Contour Lines Recognition from Raster Topographic Maps

Loh Mun San¹, Safie Mat Yatim², Noor Azam Md Sheriff³
Department of Computer Graphics and Multimedia,
Faculty Science Computer and Information System,
Universiti Teknologi Malaysia,
81310, Johor, Malaysia.
munsan81@hotmail.com¹, safie@itc.utm.my², azam@fsksm.utm.my³
+6 012-2574495  +6 03-2694692  +6 07-5532322

Abstract

Map digitization is still an ongoing research in development of GIS, photogrammetry, remote sensing, military, urban planning and etc. Majority of recent approach, either associates with human interaction or semi-automated process are insufficient in various aspects. There are usually time-consuming, inefficient, and shown inaccuracies in digitized data. This research is intended to develop an efficient approach for contour lines recognition from scanned topographic maps. A detail investigation of existing automated map digitization process based on formal, heuristic, knowledge and interactive algorithm is carried out. Several vectorization and image acquisition techniques are performed at early stage to analyze the map for ensuring all features on map being well separated, located, classified, and traced, consequently stored in vector format. The overlaid information and texture background are distinguished and segmented by color quantization approach. Due to scanning or deformation process, broken or incorrect contours can occur after recognition process. Several procedures based on local and global geometries are used to authenticate and reconstruct the digitized contour lines. Finally, the efficiency and accuracy of the result is evaluated based on computation time, cost and human interaction for the entire process compare with manual digitizing approach.

KEYWORDS
Map digitization, image acquisition, vectorization, line tracing, curve reconstruction

1. Introduction

Since the use of GIS is increasing, retrievability information from topographic maps become an essence process to be integrated into GIS. Paper-based map has become a low-cost input source data compare to other surveying technique, for instance photogrammetry, remote sensing, satellite imaging, RADAR and etc.

Conventional recognition methods for contour images, either automated or semi-automated approach have two main problem areas [1, 2, 3, 4, 5]. Firstly is automatic extraction of contour lines from maps. This problem involves developing techniques that can distinguish genuine contour lines from other geographical components such as buildings, roads, and etc. However, it is difficult to achieve because they demand high level understanding of maps and they are potential to generate high error.

The second problem is how to precisely recognize contour lines. Two technical aspects are concerned: how to consistently trace contour lines without crossing other lines; and how to precisely trace contour lines. Many advanced contour recognition methods have already been reported that deal with this problem [6, 7, 8, 9, 10, 11, 12], but they are still insufficient to establish fine results without human intervention.

2. Research Methodology

2.1 Phrase I – Preprocessing

A typical topographic map consists of different types of lines, text and symbols in color, black or
white. The pre-processing is intended to extract and clean extraneous information overlaid on map for resultant a preliminary contour map.

2.1.1 Scanning Paper-based Map

Since the evolvement of scanning technology, many color image scanner is capable to produce high resolution image with compatible cost. As an initial step, a compatible resolution should be chosen to preserve both quality and accuracy in an acceptable storage for a scanned map.

2.1.2 Color Quantization

Maps are constructed in a number of layers, where each layer is printed in a distinct color, represents a subset of particular information. The colors in map are not easily distinguishable due to complex interactions among various colors used during printing and digitization process. Color quantization is a process to reduce the number of colors by replacing them with the closest representative color. The process is divided into four phases: (1) Sampling the original image for color statistics; (2) Choosing a color based on statistics; (3) Mapping the colors to their representative in the color map; and (4) Quantizing and generating new image [13].

2.1.3 Map Layers Separation

An important task for determining intersection or overlapping features on the map is first separates each feature (e.g., annotations, contour lines, physical boundaries, land use area shadings, etc) based on their representative color.

2.2 Phrase II – Processing

Image processing techniques are applied in map digitization and contour lines recognition to improve the quality of an image. Several approaches are significant to assist subsequent interpretation process related in this work.

2.2.1 Eliminate Noise

Noise is originated from paper-based map or as a result of poor-sampling process after color quantization. Since the isolated small noise spread over desire features, it causes ambiguity in contour lines classification. Median filter is applied to reduce those undesired noise before further vectorization process.

\[
\text{new \_ \_pixel \_value} = \frac{1}{9} \left[ \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right] \left[ \begin{array}{c} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \\ n_9 \end{array} \right] 
\]

2.2.2 Binarization

Binarization reduces the map resolution into two intensities value (0 or 1). Conventional thresholding is applied to segment the map into two classes: (1) Pixels those have intensity above a given threshold value (background); (2) Pixels below a given value (contour lines).

\[
g (x, y) = \begin{cases} 1 & \text{if } f (x, y) > T \\ 0 & \text{otherwise} \end{cases}
\]

2.2.3 Edge Enhancement

Edge detection is based upon discovery of local discontinuities which mainly correspond to the boundaries of objects in the map image. Several operators and filtering approach like Sobel, Gaussian, homogeneity, low pass, high pass and etc are performed to result an improved contour map. Another important image enhancement technique that typically applied is mathematical morphology [14], including dilation, erosion, opening, closing and hit and miss transform. The appropriate operator and structuring element are applied in order to improve the edge continuity for alleviating further tedious process.

2.2.4 Character Recognition

Since characters on topographic map often come into different orientations and fonts, detecting and recognizing text has been a challenging task. This is due to the complexity and mixture of characters with vector lines, polygons or textured background. Moment invariants are features based on statistical moments of characters that applied for elevation recognition in this work.

2.2.5 Thinning

The main purpose of thinning is to reduce and simplify the data needed for further process. Since topographic map represent spatial data, the skeleton produced must preserve both topology and its geometry for accurate vectorization [15]. However, due to boundary noise in original map, noisy branches may exist and result a number of additional vectors. For this reason, single or isolated pixel is removed by examining all 8-
connected neighbors for every pixel in the thinned image. If all 8-connected pixels are white, then the examine pixel is deleted.

### 2.2.6 Endpoints Detection

From the extracted skeleton lines, an edge line has three basic properties: type, position and direction. The type can either be end points or continuous lines. The ending pixel is defined to have exactly one neighbor pixel that belongs to a defined mask. In contrast, a continuous line of pixels have exactly three neighbor pixels or more. The 3x3 mask for detecting endpoints is shown in Figure 1.

![Figure 1: 3x3 masks to detect endpoints](image)

### 2.2.7 Contour Lines Reconnection

Several criteria, for instance angle, distance and orientation are taken into consideration for reconnecting broken contour lines. Distance between two endpoints is calculated in following equation; Orientation is determined using chain code; while angle is calculated by back tracing 10 pixels from detected endpoint to form a triangle as shown in Figure 2.

\[
distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}
\]

\[
\angle \theta = \sin^{-1}\left(\frac{t}{h}\right)
\]

![Figure 2: Angle calculate using triangle property](image)

Initially, emphasis is placed on nearest endpoint detected with appropriate angle and orientation. The nearest endpoint is detected in 2h×2h square window form at each endpoint. Then, cubic spline interpolation [16] is used to reconstruct the approximating curve lines. Four control points are taken from two endpoints and two points which are back traced from 10-pixel commence from corresponding endpoint (Fig. 3).

![Figure 3: Distance, angle and orientation](image)

h: Euclidean distance estimate from endpoint
a (x1, y1): Pixel back trace from endpoint
b (x2, y2): Endpoint
c (x3, y3): Nearest endpoint detected in square window (2h x 2h)
d (x4, y4): Pixel back trace from c(x3, y3)

### 2.2.8 Construct Graph Tree

In this stage, contour map with its elevation data is successfully extracted. In order to improve the performance of data retrieval and manipulation, an efficient data structure design is indispensable. The data structure is organized in contour graph tree [17], and record in several fields of identification number, number of pixels, status of closed/ opened contour, elevation value, adjacent region’s id and chained pixels.

**Data structure of a node (inter-contour region)**

<table>
<thead>
<tr>
<th>Pointer 1</th>
<th>Node id</th>
<th>Closed</th>
<th>Elevation</th>
<th>Adjacent edges</th>
<th>Pointer 2</th>
</tr>
</thead>
</table>

**Data structure of an edge (contour line)**

<table>
<thead>
<tr>
<th>Pointer 1</th>
<th>Edge id</th>
<th>Closed</th>
<th>Elevation</th>
<th>Adjacent edges</th>
<th>Pointer 2</th>
</tr>
</thead>
</table>

![Figure 4: Data structure in a contour tree](image)

### 2.3 Phrase III – Post-Processing

Post-processing is the final stage of recognition process. It is mainly based on evaluating the accuracy of recognized data and corrects some anomalies case by operator. Although this work is intended to develop a nearly automated recognition and digitization process, the result is evaluated and inspected by an operator. Several criteria should be inspected to evaluate the precision and robustness.
2.3.1 Accurateness

The definition of the accurateness of contour lines recognition is more complicated than usual character or symbol recognition [18]. The following definition is selected, where discrimination between the total pixels of contour lines in scanned map and the contour lines which successfully extracted that overlay on the same coordinate in the map is considered in our approach. Several comparisons among previous work are performed to evaluate the percentage of accuracy obtained.

\[
\text{Accuracy} = \frac{\text{Total Pixels Of Contour Lines Extracted}}{\text{Total Pixels Of Contour Lines In Scanned Map}}
\]

2.3.2 Digitizing Cost and Computation Time

The efficiency of this approach mainly relies on the time and cost needed for automated extracting contour lines compare with manual digitization. The computation time including the pre-processing from image acquisition while scanning until acquire of result in vector format.

2.3.3 Labor Intensive Cost

Since this research put emphasis on a nearly automated approach, human interaction is get rid of compare to conventional method, either manual digitization, semi-automated or interactive approach. However, due to the occurrence of anomalies cases and errors, human is required to supervise the operation.

3. Result and Discussion

The result is shown in Figure 5. From a raster topographic map, color quantization is applied to separate features into different layers. Initial contour map is generated by overlapping two or more layers. Elevation data is then recognized from map using moment invariant approach. Then, contour image is binarized and enhanced by several filter and morphological operator before thinning process. Endpoints are detected and broken lines are connected using cubic interpolation with several constraint rules (angle, distance and orientation).

There are several novel approaches implemented in recognizing contour lines from raster map. Color quantization can generate initial contour map with low cost and computation time compare with other superior approaches like fuzzy logic and neural network. Besides, cubic spline interpolation with several define rules is capable to reconstruct almost the entire gaps exist in the map with less operational time.

4. Conclusion and Future Work

Although many raster-to-vector approaches have been developed in GIS, automated contour lines recognition and extraction from topographic map is only partially resolved recently. Majority processes are interactive or semi-automated rather than fully automated due to the intricate and time consuming image acquisition process.

This research is intended to discover an efficient approach for contour lines recognition from scanned topographic map. Future work is in progress to resolve problems in developing a fully automated digitization process with least error rates and human intervention. Therefore, appropriate algorithm with specific knowledge and generic models of particular features with their interrelationships are indispensable in upcoming development. In addition, another improve process is essential to carry out for eliminating undesired features, like road, trails, and etc which have similar color representation with our recognized feature (contour lines).

6. Acknowledgment

We express appreciate to all reviewers for their helpful criticisms and suggestions to our research and manuscript.
7. References


