

## A FAST MATCHING APPROACH OF POLYGON FEATURES

Qingxiang Meng<sup>a,\*</sup>, Zhongliang Fu<sup>a</sup>, Yan Huang<sup>b</sup>, Meicen Shen<sup>a</sup>

<sup>a</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, China –  
micromqxgis@126.com, fuzhl@263.net, mcshen@whu.edu.cn

<sup>b</sup> Wuhan Geotechnical Institute, Wuhan Wansongyuan Road No.209, Wuhan, 430022, China -  
huangyangis@hotmail.com

Commission II, WG II /3

**KEY WORDS:** Geometric, Feature, Polygon Simplification, Vector Data, Polygon Matching, Proximate-tangent Space

### ABSTRACT:

This paper proposes a new method for the matching of polygon features. Firstly, the main points depicting the shape of the polygon feature are extracted with the method of simplifying the delineation and are represented in the Proximate-tangent Space. Secondly, constructing the analogous estimate function based on the polygon's property that going along the edges of the polygon covers the minimum total area. With the help of the constructed analogous estimate function, the corresponding feature points on the matching features and objective features are found and paired. After pairing all the feature points, the interpolation on the edges and angles realize the quick matching of the polygon features. At last, the polygon features on the two maps with the scale of 1:10000 and 1:2000 respectively are put into the experiment. It turns out that this matching method not only operates with a high speed but also has a strong robustness.

### 1. INTRODUCTION

Entity matching means to identify the identical entities in different sources through some similarity measures, which is a process of establishing correspondence between spatial entities. Recognizing and matching identical entities is a key technique in map conflation, spatial data integration and updating. Besides, it is also the premise of some applications (Mantel, 2004). Discrepancy in representations of a spatial entity due to factors, such as sources, temporality, scales, accuracy, topology, semantics, has led to much complexity in matching identical spatial entities.

Existing matching algorithms can be divided into geometric matching, topological matching, and semantic matching (Chan-Hee, 2011). Spatial entity matching makes use of different kinds of similarity measurement. For computing similarity between lines, distances between nodes and poly-lines and the statistics of nodes were used, while distances and directional changes were also used as the criteria for judging the relationships between lines; for polygon entity matching, candidate matching entities are first collected according to the overlapping area between two polygon entities, then matching between two entities or entity collections is determined by fuzzy topological relationship; semantic matching may also be performed, which depends on data models and types of attributes and thus may not work in the absence of the associated attribute fields (Chen, 2011). This paper focuses on the polygon entity matching.

Lim (Lim, 2011) promotes a method based on matching by central point of maximum enclosed circles of polygon, which is of high efficiency and limited precision, so it used getting the matching candidate result set. Raveaux (Raveaux, 2011) compares the shape of polygons by the Fourier transform of image processing. Owing to uncertainty and requirement of

sampling unique-interval points in matching process, this method cannot reduce the cost of devise.

Existing matching algorithms are derived from transplanting feature matching methods and the practical applicability on vector features matching is not good enough. For multi-scale spatial vector data matching, most of methods concentrate on using spatial topological relationship and semantic information in the process of filtering matching features. In additional, the semantic information and geometric characteristics of the spatial vector polygon features are less (Huang, 2009).

The algorithm introduced in this paper focus on the geometric characteristics of vector polygon features. Firstly, the characteristic points which are critical in depicting the shape of the polygon are extracted by the method of simplifying the outlines of the polygon. Then, an analogous estimate function is constructed in similar tangent space to find each characteristic point's highest matching-probability point. At last, the corresponding relationship of the characteristic points from different polygons is built up so that the polygon features' matching relationship could be formed. The extraction of the characteristic points provides accurate depiction of the certain polygon and reduces the compute quantity of the matching. So the extraction plays a key role in the algorithm which contributes the efficiency and veracity. And then we authenticate experimentally the efficiency and veracity on multi-scale spatial vector polygon features.

### 2. METHOD

#### 2.1 Simplification of multi-scale vector polygon features

When comparing two polygons, some nodes play a key role. Such as the marked vertexes shown in Figure 1, these nodes which are called critical nodes have a great influence when

---

\* Corresponding author.

human eyes recognize polygons. For the polygon made of curves, these nodes, which play a decisive role in determining the shape of elements, are outline feature points such as sharp points, deformation points, and the curvature extreme points and so on (Ni, 2004, Zhao, 2010). And among the outline feature points of the polygon, some more decisive points to recognize shape can be extracted. The specific points (shown in Figure 1) are the articulation point of the polygon (Shao, 2010). During the process of gradient, in the process of gradual change these points can decide the geometry shape of elements, known as the principal points. The rest of the point is also characteristic point, less important than the principal points (Xie, 2010).

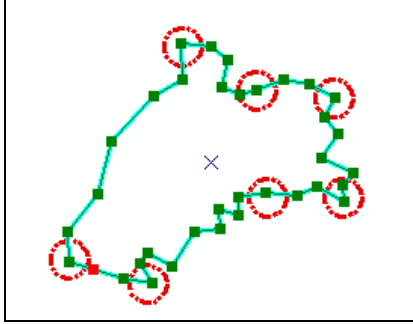


Figure 1. The characteristic points and principal points of polygon

Polygon A includes  $n$  points, namely  $A=\{A_1, A_2, \dots, A_n\}$ , the target polygon include  $m$  points, namely  $B=\{B_1, B_2, \dots, B_m\}$ . Here the vertex coordinates of matching face elements and target face elements are given, and concentration of the points is not high. For example, only less than 100 points can accurately describe the outline of polygon. These points connected end to end in turn to form a polygon that an approximation to source elements. This paper adopted the polygon simplified algorithm to acquire the polygons which contain more principal points and reasonably reflect the outline of the simplified polygons.

Calculate the correlation parameters of the  $n$  vertexes in the polygon A, take  $A_i$  for instance, the correlation parameter is defined as

$$K_{(s_{i-1}, s_i)} = \frac{\beta_{(s_{i-1}, s_i)} l_{(s_{i-1})} l_{(s_i)}}{l_{(s_{i-1})} + l_{(s_i)}} \quad (1)$$

where  $s_{i-1}$  and  $s_i$  are two adjacent edges of the Vertex A  
 $l_{(s_{i-1})}$  and  $l_{(s_i)}$  are the normalized edge length  
 $0 < l_{(s_{i-1})}, l_{(s_i)} < 1$   
 $0 \leq \beta_{(s_{i-1}, s_i)} < \pi$   
 $\beta_{(s_{i-1}, s_i)}$  is the absolute value of rotation angle of the two adjacent edges

Find out the minimum correlation parameter of the  $n$  vertexes and the corresponding vertex will be cut off, then a polygon with  $n-1$  vertexes is constituted. By such analogy, after  $j$  times simplified, polygon contains  $n-1$  vertexes. This polygon simplified algorithm firstly is used to simplify polygon to eliminate the noise effect, then remove the details information

of the edge to acquire the main body information, while the principal characteristic points is the key point to decide the boundary of the main body. The simplified polygon can explicitly express the outline of the raw element.

The simplified polygons of matching polygon and the target polygon is  $A^*$  and  $B^*$ . The time to simplify polygon is not the more the better, too much times will lead to distortion which is caused by losing principal characteristic points, while too little times will make the simplification meaningless. The simplified polygons should include much more principal characteristic points to reflect the outline of the raw element. Moreover, there is no need to require the equal of the number of  $A^*$  and  $B^*$  vertices which vary not too much. Suppose  $A^*$  contains  $n$  vertices  $\{a_1, a_2, \dots, a_n\}$ ,  $n < N$ ;  $B^*$  contains  $m$  vertices  $\{b_1, b_2, \dots, b_m\}$ ,  $m < M$ . In general, the value of  $n$  and  $m$  is not too large (in this experiment,  $10 < n, m < 40$ ). For these simple polygons, it can adopt related geometry property based on vertex to establish similarity measures function, but the effect of the existing approach is not very good, this paper presents a new form of analogous estimate function.

## 2.2 Construction of analogous estimate function under similar tangent space

Similar tangent space: If a given element C contains  $N$  vertexes  $\{C_1, C_2, \dots, C_N\}$ , this paper selects one certain vertex  $C_i (i=1, 2, \dots, N)$  as reference point and traverses the whole vertex sequence in clockwise order, with X axis representing the normalized distance between  $C_i$  and all the other vertexes along the polygon boundary and Y axis representing the angle between every edge and horizontal axis. Take Figure 2 as an example,  $C_i = C_1$ ,  $\theta_1$  represents the angle between edge  $C_1C_2$  and horizontal axis  $\theta_1 \in [-\pi, \pi]$ , and  $\varphi_1$  represents the rotation angle from edge  $C_1C_2$  to edge  $C_2C_3$ , with counter clockwise direction positive, then the angle between edge  $C_2C_3$  and horizontal axis can be calculated as  $\theta_2 = \theta_1 + \varphi_1$ . From the recurrence can the angle expressed as  $\theta_i = \theta_{i-1} + \varphi_{i-1}$ .

$i(i=3, \dots, N)$ , as well as  $L_i = \frac{|C_i C_{i+1}|}{L} + L_{i-1}$  with  $L$  representing the length sum of all the edges, and  $L_i = \frac{|C_i C_{i+1}|}{L} + L_{i-1}$  ( $i=2, \dots, N, C_{N+1}=C_1$ ). If  $\theta_i$  is used as ordinate and  $L_i$  as abscissa, the similar tangent space of polygon C can be expressed.

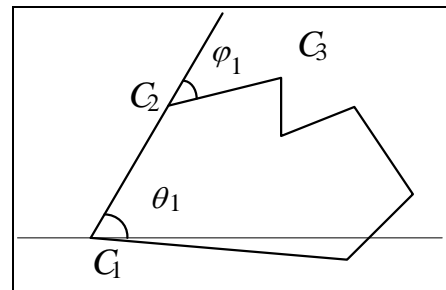


Figure 2. The tangent space of simple polygon

A graphic gradient process with good effect should be smooth and natural enough, which requires the corresponding edges, corresponding angles as well as the corresponding area among edges keep monotonic as much as possible in the gradual process. For the reason that irrational vertex is bound to cause the self-cross phenomenon among the adjacent edge and other edges in the gradual process, leading to the expansion or contraction of local area and the gradient distortion, one has to avoid the unnecessary large rotation from edge to edge(except for their required deformation). Based on the mentioned above, this paper uses the rotation area from the two adjacent edges of initial polygon vertexes to the two adjacent edges of objective polygon vertexes in similar tangent space as similarity measurement, and constructs the analogous estimate function as follows.

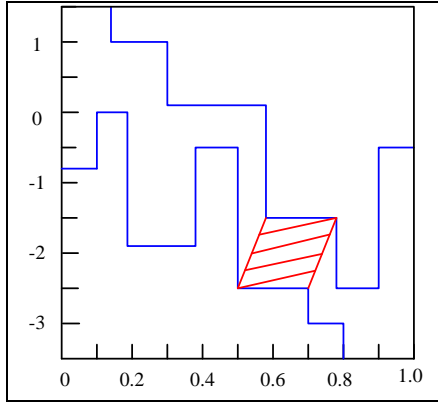


Figure 3. The geometric meaning of similarity function

First, calculate the normalized length between the vertex  $a_i$  on  $A^*$  and the two adjacent edges, which is  $l_{i-1}^1$  and  $l_i^1$  ( $i=1,2,\dots,n$ ;  $l_0^1=l_n^1$ ), then similarly calculate the normalized length about vertex  $b_j$  on  $B^*$   $l_{j-1}^2$  and  $l_j^2$  ( $j=1,2,\dots,n$ ;  $l_0^2=l_n^2$ ). Based on the similar tangent space representation, calculate the rotation angle of corresponding edges  $\theta_{i-1,j-1} = |\theta_{i-1}^1 - \theta_{j-1}^2|$  and  $\theta_{i,j} = |\theta_i^1 - \theta_j^2|$  ( $\theta_0^1 = \theta_n^1$ ,  $\theta_0^2 = \theta_m^2$ ). Then define the similarity function of  $a_i$  and  $b_j$  as:

$$Sim(a_i, b_j) = \left[ \max \left[ \frac{l_{i-1}^1}{l_{j-1}^2}, \frac{l_j^2}{l_i^1} \right] \right]^{ex} \times (l_{i-1}^1 + l_j^2) \times \left[ \frac{\theta_{i-1,j-1}}{2\pi} \right] + \left[ \max \left[ \frac{l_i^1}{l_j^2}, \frac{l_j^2}{l_i^1} \right] \right]^{ex} \times (l_i^1 + l_j^2) \times \left[ \frac{\theta_{i,j}}{2\pi} \right] \quad (2)$$

where  $ex$  is an parameter which needs artificial setting, in this paper  $ex=0,1,2$

$$\left[ \max \left[ \frac{l_i^1}{l_j^2}, \frac{l_j^2}{l_i^1} \right] \right]^{ex} \text{ is used to reduce the}$$

possibility of short edge corresponds to long edge(which needs to be avoided in the gradual process as well)

$$(l_i^1 + l_j^2) \times \frac{\theta_{i,j}}{2\pi} \text{ is the trapezoid area of a group of}$$

corresponding edges in similar tangent space. As the Figure 3

$$\text{shows, the shades area is } l_6^1 + l_7^2 \times \left[ \frac{\theta_{6,7}}{2\pi} \right].$$

Formula(2) can also be viewed as a kind of similarity function of the two similar tangent space expression figures, the higher the similarity of two ladder-like graphics, the smaller the corresponding similarity function value. The Figure 4 shows the established corresponding relationship graphic of Figure 2, and the comparison of their similar tangent space expression figures.

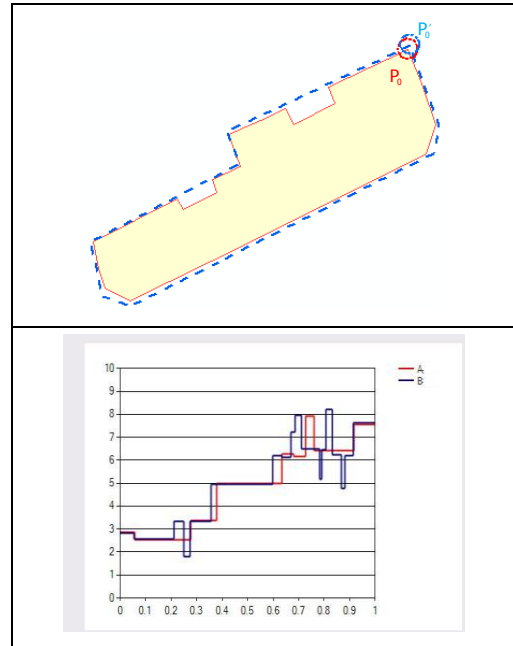


Figure 4. The comparison of shape similarity function of two polygons

### 3. EXPERIMENTS

#### 3.1 Matching accuracy

The data of the experiments is shown in Figure 5(a) to 5(d) and we use the polygon elements of solid curves and dotted curves to match. Table 1 is the matching results of similarity calculation through Formula (2) in 2.2.

The results show that the geometric matching method of shape similarity based on tangent space in this paper has high efficiency and fast speed. It adapts to not only most of simply or complex polygon entities, but also multi-scale one (Figure 5(c) and 5(d)). But in the multi-scale matching, it is hard to get the similarity threshold. The threshold in Exp 1 is 0.8. When the

threshold in-creases, it may not be matched in Figure 5(c) and 5(d).

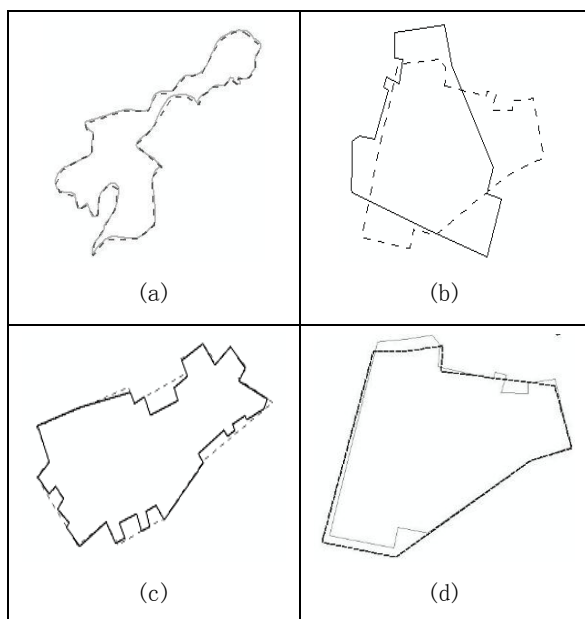


Figure 5. The matching of polygons in proximate-tangent space

Matching entities	Similarity of features	Process cost (ms)	Matched result	Correctness of result
Figure 5(a)	0.935	13.12	matched	corrected
Figure 5(b)	0.385	10.14	not	corrected
Figure 5(c)	0.795	19.15	matched	corrected
Figure 5(d)	0.825	9.63	matched	corrected

Table 1. Matching results of polygon features

### 3.2 Multi-scale matching

This article selected two scales of hydrographical data as the experimental data, the scale of 1:2000 and the scale of 1:10000, which belong to the same area but in different times (Figure 6).

The 1:2000 scale hydrographical data contains 3609 entities, and the 1:10000 contains 143 entities. While performing the experiments, we could take the 1:10000 scale hydrographical data as the source data and the 1:2000 as the target data. After doing the entity matching experiments, Figure 7 is the cartogram of matching types of hydrographical data.

The experiments show that two scales of hydrographical data are quite different. The amount of entities which cannot be matched in the source data accounts for 30.07% of the total, and in the target data it accounts for 95.93%. If the threshold of homo-entities is 0.36, the recall ratio and precision ratio are 83.9% and 97.8%. If the threshold of homo-entities is 0.42, the recall ratio and precision ratio are 82.5% and 100%.

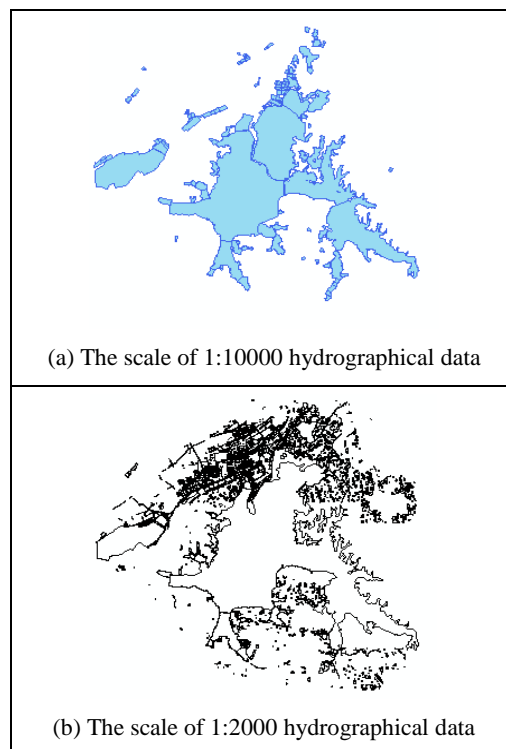


Figure 6. The matching of multi-scales hydrographical polygons

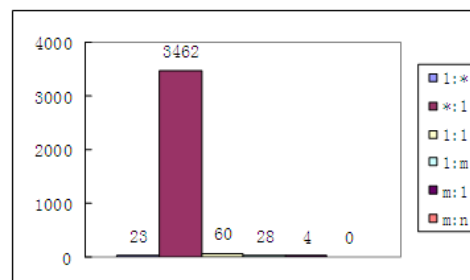


Figure 7. The cartogram of matching types of hydrographical data

## 4. CONCLUSIONS

Compared with other algorithms, the method mentioned in this paper reduces the number of the corresponding points by simplifying the features so that the complexity of the matching is well reduced. At the same time, simplifying the polygon also reduces the effect of the noise caused by the original feature and improves the robustness. Because of the algorithm builds reasonable corresponding relationship for the characteristic points, the gradual change process of the polygon feature avoids the distortion. According to the results of experiments, the proposed algorithm is able to build excellent corresponding relationship of the characteristic points under the condition of rotation, expansion and distortion. Additionally, this algorithm has a broader adopted extent, it well functions on the points pairing of similar polygons and on vertex pairing which extracted from pixel outline.

This algorithm regards a condition as a pre-requirement that it is not a huge change between the two polygons. It requires further study under the condition of huge change happen. And getting

rid of the manual intervention should be taken into consideration.

### Acknowledgements

The research in this paper is funded in part by the Central University Independent Research of Ministry of Education, China. Furthermore, I would like to thank PhD Liu siyuan and PhD Zhou fan for their valuable advice and significant support on providing the amount of test data.

### References

- Chan-Hee H, Yun-Ho K, 2011. Fast Shape Matching Using Statistical Features of Shape Contexts. *IEICE Transactions on Information and Systems*, pp. 2056-2058.
- Chen W, Long Y, Shen J, et al, 2011. Structure recognition and progressive simplification of the concaves of building polygon based on constrained D-TIN. *Geomatics and Information Science of Wuhan University*, 36(5), pp. 584-587, 592.
- Huang J, 2009. A new model for general polygon matching problems. *Precision Engineering*, 33(4), pp. 534-541.
- Lim Moon-Jai, Han Chan-Hee, Lee Si-Woong, Ko, Yun-Ho, 2011. Fast shape matching using statistical features of shape contexts. *IEICE Transactions on Information and Systems*, E94-D(10), pp. 2056-2058.
- Mantel D, Lipeck U W, 2004. Matching Cartographic Objects in Spatial Databases. *Archives of Photogrammetry, Remote Sensing and Spatial Sciences*, 35, pp. 172-176.
- Ni M, Reichenbach S E, 2004. Pattern matching by scan-converting polygons. *Proceedings of SPIE-The International Society for Optical Engineering*, 5438, pp. 101-110.
- Raveaux R, Burie J, Ogier J, 2011. A local evaluation of vectorized documents by means of polygon assignments and matching. *International Journal on Document Analysis and Recognition*, pp. 1-23.
- Shao S, Tong C, 2010. A matching method for multi-characteristic vector elements of complex polygon. *2010 International Conference on Multimedia Technology (ICMT)*, pp. 1-4.
- Xie J, Xu G, Zhang F, et al., 2010. A local tangent space based approach for single-trial representation of event-related potentials. *2010 5th Cairo International Biomedical Engineering Conference, CIBEC 2010*, pp. 184-187.
- Zhao Z S, Tian Q J, Wang J Z, et al., 2010. Hierarchical hybrid multi-scale feature match. *2010 3rd International Congress on Image and Signal Processing (CISP)*, pp. 1653-1656.