

# Perception Recognition Study on Long and Narrow Arcs of Polygons in Maps using Intelligent Skeleton Decomposition

Han Zhang<sup>1,3\*</sup>, Xiaoyong Chen<sup>2</sup>, Zhihui Yi<sup>3</sup>, Chao Xu<sup>3</sup>

<sup>1</sup> School of Information Engineering, East China University of Technology, China

<sup>2</sup> State Key Laboratory of Nuclear Resources and Environment, East China University of Technology, China

<sup>3</sup> Geographic Information Engineering Bridge, Jiangxi Provincial Bureau of Geology, China  
119970184499@163.com, 200960002@ecut.edu.cn, 309224038@qq.com, 18807911230@139.com

## Abstract

The paper presents an automated method for identifying elongated and narrow sections of polygons, which may arise due to uncertainties inherent in map data production and updates. It leverages a skeleton-based shape decomposition approach, encompassing the creation of skeletons and end-point sub-polygons based on the polygon boundary-constrained triangulation. Subsequently, it identifies a pair of break points using concave vertices of polygon boundaries based on the nearest neighbor principle. The optimal segmentation of the end-point sub-polygons is then determined by evaluating weighted base-height ratios, which are added to a set of candidates long and narrow arcs. Finally, the model selects the long and narrow arcs based on their base-height ratios and shape indexes. The paper reports the results of an experiment using land parcel data from Jinjiang in Fujian Province, demonstrating that the proposed method produces results that align with human perception and outperforms the internal and external buffering method. In summary, the innovative method exhibits promise in accurately identifying elongated and narrow sections of polygons. Its applicability spans diverse domains such as map data analysis and updates, where precision in delineating such features is crucial.

**Keywords:** Long and narrow arcs, Shape decomposition, Uncertainty of geospatial data

## 1 Introduction

Vector maps are widely used in various industries related to administration, land use, environmental monitoring, and spatial decision-making. The quality of vector map data directly affects the effectiveness of their applications and has been a long-term hotspot in the fields of surveying, mapping, and geographic information science [1]. Polygons are a basic type of graphic element in two-dimensional vector maps that correspond to real-world geographic objects such as administrative regions and land use parcels. However, due to various uncertainties in the process of map data production, processing, and application [2-4], the polygon data that is waiting to be (or has already been) stored in the database

often contains topological inconsistencies and geometric conflicts. While practical and robust algorithms [5-10] have been developed to solve the former issue (topological inconsistencies), which have been integrated into commercial geographic information system platforms, the latter issue still lacks mature and efficient solutions. It remains an important technical problem that needs further research and resolution in the fields of surveying and geographic information.

For a single polygon, there may be three basic types of geometric conflicts, namely, size conflict caused by too small polygon size, distance conflict caused by too small distance between consecutive vertices, and distance conflict caused by too small polygon size. The distance between discontinuous vertices is too small, which results in a narrow conflict [11]. Among them, all sorts of conflicts can be identified and controlled by tolerances, while the identification of narrow conflicts that this article focuses on is more complicated. In order to detect narrow conflicts, Bader and Weibel [12] proposed two methods, one is based on the rolling ball principle, similar to the  $\epsilon$ -band proposed by Perkal [13], and the other is based on consistent Delaunay triangulation. The method based on Delaunay triangulation is more effective and can provide a basis for subsequent polygon boundary modification. However, consistent triangulation needs to insert sufficiently dense vertices at the polygon boundary, so that the result of the mesh strictly meets the Delaunay triangulation condition. In order to avoid inserting extra vertices and reduce the amount of calculation, it is allowed to violate the Delaunay triangulation condition in a local area and construct a boundary constrained triangulation network [14]. Nevertheless, the process of detecting narrow conflicts based on triangulation is still more complicated. After the triangulation is generated, it is necessary to compare the height of each triangle with the distance threshold, identify the neighboring relationship of triangles with smaller heights, and combine adjacent triangles into Narrow area. After that, Peter [15] proposed a simpler method of identifying the narrow part of a polygon, that is, generating a buffer inside the polygon with a certain distance threshold, and judging whether there is a narrow part according to the number of generated core regions, and determining whether the polygon has a narrow part. Partially, then solve its exact position. According to the positional relationship between the narrow part and the core area, it can be divided into two situations.

One is the corridor-like part between the core areas, and the other is the columnar part of the polygon edge. They are called narrow corridor-like sections conflict and jut-like section conflict, shown in Figure 1. Inward buffering can effectively identify narrow corridor conflicts, but the effect of narrow column conflict recognition is not ideal. If the polygon with narrow column conflicts has only one core area after inward buffering, this method will not regard it as a wrong polygon. Gao et al. [11] extended Peter's ideas and designed an internal and external buffer method for detecting and locating narrow corridor and narrow column conflicts. This method is similar to the combination of erosion and expansion operations for raster data, targeting narrow column conflicts. The detection effect is better than the inward buffering method proposed by Peter, but it is highly dependent on the buffer width and area threshold setting, and the buffer generation method will cause the detection result to be inaccurate to a certain extent.

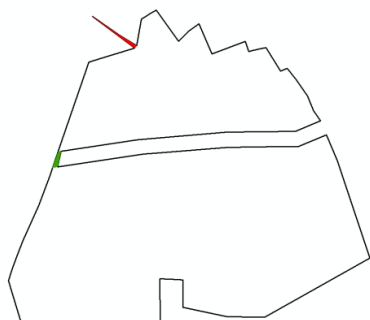


Figure 1. Corridor-like sections (green) and Jut-like sections (red)

Aiming at the narrow and long arcs in the vector polygon map, that is, the narrow column conflict problem mentioned above, this paper proposes a recognition method based on skeleton line decomposition to explore new ways to solve such problems.

## 2 Research Area and Data

The control area for this study is the Jinjiang City in Fujian Province. We used plot data from 2018, following land-use adjustments and improvements, as our experimental samples. The data comprises 18,357 plot polygons recorded in vector format, referred to as the Map, covering an area of 744.6179 square kilometers and 42MB in size. The Map's scale is 1:50,000, it depicts in Figure 2). After inspection, we found no topological errors in the plot data's polygons. However, we identified narrow and long arcs in some map spots that do not conform to cartographic standards, it shows in Figure 3.

## 3 Research Methods

The long and narrow arc has the following characteristics: Generally located on the branch structure of the polygon; At least one of the two end points is a concave vertex of

the polygon; The shape is long and narrow. In this section, based on the characteristics of the long and narrow arc, the structure of the polygonal skeleton line, the boundary point information and the geometric shape of the boundary line are comprehensively used to realize the automatic identification of the long and narrow arc. The branch points and their associated boundary points divide the polygon to obtain end-point sub-polygons; Separation line solution: solve the polygon boundary concave point and the adjacent points on the opposite side of the skeleton line, and connect the two as the dividing line; The optimal segmentation result is determined: Introduce the weighted base height ratio as a measure of the saliency index of the arc segment to filter the optimal segmentation in each sub-polygon; Narrow arc segment recognition: extract the base-height ratio and shape index of the arc segment, from the optimal segmentation of each sub-polygon to determine the final recognition result. They are shown in Figure 4.

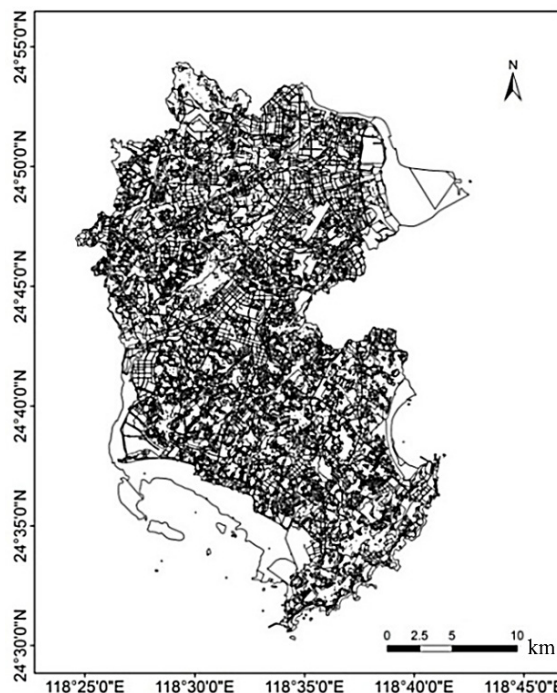


Figure 2. Control area data of Jinjiang

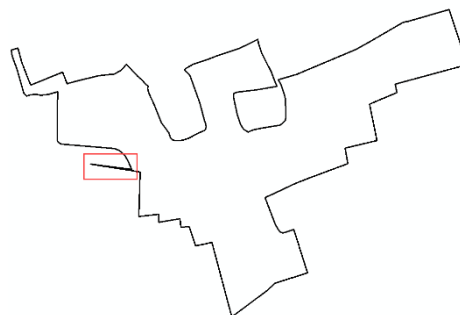


Figure 3. Illustration of a long and narrow arc



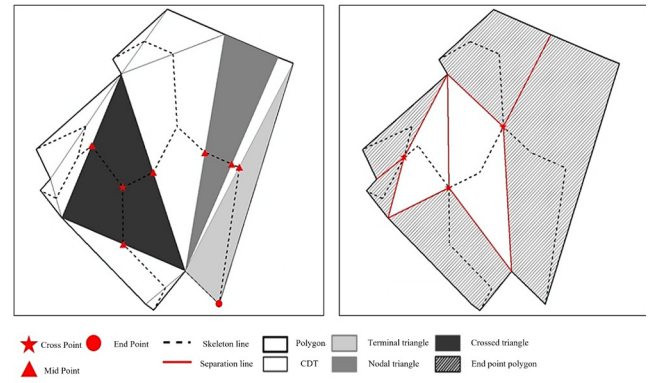
**Figure 4.** Recognition of long and narrow arcs by skeleton based decomposition

### 3.1 End-point Sub-polygon Acquisition

The narrow and long arcs typically exist in the branch structures of polygons. The skeleton line can effectively reflect the surface elements' extension direction and shape characteristics, enabling it to describe the hierarchical relationship between the shape's main body and the branch structure [16]. To address the recognition problem, we propose a novel approach that involves using the skeleton line to extract the branch structure of the polygon and identify the narrow and long arcs. Whenever a branch node appears on the skeleton line, the segment of the line from the node to its endpoint represents the shape's convex area. As a result, the skeleton line's branch nodes contain information on positions that may result from shape decomposition, and this information can guide the process [17].

This section is based on Delaunay triangulation to obtain a boundary-constrained triangulation, ensuring that all triangles are contained within the polygons. Depending on the number of common edges between triangles and polygons, the triangles are classified into end triangles, node triangles, and crossing triangles. Different strategies are used to extract the skeleton lines of different types of triangles. As shown in the left part of Figure 5, the endpoint triangle's skeleton line connects the endpoint of the skeleton line and the midpoint of the opposite side; the node triangle's skeleton line connects the midpoint of the two sides not on the polygon outline; and the crossing triangle's skeleton line connects the intersection point of the skeleton line with each side, where the skeleton line intersection point is the skeleton line node included in the intersecting triangle and also the center of the triangle.

The polygon is initially divided by the line between the intersection of the skeleton line and the vertices of the corresponding triangle, and the sub-polygon containing the end points of the skeleton line is defined as the end-point sub-polygon, shown as the shaded polygon in the right part of Figure 5. End-point sub-polygons can be considered the branch structure of the polygon. Long and narrow arc segments in the polygon are often included in the end-point sub-polygons.



**Figure 5.** Obtaining the end point sub-polygons

### 3.2 Solving Dividing Line

Determining the dividing line is the key to shape decomposition [18]. However, due to the highly complex and varied shapes of real-world objects, finding a single criterion for all shape decomposition problems is difficult. A common approach for shape decomposition is to first select candidate points using the "minimum curvature criterion" [19], and then use the "shortest dividing line criterion" [20] to determine the dividing line (hereinafter referred to as the classic method). The "minimum curvature criterion" involves selecting the point with the minimum curvature (depressed point) on the shape as the end point of the dividing line, while the "shortest dividing line criterion" uses the shortest possible dividing line to decompose the shape. Shape decomposition is a fundamental problem in computer vision and computer graphics, with raster images being the primary focus of study [21-22]. Considering the differences in the representation of raster and vector data, this section proposes a "nearest neighbor criterion" combined with the "shortest dividing line criterion" to determine the dividing line at the concave point of the polygon boundary.

Since the end of the narrow arc segment has at least one concave vertex, the boundary concave vertex can be used as the key information to guide the segmentation. Use the vector product method to judge the concavity and convexity of the vertices of the polygon, solve the adjacent boundary points on the other side of the polygon skeleton line based on the concave vertices of the polygon boundary, and use the concave point and its nearest neighbor as the split point pair, and the two connected as the split line. As shown in Figure 6, suppose point a is the current concave point to be processed on the boundary of the polygon, and find the neighboring points of this point on the other side of the skeleton line (AB and BC) respectively. The vertical line of the boundary arc is drawn from the concave point. If the vertical foot is on the arc, the vertical foot is regarded as the adjacent point of the concave point, such as the point a1 of the AB section; if the vertical foot is on the extension line of the arc, the vertical nearest end of the point is regarded as the neighboring point, such as point B (a2) in the BC section. The nearest neighbor point is the point closest to the concave point among all neighboring points, such as a1. Connecting it to the concave point is the dividing line at the concave point.



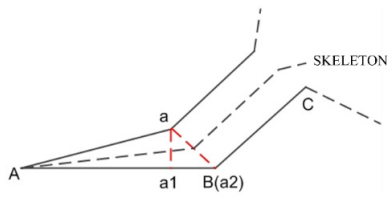


Figure 6. Determining the cutting line

### 3.3 Optimal Segmentation and Screening

In cases where multiple concave points generate segmentation lines in the endpoint sub-polygon, the candidate long and narrow arc segment must be selected based on a specific criterion to determine the optimal segmentation result. Narrow and long arc segments generally exhibit obvious outward convex characteristics, and the shape significance index is often used to describe the convex strength of the shape branch structure [22]. In this study, the weighted base height ratio is used as the shape saliency index. The arc with the highest degree of shape saliency within each sub-polygon is chosen as the optimal segmentation outcome and subsequently incorporated into the collection of candidate narrow and elongated arcs.

$$P_0 = \frac{L}{W'} \quad (1)$$

$$P_w = \frac{L^2}{W} \times P_0 = \frac{L^3}{W^2} \quad (2)$$

In the formula,  $P_0$  represents the ratio of the base height,  $L$  is the length of the skeleton line in the segmentation result, and  $W$  is the width of the base (division line). The weighted base height ratio takes the square of the length of the skeleton line to the width of the base as the weight, comprehensively considering the influence of both the length of the skeleton line and the width of the dividing line. A larger significance index indicates a stronger outward protrusion strength of the decomposition result.

### 3.4 Recognition of Long and Narrow Arcs

The existence of concave points is a necessary and insufficient condition for the existence of long and narrow arcs. The candidate arcs obtained by only using concave points and the significance index are likely to contain a large number of non-narrow and long arcs, which need to be further screened. Considering the characteristics of the narrow and long shape of the area enclosed by the long and narrow arc, this section introduces  $SI$  (The Shape Index) [23] as a quantitative basis for screening, which is defined as follows:

$$SI = \frac{P}{2\sqrt{\pi A}} \quad (3)$$

Where  $A$  is the area of the part enclosed by the arc and is the perimeter of the area enclosed by the arc.  $SI$  represents

the compactness of the shape, and its value is greater than or equal to 1. The closer the  $SI$  is to 1, the closer the shape is to the circle; the larger the value of  $SI$ , the greater the difference between the shape and the circle, and the less compact the shape. Baiqing Wu et al. [24] pointed out in the study of the method of removing error polygons in the land use spatial database, when the threshold of  $SI$  is set to 2.587602, it can distinguish between long and narrow spots and correct ones, that is, the shapes with the shape index greater than the threshold are considered to be long and narrow. Figure spots, this section sets the  $SI$  threshold for distinguishing long and narrow arcs from non-narrow and long arcs to 2.6.

Based on the geological data analyzed, it has been found that among the candidate arcs, some exceed the  $SI$  threshold and can be classified into two types: long and narrow arcs, and normal boundary arcs with small turning angles. Take Figure 7 as an example, the former is characterized by a larger base height ratio, while the latter is known as a straight arc due to its smaller base height ratio. The straight arc can be identified by a base width greater than the length of its skeleton line, resulting in a base height ratio less than 1. Conversely, the long and narrow arc can be distinguished by a base width less than the length of its skeleton line, resulting in a base height ratio greater than 1. In this paper, candidate arcs with a base-to-height ratio greater than 1 are considered to be long and narrow arcs. This method provides a reliable means of distinguishing between the two types of arcs, and it can be applied to further study the geological evolution of the region.

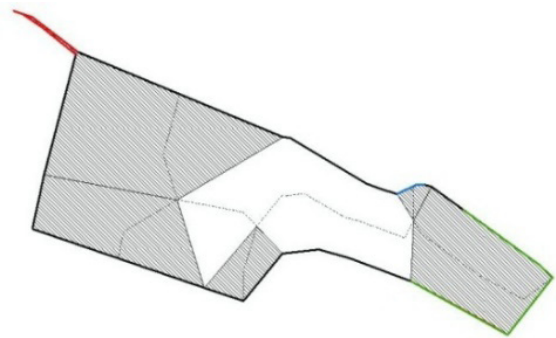


Figure 7. Long and narrow arcs in the candidate set (blue and green arcs)

## 4 Experiment and Result Analysis

This study utilizes land use control area data from Jinjiang City, Fujian Province, and employs ArcMap for database construction and data visualization. Python and its ArcPy library are utilized to recognize narrow and long arcs by employing the shape decomposition of the skeleton line algorithm. The experimental environment includes a Microsoft Win7 64-bit operating system, Intel Core I7-6700 CPU, running at 3.4GHz with 8GB of memory.

### 4.1 Analysis of the Effect of Dividing Line Solution

The experiment produced representative results, which were compared to the classic method and the method of

drawing a circle to find the intersection described in literature [21]. Figure 8 illustrates the end sub-polygon boundary containing long and narrow arcs represented by the thick solid line, and the thin solid line indicating the skeleton line of the end sub-polygon. The branch point of the skeleton line is marked by a star, while the concave vertices of the polygon boundary are denoted by circle and triangle marks. The triangles indicate the example points used to generate the dividing lines in this study. The dashed lines in the figure depict the dividing lines generated at the instance points, and the shaded areas represent the regions enclosed by the dividing lines and the long and narrow arcs.

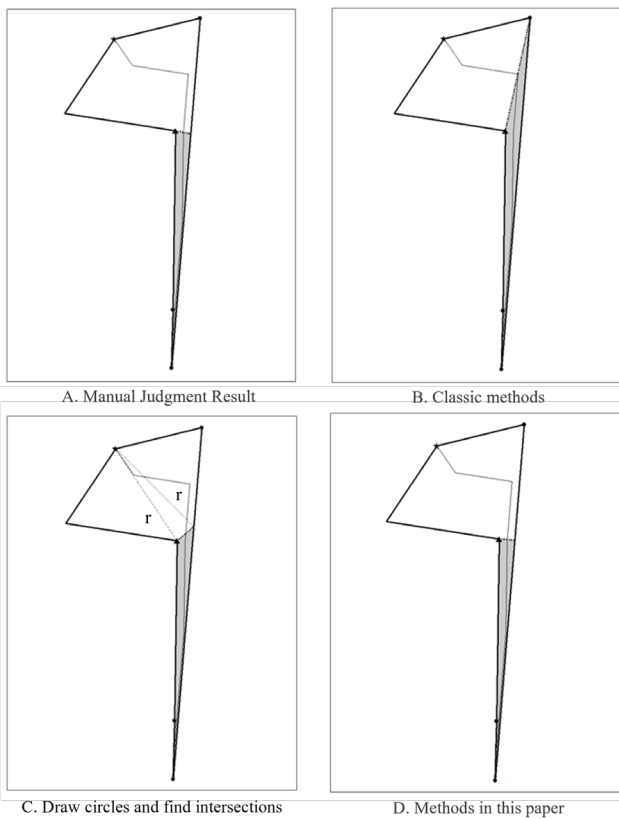


Figure 8. The cutting lines obtained from different methods

Comparing the results obtained by the three automatic recognition methods with the results of manual judgment, it can be seen that the method in this paper has obtained the long and narrow arc that is most in line with human cognition. For vector polygons, the “minimum curvature criterion” of the classic method is to use the concave vertices of the polygon as candidate endpoints. If the candidate point on the other side of the skeleton line is not in the vicinity of the concave point or there is no candidate point, the dividing line is unreasonable or even the dividing line cannot be generated. The accuracy of the dividing line obtained by drawing a circle to find the intersection point depends on the position between the branch point, the concave point, and the narrow arc of the skeleton line. As shown in Figure 9, for the same long and narrow arc and the boundary concave point C, the skeleton line branch points are in different positions, such as B1, B2, B3, and the corresponding division points generated by this method are also in different positions,

such as A1, A2, and A3. The method in this paper is to find the existing neighboring points (end points of the boundary segment) or inserting new neighboring points according to the boundary concave points and is not affected by the branch points of the skeleton line. The generated dividing line is obviously better than the above two methods.

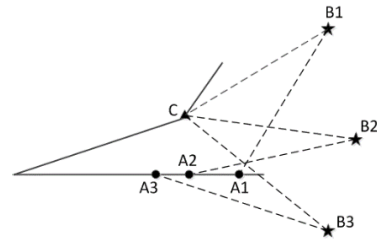


Figure 9. Uncertainty of the intersection between circle and polygon edge

### 4.2 Performance Evaluation of the Saliency Index

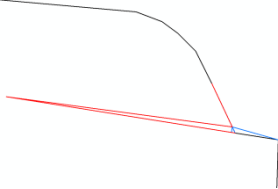
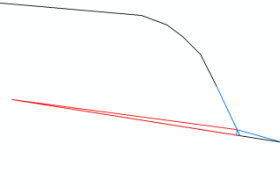
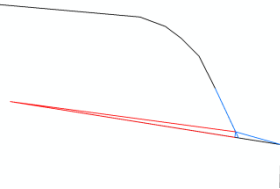
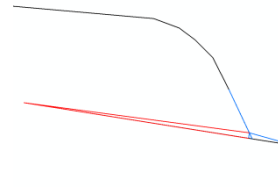
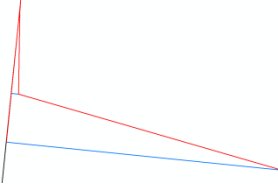
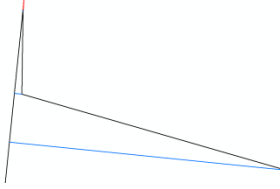
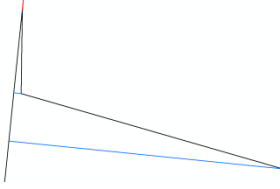
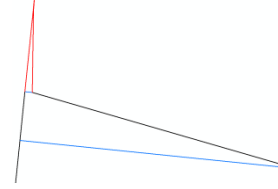
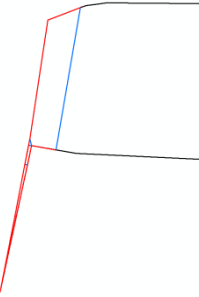
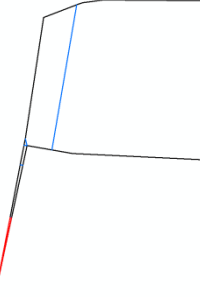
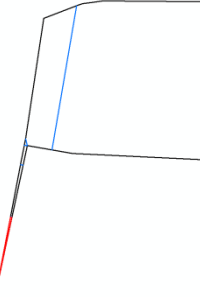
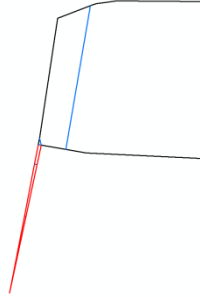
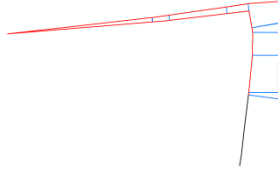
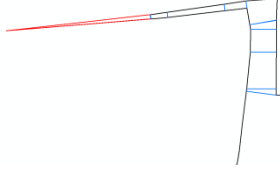
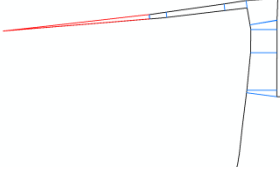
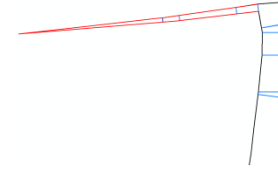
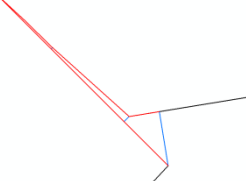
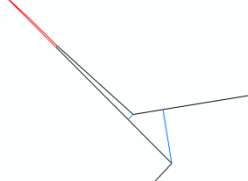
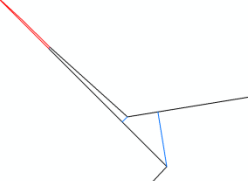
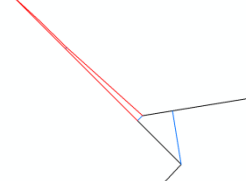
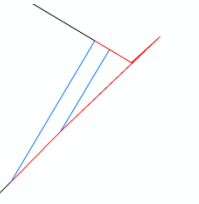
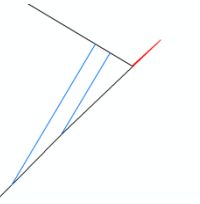
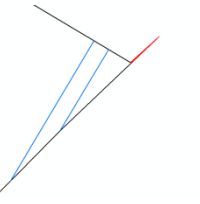
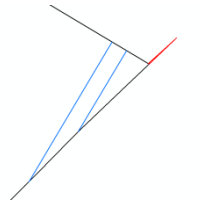
The basic idea of the performance evaluation scheme for the shape saliency index is to evaluate the consistency of the objective measurement results and the subjective experience-based comparison results to evaluate the performance of the shape saliency index for branch structures [25]. In order to analyze the advantages of the weighted high-contrast visual saliency index proposed in this paper, a comparison was made with existing objective saliency measurement indices, including: the contour length of the branch structure (the length of the contour segment between the two endpoints of the dividing line), the outstretching degree of the branch structure (the ratio of the length of the branch structure contour segment to the length of the base), and the base-to-height ratio of the branch structure (the ratio of the length of the branch structure skeleton line to the width of the base). The experiment statistically compared the cumulative difference, mean difference, and root mean square error of the optimal segmentation results selected by different indices and the reference data arc length obtained by manual experience-based judgment. All three indices measured the difference between the optimal segmentation results and the reference data. The smaller these statistical values are, the smaller the difference between the two sets of data, indicating that the results obtained through objective measurement indices are more consistent with the results of manual experience-based judgment.

It can be seen from Table 1 that the optimal segmentation results screened by the weighted base height ratio significance index are more in line with manual judgment than the results obtained by other objective significance indexes. Combining the three sets of statistical data in Table 2, it can be seen that the contour length will screen out candidate arcs with the largest contour length, and lack consideration of the overall narrowness of the segmentation results, which has the greatest deviation from the reference data; the degree of extension and base height The ratio has similar measurement performance, but lacks the consideration of the extensibility of the segmentation results, and tends to filter out the narrow part of the arc; The ratio of the square to the base width is used as the weight, which improves the performance

of the indicator for measuring the extensibility of the arc. The performance is better than the other three objective

measurement indicators, and the results obtained are more reliable.

**Table 1.** The optimal segmentation selected by different significance index

| Contour length  | Extended degree   | Base-High ratio  | Weighted base height ratio  |
|---|---|--|---|
|    |    |    |    |
|    |    |    |    |
|   |   |   |   |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

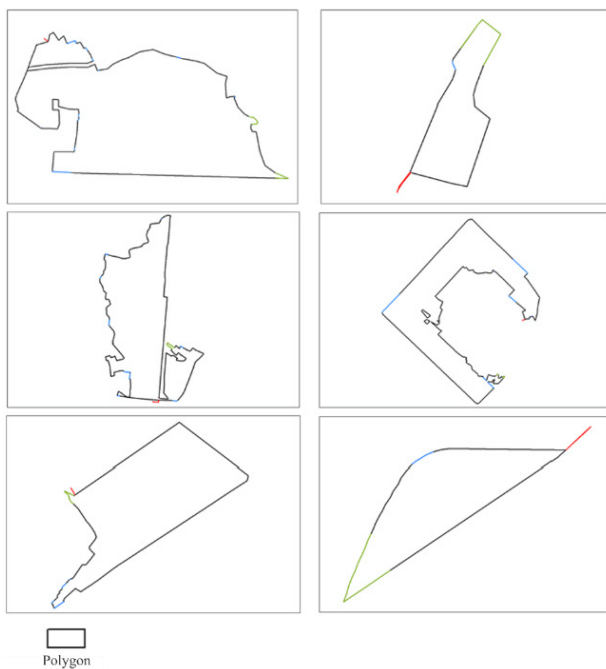
*Note.* The black arcs in the table represent part of the boundary arcs of the polygon, the blue lines represent the dividing lines at the concave points, and the red arcs represent the narrow and long arcs of the optimal segmentation results screened by different saliency indexes.

**Table 2.** Statistical data of arc length difference between optimal segmentation and reference data

| Statistics             | Significance index |        |                      |                            |
|------------------------|--------------------|--------|----------------------|----------------------------|
|                        | Profile length     | Reach  | Base to height ratio | Weighted base height ratio |
| Cumulative difference  | 284.801            | 91.682 | 91.682               | <b>2.504</b>               |
| Average difference     | 15.822             | 5.093  | 5.093                | <b>0.139</b>               |
| Root mean square error | 22.185             | 9.016  | 9.016                | <b>0.321</b>               |

### 4.3 Verification of the Rationality of Narrow and Long Arc Segment Identification

To verify the rationality of the algorithm used in this paper for identifying narrow and long arcs from the candidate arcs by combining the base-to-height ratio and the shape index, we compared the performance of the base-to-height ratio and the shape index separately and combined. The experimental results are presented in Figure 10, where the green line represents the base-to-height ratio identified arcs, the blue line represents the shape index identified arcs, and the red line depicts the combined identification result. It can be seen that the narrow and long arcs (red arcs) in the polygon can be more accurately identified using either the base-to-height ratio or the shape index alone, or both combined. However, only using the base-to-height ratio cannot effectively distinguish the long and narrow arcs from the normal arcs (green arcs) that protrude significantly, while only the shape index cannot effectively distinguish the long and narrow arcs from the straight arcs (blue arcs) that meet the threshold condition. Combining the two methods can avoid these types of misidentifications and improve the accuracy of narrow and long arc recognition.

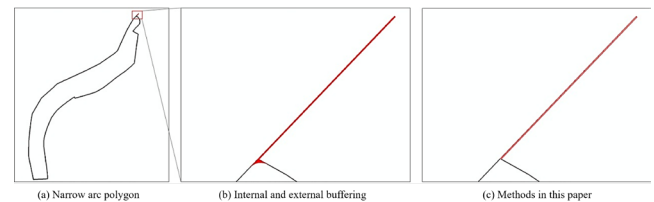

**Figure 10.** Recognition results using different indices

### 4.4 Comparative Analysis with Existing Methods

This section analyzes the advantages of the method designed and implemented in this paper over the internal and

external buffering method proposed by Gao et al. through comparative experiments, as indicated in Table 3.

The internal and external buffer method, as well as the method proposed in this paper, were utilized to identify the representative long and narrow arcs. However, due to the limitations of the polygon buffer generation method, the edges and corners of the original polygon boundary tend to become smoother after the inward and outward buffer analysis, resulting in a deviation between the detected long and narrow arcs and the visually judged results. On the other hand, the proposed method in this paper uses the nearest point pair of the concave point to determine the segmentation point pair and the segmentation line, resulting in a more visually accurate result, as depicted in Figure 11. The cumulative difference, average difference, and root mean square error between the identification results of the proposed method and the reference data are all smaller than those of the internal and external buffer method, as indicated in Table 4.


**Figure 11.** Long and narrow arcs recognized by different methods

To further compare the effectiveness of the method proposed in this paper with the internal and external buffer method, two sets of sample data were randomly selected from the Jinjiang control area dataset. As depicted in Figure 12, each dataset comprises polygons that include long and narrow arcs as well as those without such arcs, refers to Table 5.

In the experiment, the performance of the two methods was quantitatively evaluated by statistically analyzing the number of accurately recognized, misrecognized, and missed narrow arc segments. Accurate recognition refers to narrow arc segments that are present in both the recognition results and reference data. Misrecognition refers to narrow arc segments that are identified in the recognition results but not present in the reference data. Missed recognition refers to narrow arc segments that are present in the reference data but not identified in the recognition results.

According to Table 6, when the area threshold of the internal and external buffer method is set to 0.2, 0.5, and 1.0, the accurate recognition and false recognition decrease while the missed recognition increases with the increase of the area threshold. The recognition results heavily depend on the set

area threshold, and it is difficult to obtain results with a high accuracy of recognition and a low amount of false and missed recognition. The shaded area represents the statistical data of the recognition results obtained by the method proposed in this paper. Although there are some missed recognitions,

it has a higher number of accurate recognitions and a lower number of false recognitions. The method proposed in this paper performs better than the internal and external buffer method.

**Table 3.** Detailed steps of the internal and external buffering method

| Internal and external buffering method: Take polygon aPolygon as an example |  |
|---|--|
| <b>Step1:</b>   | Buffer the Polygon inward according to the predefined buffer width to get the result 1stBufferPolygons. If the 1 <sup>st</sup> Buffer-Polygons are empty in the inward buffering, it means that the entire Polygon is long and narrow. The algorithm ends, and a new polygon is taken and calculated from Step1 until all the polygons in the layer are processed; |
| <b>Step2:</b>   | Based on 1stBufferPolygons, use the same buffer width to buffer outward to get the result 2ndBufferPolygons;   |
| <b>Step3:</b>   | Calculate the difference between aPolygon and 2ndBufferPolygons, and record the result as narrow-sections;   |
| <b>Step4:</b>   | Use the area threshold to filter out small protrusions from narrow-sections, and obtain narrow conflicts (including long and narrow arc segments) in the polygon;  |
| <b>Step5:</b>   | Take a new polygon and calculate from Step1 until all the polygons in the layer are processed.   |

**Table 4.** Statistical data of arc length difference of different methods

| Statistics             | Recognition methods             |                        |
|------------------------|---------------------------------|------------------------|
|                        | Internal and external buffering | Method of this article |
| Cumulative difference  | 12.076                          | <b>2.504</b>           |
| Average difference     | 0.671                           | <b>0.139</b>           |
| Root mean square error | 0.747                           | <b>0.321</b>           |

**Table 5.** Composition of sample data

| Sample number | Number of polygons | Number of long and narrow arcs |
|---------------|--------------------|--------------------------------|
| 1             | 100                | 24                             |
| 2             | 100                | 25                             |

**Table 6.** Accuracy statistics of recognition results

| Sample number | Area threshold | Accurately identify the number | Number of misidentifications | Number of missed identifications |
|---------------|----------------|--------------------------------|------------------------------|----------------------------------|
| 1             | 0.2            | 16                             | 30                           | 8                                |
|               | 0.5            | 12                             | 10                           | 12                               |
|               | 1.0            | 8                              | 7                            | 16                               |
|               | —              | 20                             | 0                            | 4                                |
| 2             | 0.2            | 21                             | 36                           | 4                                |
|               | 0.5            | 17                             | 16                           | 8                                |
|               | 1.0            | 14                             | 9                            | 11                               |
|               | —              | 21                             | 1                            | 4                                |



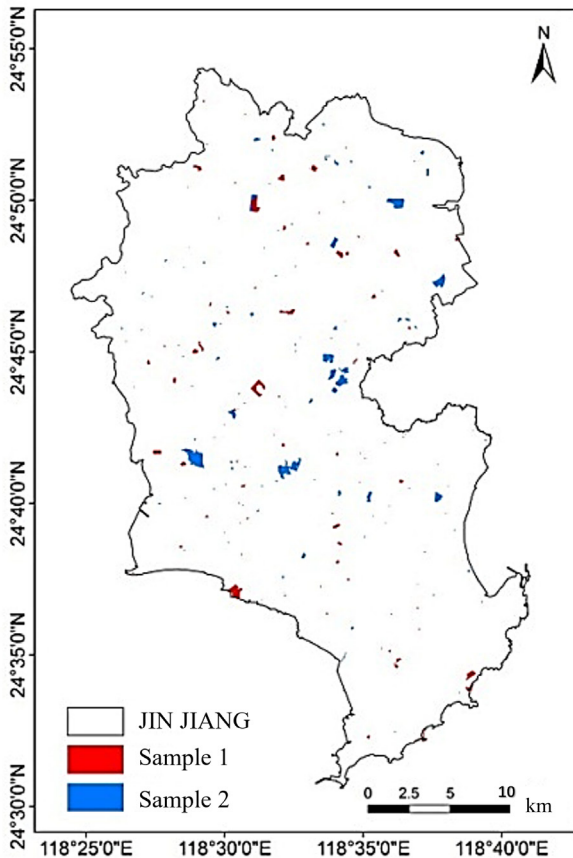


Figure 12. Distribution of sample data

## 5 Conclusion

To remove the narrow and long arcs from the vector map polygon elements is a fundamental problem in map data creation and maintenance. Accurate and efficient identification and extraction of narrow and long arcs is a core issue, which is of great significance for improving the quality of vector map data. Compared to similar mainstream research results such as the internal and external buffer method, the method proposed in this paper based on skeleton line shape decomposition for identifying narrow, long arcs in polygon elements has the advantages of being conceptually intuitive, computationally simple, which is consistent with human perception of results. The method includes four main steps: obtaining endpoint sub-polygons, solving segmentation lines, filtering optimal segmentation, and removing non-narrow and long arcs. Solving segmentation lines and filtering optimal segmentation is the tough task in the method, and the solution proposed in this paper considers both the consistency with human shape perception, computational efficiency. Narrow and long arcs are a relative concept, and it is necessary to further explore how to determine accurate and quantitative evaluation criteria and reasonably incorporate them into the implementation of the narrow and long arc identification and extraction algorithm, which is also the direction of future research.

## References

- [1] R. Devillers, A. Stein, Y. Bedard, N. Chrisman, P. Fisher, W. Shi, Thirty years of research on spatial data quality: achievements, failures, and opportunities, *Transactions in GIS*, Vol. 14, No. 4, pp. 387–400, August, 2010. <https://doi.org/10.1111/j.1467-9671.2010.01212.x>
- [2] Y. Chun, M.-P. Kwan, D. A. Griffith, Uncertainty and context in GIScience and geography: challenges in the era of geospatial big data, *International Journal of Geographical Information Science*, Vol. 33, No. 6, pp. 1131–1134, 2019. <https://doi.org/10.1080/13658816.2019.1566552>
- [3] D. Gaeuman, J. Symanzik, J. C. Schmidt, A map overlay error model based on boundary geometry, *Geographical Analysis*, Vol. 37, No. 3, pp. 350–369, July, 2005. <https://doi.org/10.1111/j.1538-4632.2005.00585.x>
- [4] K. Rybaczuk, Using information based rules for sliver polygon removal in GISs, in: M. M. Fischer, P. Nijkamp (Eds.), *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, Springer Berlin Heidelberg, 1993. [https://doi.org/10.1007/978-3-642-77500-0\\_6](https://doi.org/10.1007/978-3-642-77500-0_6)
- [5] L. Barazzetti, Sliver removal in object-based change detection from VHR satellite images, *Photogrammetric Engineering & Remote Sensing*, Vol. 82, No. 2, pp. 161–168, February, 2016. <https://doi.org/10.14358/PERS.82.2.161>
- [6] Z. Xie, Z. Ye, L. Wu, A design for polygon overlay algorithm in the simple feature model, *2008 Seventh International Conference on Grid and Cooperative Computing. IEEE*, Shenzhen, China, 2008, pp. 680–685. DOI: 10.1109/GCC.2008.41
- [7] S. Cho, M. X. Punithan, J. Gim, Y. Huh, Tagging-the-triangle algorithm for partitioning features with inconsistent boundaries, *International Journal of Geographical Information Science*, Vol. 28, No. 12, pp. 2533–2550, June, 2014. <https://doi.org/10.1080/13658816.2014.937716>
- [8] M. Yang, T. Ai, P. Liu, X. Cheng, The matching and consistency correcting in the integration of contour and river network, *Acta Geodaetica et Cartographica Sinica*, Vol. 41, No. 1, pp. 152–158, February, 2012.
- [9] X. Chen, R. Guo, H. Yan, Creation and Consistency Maintenance of Topological Relationships Between Patches in Landuse Database Generalization, *Geomatics and Information Science of Wuhan University*, Vol. 30, No. 4, pp. 370–373, April, 2005.
- [10] T. Ai, H. Wu, Consistency correction of shared boundary between adjacent polygons, *Geomatics and Information Science of Wuhan University*, Vol. 25, No. 5, pp. 426–431, October, 2000. DOI: 10.3969/j.issn.1671-8860.2000.05.010
- [11] W. Gao, J. Gong, L. Yang, X. Jiang, X. Wu, Detecting geometric conflicts for generalization of polygonal maps, *The Cartographic Journal*, Vol. 49, No. 1, pp. 21–29, 2012. DOI: 10.1179/1743277411y.0000000016
- [12] M. Bader, R. Weibel, Detecting and resolving size and proximity conflicts in the generalization of polygonal

maps, *Proceedings 18th International Cartographic Conference*, Stockholm, Sweden, 1997, pp. 1525-1532.

- [13] J. M. Ware, C. B. Jones, Conflict reduction in map generalization using iterative improvement, *GeoInformatica*, Vol. 2, No. 4, pp. 383–407, December, 1998. <https://doi.org/10.1023/A:1009713606524>
- [14] L. P. Chew, Constrained delaunay triangulations, *Proceedings of the third annual symposium on Computational geometry*, Waterloo, Ontario, Canada, 1987, pp. 215–222. <https://doi.org/10.1145/41958.41981>
- [15] B. Peter, Measures for the generalization of polygonal maps with categorical data, *Fourth ICA Workshop on Progress in Automated Map Generalization*, Beijing, China, 2001, pp. 1–21.
- [16] X.-F. Liu, Y.-L. Wu, H. Hu, A method of extracting multiscale skeletons for polygonal shapes, *Acta Geodaetica et Cartographica Sinica*, Vol. 42, No. 4, pp. 588-594, August, 2013.
- [17] D. Reniers, A. Telea, Patch-type segmentation of voxel shapes using simplified surface skeletons, *Computer Graphics Forum*, Vol. 27, No. 7, pp. 1837–1844, October, 2008. <https://doi.org/10.1111/j.1467-8659.2008.01330.x>
- [18] J. Jiang, D. Zhou, S. Hao, Y. Guo, S. Zhan, Planar shape decomposition combining skeletal and boundary features, *Journal of Image and Graphics*, Vol. 17, No. 11, pp. 1425-1430, November, 2012. DOI: 10.11834/jig.20121112
- [19] D. D. Hoffman, M. Singh, Saliency of visual parts, *Cognition*, Vol. 63, No. 1, pp. 29–78, April, 1997. doi: 10.1016/s0010-0277(96)00791-3
- [20] M. Singh, G. D. Seyranian, D. D. Hoffman, Parsing silhouettes: The short-cut rule, *Perception & Psychophysics*, Vol. 61, No. 4, pp. 636–660, January, 1999.
- [21] J. Zeng, R. Lakaemper, X. Yang, X. Li, 2d shape decomposition based on combined skeleton-boundary features, *Advances in Visual Computing: 4th International Symposium, ISVC 2008*, Las Vegas, NV, USA, 2008. pp. 682–691. [https://doi.org/10.1007/978-3-540-89646-3\\_67](https://doi.org/10.1007/978-3-540-89646-3_67)
- [22] Y. Ma, M. Lu, S. Teng, J. Zhang, A visual saliency based hierarchical shape decomposition algorithm, *Journal of Computer-Aided Design Computer Graphics*, Vol. 26, No. 6, pp. 914–922, June, 2014.
- [23] J. Li, Y. Gao, J. She, Method for generating forest landscapes based on the characteristics of plant distribution, *Acta Geodaetica et Cartographica Sinica*, Vol. 47, No. 8, pp. 1133-1140, August, 2018. doi: 10.11947/j.AGCS.2018.20180102
- [24] B. Wu, Z. He, H. Xu, Z. Li, Y. Zhang, W. Qiu, L. Jiang, Eliminating methods to false polygon in land use database, *Journal of Henan Agricultural Sciences*, Vol. 32, No. 5, pp. 62-65, May, 2008.
- [25] G.-S. Xia, J. Huang, N. Xue, Q. Lu, X. Zhu, Geosay: A geometric saliency for extracting buildings in remote sensing images, *Computer Vision and Image Understanding*, Vol. 186, pp. 37–47, September, 2019. <https://doi.org/10.1016/j.cviu.2019.06.001>

## Biographies



**Han Zhang**, Male, born in 1984, senior engineer of surveying and mapping engineering, PhD student in East China University of Technology, won the Young Science and Technology Talent Award of Chinese Society of Surveying and Mapping in 2021, mainly engaged in the research of geographic information system, surveying and mapping engineering, remote sensing, geological disaster monitoring and territorial space planning.



**Xiaoyong Chen**, Male, born in September 1961, professor, doctor, doctoral supervisor, mainly engaged in theoretical research and application technology development of geographic information science.



**Zhihui Yi**, Male, born in 1985, from Xinyu, Jiangxi province, senior engineer of Geographic Information Engineering Team of Jiangxi Geological Bureau, has won the second prize of Geographic Information Technology Progress Award, the first prize of Jiangxi Provincial Surveying and Mapping Geographic Information Engineering Award, and the second prize of Fujian Geographic Information Technology Progress Award.



**Chao Xu**, Male, born in 1991, Shangrao, Jiangxi province, intermediate engineer. Engaged in 5/6 G network technology research, smart city, time and space sequence analysis, won the first prize of surveying and mapping geographic information engineering in Jiangxi province.