Shape Similarity Measures of Linear Entities

ZHANG Qiaoping LI Deren GONG Jianya

1 Introduction

Feature matching, as the name indicates, involves the identification of features from different maps as being representations of the same geographic entity. Matching criteria can include properties related to distance measures, geometry, topology, graph properties, neighborhood groupings and attribute similarity [1].

Among all the possible similarity measures, the shape similarity measure is one of the most important measures because it is easy to collect the necessary parameters and it is also well matched with the human intuition. The shape similarity measures not only concern the area of feature matching but also a lot of other fields such as map generalization, pattern recognizing, intelligent spatial inquiry, etc. Many shape similarity measures were presented, but, none of them could be well applied to the area of feature matching. Some of them are only qualitative measures, which are not suitable for comparison, others are too complex to be calculated or having dependence on translation, scale, rotation etc. In this paper a new shape similarity measure of linear entities based on the differences of direction change along each line is presented and its effectiveness is justified by a lot of tests.

This paper is organized as follows: a detail investigation on existent shape similarity measures is presented in the following section. Then we present a new shape similarity measure in Section 3, and its effectiveness is also illustrated in the same section. At last some conclusions will be drawn in the Section 4.

2 Shape similarity measures of linear entities

The shape measure is based on the technology of shape anal-
analysis, it could be used for matching linear or areal entities. The shape analysis is very useful in the area such as CAD/CAM, Pattern Recognizing, Spatial Analysis, etc., and it has held research interests from many fields such as computer science, computational geometry, mathematics, statistics, cognitive science and GIS[2].

Intuitively, shape seems like a simple concept, but in actuality, the human visual/perceptual mechanism for distinguishing and recognizing shape is complex. Consequently, in a mathematical context, shape is also complex. Unlike measurements for area or perimeter, the definition of shape is based on linguistic expressions and human intuition. Presently, mathematics and statistics are not in a position where shape can be reduced to a single number. Due to complexity of the concept, a single numerical representation of shape may be impossible to achieve[2]. On the basis of the assessment of shape analysis in variant areas, Wentz concluded that though there is a great need for shape analysis capability in a GIS, there is no suitable measure existing. At the same time he pointed out that there is a potential approach that will allow for a more powerful shape analysis technique.

An index to describe shape and allow for comparisons must meet several specific criteria. MacEachren states that "the first criterion is to develop a measure of shape uniqueness by which any shape can be distinguished from all other shapes and similar shape result in similar descriptions"[3]. The criteria for a shape index for this research are:
1) each unique shape represented with a unique number;
2) independent from translation, rotation, size, and scale change;
3) match human intuition;
4) deals with regions that contain holes;
5) easy to calculate and interpret the results.

The ideal shape index would meet all these criteria[2].

The primary reason that shape analysis is not currently part of the GIS toolkit is that no single satisfactory method has been developed[4]. Nevertheless, numerous indices have been suggested varying from simple area and perimeter calculations to complex indices using sophisticated mathematical functions[2].

In the area of artificial intelligence, people have designed many approaches to shape analysis and object identification. Chain code is one of the most common technologies to describe a curve line, which is designed by Freeman. In order to get the chain codes, the line is rasterized with a given grid size. Every two neighboring pixels on the line are coded according to their forward direction following the rules illustrated in Fig. 1(a). And a line then could be described by a string of chain codes. For example, the line in Fig. 1(b) could be described as 0,1,2,2,3,3,2[5].

If two lines a, b are described by their chain code {a1, a2, ..., an}, {b1, b2, ..., bn} respectively, then the shape similarity measure SSMab is defined as follows:

$$SSM_{ab} = \frac{1}{n} \sum_{i=1}^{n} a_i b_i$$

In which $a_i b_i = \cos \alpha_i - \cos \beta_i$, and $\alpha_i$, $\beta_i$ is the direction angle determined by the code value of $a_i$, $b_i$. Certainly the $SSM_{ab}$ value could present some information of shape similarity of the two concerned lines. But it is not independent under scale, rotation. 

![Chain code of pixel P based on 8-neighborhood](image1)

![Chain code of a curve](image2)
change. And it is difficult to compare two lines having different lengths\[^5\]. Moreover the measure is not exact as it divides all the possible direction changes into only 8 groups. These limit its potential application for linear entities matching.

Other measures are based on the segmentation of the lines. The main idea is to divide the lines into a series of sub-lines, which could be represented by some kind of simpler approximate lines. The sub-approximate-lines are described by a number of parameters according to what kind of approximate curve function it uses. The most useful methods are linear segmentation and spline function segmentation\[^3\]. L. Cinque, et al. present a method for shape description consisting of an approximation of a shape by a variable number of Bezier curve segments. In their method the user can control the accuracy of the Bezier approximation by a parameter thus controlling the complexity and resolution of the approximation process. The technique described by them is suited to a variety of shape-based image retrieval applications and matching process\[^6\].

However, these measures are still not suited to linear feature matching in the context of GIS because they are not independent under scale and rotation change either. Moreover if the two concerned lines are not divided in the same way (it happens if the two lines were captured by different people, with different accuracy), the comparison will be sure to fail.

There are also some other measures from shape analysis, such as the fractal dimension of the line, the compactness of the area that the line encompasses. The fractal dimension of the line is difficult to compute and the compactness of a region is not suitable to be used as comparing criterion because two similar shapes can have different numerical representations and two different shapes can have similar numbers\[^2\].

In feature matching context, both linear and areal entities need shape measures to identify the final matched entity, usually from the candidate entities set after distance measure comparing because in some cases there is no other measure you can resort to. On the other hand if we could gain some linear shape similarity measures it will surely be helpful to matching areal entities because an areal entity could be usually described by its boundary lines.

This paper will present a shape similarity measure of linear entities based on the differences of direction change along each line.

3 Shape similarity measure based on the differences of direction changes

A line differs from other lines in many aspects. The most important one may be the differences of direction changes along the line. If the line turns the same direction as the other at the same position, we can safely say that the two lines are similar in shape. However the problem lies in the fact that we are difficult to determine at which position along the line we will compare the direction changes because the two lines are usually failed to match by vertex to vertex. In the context of feature matching from different spatial databases, the matched entities will differ greatly in the number of vertexes, the length of each segment and the direction changes on the vertexes because of the differences between the two source databases in the scale, the accuracy, and the application intention. On the other hand, when we speak of two lines taking similar shapes we usually consider it in some comparing scale. For example, two squares are certainly similar, but a square is only similar to a rectangle to some degree. The degree is dependent on the detail of comparison or the scale of comparison. This consideration leads us to introduce a comparing scale (\(S\)) with which we will decide whether two lines are similar. And then we will introduce a new strategy to reach a new shape similarity measure based on the differences of direction changes of linear entities.

Firstly the vertex points are mapped onto the horizontal axis (that is \(X\)-axis) according to their distances along the line compared to the total length of the line. Then every direction change along the line (it certainly happens on the vertex point) could be illustrated by a vertical line drawn from the point \((x_i,0)\) to the end point \((x_i,y_i)\), in which \(y_i\) is the direction change value on the \(i\)th vertex point, \(x_i\) is
the distance index value of the \(i\)th vertex point. If the direction changes clockwise the direction change line will be upwards, otherwise downwards. The same process is applied to the other line (Fig. 2).

\[
(a) \quad \text{Direction change along the } L_1 \\
(b) \quad \text{Direction change along the } L_2
\]

Fig. 2 Direction change along two lines

It is easy to see that if the two lines are strictly similar, the differences of the direction change along the two lines will be zero. But usually the two concerned lines will not be consistent with each other in the context of direction changes. Then we define the comparing scale \(S_c\) as a length under which we will perceive the shape similarity of lines, that is the degree of detail. \(S_c\) has a domain from 0 to 1. 0 because we have mapped all the vertex points according their distances along the line onto \([0,1]\). The direction change lines within the length will be merged into a new direction change difference line (illustrated with heavy solid lines or heavy points in Fig. 2). The coordinates of the end point of the new line are calculated as:

\[
X_k = \frac{\sum_{i=1}^{\text{num}} x_i}{\text{num}}, \quad j = 1 \text{ to } 2
\]

\[
Y_k = \sum_{i=1}^{\text{num}} (-1)^{-1} y_i, \quad j = 1 \text{ or } 2
\]

if the direction change line is from line \(L_1\) then \(j = 1\), otherwise \(j = 2\). \((x_i, y_i)\) is the coordinate of the original end point of the \(i\)th direction change line along the \(X\)-axis. In order to process merging, the direction change lines of the two concerned lines are ordered by their positions along the \(X\)-axis. It is easy to do this because when we traverse the first line to map its vertex point onto \(X\)-axis, we would also get its direction change lines in \(X\)-increasing-order, and when mapping the other line, the only thing we need to do is either to insert the new direction change line into the existing queue or to append to the end of the existing queue.

Then we can define the shape similarity measure (SSM) in the scale of comparing \(S_c\) as the following:

\[
SSM(S_c) = 1 - \frac{\sum_{k=1}^{N(S_c)} \text{ABS}(Y_k)}{\sum_{i=1}^{M(S_c)} \text{ABS}(y_i)}
\]

\(N(S_c)\) denotes the total number of direction change difference lines after merging. It depends on the comparing scale \(S_c\) used. \(M\) denotes the total number of direction change lines of the two lines before merging. It could be easily seen that if the \(S_c\) is zero then \(SSM(S_c) = 0\), unless two lines are strictly similar. In the latter case the \(SSM(S_c)\) will always be 1. 0 with whatever \(S_c\) values. With the increasing of the comparing scale, the shape similarity measure of the two concerned lines will also increase. If the \(S_c\) equals 1, then \(SSM(S_c)\) will reach its biggest value, which indicates the most possible shape similarity measure of the two lines. If two lines are totally impossible to be considered as similar, for example one of them turns clockwise when the other turns anti-clockwise, the \(SSM(S_c)\) will still be very small even \(S_c\) equals 1 (for example, two lines showed in Fig. 4).

In Fig. 2, the comparing scale value \(S_c\) equals 0. 04 and the shape similarity measure \(SSM(S_c)\) value is 0. 97 which means the two lines are similar under comparing scale equal or greater than 0. 04.
But if we let $S_c = 0.03$, the SSM($S_c$) value decreases to 0.126. If $S_c = 0.02$, then SSM($S_c$) = 0.
These show that the two concerned lines are not similar if the comparing scale is less than 0.03.
Some other examples are illustrated in Fig. 3 and Fig. 4.

In Fig. 3, heavy solid lines or heavy points represent the direction change difference lines. The comparing scale value $S_c$ equals 0.1 and the shape similarity measure value (SSM($S_c$)) is 0.965 which means the two lines are similar under comparing scale equal or greater than 0.04. But if we let $S_c = 0.01$, the SSM($S_c$) decreases to 0.724. If $S_c = 0.001$, then SSM($S_c$) = 0.249. These show that we can safely say that the two concerned lines are not similar if the comparing scale is less than 0.01.

In Fig. 4, $L_2$ turns completely different from $L_1$ and the SMM($S_c$) are equal to 0 under any $S_c$, thus they are not similar in any degree. Of course if we invert one of the lines they could also be considered as similar under some comparing scale.

From above observations, we can conclude that the SSM($S_c$) could provide information on the shape similarity of two lines under the comparing scale $S_c$. And the linear shape similarity measure based on the differences of direction change has some obvious advantages. It is independent under translation, rotation, size, and scale change. It matches human intuition. It is not sensitive to tail of a line because we map all the vertex points into $[0,1]$ according to their distances comparing to the total length of each line.

The shape similarity measure defined above is also easy to be applied for areal entities comparison. The only modification may be that before mapping vertex points into X-axis we should add the second vertex point to the end of the vertex points string so that they would reflect the whole direction...
change along the boundary line.

When applying the linear shape similarity measure based on the differences of direction change to linear or areal entities comparison, the difficulty maybe lies in how to determinate the beginning vertex point of each line. However in the context of feature matching in GIS it can be solved by the determination of the closest end point pair of the concerned two lines. If the starting point of one line is closer to the ending point of the other line, then reverse one of the lines.

4 Conclusion

The shape similarity measures to two linear entities are important to many kinds of applications. There are some existing approaches to calculate the shape similarity of lines. But none of them could satisfy all the required criterions. This paper gives a detail investigation into the shape similarity measures between two linear entities in the context of feature matching in spatial databases. We conclude that the new shape similarity measure based on the differences of direction changes is effective to compare two linear entities. The measure is easy to calculate, independent under translation, rotation and scale change, well matched with human intuition. Moreover it can be easily applied for measuring shape similarity of two areal entities.

References

1 Cobb M., Chung M., Foley H. (1998) A rule-based approach for the conflation of attributed vector data. GeoInformatica, 2(1):7-35
2 Wentz E. A. (1997) Shape analysis in GIS. ACSM/ASPRS, 1997, Seattle, Washington. 5 (Auto-Carto 13): 204-213
3 MacEachren A. M. (1985) Compactness of geographic shape: comparison and evaluation of measures. Geografiska Annaler, 53-67
4 Ehler G. B., Cowen D. J., et al. (1996) Development of a shape fitting tool for site evaluation. Spatial Data Handling, 1:4A1-4A, 12
5 Cai Z, Xu G (1996) Artificial intelligent and its application. Beijing: Tsinghua University Press (in Chinese)
6 Cinque L., Levialdi S., Malizia A. (1998) Shape description using cubic polynomial Bezier curves. Pattern Recognition Letters, 19:821-828
7 Ebdon D. (1988) Statistics in Geography. Worcester, Billing and Sons Ltd.
8 Walter V., Fritsh D. (1999) Matching spatial data sets: a statistical approach. IJGIS, 13(5):445-473