic maps with perfect conditions are still insufficient to provide satisfactory results in achieving accurate contour data. There are four main difficulties in this field: the first is automatic contour lines extraction from topographic maps; the second is how to recognize the contour lines accurately; the third is thick lines; and the fourth is gap. These criteria have to be examined to further the research on contour data. Our paper proposes a method to extract contour lines from scanned neutral or color topographic maps with poor conditions, based on cartography and graphics knowledge.

Our method consists of four steps. First, color space is converted to CMKY space to get a binary image. Second, the binary image is tracked and thinned by knowledge of cartography and graphics of contour lines. Third, other symbols are removed and the contour lines are saved according to knowledge of the different characteristics. Finally, corresponding methods, presented by previous research, are adopted to solve the different cases of gaps and thick lines.

2. Knowledge of cartography and graphics

There are many kinds of symbols on a topographic map, such as contour lines, houses, roads, characters, and numbers. Each kind of symbol has characteristics that provide information of cartography different from the information provided by contour lines. These scanned topographic maps are displayed on computers using graphic technology.

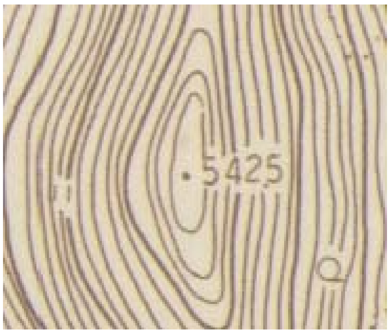
2.1 Cartography knowledge

There is much useful knowledge on the mathematics and introduction of the topological map, such as, scale, coordinates system, symbol system. Every kind of symbols has its graphical feature. For example, a contour line is composed of different parcel units, and any parcel unit of a dashed and dotted contour line should have promissory size (length and wide) with standard color. Arcs of a contour lines are represented by a B-spline curve which mean the grade orientation must be consecutive. The obliquity of consecutive lines constituted by several points should be almost the same. Every kind of symbols has its topological feature. For example, a valid contour line must not contain a self-intersecting arc. Intersections of contour line geometries must be avoided, since they prohibit the creation of a topologically consistent contour line subdivision. And on the whole, the contour lines should be approximately parallel curves.

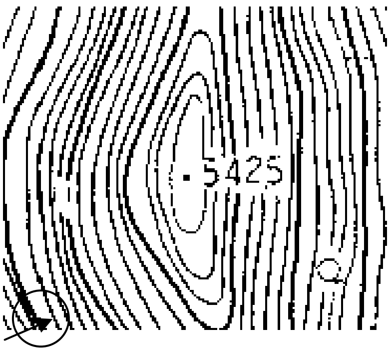
3. Binary Image Conversion

In order to obtain a binary image in the CMYK space, the scanned topographic map is first converted into CMYK space, and then the value of the K channel is used to perform binary segmentation. The image can be divided into objects and background by applying the following equation:

$$gray = \begin{cases} 255 & key = 0 \\ 0 & others \end{cases} \quad (1)$$



(a) Color topographic map sample.



(b) Segmentation result of (a).

Fig.1 Binary for a color topographic map.

Fig. 1(a) shows a color topographic map sample. Fig. 1(b) is the resulting segmentation the image of the color topographic map sample using the value of the K channel in the CMYK space. It is a good segmentation.

4 Tracking for vectorization based on knowledge

In order to track a contour line, three requirements should be fulfilled:(i) The contour lines must be centered, which mean the contour line is a centerline, (ii) The contour must appear smooth, and (iii) The characteristic of the contour must be retained. Line tracing is performed as follows.

4.1 Calculate the edge point for tracking

The Sobel is adopted to calculate the gradient field ∇f .

In accordance with the following equation:

$$\nabla f = f(x,y) \otimes G(x,y) \quad (2)$$

, the orientation of a contour line passing through the point is obtained. The next edge point for the track is selected on the following conditions:

(i) The gradient of the edge point is identical with the direction of part of the contour line.

(ii) The reverse must be avoided.

(iii) The contour line must be a single pixel.

(iv) Contour lines must be consecutive.

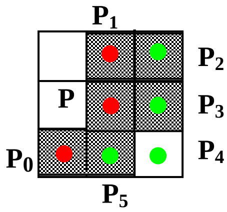
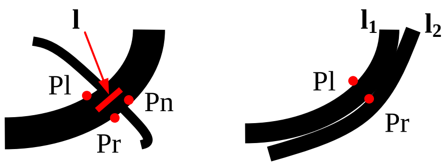


Fig. 2 Track sample



(a) junction (b) thick lines

Fig. 3 Junction and thick lines sample

The next point for tracing must be an edge point (P_1 in Fig. 2). Equation (3) is adopted to obtain the edge point. For a contour line with a single pixel, the relation of the last edge point for tracing (P_0 in Fig. 2), the current edge point, and the next point must be considered.

$$edgeflag(x,y) = \begin{cases} 0 & , f(x_i, y_j) = 0 \\ 1 & , others \end{cases} \quad (3)$$

Where , $i = x-1, x, x+1;\ j = y-1, y, y+1$.

In Fig. 2, The P and its eight adjacent points are calculated using equation (4). Labels are given to the results of the calculations. Label 1 shows that the point is not a single pixel. Label 2 shows that the reverse is produced. Label 3 shows that the point is either a non-single pixel or a reverse point. Label 0 shows that the point is the next edge point for tracing. If there is not a point labeled 0, Section 3.4 provides the method to resolve the problem.

$$nextflag(x, y) = \begin{cases} 1 & , A_1 = 1 \\ 2 & , A_2 = 1\ or\ A_3 = 1 \\ 3 & A_4 = 1 \\ 0 & , others \end{cases} \quad (4)$$

where

$$\begin{aligned} A_1 = {} & \max(|x-x_0|, |y-y_0|) = 1 \&\& \min(|x-x_0|, |y-y_0|) = 0 \&\& \\ & \max(|x-x_1|, |y-y_1|) = 1 \&\& \min(|x-x_1|, |y-y_1| = 1) \end{aligned},$$

$$\begin{aligned} A_2 = {} & \max(|x-x_0|, |y-y_0|) = 1 \&\& \min(|x-x_0|, |y-y_0|) = 1 \&\& \\ & \max(|x-x_1|, |y-y_1|) = 1 \&\& \min(|x-x_1|, |y-y_1| = 0) \end{aligned},$$

$$A_3 = (x = x_1 \&\& y = y_1) \quad \text{,and}$$

$$\begin{aligned} A_4 = {} & \max(|x-x_0|, |y-y_0|) = 1 \&\& \min(|x-x_0|, |y-y_0| = 0) \&\& \\ & \max(|x-x_1|, |y-y_1|) = 1 \&\& \min(|x-x_1|, |y-y_1| = 0) \end{aligned}.$$

Furthermore, for the influence of mapping and scanning, the gradient of the points may not be accurate. A sector based on orientation θ is produced for the next edge point search. The next edge point can be obtained by using equation (5).

$$\theta = \theta_1\ or\ \theta_2 \quad , \theta_1 \le \theta \le \theta_2 \qquad (5)$$

where $\theta_1, \theta_2 \in \{n * pi/4 | n = 0 \sim 8\}$.The influence of junction lines and thick lines may cause the gradient of the edge point to be different from the contour line, and the next edge point may be false.

A Junction

The next edge point is searched from both sides of a contour line. In Fig. 3(a), P1 and Pr do this at the same time. The gradients of Pl and Pr should be similar. If the gradient of the next edge point of Pl or Pr changes greatly, a junction occurs. If the gradient of Pl(Pr) changes, only Pl(Pr) is influenced. If the gradients of Pl and Pr change, the next edge point belongs to the junction part of the contour line and should not be searched; only the center line is searched. After the junction is processed, the last part of the contour line is used to search the next edge point. Three consecutive points are selected, and the gradient k of the contour line is calculated, using the least squares estimation in equation (6). A new edge point for tracing is obtained (Pn in Fig. 3(a)).

$$\mathrm{K} = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \qquad (6)$$

B Thick lines

Thick lines make it difficult to search the next edge point. (Fig.3(b) point Pr.) The width of the contour line helps with the decision. If the width of Pl is doubled, the edge point search is done only on one side of the contour line. Other thick lines should be deleted and wait for restoration.

4.2 Search the centerline

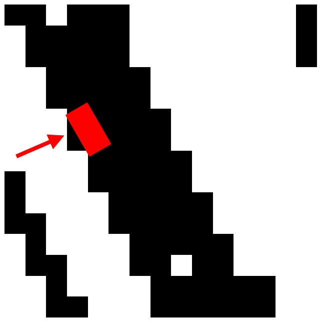
Line scanning is adopted to get the normal of the edge point according to the gradient. Then, the middle point can be searched under the following conditions:

(i) Two middle points searched consecutively may not be consecutive, and the gap appears and must be resolved with a beeline.

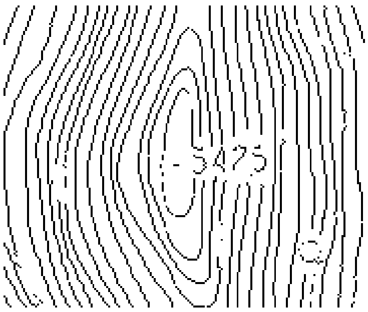
(ii) Skin the contour line

The traditional skin algorithm wastes time and is unstable [8]. According with the knowledge of graphics, the seed fill algorithm is used to skin the contour lines. The normal of the current edge point, the last edge point, and the edges compose a region. Non-center points are removed and the center contour line can be obtained. The skin of the thick lines must be resolved.

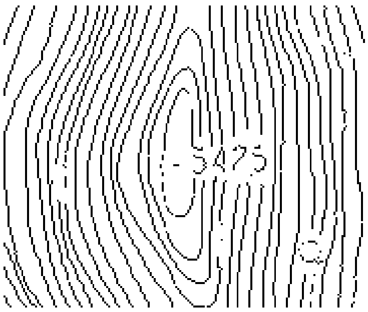
Fig. 4(a) is a zoomed-in view of a thick line sample from Fig. 4(b). Contour lines may be lost in the thickness. The width is adapted to help resolve it. The equation (7) is used to calculate the width of the contour line. The least wide section of the contour line is represented by n1; the widest section of the contour line is represented by n2; and σ is a slack field. The centerline can be obtained according to the wide and normal. Fig. 4(c) is the result of resolving thickness. If the width of the thick lines exceeds two contour lines, the thick lines have to be removed and wait for contour line reconstruction.



(a) Thick lines sample.



(b) Tracked result.



(c) Resolved result.

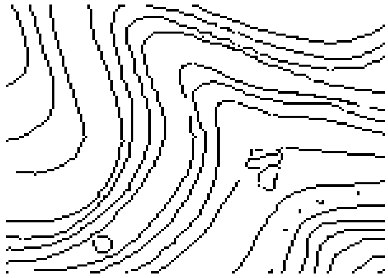
Fig.4 Thick lines of two contours are removed.

$$N=\begin{cases} 1 & wide \in (n_1, n_2) \\ 2 & wide \in (2n_1+\sigma, 3n_2+\sigma) \\ n(n>2) & others \end{cases} \qquad (7)$$

4.3 Removing symbols other than contour lines

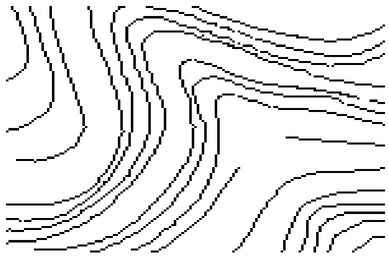
Fig. 5(a) is an image of contours with lettering and symbols of a grassplot. The parcel unit composed of the grassplot is a line segment, and the length and interval of the parcel units, which are parallel lines and uprights, can be obtained. The grassplot differs from contour lines and can be removed. Fig. 5(b) is an image of contours with the symbol of a house. Figs. 5(c) and (d) show the results after removal of lettering (a) and the house symbol (b).

(a) Lettering overlays the contours.



(b) House breaks the contours.

(c) Result after lettering is removed.

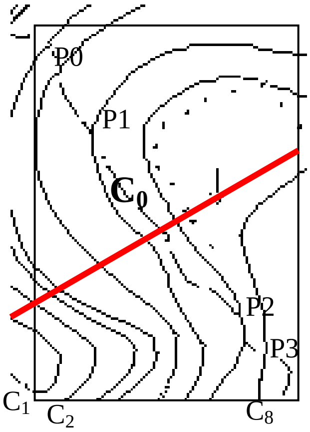


(d) Result after the house is removed.

Fig.5 Samples of removing other symbols.

Roads are easily extracted because of their shape, which is similar to contours. However, they do not come together in a group like contours, and they generally intersect a group of contours. Fig. 6(a) is a topological map. Fig. 6(b) is the vectorization result. For recognition of roads, a window is used in which a line intersects many curves (Fig. 6(b) C_1~C_8) that is orthogonal to the current curve (Fig. 6(b) C_0). If C_0 approximately parallels C_1~C_8, and C_0 does not intersect two or three curves of C_1~C_8, it is a contour line; otherwise, it is a road. Fig. 6(c) shows the result of removing the roads.

(a) Binary image.



(b) Vectorization result.

(c) Removing the road.

Fig.6. Sample of the road being removed.

4.4 Reconstruction of contour lines

Because contour lines are always intersected or overlapped with other features and discontinuous supplementary contour lines are used, contour lines disconnect with a gap. Fig. 5(c), Fig. 5(d), and Fig. 6(c) show the gap. Many methods can be adopted to solve the different cases of gaps.

Eikvil et al. solve the problem of gap reconstruction by assuming that there is only one possible continuation from an end point [8]. Chen et al. have recently used this approach [7]. Spinello and Guitton propose that contour lines should be vectorized using a Delaunay triangulation and connect with strictly geometric rules [6], however the global approach they use is not optimal. Pouderoux and Spinello propose a novel parameterless reconstruction scheme based on the extrapolation of the gradient orientation field from the available pieces of thinned contours and orientation flow to fill the gaps with a smooth curve that respects the tangents at the end points by using the method offered by Chessel et al. and Debora and Petia[9][10][11].

In this paper, the binary map made of spaghetti contours was obtained from the above algorithm. Each contour line was interpolated with a B-spline, and normal orientation of each contour line could be obtained [9]. AMLE interpolation operator was adopted to estimate the orientation of the normal at every point of the image [10]. For each pair of end-points, a weight given by an energy function (Eq.(8),G(P) is the orientation of B-spline in point P, F(P) is the normal orientation of P by interpolation) was associated to construct a matrix W by calculating the function for each pair of end-points. The contour line was interpolated with a B-spline, and the orientation of point P was obtained. Fig. 7 illustrates the reconstruction scheme: the original contour lines C1 and C2 are in black, the B-spline is in green, and the pixels of the curve from the endpoints of C1 and C2 reconstructed using the algorithm described by Pouderoux are in blue[9]. Then, global matching of the end-points is performed to find a perfect maximum weight match. Once the match is obtained, the gap between these end-points is filled [9]. Fig. 8(a) is a topological map sample, and Fig.8 (b) is the reconstructed contour lines.

$$E(\mathrm{N}_1,\mathrm{N}_2)=\sum_{t=0}^{1}\left|\overrightarrow{G(p)}\cdot\overrightarrow{F(p)}\right| \quad (8)$$

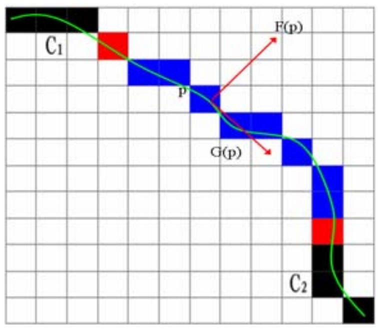
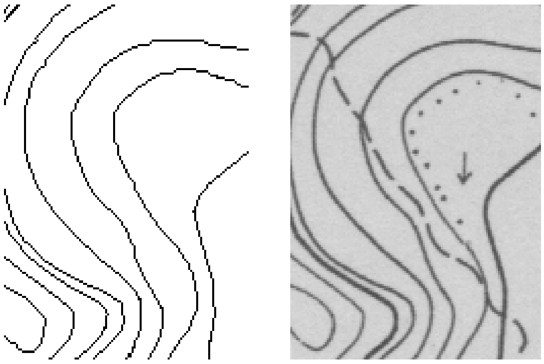


Fig. 7 Reconstructed contour between a pair of end-points.



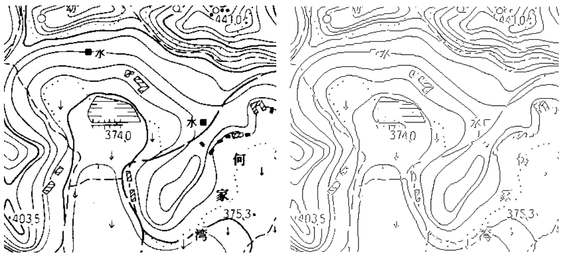
(a) Image sample. (b) Reconstruction result.

Fig. 8 Contour lines reconstruction.

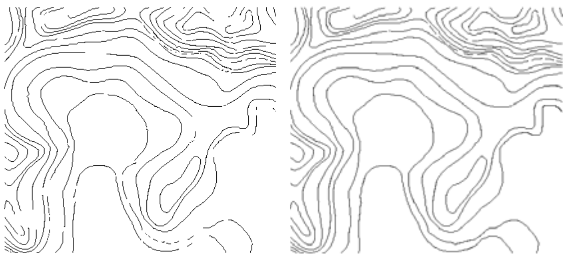
5. Result

To evaluate our method, we give a full procedure of experiments and results as shown in Fig.9 .Fig.9 (a) is a color topological map with size of 410 × 372 pixels. The value of the K channel in the CMYK space is shown in Fig. 9(b). The binary image, after binary segmentation, is shown in Fig. 9(c). After tracing and thinning, the result is shown in Fig.9 (d). Other symbols are removed in Fig. 9(e). After reconstruction, the contour lines are obtained in Fig. 9(f).The total computing time is about 30 seconds. It can be seen that the contour lines were correctly extracted and the contours are smooth and natural-looking.

(a) Brown topological map (b) K channel image



(c) Binary segmentation (d) Tracing and thinning



(e) Remove symbols (f) Contour lines reconstruction

Fig .9 Full procedure of extracting contour line

6. Conclusion

Although many raster-to-vector approaches have been developed, automated contour line extraction from scanned topographic maps still poses a problem. In general, the process is semi-automated, rather than fully automated, due to the intricate and time consuming image acquisition process. We have presented an efficient technique to extract contour lines. Our method is based on the knowledge of cartography and graphics in scanned topological maps. The results obtained on different maps are very satisfying in terms of both error rate and appearance.

Our method has been applied to extract topographical contour lines of neutral and color maps with poor conditions. But it could be applied to the extraction of other symbols on a topographical map. Furthermore, it could be used to create a special data base for GIS. However, the method should be optimized to save computation and improve speed.

Acknowledgements

This work has been supported by the National Nature Science Foundation of China through grant No.40771144, and the National 863 Plans Projects of China through grant No.2009AA12Z140. Finally, the authors express their thanks to C.H. Zhou for contributing some spatial modules of the software and the maps to this study.

References

[1] San, L.M., Yatim, S.M., Sheriff, N.A.M.,Extracting contour lines from scanned topographic maps. Proceedings of International Conference on Computer Graphics, Imaging and Visualization 2004, pp.187- 192.

[2] Dori, D., Liu, W.Y.,Sparse pixel vectorization: An algorithm and its performance evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.21, No.3, pp.202-215.

[3]Song, J.Q., Su, F., Chen, J.B., Tai, C.L., Cai, S.J.,Line Net Global Vectorization: an Algorithm and Its Performance Evaluation. Proceedings of the 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South Carolina, vol. 1, 13-15 June, 2000, pp. 383-388.

[4]Kök, E.H.,Developing An Integrated System For Semi-Automated Segmentation of Remotely Sensed Imagery. MSc., Thesis, Geodetic and Geographic

Information Technologies (GGIT), Middle East Technical University, Turkey, pp.146.

[5]Liu, R. Z., Huang, W. H., Tan, C. L.,Extraction of Vectorized Graphical Information from Scientific Chart Images. Proceedings of 2007 Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Vol. 1, 23-26 Sept, 2007, pp. 521-525.

[6] Spinello, S., Pascal, G., Contour line recognition from scanned topographic maps. Journal of WSCG, 12 (1-3), pp.419-426.

[7]Chen, Y., Wang, R. S., Qian, J., Extracting contour lines from common-conditioned topographic maps. IEEE Transactions on Geoscience and Remote Sensing, Vol.44, No.4, pp. 1048-1057.

[8] Eikvil, L., Aas, K., Koren, H.. Tools for interactive map conversion and vectorization. Proceedings of the Third International Conference on Document Analysis and Recognition, vol. 2, 1995, pp. 927-930.

[9] Khotanzad, A., Zink, E.,Contour line and geographic feature extraction from USGS color topographical paper maps. IEEE transactions on pattern analysis and machine intelligence, Vol.25, No.1, pp. 18-31.

[10]Pouderoux, J., Spinello, S.,Global Contour Lines Reconstruction in Topographic Maps. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) .Vol. 2, pp. 779-783.

[11]Chessel, A., Fablet, R., Cao, F., Kervrann, C., Orientation Interpolation and Applications. IEEE International Conference on Image Processing, 8-11 Oct. 2006 Atlanta, GA., pp. 1561-1564.

[12]Debora, G., Petia, R.,Extending anisotropic operators to recover smooth shapes. Computer Vision and Image Understanding archive,Vol. 99,No.1,pp.110-125.

Received: May 2009. Accepted: July 2009