Research on Vectorization Method of Complex Linear Image Data

Jinbao Shan[1(✉)] and Weiwei Jiang[2]

[1] College of Information Technology and Art Design, Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China
shanjbl980@sina.com

[2] College of Food and Drug, Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China

Abstract. The traditional data extraction method of complex linear pixel image data can not cope with the vibration and noise of data in the process of data vectorization, which causes the problem of low accuracy of the extraction results. To solve this problem, a complex vector image extraction method is proposed. The MATLAB method is used to remove the noise of complex linear pixel images. In this way, the preprocessing of complex linear pixel image data is provided as the condition of segmentation. The two value algorithm is used to segment the complex linear pixel image data, and the minimum value of the target function is calculated. The data curves are drawn according to the calculation results. Vectorization of image data. The experimental results of the simulated application environment design show that the accuracy of the extraction result is about 45% compared with the traditional extraction method when the same image data is used.

Keywords: Complex line element image · Data vectorization extraction · Image preprocessing · Image segmentation

1 Introduction

Because the traditional extraction method of complex linear pixel image data extraction has the problem of low accuracy and poor integrity, this paper uses MATLAB to preprocess and calculate images, divide the images through two values algorithm, and designs a vector extraction method of complex heterotin image data based on these two algorithms. Experimental results show that the method designed in this paper is more suitable for image data extraction than traditional methods.

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 315–321, 2019.
https://doi.org/10.1007/978-3-030-19086-6_35

2 Design of Vectorization Method for Complex Linear Image Data

2.1 Preprocessing of Complex Linear Pixel Image Data

In order to extract the vectorization data of complex linear pixel images, it is necessary to read the image first and preprocess it, and normalize the image information for subsequent processing. A complex linear pixel digital image can be expressed as a real matrix structure. Most complex linear images can be expressed as a two-dimensional matrix by using MATLAB software. The matrix expression of the image is as follows [1, 2]:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix} \tag{1}$$

In this formula: $f(x,y)$ represents the image, M is the width of the image, N is the height of the image, x represents the row number, and y represents the number of columns. Each element in the matrix corresponds to a single pixel point in the image. There is a one-to-one correspondence between the coordinates of the corresponding pixels in the image and the index of the matrix, and the element values represent different luminance or gray [3]. In the spatial coordinates of the pixel points of the image, the pixels are regarded as discrete unit and represented by the corresponding coordinate pairs, which is similar to the plane right angle coordinate system, and that the Y coordinates are increasing from top to bottom. In MATLAB, the image can be read by calling the imread function. The specific information of the image or the element value of the pixel point corresponding to the digital matrix can be viewed through the imfinfo or imview function. The subsequent operation of the image is actually the operation of the digital matrix.

Image noise refers to the discrete-time pixel points which are deviating from the original image and interfering with the subsequent recognition process because of the influence of the acquisition or transmission of the image. It is the non source information in the image, such as the noise of the outer camera itself and the noise of the transmission circuit. Noise is often visually different from adjacent pixels. It can be regarded as a random variable and can be described by probability density function. The basic model used for image noise processing is as follows [4]:

$$y = Hx + \tau \tag{2}$$

In this formula: y is a complex linear image with noise τ, and H is known, unknown or only partially statistical information.

To sum up, when we are preprocessing a complex line image, it should be extracted and processed first, and the denoising process is carried out. The calculation equation of the image preprocessing is as follows:

$$f(x, y) = f(M-1, N-1, aC_1 + bC_2 + cC_3), \quad y = Hx + \tau \tag{3}$$

2.2 Complex Linear Pixel Image Data Segmentation

Image segmentation means the binarization of an image, which means that the pixels of an image are divided into two similar parts according to the gray level. The process is similar to the data according to the approximate attribute of the data, which is similar to [5–7].

A complex line element image is divided into two classes S_1 and S_2, m_1 and m_2 are two categories of gravity centers, and i represents the number of image segmentation. The two valued problem of the image is divided into S_1 and S_1, so that the following objective functions are minimum:

$$f(m_1, m_2) = \sum_{i=1}^{2} \sum_{l \in S_l} n_l (l - m_i)^2 \tag{4}$$

Threshold t can be obtained from the final division results S_1 and S_2.

Here, the function $f(m_1, m_2)$ is called a weighted error square sum function, and the l in the formula represents the gray level number of the image, and n_l represents the number of pixels with the gray level of l, that is, the weighting factor, and the center of gravity is calculated by the lower formula [8, 9]:

$$m_i = \tfrac{1}{d_i} \sum_{l \in S_l} n_l l, \quad i = 1, 2 \tag{5}$$

$$d_i = \sum_{l \in S_l} n_l, \quad i = 1, 2 \tag{6}$$

In this formula, d represents the dependent variable. For the following analysis, the square error of the weighted error of class S_i is as follows:

$$f_i = \sum_{i \in S_l} n_l (l - m_i)^2 \tag{7}$$

A new iterative algorithm is derived below, whose purpose is to find a partition that can make the objective function $f(m_1, m_2)$ minimum [10].

That is, suppose that a gray level k is in class S_i, and if it is moved to class S_j, then m_j becomes m_j^*.

$$m_j^* = m_j + \frac{(k-m_j)n_k}{d_j + n_k} \tag{8}$$

In this formula, n_k represents the number of pixels with a gray level of K. At the same time, f_j increased to:

$$f_j^* = f_j + \frac{d_j n_k (k-m_j)^2}{d_j + n_k} \tag{9}$$

When $d_i \neq n_k$, by similar deduction, we can get the formula of m_i and f_i, by similar deduction.

$$m_i^* = m_i - \frac{(k-m_j)^2}{d_j + n_k} \tag{10}$$

$$f_i^* = f_i - \frac{d_i n_k (k-m_j)^2}{d_i - n_k} \tag{11}$$

To sum up, the calculation formula of complex linear image segmentation is as follows:

$$f_{\min}(m_1, m_2) = (m_j - \frac{(k-m_j)n_k^2}{d_j + n_k})(f_i - \frac{d_i n_k (k-m_j)^2}{d_i + n_k}), \ \frac{d_i n_k (k-m_i)^2}{d_i - n_k} > \frac{d_j n_k (k-m_j)^2}{d_j - n_k} \tag{12}$$

2.3 Vector Extraction of Image Data

The image preprocessing algorithm designed by 1.1 and the 1.2 design image segmentation algorithm (the image two value algorithm) are integrated, and then the algorithm model which can extract the data of the complex linear pixel image is summarized, namely:

$$f(x, y) = f(M-1, N-1, aC_1 + bC_2 + cC_3), \quad y = Hx + \tau \tag{13}$$

$$f_{\min}(m_1, m_2) = (m_j - \frac{(k-m_j)n_k^2}{d_j + n_k})(f_i - \frac{d_i n_k (k-m_j)^2}{d_i + n_k}), \frac{d_i n_k (k-m_i)^2}{d_i - n_k} > \frac{d_j n_k (k-m_j)^2}{d_j - n_k} \tag{14}$$

First, the complex linear pixel images which need to be extracted from the data are brought into the Eq. (13) to extract and denoise. In the process of processing, only the useful feature information in the image is processed. After the preprocessing, the image affected by the noise can be restored. Then, the reconstructed image data is brought into (14), and the image data is segmented and converted into a two value black-and-white map based on the image data. The calculated $f(m_1, m_2)$ values are the proportional factor of the coordinate axis, and all the features except the image are set to the background color, and then the pixels of the image are searched in the gray matrix. Coordinates, multiplied by the scale factor, can get the true coordinates of each point of the curve. For the pixels with multiple Y values, the mean value of the Y value is processed to get the coordinates of each point of the image. In order to draw the data graph with image data as the horizontal and vertical coordinates, the vectorization of image data is realized.

3 Experimental Analysis

3.1 Experimental Data

In order to verify the accuracy of the method, we extract the experimental image from the complex linear image with a function of $y = \sin x + 25\cos(5x) + e^{x/2} + 200$, as shown in Fig. 1. The data extraction method designed in this paper is used to extract the JIN line data of the graph, and the horizontal and vertical coordinates of the image data curve are calculated, and the image data coordinates are drawn.

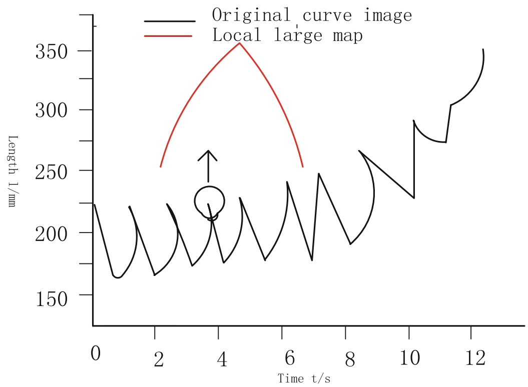

Fig. 1. Original curve image and local large map

3.2 Experimental Results and Analysis

The data is calculated by using the method and the traditional method respectively, and the graph of the image data is drawn according to the calculation results. See Fig. 2 for details.

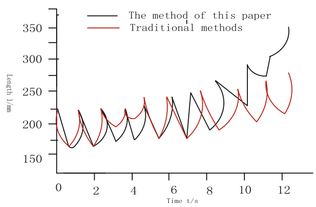

Fig. 2. Data graph

From Fig. 2, it can be seen that the curves drawn according to the results of this method are more consistent with the original ones, while the curves drawn according to the traditional algorithms are less similar to the original ones, that is to say, this algorithm is more accurate for the image data extraction than the traditional algorithm.

Image enlargement with a horizontal coordinate of 3.5 to 4 and a vertical coordinate of 210–235 in Fig. 2, observe the pixel coordinates and the curves obtained after calculation. See Fig. 3 for details.

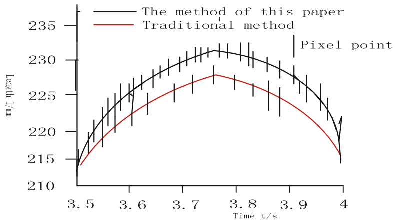

Fig. 3. Comparison of local large map and pixel points

From Fig. 3, it can be seen that the data extraction algorithm designed in this paper can take the mean value at the pixels of the multi Y value, and can obtain more pixels and more complete representation of the image data. Therefore, this method is more complete than the traditional method.

The results of this method, the error between the calculated results and the actual values of the traditional methods are shown in Fig. 4.

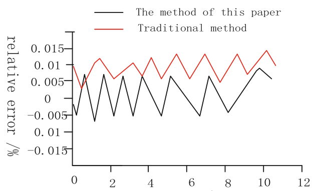

Fig. 4. Error contrast diagram

As can be seen from Fig. 4, the relative error of the calculation results in this method fluctuates in the range of −0.7%–+1.2%, and the mean of the relative error is 0.0018, indicating that the overall deviation is 0.18%; the standard deviation is 0.0084, indicating that the deviation of the overall data is very small. The relative error of the traditional algorithm fluctuates in the range of 0.005%–+1.5%, and the mean value of the relative error is 0.0068, indicating that the overall deviation is 0.68%, and the standard deviation is 0.0164, which indicates that the deviation of the overall data is larger. That is, compared with the traditional methods, the accuracy of the calculation results is higher and the extracted data are more complete.

Many experimental results show that the data extraction integrity and accuracy of this method are more than 45% of the traditional method, so this method is more suitable for the extraction of vectorization data of complex linear pixel images.

4 Concluding Remarks

In this paper, a new method of image data extraction is proposed. The method is based on the image preprocessing algorithm and image segmentation algorithm. By removing noise and image two value calculation, the accuracy of the calculation results is improved. The test data show that the accuracy of this method is improved by about 45% compared with the traditional method, and it has high effectiveness. It is hoped that this study can provide useful help for the vectorization of complex linear image data.

References

1. Chang, J., Liu, H., Zhang, Y.: An image feature extraction algorithm based on big data. Mob. Commun. **41**(4), 79–83 (2017)
2. Jiang, D.: A data extraction method based on image recognition technology and its mobile terminal: CN106372198A[P]. **1**(01), 10–11 (2017)
3. Yu, X., Chen, E., Ji, P., et al.: Data collection and clustering analysis of crowdsourcing images. Geospatial Inf. **15**(11), 16–17 (2017)
4. Tang, J.: Design of image acquisition and processing system. Sci. Consult **11**(44), 109–111 (2017)
5. Li, L.: Design of high-speed acquisition system for fuzzy image information data. Mod. Electron. Tech. **40**(8), 110–113 (2017)
6. Duan, S., Zhu, F., Yan, X.: Study of multi-window binarization algorithm for image processing. Comput. Eng. Appl. **53**(17), 212–217 (2017)
7. Cao, Y., Baojie, X.V., Xiaoli, X.V., et al.: Two value research of image based on improved Bernsen algorithm. Plant Maint. Eng. **14**(18), 26–28 (2017)
8. Zhou, L., Jiang, F.: A survey of image segmentation methods. Appl. Res. Comput. **34**(7), 1921–1928 (2017)
9. Liang, J., Liang, L.: Optimization of feature data extraction in large data environment. Comput. Simul. **11**(12), 345–348 (2017)
10. Liu, H., Yang, L., Hou, X., et al.: An improved fuzzy c-means algorithm for image segmentation. J. Zhengzhou Univ. (Nat. Sci. Ed.) **49**(2), 66–71 (2017)