Improved Ant Colony Algorithm Based on Binarization in Computer Text Recognition

Zhen Li
School of Forensic Science and Technology, Criminal Investigation Police University of China, Shenyang, 110854, China

***Abstract*—Pheromones, path selection, and probability transfer functions are the main factors that affect the performance of computer text recognition. The path selection function is the most important factor affecting the recognition rate. In response to the difficulties in path selection and slow algorithm convergence in the text recognition, an edge detection algorithm based on improved ant colony optimization algorithm is proposed. The strong denoising performance of the ant colony optimization algorithm reduces the interference of textured backgrounds. The edge extraction effect is analyzed in the connected domain to overcome complex effects. Finally, the improved Otsu binarization algorithm is used to recognize the text. According to the results, the proposed method could effectively preserve the edge information of characters in images. The positioning effect of the text area was good. The accuracy rate reached around 85%. The tuned threshold improved the binarization effect. The text recognition rate of the improved ant colony algorithm proposed in the research has generally reached 80%, with good text positioning accuracy and recognition rate, which has great practical significance in computer text recognition.**

Keywords—Binarization; ant colony algorithm; text recognition; edge detection; Otsu algorithm

I. Introduction

The core of computer text recognition is to recognize characters. Therefore, binarization of character images is the prerequisite and foundation for it [1-2]. The high degree of deformation and discontinuity in text images makes binarization techniques in image processing very difficult. Most existing character recognition algorithms analyze character images to determine the corresponding character category for each character. The main steps include image processing, feature extraction, character segmentation, and character recognition. The feature extraction part is the most important part of the entire algorithm. Its main task is to transform the text image into a series of feature vectors for text recognition based on the analysis of the text image. In the text recognition, the image binarization quality directly affects the quality of text recognition result. Therefore, how to achieve binarization is a key link in the entire text recognition process [3]. The existing binarization algorithms include maximum inter class variance method, threshold segmentation method, and threshold division method. These algorithms have certain limitations in practical applications. They cannot meet the binarization requirements of text and images in complex backgrounds [4-6]. Due to factors such as pheromone concentration, path selection, and probability transfer function, there are redundant pheromones and invalid paths in text images before binarization, resulting in noisy or blurry areas in the binarized text image [7-8]. To obtain high-quality binarization results, an adaptive method must be used to smooth the binarized image. However, the smoothing methods used in existing algorithms to some extent increase the noise points or blurry areas of text images. In the multimedia information retrieval, the image and video search in the search engine is still based on keywords, and the image is manually annotated first. When the user enters the keyword to search, in fact, the search engine only maps the results to the corresponding pictures. However, for complex and changeable video and various kinds of text, these methods have limitations, and are difficult to achieve a practical application level. In order to extract text in multimedia quickly and effectively, a method of "image search" is proposed, that is, by extracting multi-dimensional information such as texture, color and shape of the target image, the most similar image from the image library is find according to a specific algorithm. Texts are extracted from the video stream to realize content-based video retrieval.

The improved ant colony algorithm (ACA) is combined with the edge detection algorithm to solve the problem of slow convergence and local optimal solution. In the binarization stage, an improved Otsu binarization algorithm is proposed on the basis of the traditional binarization algorithm, and the corresponding binarization threshold is obtained by using the traditional Otsu algorithm. The threshold is fine-tuned on the basis of the traditional algorithm to improve the binarization effect.

The article conducts research from six sections. Related works is given in Section II. Section III is a review of research on improved ACA in computer text recognition. Section IV constructs the text recognition method based on improved ACA. Section V is to verify the performance of the proposed method. Section VI is the conclusion.

II. Related Works

ACA is extensively applied in combinatorial optimization problems, such as travel salesman problems, vehicle routing problems, graph coloring problems, and network routing problems. To promote economic development, many scholars have conducted research on this. Yi et al. proposed an improved ACA for task scheduling problems in information physics systems. The adaptive and mutation strategies were adopted to reduce solution time and accelerate the convergence of information physics systems. After numerical simulation, the improved algorithm could effectively improve the local optimization ability. It had good adaptability and stability [9]. To solve the long time consumption and multiple intermediate nodes in traditional path guidance methods in path planning results, Tang et al. optimized the potential field of ant colony from three aspects: potential field function, pheromone update process, and heuristic function. The ability to avoid obstacles and optimize was enhanced. The results showed that this method had fewer midpoints and the shortest path on each route [10]. Zhu et al. fused the artificial potential field method with the ant algorithm to overcome the slow convergence speed and local optimal problem of the ACA. The induction heuristic factor was used to dynamically adjust the state transition law, resulting in higher global search ability and faster convergence speed of the algorithm. After verification, this method could effectively determine whether a collision is occurred and take obstacle avoidance measures [11]. Yu et al. introduced a new heuristic clustering algorithm called co-evolutionary chain to improve the accuracy and stability of ACA. Ant colony clusters were divided for the balance of convergence and speed. The combination of group co-evolution and link dimension reduction improved the diversity and stability, separating it from the local optimal problem [12]. Hwang et al. proposed an ACA bidirectional Long Short Term Memory (LSTM) network model. The existing physiological signal information was fully utilized. The ACA was applied to find the optimal emotion recognition features, improving the emotion recognition ability of existing LSTM cell states. The final results indicated that the model had good valence performance [13].

Text recognition is widely used in daily life. Currently, many applications have been put into practice, such as documents conversion, license plate recognition, photo search, target translation, etc. Guptha et al. proposed a new deep learning based automatic character recognition model to address the difficulty of handwritten character recognition. Gaussian filtering and tilt detection techniques were used for preprocessing handwritten images. Projection contour and threshold segmentation techniques were used to segment individual lines and characters from denoised images. Finally, characters were classified using the LSTM [14]. Liu et al. developed a recognition algorithm to address the low character recognition efficiency and accuracy. A feature model was developed by combining histogram Gabor features with grid level features. Then, a deep belief network was applied to train the feature model. Finally, a probability model was used to judge the recognition symbols. After verification, it had higher accuracy and better performance [15]. According to the image information obtained by a 3D camera, Alam et al. proposed a new method for character recognition based on finger joint tracking system. The distance between the tip of the thumb and the joint of another finger was used for calculation. The Euclidean distance threshold and geometric slope technology were used to recognize numbers, letters, characters, special keys, and symbols. After verification, the overall recognition accuracy was over 90%. The recognition time for each character was less than 60 milliseconds [16]. In response to the low accuracy of character recognition in natural scenes, Chandio et al. extracted image features by cutting characters. Then, the obtained features were transmitted to the machine learning classifier for classification and character recognition through directed gradient histograms. The accuracy of this method reached 78.52% [17]. Lee et al. proposed a real-time character recognition algorithm. Based on the architecture of an improved local binary mode shallow depth convolutional neural network (CNN), it combined the manual feature preprocessing and character learning in CNN supervised advanced functions. Networks with different depths were applied for learning. The learned features were used for classification. The algorithm had good performance [18].

In summary, researchers have proposed many methods to improve the accuracy of computer text recognition. They have also achieved certain results. However, accuracy and efficiency are still lacking. Therefore, by improving the ACA based on binarization, it is expected to quickly and effectively increase the accuracy of text recognition.

III. Text Recognition Method Based on Improved ACA

A text recognition method based on improved ACA is proposed, which optimizes the edge detection algorithm and improves the accuracy of character region localization in images. Then, the Otsu algorithm is improved to find the optimal binarization threshold and enhance the binarization effect.

A. Character Localization Based on ACA

ACA is a simulation optimization method that mimics ants' foraging behavior. It has been widely applied in practical applications, especially in combination optimization, which has achieved great achievements. This algorithm utilizes pheromones to remember the best route, thereby strengthening the route and finding the best one. The study transforms the problem of text recognition into solving the optimal path selection problem. The improved ACA is applied to solve linear optimization problems. The characteristics of the optimal solution are determined based on ACA. The movement path of ants between pixels in the image is used to represent the edges of the image. When an ant moves from one pixel to another, it leaves pheromones along the way (the release rules of pheromones include a series of information such as image gradient and image color). These heuristic information are used to determine pixel regions with obvious edge features between the paths passed.

Edge detection has important implications in image processing. Its results directly affect image detection and content recognition [19]. ACA is used for edge detection of text images. The movement path of ants between pixels is used to represent the edges of the image. These heuristic information can determine the pixel regions with obvious edge features between the paths passed. When the initial value of pheromone concentration $\tau_{i,j}$ is not 0 and it moves, it will volatilize with a volatilization rate ρ over time. The relationship is shown in Eq. (1).

$$\tau_{i,j}(t+1) = (1-\rho)\tau_{i,j}(t) + \sum_{k-1}^{m} \Delta\tau_{i.j}^{k} \tag{1}$$

In Eq. (1), $\rho \in (0,1]$. $\Delta\tau_{i.j}^{k}$ represents the amount of pheromones added during ant movement, as shown in Eq. (2).

$$\Delta\tau_{i,j}^{k}=\begin{cases}\dfrac{\eta_{i,j}}{255}, k\in(i,j)\cup\eta_{i,j}>T\\ 0,\ \text{Or else}\end{cases} \tag{2}$$

In Eq. (2), η represents heuristic information. T represents the threshold of binarization, which affects pheromone map updates and detection. The transfer probability for ant movement is expressed in Eq. (3).

$$p_{(i_0,j_0)(i,j)}^{n}(i,j)=\frac{\left(\tau_{i,j}^{n-1}\right)^{\alpha}\left(\eta_{i,j}\right)^{\beta}}{\sum_{(i,j\in allowed_k)}\left(\left(\tau_{i,j}^{n-1}\right)^{\alpha}\left(\eta_{i,j}\right)^{\beta}\right)} \tag{3}$$

In Eq. (3), (i,j) represents a pixel. α represents the pheromone heuristic factor. A higher value indicates a greater likelihood of repeated searches. β stands for the expected heuristic factor. The numerical value determines the probability of ants choosing the local shortest path. $allowed_k$ represents the pixels that ants can pass through.

The implementation process of the edge detection ACA is displayed in Fig. 1. Firstly, initialization is performed. On this basis, each ant generates heuristic information based on the pheromone values of adjacent pixel points. At the same time, the transition probability is calculated. The pixel points that the next path will experience are determined based on the transition probability. The pheromone value is updated after entering a new pixel. When the ant colony moves to a new pixel position, it can update the overall pheromone map. When searching for paths, taboo table permissions are introduced to avoid ant colonies from repeatedly reaching the same pixels on the same path. If the next pixel entered by the ant colony is empty, it is randomly assigned to a pixel data to start over. Otherwise, the algorithm ends and the input grayscale image is converted into a binary image.

In the edge detection, ACA always has the slow convergence speed and the problem of local optimal solutions. The main difficulties in algorithm convergence are as follows. The distribution of initial points is random, without considering initial pheromones, resulting in slow search speed, high computational complexity, and long time consumption [20]. If it is trapped in a local optimal solution, it will cause the search process to pause and make it difficult to obtain a global optimal solution. Traditional ACAs do not limit the initial position of ants. There is no restriction on the initial layout of ants, but a random selection method is used. Therefore, ants continuously iterate during the search process, ultimately obtaining the optimal path [21]. However, using a random distribution method can lead to a large number of ants blindly searching under uncertain boundaries, thereby reducing search efficiency and lacking targeted edge detection. In view of this, a new 5×5 neighborhood based ACA is proposed to address the problems in ACA. This algorithm can better reflect the true boundary characteristics and effectively achieve the optimal localization of ant colonies.

The method for obtaining gradient information is as follows. The pixel points in the $allowed_k$ image are set to (i,j). Their grayscale value is $f(i,j)$. Then the grayscale gradient $\psi_T(i,j)$ of the pixel can be expressed as Eq. (4).

$$\psi_T(i,j)=\frac{\sqrt{\left[f(i+1,j+1)-f(i-1,j-1)\right]^2+\left[f(i+1,j-1)-f(i-1,j+1)\right]^2}}{255} \tag{4}$$

The gradient information of the image is used to extract pixels with significant differences. Meanwhile, based on pixel (i,j), 5×5 neighborhoods are established, as shown in Fig. 2. The neighborhood is divided into two regions according to different directions θ. Four directional angles are used to ensure edge diversity.

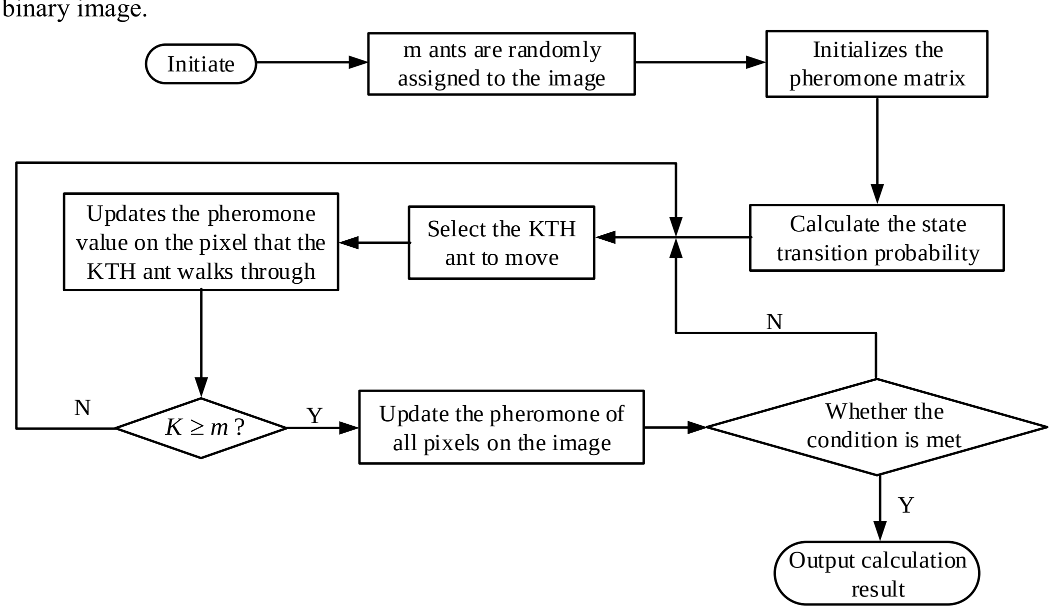


Fig. 1. Flowchart of edge detection based on ACA.

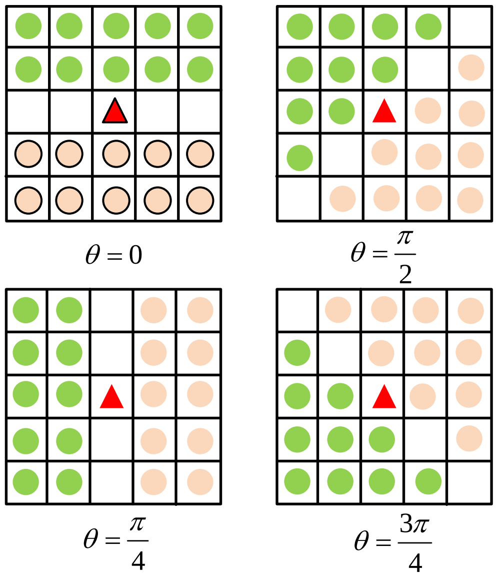


Fig. 2. Regional gray mean difference.

The mean value $B_{\theta_n}^{r_m}$ of regional grayscale can be expressed as Eq. (5).

$$B_{\theta_n}^{r_m} = \frac{\sum_{i,j \in D_m} f(i,j)}{N(i,j \in D_m)}, m = 1,2; n = 0,1,2,3 \tag{5}$$

In Eq. (5), $N(i,j \in D_m)$ represents the total pixels in the region D_m . θ_n stands for the direction angle. The mean difference ΔB of regional grayscale can be expressed as Eq. (6).

$$\Delta B = \max\left(\Delta B_{\theta_n}\right) \tag{6}$$

Based on the difference in regional grayscale mean ΔB, the grayscale difference on both sides of the short line segment where pixel point (i,j) is located is significant. It indicates that the position of the short line segment may be the boundary of the image. The 5×5 neighborhood can generate more short line segments compared with traditional structures, thereby avoiding rough estimation of line segments in the 7×7 neighborhood and improving the accuracy of boundary information. After removing the influence of noise points, a fusion gradient $H(i,j)$ is constructed by combining the grayscale gradient with the average regional grayscale value, which has high edge detection accuracy. Eq. (7) displays the calculation method.

$$H(i,j) = x\psi_T(i,j) + y\Delta B(i,j) \tag{7}$$

In Eq. (7), x and y are all relationship coefficients, $x + y = 1$. To obtain the noise reduction effect and ensure the accuracy of edge computing, $x = 0.5$ and $y = 0.5$ are set as the weight of the difference between the gray gradient and the regional gray mean. By using stroke width transformation, the edge extracted image is transformed into a stroke width transform image containing stroke width information. Then it is integrated and classified based on stroke width information. The corresponding connected domains are extracted. On the left side of Fig. 3 is a typical stroke image. The red dots represent pixels and white dots represent the background. The boundary extraction of the original stroke map can obtain the stroke boundary map, as shown on the right side of Fig. 3.

B. *Text Extraction and Recognition Based on Binary Processing*

After accurately locating the position of characters in the image using edge detection ACA, the determined character area is segmented and recognized. In the template library, there are only single characters. The existing methods can only recognize single characters. Therefore, to accurately recognize characters, the image needs to be segmented [22]. To grayscale a color image, the grayscale of each point in the image is changed to 0 or 255. The processed color image is presented as a gray image. To binarize the obtained gray image, a specific algorithm is used to obtain a corresponding threshold. Then, the grayscale of each point in the image is compared with this threshold. It is divided into object and background. After image binarization, the grayscale of each pixel in the image is only 0 and 255, excluding other grayscale sizes. In the binarization, all points in the image with grayscale above the threshold are targeted, and the grayscale is set to 255. On this basis, points with a grayscale lower than this threshold are used as backgrounds. Their grayscale is set to 0. The binary segmentation method is shown in Eq. (8).

$$g(x,y) = \begin{cases} 255, f(x,y) \geq T \\ 0, f(x,y) \leq T \end{cases} \tag{8}$$

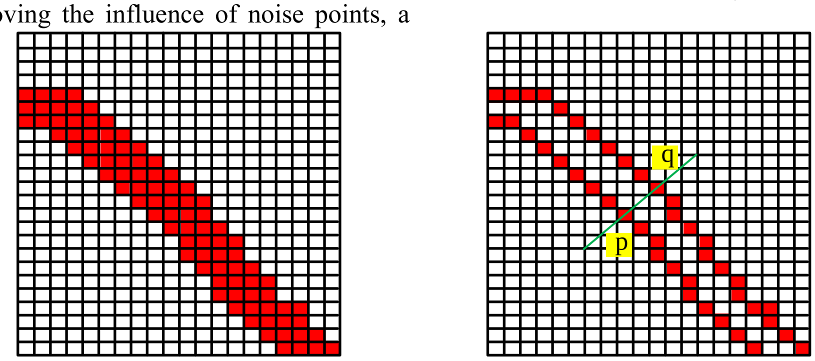


Fig. 3. Construction of edge point pairs.

In Eq. (8), $f(x,y)$ stands for the pixel value at the coordinate (x,y). $g(x,y)$ is the pixel value at the (x,y) coordinate in the image processed by the binarization method. A low-pass filter preprocesses the collected images to reduce or eliminate noise. The algorithm determines the optimal threshold, ensuring that the image can be effectively segmented into the target and background at the boundary of the optimal threshold. All points in the image with grayscale greater than the calculated threshold are set to 255, and points below the calculated threshold are set to 0. The obtained image only has two colors, black and white, which means dividing the image into target and non-target regions to achieve image binarization.

Binarization is a very important image processing technique. Selecting the appropriate threshold is an important step in image binarization. The Otsu algorithm uses a special discriminant function to determine the threshold size of binarization. The binarization algorithm divides the grayscale values of points on the entire image into two types at the threshold t. $C_0 = \{0,1,2,\cdots,t\}$ represents the background area [23]. If the total pixels in the character area after grayscale processing are N, and each pixel has the highest grayscale level L, the grayscale size of the entire image is within $[0, L-1]$. The pixel with a grayscale value of i is n_i. The probability calculation method for i is expressed as Eq. (9).

$$p_i = \frac{n_i}{N} \tag{9}$$

Assuming the threshold is T, the grayscale is $[0, T-1]$ for the C_0 region, and the grayscale is $[T, L-1]$ for the C_1 region, the probability of C_0 and C_1 occurring is shown in Eq. (10).

$$\begin{cases} W_0 = \sum_{i=0}^{T-1} p_i \\ W_1 = \sum_{i=T}^{L-1} p_i = 1 - W_0 \end{cases} \tag{10}$$

The grayscale mean of S and C_1 can be expressed as Eq. (11).

$$\begin{cases} u_0 = \sum_{i=0}^{T-1} i p_i \\ u_1 = \sum_{i=T}^{L-1} i p_i \end{cases} \tag{11}$$

The inter class variance σ^2 is displayed in Eq. (12).

$$\begin{cases} \sigma^2 = W_0 (u_0 - u)^2 + W_1 (u_1 - u)^2 \\ u = W_0 u_0 + W_1 u_1 \end{cases} \tag{12}$$

In Eq. (12), u represents the average grayscale. T increases in steps of 1 within the range of $[0, L-1]$. When σ^2 is the maximum, the corresponding T value is the optimal binarization threshold.

The conventional Otsu algorithm often has unsatisfactory results when conducting binarization to images with slight differences between the target and background. Moreover, there are many parts that belong to the key regions of characters that have not been extracted. This study improves the effectiveness of the algorithm by adjusting the size of the threshold. The threshold size that needs to be adjusted for binarization is directly proportional to the average grayscale. The relationship between the required adjustment size w and the average grayscale value E is shown in Fig. 4.

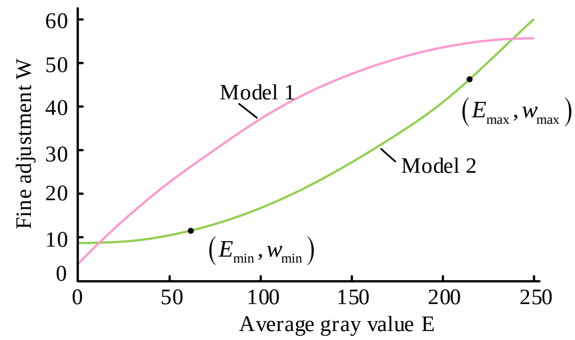


Fig. 4. Relationship between adaptive fine-tuning amount w and average gray value E.

In Fig. 4, the curve in the figure is an arc. Model 1 is a monotonically smooth curve fitting that is convex upwards. Model 2 is fitted with a monotonic smooth curve that is concave downwards. The downward concave curve fitting effect is better, that is, mathematical model 2. The following is a specific analysis for mathematical Model 2. Assuming (x_0, y_0) is the center of a circle, the relationship between x_0 and y_0 can be expressed as Eq. (13).

$$y_0 = -\frac{E_{max} - E_{min}}{w_{max} - w_{min}} \left(x_0 - \frac{E_{max} + E_{min}}{2} \right) + \frac{w_{max} + w_{min}}{2} \tag{13}$$

In Eq. (13), E_{max} and E_{min} respectively represent the maximum and minimum of the average grayscale values for a single column in the image. w_{max} and w_{min} represent the maximum and minimum of the required adjustment amount in the entire image, respectively. x_0 determines the curvature of a circle. If x_0 is small, then the curvature is small. If x_0 is large, then the curvature is large. To ensure a moderate curvature of the fitted curve, the value between $E_{max} - 50$ and $E_{min} - 60$ is the most appropriate for x_0. The equation for the final fitted circle is shown in Eq. (14).

$$(w - x_0)^2 + (E - y_0)^2 = (x_0 - E_{min})^2 + (y_0 - w_{min})^2 \tag{14}$$

Therefore, the relationship between the fine-tuning amount w and the average grayscale value E of any column can be expressed as Eq. (15).

$$w = -\sqrt{\left(x_0 - E_{\min}\right)^2 + \left(y_0 - w_{\min}\right)^2 - \left(E - y_0\right)^2 + x_0} \tag{15}$$

The relationship between fine-tuning amount and average grayscale value can be obtained from Eq. (15). On this basis, the traditional Otsu algorithm is improved using the above formula. The threshold after binarization is adjusted to achieve ideal segmentation results.

IV. The Effectiveness Analysis of Text Recognition Methods

The experiment was carried out on an Intel i5-8250U 1.6GHZ processor and 8GB of memory. The operating system was Windows 10. Matlab 2015b served as the operating platform. The algorithm performance was validated through the ICDAR2017RCTW database, including edge detection performance, text region localization performance, binarization performance, and image text recognition rate.

A. Analysis of Character Positioning Effect

The research was based on the ICDAR2017RCTW database. Color text images and texture physical background images that meet the requirements were used to construct a data set. The text recognition effect was compared. The model was trained using TensorFlow 1.3. The programming language was Python. The server configuration was NVIDIA TESLA P4 graphics card. Fig. 5 was one of the images in the data set. It was used for subsequent recognition analysis.

To demonstrate the advantages of the Mallat wavelet fast decomposition algorithm in license plate edge detection, this study compared the improved algorithm with traditional edge extraction algorithms. Fig. 6 showed the comparison results. Fig. 6 (a) displayed the edge extraction effect of traditional algorithms. The text area was successfully separated from other areas. There was a clear difference between the white rectangular area and other connected areas. The text area was longer than other areas. Fig. 6 (b) showed the edge extraction effect of the improved algorithm. It could effectively preserve the edges of characters in the image. However, the traditional algorithms were prone to losing edge information of character regions during edge detection.

Fig. 7 showed the results of using the proposed algorithm to locate text regions in the image. From the figure, the improved algorithm located a total of seven main text regions in the image, basically including all the text in the image. The text regions could be well distinguished, providing a significant non-interference edge extraction effect.

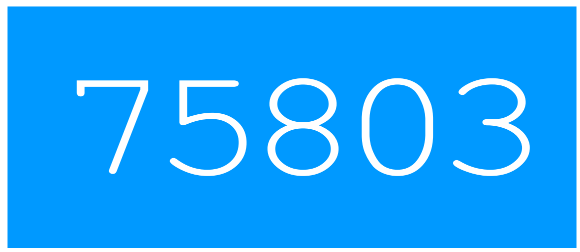
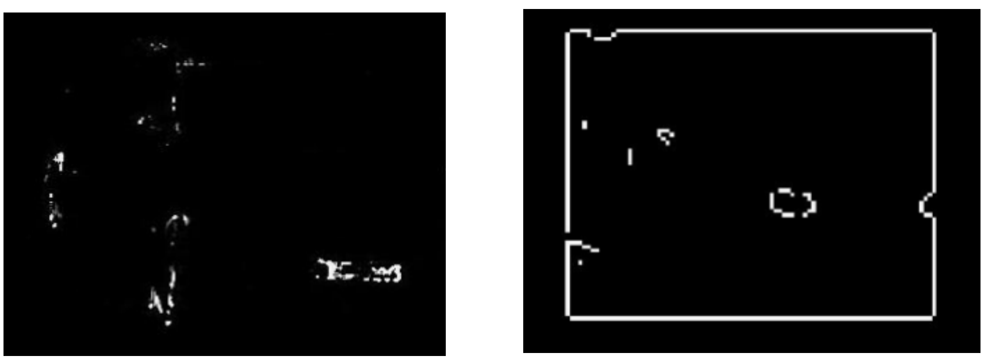


Fig. 5. Original image.



(a) Edge detection graph

(b) Edge detection graph based on ant algorithm

Fig. 6. Comparison of processing effects of edge detection algorithms.

Fig. 7. Character positioning effect.

In order to test the effectiveness of ACA for text localization, the research selected four typical images from the data set for text detection and localization, and analyzed them through five indexes including correct detection rate, missed detection rate, false detection rate, recall rate, and precision rate. The positioning results are shown in Table I. Among them, the number of key frames and the total number of text lines were the total character area of the image. The number of correct detected lines and the number of missed lines represented the positioning accuracy of the algorithm, and the recall rate and precision rate represented the recognition accuracy of the algorithm. From Table I, the text localization effect of the proposed method was good. The text recall rate reached 95%, and the accuracy rate was around 85%. The accuracy rate was relatively low due to complex background interference, which couldn't exclude all factors, but it is already higher than traditional detection methods. The extracted text features were targeted. The ACA was effective, and the regional posterior method also reduced the false detection rate.

To verify the advantages of the proposed algorithm in processing text information detection, the study compared the edge point pair construction effects of the two algorithms. The results were shown in Fig. 8. From the projection of Fig. 8 (a) and Fig. 8 (b), the text area was more prominent because the irrelevant backgrounds were removed. It could effectively improve the character recognition rate in complex backgrounds and filter out interference in the background, thereby extracting characters more efficiently. In summary, this method could effectively solve the image segmentation in complex backgrounds, effectively reducing the impact of non character regions on characters, and reducing the character stroke missing.

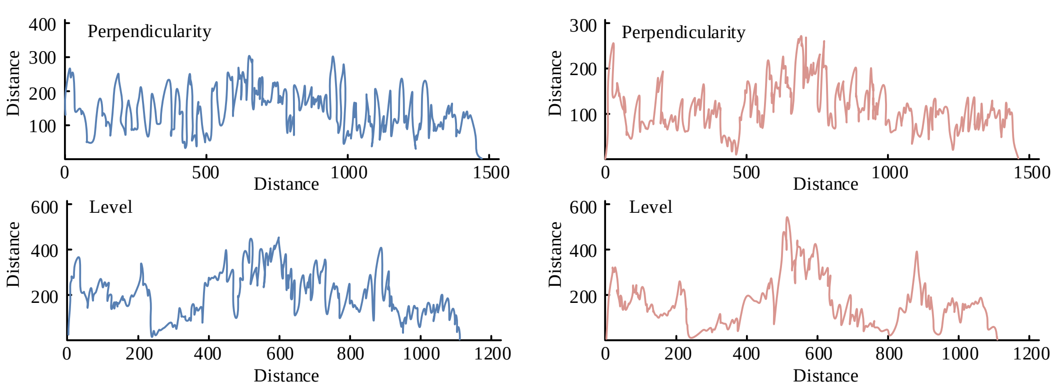
B. Analysis of Text Recognition Effect

The method of combining vertical and horizontal projection was used to segment the license plate area of vehicles. Taking Fig. 5 as an example, first, the license plate area was projected vertically. There were no peaks in the middle of each character. Based on this feature, the license plate was divided into individual characters to obtain a vertical text projection image, as shown in Fig. 9. From Fig. 9, the character area had more peaks, while the non-character area had no peaks. The area where the text was located had many peaks and valleys, and the peaks and valleys changed frequently. Therefore, the peak of the vertical projection curve had significant changes.

The study compared the binarization effects of text images. The binarized images obtained after processing with the Otsu algorithm were shown in Fig. 10. Fig. 10 (a) showed the Otsu binarization effect of intra class variance. In the figure, for color text images, the Otsu binarization effect of inter class variance was not very ideal. There were phenomena such as unclear characters in the text area. Fig. 10 (b) showed the Otsu binarization effect of intra class variance. Compared to the inter class variance algorithm, there was no significant improvement in the performance of color text images. Moreover, this algorithm was more complex than the inter class variance Otsu algorithm, and the computational efficiency significantly decreased. Fig. 10 (c) made appropriate adjustments to the threshold obtained by the Otsu algorithm. On the basis of the above threshold, the threshold was reduced by 15. In the figure, the binarization effect after reducing the threshold by 15 was better than before, especially for the binarization effect of the digital part.

TABLE I. CHARACTER EXTRACTION RESULTS

Index	Key frame count	Total lines of text	Correctly detect the number of rows	The missed rows	Number of misdetected rows	Recall factor	Precision ratio
Image 1	4	1	1	0	0	100.0%	100.0%
Image 2	23	18	17	1	3	94.4%	85.0%
Image 3	14	10	10	0	3	100.0%	77.0%
Image 4	24	15	14	1	2	93.3%	87.5%
Total	65	44	42	2	8	95.3%	84.0%



(a) SWT algorithm projection diagram (b) Improved algorithm projection diagram

Fig. 8. Comparison of algorithm effects before and after improvement.

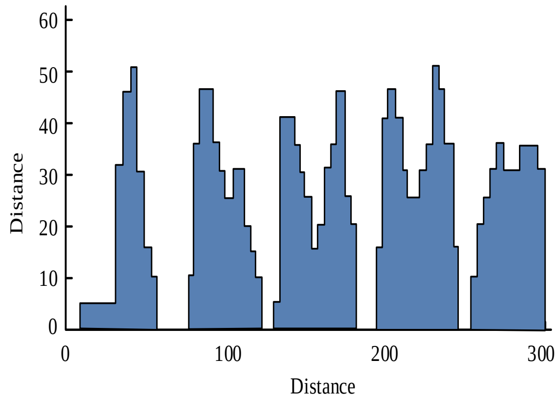


Fig. 9. Projection of license plate in vertical direction.

(a) In-class variance Otsu binarization rendering

(b) Otsu binarized rendering of inter-class variance

(c) Threshold reduction by 15 binary effect

Fig. 10. Comparison of binarization effects.

The images in the data set were located and the text area map was binarized, followed by text recognition. The recognition results were displayed in Table II. The recognition rate of Image 1 was 70.27%, which may be due to the low image clarity. The text recognition rate of the improved ACA proposed in the research generally reached over 80%, verifying the effectiveness of the binarization method. There was still room for further improvement in recognition rate. Overall, it had certain advantages compared to other algorithms.

TABLE II. TEXT RECOGNITION RESULTS

Index	Total number of characters in the location area	Correct word count	Recognition rate
Image 1	37	26	70.27%
Image 2	356	293	82.3%
Image 3	261	228	87.36%
Image 4	348	291	83.62%
Total	1002	836	83.63%

V. RESULTS AND DISCUSSION

Shot detection is the premise of key frame extraction (i.e., subtitle frame), and subtitle frame detection is the basis of subsequent subtitle positioning and recognition [24]. The wavelet decomposition algorithm is used to divide the video image into high frequency part and low frequency part. The high frequency part is the region where the characters are located. By comparing the wavelet decomposition algorithm with the traditional Sobel operator edge detection algorithm, it can be seen that the text region obtained by the wavelet fast decomposition algorithm is more complete. The localization of text area is divided into two processes, namely preliminary localization and precise localization. The character part is extracted from the whole video image by using the open and close operation function of MATLAB toolbox. The precise positioning of the character part refers to removing the edge of the extracted image character region, leaving only the part containing characters. With horizontal and vertical projection methods, using the character area projection gray histogram crest and trough continuous jump characteristics, the character is accurately located. As described in the review, the proposed algorithm can accurately determine the region where the characters are located in the video image, remove other background regions in the image, and the processed character region is clearly visible, and the noise is significantly reduced, which is conducive to the next step of character research.

After the text of the target area is located, in order to extract effective text information, the semantic part of the text needs to be extracted from the complex background, and the threshold segmentation of the text field is carried out again [25]. Subtitle extraction is actually a binary process, which requires that the resulting graph can basically maintain the original character features, without blank, hollow, and broken. Character extraction is mainly divided into two parts, including picture binarization and character segmentation. In this paper, an improved Otsu algorithm is proposed for the binarization of character images, and the threshold is fine-tuned on the basis of the traditional Otsu algorithm. The amount of fine tuning is determined by the size of the average gray value. There is a curve function relationship between the average gray value and the fine-tuning amount [26]. Comparing the improved binarization effect with the traditional binarization effect, the improved binarization effect is significantly improved. The character area is clearer, the contrast with the background area is significantly improved, and the improved Otsu algorithm has higher processing accuracy under different lighting conditions. In the character segmentation stage, the research combines vertical projection and horizontal projection, and uses the characteristics of continuous peak value of character area and no white pixels in non-character area to remove the background area around characters in the image, so as to accurately locate the region where characters are located. In the character recognition stage, a character recognition algorithm based on template matching is used. The results showed that the text recognition rate of the improved ACA generally reached more than 80%, which verifies that the binarization method in this paper is effective.

VI. Conclusion

ACA has the advantages of random search, fast convergence speed, and strong local search ability. This algorithm is used to binarize text images. The study combines the pheromone update mechanism and path selection mechanism in traditional ACA. The sensitivity to initial pheromone concentration and path selection strategies is improved. This to some extent enhances the local and global search capabilities. Then, the threshold obtained from the traditional Otsu algorithm is adaptively fine-tuned. The functional relationship between the fine tuning amount and the average grayscale value is found, completing the binarization of the image. The improved ACA could effectively preserve the edges of characters in images. Traditional algorithms were prone to losing edge information of character regions during edge detection. The text localization effect based on the improved ACA was good, with a text recall rate of 95% and a precision rate of about 85%. The regional posterior method also reduced the error detection rate. The binarization effect after reducing the threshold by 15 was better than before, indicating that fine-tuning the threshold improved the binarization effect. The text recognition rate of the improved ACA proposed in the research reached over 80%, verifying the effectiveness of the binarization method. Although the ACA is used in this study to avoid the positive and negative sample selection training of the classifier, the iterative nature of the ACA is bound to reduce the efficiency of the algorithm. In the future, it is necessary to study how to improve the efficiency of the algorithm, or develop an alternative method.

Acknowledgment

The research is supported by Special project of basic work for strengthening police with science and technology of the Ministry of Public Security. Project number: No.2020GABJC03; Science and Technology Program Project of Ministry of Public security; Basic Scientific Research Projects of Education Department in 2023; Project number: No. JYTZD2023146.

References

[1] Cao H, Wu Y, Bao Y, Feng X, Wan S, Qian C. UTrans-Net: A Model for Short-Term Precipitation Prediction. Artificial Intelligence and Applications. 2023, 1(2): 106-113.

[2] Wu Z, Zhao Y, Zhang N. A Literature Survey of Green and Low-Carbon Economics Using Natural Experiment Approaches in Top Field Journal. Green and Low-Carbon Economy, 2023, 1(1): 2-14.

[3] Marchi S D, Erb W, Marchetti F, Perracchione E, Rossini M. Shape-Driven Interpolation with Discontinuous Kernels: Error Analysis, Edge Extraction, and Applications in Magnetic Particle Imaging. SIAM Journal on Scientific Computing, 2020, 42(2):472-491.

[4] Xiao H, Xiao S, Ma G, Li C L. Image Sobel edge extraction algorithm accelerated by OpenCL. Journal of supercomputing, 2022, 78(14):16236-16265.

[5] Lin F, Yu Z, Jin Q. Semantic Segmentation and Scale Recognition–Based Water-Level Monitoring Algorithm. Journal of Coastal Research, 2020, 105: 185-189.

[6] Zhang J, Li C, Rahaman M M, Yao, Y D, Ma P L, Zhang J H, Zhao X, Jiang T, Grzegorzek M. A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches. Artificial intelligence review, 2022, 55(4):2875-2944.

[7] Usman A M, Abdullah M K. An Assessment of Building Energy Consumption Characteristics Using Analytical Energy and Carbon Footprint Assessment Model. Green and Low-Carbon Economy, 2023, 1(1): 28-40.

[8] Gan J, Wang W, Lu K. Compressing the CNN architecture for in-air handwritten Chinese character recognition - ScienceDirect. Pattern Recognition Letters, 2020, 129(6): 190-197.

[9] Yi N, Xu J, Yan L, Huang L. Task optimization and scheduling of distributed cyber–physical system based on improved ant colony algorithm. Future Generation Computer Systems, 2020, 109(23): 134-148.

[10] Tang Z, Ma H. Path guidance method for unmanned vehicle based on improved potential field ant colony algorithm. International Journal of Vehicle Design, 2022, 89(2): 84-97.

[11] Zhu S, Zhu W, Zhang X, Cao T. Path planning of lunar robot based on dynamic adaptive ant colony algorithm and obstacle avoidance. International Journal of Advanced Robotic Systems, 2020, 17(3): 4149-4171.

[12] Yu J, You X, Liu S. Dynamically induced clustering ant colony algorithm based on a coevolutionary chain. Knowledge-based systems, 2022, 251(5): 336-351.

[13] Hwang W, Kang D, Kim D. Brain lateralisation feature extraction and ant colony optimisation-bidirectional LSTM network model for emotion recognition. IET Signal Process, 2021, 16(1): 45-61.

[14] Guptha N S, Balamurugan V, Megharaj G, Sattar K N A, Rose J D. Cross lingual handwritten character recognition using long short term memory network with aid of elephant herding optimization algorithm. Pattern recognition letters, 2022, 159(7): 16-22.

[15] Liu Z, Pan X, Peng Y. Character Recognition Algorithm Based on Fusion Probability Model and Deep Learning. The Computer Journal, 2020, 64(11): 1705-1714.

[16] Alam M S, Kwon K C, Kim N. Implementation of a Character Recognition System Based on Finger-Joint Tracking Using a Depth Camera. IEEE transactions on human-machine systems, 2021, 51(3): 229-241.

[17] Chandio A A, Leghari M, Kakepoto I, Leghari M, Hussain G. Multilingual Natural Scene Character Classification and Recognition with Histogram of Oriented Gradients. Fourrages, 2020, 243(9): 86-94.

[18] Lee S H, Yu W F, Yang C S. ILBPSDNet: Based on improved local binary pattern shallow deep convolutional neural network for character recognition. IET image processing, 2022, 16(3): 669-680.

[19] Liu Y, Gao T, Song B, Huang C. Personalized Recommender System for Children's Book Recommendation with A Realtime Interactive Robot. Journal of Data Science and Intelligent Systems, 2023, 2(1): 1-6.

[20] Liu Y, Ma X, Li X, Zhang C. Two-stage image smoothing based on edge-patch histogram equalisation and patch decomposition. IET Image Processing, 2020, 14(6):1132-1140.

[21] Kim B, Kim J, Ye J C. Task-Agnostic Vision Transformer for Distributed Learning of Image Processing. IEEE Transactions on Image Processing, 2023, 32(20):2023-218.

[22] Garai S, Paul R K, Kumar M, et al. Intra-Annual National Statistical Accounts Based on Machine Learning Algorithm. Journal of Data Science and Intelligent Systems, 2023, 2(2): 12-15.

[23] Huang B, Lin J, Liu J, Chen J, Zhang J, Hu Y, Chen E, Yan J. Separating Chinese Character from Noisy Background Using GAN. Wireless Communications and Mobile Computing, 2021, 54(1): 1-13.

[24] Ding C, Zhou A, Liu X, Ma X, Wang S G. Resource-aware Feature Extraction in Mobile Edge Computing. IEEE Transactions on Mobile Computing, 2020, 21(1):321-331.

[25] Clausner C, Pletschacher S, Antonacopoulos A. Flexible character accuracy measure for reading-order-independent evaluation - ScienceDirect. Pattern Recognition Letters, 2020, 131(5): 390-397.

[26] Wang R, Cao W, Wu S, Jia M, Wang X P. Optical character correction of large-curvature annular sector text in polar coordinate system. Pattern recognition letters, 2023, 167(5): 157-163.