

Image Edge Analysis and Application Based on Least Squares Fitting Model

YuYang Su^a, BingXuan Yu^b, Ruixin Peng^c, HongShu Yan^d

School of Mathematics and Statistics, Yunnan University, Kunming, China

^asuyuyang@mail.ynu.edu.cn

^btomorinao@mail.ynu.edu.cn

^cprx@mail.ynu.edu.cn

^dyanhongshu@mail.ynu.edu.cn

YuYang Su, BingXuan Yu, Ruixin Peng, HongShu Yan are co-first authors

Abstract. This paper presents an effective method to automatically segment and fit the edge contour curve data of a picture into straight line segments, circular arc segments or elliptical arc segments. Firstly, the curvature of each point is calculated, and the curvature change breakpoints are found to distinguish between straight lines, circular arcs and elliptical arcs; then, circular arc segments and elliptical arcs are distinguished by the change degree of curvature of adjacent points, and the automatic segmentation of contour edges is realized. Finally, the circular arc segment and elliptic arc segment are fitted by least squares method, based on which the length of the straight line, the direction angle of the circular arc, the rotation angle of the elliptic arc and other related information are calculated. The results show that our model can better identify the feature texture information of image edges.

Keywords: subpixel; contour extraction; least-squares fitting; curvature.

1. Introduction

Vision measurement mainly relies on the edges in the image, and edge detection is one of the most important aspects of image processing. In practical image processing problems, the edge of an image, as a basic feature of an image, is often used in higher-level processing and analysis of feature description, image recognition, image segmentation, image enhancement, and image compression, so that the image can be further analyzed and understood. Traditional contour extraction methods mainly use edge detection operators to extract edges, and then remove spurious and redundant edges and perform edge repair according to the contour characteristics of the target object. With the deepening of research and development of technology in this area, various new contour extraction methods have emerged.

In dimensional measurement systems, the measurement accuracy has a great relationship with the accuracy of edge detection. At present, the more classical operators in edge detection include Canny, Sobel, etc. However, the detection accuracy of these operators only stays at the pixel level. With the increasing requirements for product accuracy in industrial production, the pixel-level detection accuracy can no longer meet the actual production requirements. Therefore, sub-pixel level detection operators have attracted more and more attention. The sub-pixel edge detection method is developed from the classical edge detection method, which first finds the edge pixel points with single-pixel accuracy using the classical edge detection method, and then uses the grayscale values around these pixel points as supplementary information to locate the edges of the detected image at a more accurate position. Based on this, we apply the idea of least squares to the given two sub-pixel contour edge data (see Figure 1 for the shape), and give a method to automatically segment and fit the edge contour curve data of the image into straight lines. Efficient method for segments, circular arc segments, or elliptical arc segments.

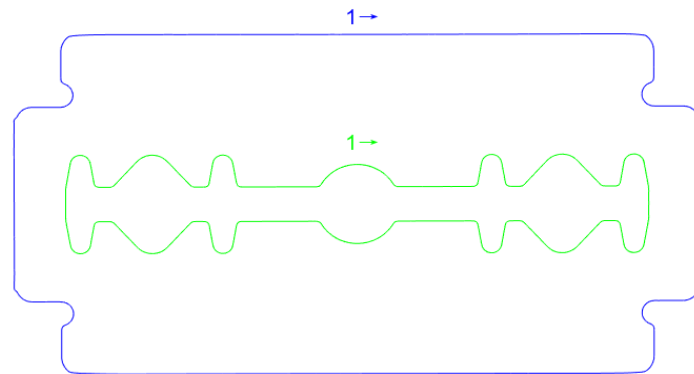


Fig 1 edge contour image data

2. Basic Method

For a curve, the curvature is the rotation rate of the tangential direction angle to the arc length at a point on the curve, which also indicates the degree of curve deviation from the straight line. Let the parametric equation of a curve in the two-dimensional plane be

$$\begin{cases} x = u(t) \\ y = v(t) \end{cases}$$

where both functions $x = u(t)$ and $y = v(t)$ contain second-order derivatives, the curvature expression of the curve is

$$K = \frac{u'(t)v''(t) - u''(t)v'(t)}{[u'^2(t) + v'^2(t)]^{\frac{3}{2}}}$$

For the curve expression, the derivative can be directly performed to calculate the curvature at a certain point, but for discrete data, the first-order derivative and second-order derivative calculated by discrete points will cause relatively large errors if the difference method is used directly to calculate them. Therefore, the curvature of the

quadratic curve fitted using three points is used as the curvature estimate K , as shown in Fig 2 below.

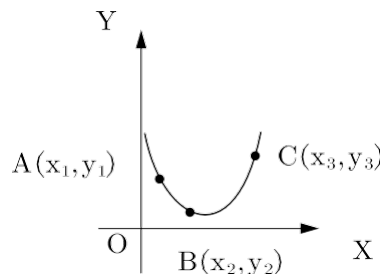


Fig 2 The quadratic curve fitted by three points

There exist three points $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$ in the right-angle coordinate system, let the parametric equation of the quadratic curve fitted by the three points be

$$\begin{cases} x = a_1 + a_2 t + a_3 t^2 \\ y = b_1 + b_2 t + b_3 t^2 \end{cases}$$

where t is the parameter of the curve. The values of the unknowns a and b can be solved by substituting the coordinates of the three points. According to Hypothesis 5, because the coordinates of the contour data points are uniformly distributed, the lengths of the two vectors are used as the range of values of the curve parameter t .

$$t_a = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$t_b = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$

and this set of data satisfies a certain condition

$$(x, y)|_{t=-t_a} = (x_1, y_1)$$

$$(x, y)|_{t=0} = (x_2, y_2)$$

$$(x, y)|_{t=t_b} = (x_3, y_3)$$

then there are

$$\begin{cases} x_1 = a_1 - a_2 t_a + a_3 t_a^2 \\ x_2 = a_1 \\ x_3 = a_1 + a_2 t_b + a_3 t_b^2 \end{cases}, \begin{cases} y_1 = b_1 - b_2 t_a + b_3 t_a^2 \\ y_2 = b_1 \\ y_3 = b_1 + b_2 t_b + b_3 t_b^2 \end{cases}$$

Its matrix form is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & -t_a & t_a^2 \\ 1 & 0 & 0 \\ 1 & t_b & t_b^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

and

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & -t_a & t_a^2 \\ 1 & 0 & 0 \\ 1 & t_b & t_b^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Let

$$\begin{cases} X = (x_1, x_2, x_3)', \\ Y = (y_1, y_2, y_3)', \\ A = (a_1, a_2, a_3)', \\ B = (b_1, b_2, b_3)', \\ M = \begin{pmatrix} 1 & -t_a & t_a^2 \\ 1 & 0 & 0 \\ 1 & t_b & t_b^2 \end{pmatrix}. \end{cases}$$

The above linear equations can be solved by matrix inversion as follows.

$$\begin{cases} A = M^{-1} X, \\ B = M^{-1} Y. \end{cases}$$

Calculate the derivatives of the variables from the results of solving the equation

$$x' = \left. \frac{dx}{dt} \right|_{t=0} = a_2$$

$$x'' = \left. \frac{d^2x}{dt^2} \right|_{t=0} = 2a_3$$

$$y' = \left. \frac{dy}{dt} \right|_{t=0} = b_2$$

$$y'' = \left. \frac{d^2y}{dt^2} \right|_{t=0} = 2b_2$$

Substituting this into the formula for curvature

$$\hat{K} = \frac{u'(t)v''(t) - u''(t)v'(t)}{[u'^2(t) + v'^2(t)]^{\frac{3}{2}}} = \frac{2(a_3b_2 - a_2b_3)}{(a_2^2 + b_2^2)^{\frac{3}{2}}}$$

The curvature of the three points at the midpoint is estimated. Since the curvature of a straight line is 0 and the curvature of a circle and an ellipse is not 0, the curvature of a series of points calculated above can be used to determine the location of the intersection of a straight line segment and a curve segment in the contour edge, i.e., the breakpoint region. In an ideal state, the curvature of a circle of radius is , and the curvature of two adjacent points of the ellipse is not the same. According to Hypothesis 5 and 6, when the points of the contour edge are uniformly arranged and the contour can be combined by a straight line segment, a circular arc segment and an elliptical arc segment, it is possible to distinguish a circular arc segment from an elliptical arc segment by comparing the change in curvature of two adjacent points.

Based on this, the automatic segmentation of the two sets of contour data can be achieved by setting and adjusting the threshold values for judging the straight line segment and the curved segment and the circular arc segment and the elliptical arc segment. Among them, the frequency distribution of the curvature calculated for the two sets of data is as follows in Fig 3 and Fig 4.

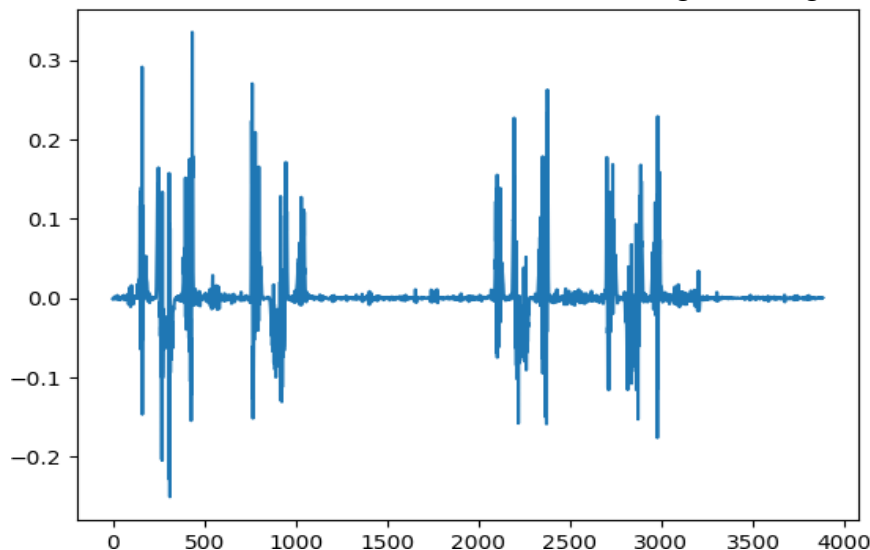


Fig 3 Frequency distribution of the curvature of the first set of data

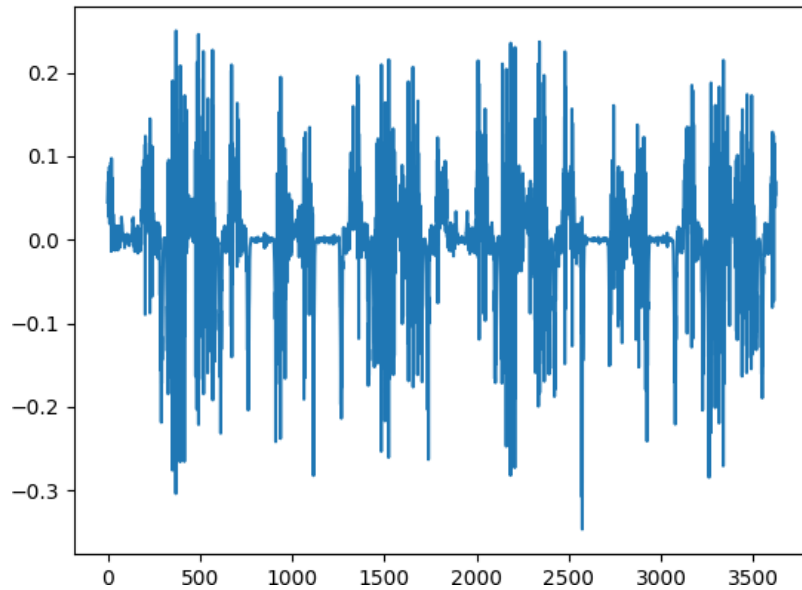


Fig 4 Frequency distribution of the curvature of the second set of data

According to the curvature of the straight line is 0 and the curvature of the curve is not 0. Observing Fig 2 and Fig4, it can be found that the first group of data represents a contour in which the straight line segment accounts for a relatively large number of points, and the number of breakpoints in which the curvature value is mutated is within the interval [8,16]; while the second group of data represents a contour in which the curve segment accounts for a relatively large number of points, interspersed with some straight line segments.

Then, the changes in curvature of the two groups of data were plotted separately by calculating the difference between adjacent curvatures, as follows in Fig 5 and Fig 6

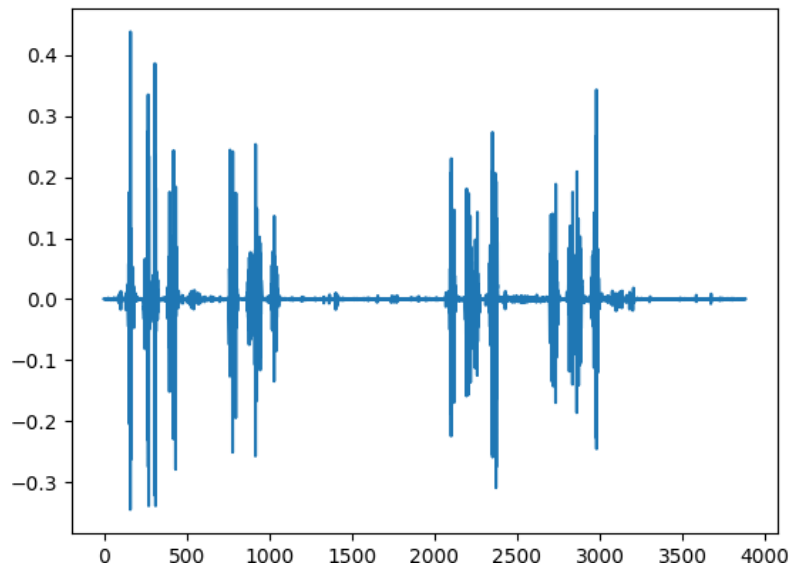


Fig 5 Frequency distribution of curvature change of the first set of data

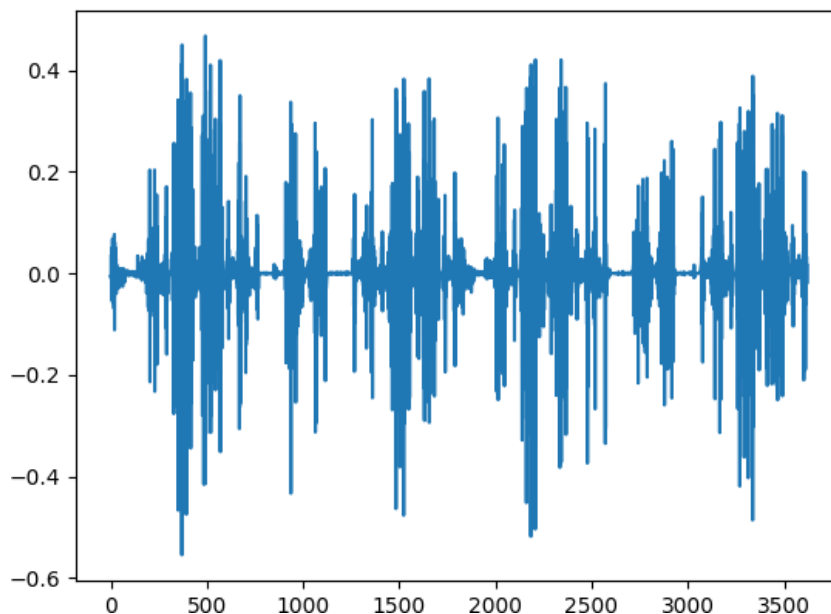


Fig 6 Frequency distribution of curvature change of the second set of data

The location of the breakpoint can be seen more clearly from Fig 3.4, and the place where the curvature changes a lot indicates the existence of elliptical arc. Set the threshold t_1 for judging straight line segment and curve segment, and t_2 for judging circular arc segment and elliptic arc segment, when the curvature of the set of points of a series of adjacent points is greater than t_1 , it indicates that the region is a curve segment, otherwise it is a straight line segment, when the curvature change value of the set of points of a series of adjacent points in a curve segment is greater than t_2 , it indicates that the curve segment is an elliptic arc segment, otherwise it is a circular arc segment. Based on this, by adjusting the size of the threshold t_1 and t_2 to find out the breakpoint region of the two sets of data, the adjusted threshold values are shown in Table 1 below.

Table.1 Adjusted threshold value

Data	Threshold	size
the first set of data	t_1	0.032
	t_2	0.0025
the second set of data	t_1	0.035
	t_2	0.0001

Scatter plots of the two data breakpoint regions are plotted as follows in Fig 7, Fig 8.

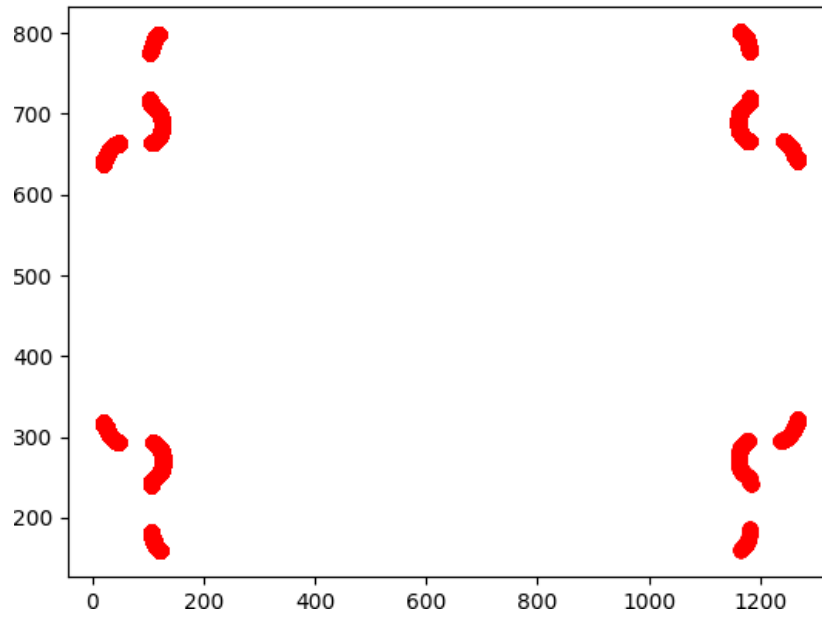


Fig 7 Breakpoint area for the first set of data

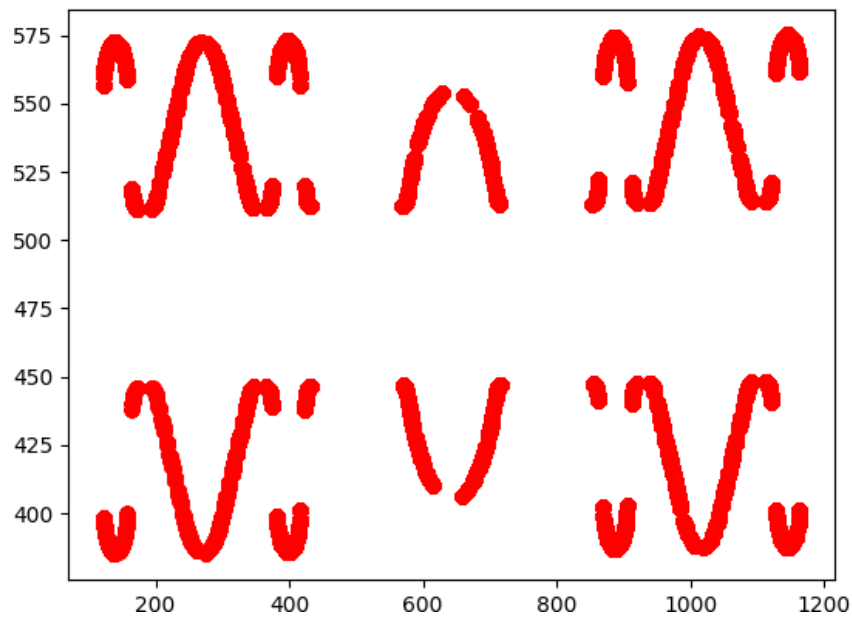


Fig 8 Breakpoint area for the second set of data

Combined with the contour plot given in the question, it can be clearly observed that the area shown in the scatter plot is roughly the area of the curved segment in the contour. Based on the threshold t_1 , the regions of the straight-line segments are plotted for both sets of data, as shown in Fig 9 and Fig 10 below.

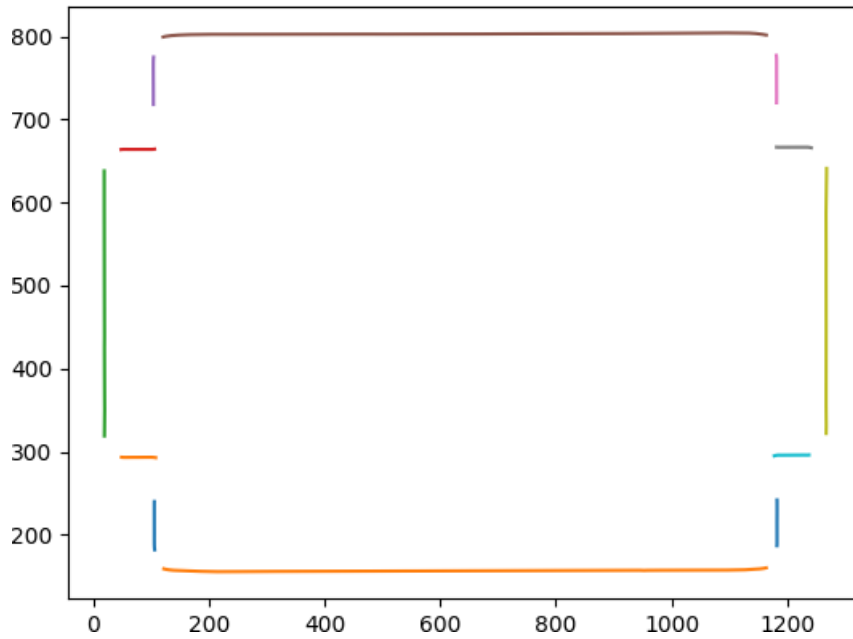


Fig 9 Straight line segment area for the first set of data

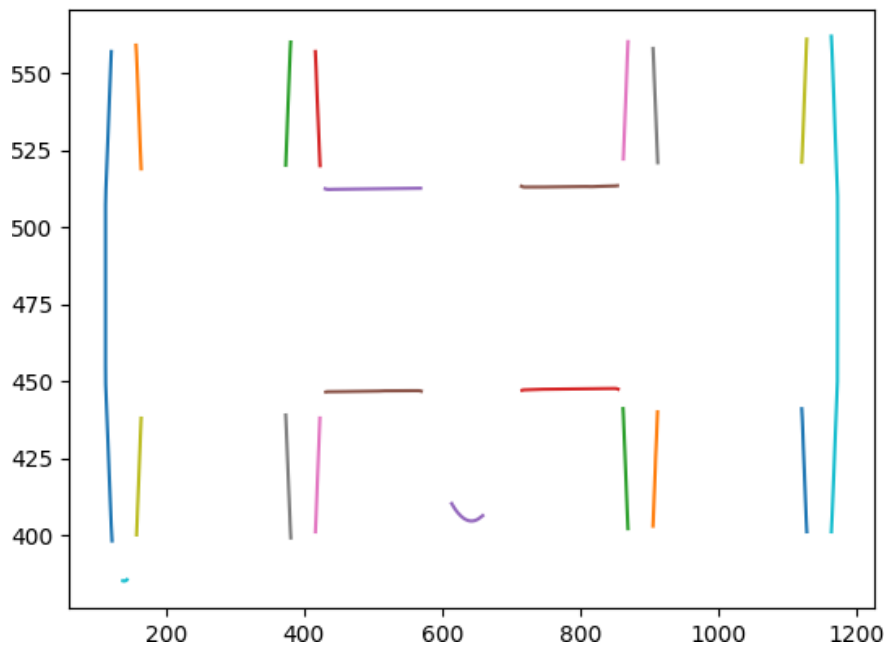


Fig 10 Straight line segment area for the second set of data

It is observed that in Fig 10, except for a small section of curve in the lower middle part and a small section of curve in the lower left corner that is incorrectly fitted, the results of the straight line segments fitted to both sets of data are relatively good, probably due to the problem of data point arrangement.

3. Fitting Model

For a series of data points (x_i, y_i) , that fall approximately on a circle, let the equation of the circle to which they are fitted be

$$(x - x_c)^2 + (y - y_c)^2 = R^2$$

The sum of the squares of their distances is

$$f = \sum (\sqrt{(x - x_c)^2 + (y - y_x)^2} - R)^2$$

The usual least squares fit is to fit the minimum value of this sum of squares, and to reduce the computational effort, we define the least squares solution equation

$$f = \sum ((x - x_c)^2 + (y - y_x)^2 - R^2)^2$$

Let

$$g(x, y) = (x - x_c)^2 + (y - y_x)^2 - R^2$$

That is, solving the equation

$$f = \sum g(x_i, y_i)^2$$

Its minimum value, which is ultimately reduced by a least squares fit, is

$$R^2 = \sum \frac{(x_i - x_c)^2 + (y_i - y_c)^2}{N}$$

Ideally, the sum of the distances of the points on the ellipse to the two focal points is a constant F . Let a point set $D_i(x_i, y_i), i = 1, \dots, n$, containing n sample points, the

coordinates of the two foci of the fitted ellipse are $f_1(x_{f_1}, y_{f_1}), f_2(x_{f_2}, y_{f_2})$ and the distances from each sample point to the two foci of the ellipse are

$$F_i = \sqrt{(x_i - x_{f_1})^2 + (y_i - y_{f_1})^2} + \sqrt{(x_i - x_{f_2})^2 + (y_i - y_{f_2})^2}$$

Calculate the variance of F and F_i

$$I = \sqrt{\frac{\sum_{i=0}^n (F_i - F)^2}{n - 1}}$$

The minimum value of its variance by least squares fitting is the resulting optimal solution. A graphical representation of the two data curve areas is shown in Fig 11 and Fig 12 below.

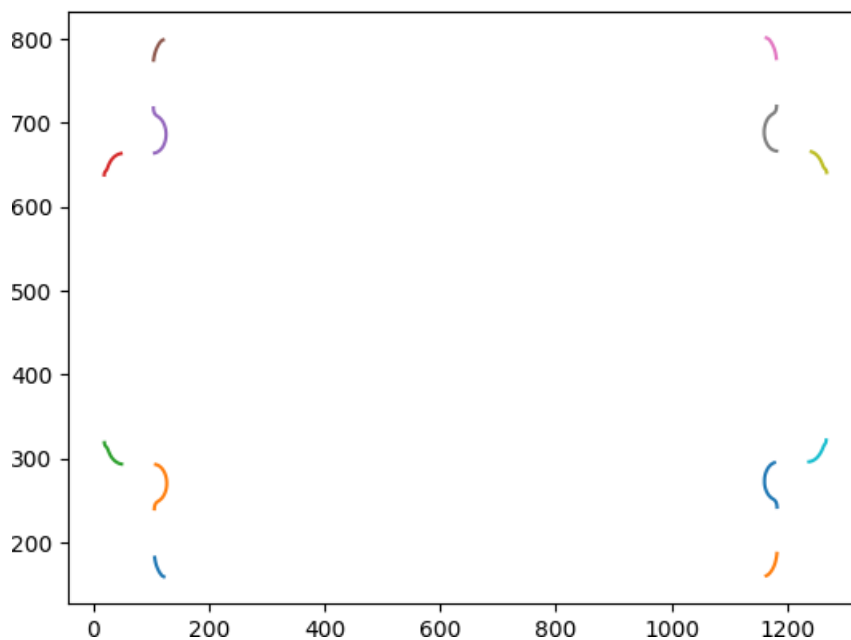


Fig 11 Curved area for the first set of data

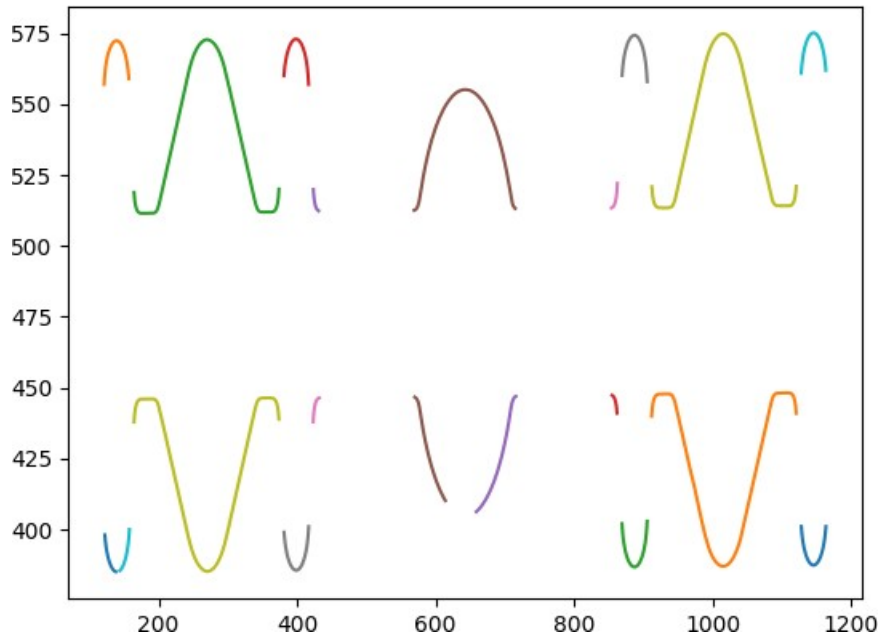


Fig 12 Curved area for the second set of data

4. Results

Straight segment, circular arc segment and elliptical arc segment to be sought attribute feature results are saved, and some of the data are shown in the following table 2.

Table 2 Part of the results show

Edge Contour 1						
NO	TYPE	PARAM				
S1	Line	(106.18,182.03)	(105.61, 240.02)	59		
S2	Line	(107.98,292.86)	(48.95, 293.41)	60		
S3	Line	(19.21,319.01)	(18.92, 638.0)	320		
S4	Line	(48.03, 663.69)	(105.97, 664.23)	59		
S5	Line	(104.05, 718.0)	(104.32, 774.96)	58		
S13	CircularArc	(272.63,1406.05)	(122.03, 159.12)	(106.18, 182.03)	1	
S14	CircularArc	(272.63,1406.06)	(48.95, 293.41)	(19.21, 319.01)	2	
S15	CircularArc	(272.63,1406.07)	(18.92, 638.0)	(48.03, 663.69)	1	
S16	CircularArc	(272.63,1406.08)	(104.32, 774.96)	(120.96, 799.19)	1	
S17	CircularArc	(272.63,1406.09)	(1164.09, 801.51)	(1181.56, 776.94)	1	
Edge Contour 2						
NO	TYPE	PARAM				
S1	Line	(122.29, 398.08)	(121.26, 556.95)	160		
S2	Line	(156.99, 559.0)	(164.46, 518.87)	41		
S3	Line	(373.8, 520.04)	(380.98, 560.01)	41		
S4	Line	(416.67, 556.94)	(423.51, 519.89)	38		

S5	Line	(431.04, 512.39)	(569.03, 512.56)	139		
S20	CircularArc	(140.08, 403.47)	(138.01, 385.13)	(122.29, 398.08)	67	
S21	CircularArc	(140.08, 403.48)	(121.26, 556.95)	(156.99, 559.0)	13	
S22	CircularArc	(140.08, 403.49)	(164.46, 518.87)	(373.8, 520.04)	52	
S23	CircularArc	(140.08, 403.50)	(380.98, 560.01)	(416.67, 556.94)	4	
S24	CircularArc	(140.08, 403.51)	(423.51, 519.89)	(431.04, 512.39)	2	

The two groups of data finally segmented out of the graph in Fig 13, Fig 14 below.

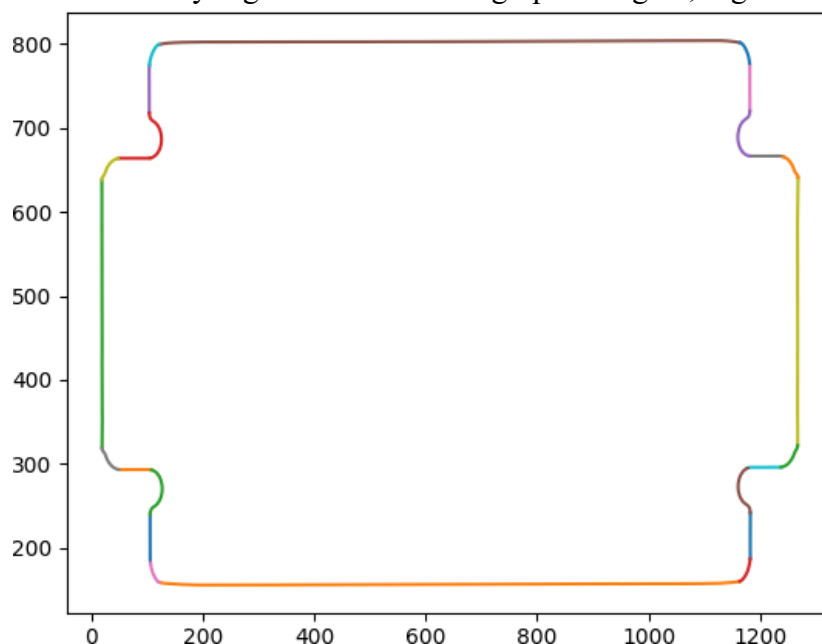


Fig 13 The first set of data segmentation chart

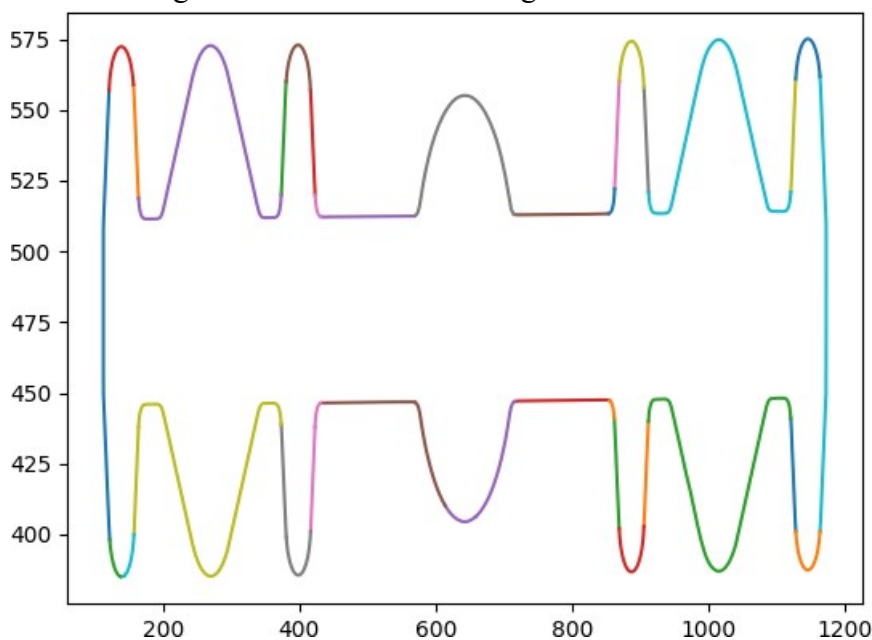


Fig 14 The second set of data segmentation chart

In summary, the segmentation results of contour data by judging the breakpoint location through curvature and fitting the curve by least squares are better, and the segmentation effect of data sets

with fewer curve contours is significantly better than that of data sets with more and more complex curve contours.

References

- [1] Li Che, Ji Tianyu, Li Yuche. Curvature attribute analysis and its application in seismic data interpretation [J]. Contemporary Chemical Industry, 2014, 43(04): 558-560. DOI: 10.13840/j.cnki.cn21-1457/tq.2014.04.002.
- [2] Wang Wei, Cai Wensheng, Deng Ke. Discussion on fitting water level-discharge relationship curve by least square method [J]. Shaanxi Water Resources, 2021(11):51-52+58.DOI:10.16747/j.cnki.cn61 - 1109/tv.2021.11.017.
- [3] Shen Y , Zhang X , Cheng W , et al. Quasi-eccentricity error modeling and compensation in vision metrology[J]. Measurement ence and Technology, 2018, 29(4).
- [4] Agustín Trujillo-Pino, Karl Krissian, Miguel Alemán-Flores, & Daniel Santana-Cedrés (2013). Accurate subpixel edge location based on partial area effect Image and Vision Computing.
- [5] Weihang Zhang, Xue Wang, Wei You, Junfeng Chen, Peng Dai, & Pengbo Zhang (2020). RESLS: Region and Edge Synergetic Level Set Framework for Image Segmentation IEEE Transactions on Image Processing.
- [6] Stanley Osher, & James A. Sethian (1988). Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations Journal of Computational Physics.
- [7] Weihang Zhang, Xue Wang, Wei You, Junfeng Chen, Peng Dai, & Pengbo Zhang (2020). RESLS: Region and Edge Synergetic Level Set Framework for Image Segmentation IEEE Transactions on Image Processing.
- [8] Shengjie Zhu, Garrick Brazil, & Xiaoming Liu (2020). The Edge of Depth: Explicit Constraints Between Segmentation and Depth computer vision and pattern recognition.