A New Method for Text Location in News Video Based on Ant Colony Algorithm

Ming Jiang[1,2] , Taotao Zha[1,2] , Xingqi Wang[1,2] , Jingfan Tang[1,2] , Chunming Wu[3]

(1. Institute of Software and Intelligent Technology, Hangzhou Dianzi University, Hangzhou, 310018, China)
(2. Zhejiang Provincial Engineering Center on Media Data Cloud Processing and Analysis, Hangzhou Dianzi University, Hangzhou, 310018, China)
(3. College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China)

Email: jmzju@163.com, zhataotao1989@126.com, xqiwang@163.com, tangjf@hdu.edu.cn, wuchunming@zju.edu.cn

***Abstract*—Text in video is a very compact and accurate clue for video indexing and summarization. The paper presents a new method for text location in news video with ant colony algorithm. Three features of characters are extracted as a basis for the formation of heuristic function. In order to balance the weight of the three features, three functions are introduced to transform them. The ants will be randomly put on sub-blocks of video frames for searching text areas. Therefore, ants would leave pheromone in each sub-block. After the ant colony algorithm is finished, it produces a pheromone matrix. By binarizing the pheromone matrix, the text blocks can be located. The result has proved that the binarization method proposed in this paper is more accurate than otsu's method. At last, to reduce the false detection rate, the different directions of edge intensity ratio of text areas are computed, as the real text areas' edge intensity ratio is much smaller than the false one.**

***Index Terms*—video indexing, text location, ant colony algorithm, binarization, edge intensity ratio.**

I. INTRODUCTION

With the widely application of the digital information technology and multimedia technology, all walks of life have a lot of digitized information. In the face of massive video data, how quickly and easily to obtain the required information has become a research focus in the field of image analysis, data mining etc. To extract the text from image, first the text area should be located in image with complex background. Existing text detection techniques is categorized into four types: the method based on edge, the method based on region, the method based on texture, the method based on machine learning. *The method based on edge*[1-2] takes the advantage of the feature that text areas always have high sharpness, whose edge components is more than the non-text area. This method has a high speed as well as high false detection ratio. *The method based on texture*[3-4] takes the characters as a special kind of texture. The characters are usually made up by many fine stroke, hence the area with more strokes are the area rich in texture, which can be used to decide the text blocks. *The method based on region*[5-6], which is about connecting the similar or the same color's pixels and filtering these areas with knowledgeable rules, has high speed in processing but not fitting the case that text words have different colors. There are many problems to locate the text in video, for example the different size of characters, the diversity of character styles, the complex of the background. *The method based on machine learning*[7] is a new algorithm that can deal with these unstable factors. But the only disadvantage is that the choosing of the training samples and testing samples have a great affect on the result, while we haven't had a good system produce the reasonable samples until now.

According to the analysis above, most of the recent methods have their own advantages as well as disadvantages. To solve the problem of the text location in video with complex background in a better way, this paper proposes a method based on ant colony algorithm. The ant colony algorithm has its advantages of intelligent searching, global optimization, robustness, positive feedback, distributed computing and so on. It has reduced the influence of human factors because this method neither needs to set experiential threshold nor needs to choose test samples. Three typical features about characters are selected as the reason to update the pheromone in ant colony algorithm. The pheromone matrix has the same size of the test image. By binarizing the pheromone matrix with the new binarization algorithm, the text blocks and non-text blocks in an image can be divided. The effectiveness of this method has been proved by the experimental results.

II. OUR METHOD

The primary process of the method of text location using ant colony algorithm is as follows:

doi:10.4304/jcp.9.5.1090-1095

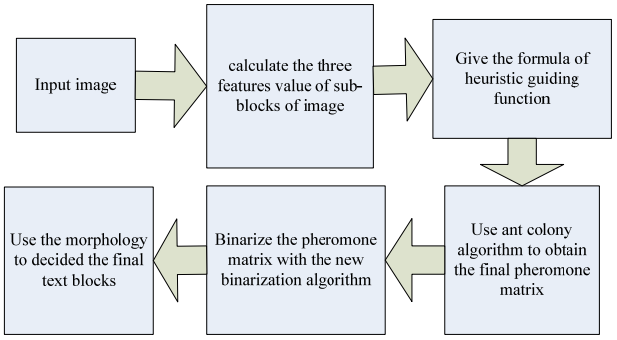


Fig 1. The process of ant colony algorithm

A. The Three Features of Characters

1. Gray-level Co-occurrence Matrix

GLCM, a joint distribution of two pixels' grayscale in different directions within an image, is a common method to analysis texture, which can well reflect the correlation law of grayscale about texture. According the 'Textural Features for Image Classification' written by Haralick[12], there are 11 characteristics of GLCM to describe texture. Here one of these, the variance whose value indicates the changing speed of texture and the length of period of texture, is adopted in our formula. The experimental result has proved that it is more effective and efficient than other features.

The formula to calculate the variance is

$$\sum_{k=2}^{2N}\sum_{i=1}^{k-1}(2i-k)^2 P_\delta \,. \tag{1}$$

where N is the grayscale of image, P_δ is the GLCM of image.

2. Wavelet Transform

A single can be decomposed into sub-bands at various scales and frequencies through wavelet transform. In the case of images, the wavelet transform is useful to detect edges with different orientations. The wavelet transformation can be implemented using filter banks consisting of high-pass and low-pass filters. The application to an image consists of a filtering process in horizontal direction and a subsequent filtering process in vertical direction. For example, when applying a 2-channel filter bank (L: low pass filter, H: high-pass filter), four sub-bands are obtained after filtering: LL, HL, LH and HH. The three high-frequency sub-bands (HL, LH, HH) enhance at most edges in horizontal, respectively vertical or diagonal direction. Since text areas are commonly characterized by having high contrast edges, high valued coefficients can be found in the high-frequency sub-bands.

Gllavata[4] proposed 2 characteristics from wavelet transform in both sub-band HL, LH: variance and histogram variance of wavelet coefficient.

$$VarCoef = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}(Coef(i,j)-MeanCoef)^2 \tag{2}$$

$$HistVarCoef = \sum_{\min_{coef}}^{\max_{coef}}(Bin\_new(j)-Mean_{Bin\_new})^2$$
$$\begin{cases} if\ Bin(j)>0, Bin\_new(j)=j; \\ if\ Bin(j)=0, Bin\_new(j)=0; \end{cases} \tag{3}$$

where *Coef(i,j)* is the wavelet coefficient, *MeanCoef* is the average of all coefficients. *Bin(j)* is the *jth*'s number of wavelet coefficient; *M, N* is the number of the ranks of sub-band HL, LH. These two sub-band focus on the horizontal and vertical texture of the characters. As the great correlation between the variance and histogram variance, we here just pick up the variance of coefficient with higher value between sub-band HL and LH.

3. The Corners

The corners, existing in the edge or outline of object, can reflect the important information of the local image with little redundancy. The text words in image have more and intenser corners, especially Chinese characters. Text in video or image always has sharp contrast with the background in color and brightness. Therefore, the text area will contain more corners. Harris[13], using the method of differential operation and sub-matrix correlation, is a common corner detection algorithm. This algorithm has the advantage of simple calculation, extracting the feature points with predefined number and stable operator. This paper used Harris algorithm for corner detection.

B. Text Detection in Video with Ant Colony Algorithm

The ant colony algorithm has been used into edge detection of image that takes images as undirected graphs [14] .The principle of ant colony algorithm used in text blocks detection is based on it: Split the input image into blocks with the size of 8×8, then put the ants into these blocks randomly. Each ant will search the image blocks according to the intensity of the pheromone in blocks. At the end of the algorithm, the text blocks will have more pheromone than the non-text ones. The specific process can be summarized as the following four steps:

1) Initialization. Image I is taken as a graph, the splited blocks can be taken as nodes of graph. *m* ants are randomly put into nodes ; *m* is $\sqrt{\frac{M}{8}\times\frac{N}{8}}$, where *M, N* is the rows' and columns' number of *I*. Initialize parametric variables and the pheromone in each block. The pheromone τ should be set with a positive that tends to 0.

2) Path Finding. The probability of an ant move from the *ith*'s to the adjacent *jth*'s block is:

$$p_{(i,j)}(t)=\frac{\tau_{i,j}(t-1)^{\alpha}{\eta_{i,j}}^{\beta}}{\sum_{j\in\Omega_i}\tau_{i,j}(t-1)^{\alpha}{\eta_{i,j}}^{\beta}}. \tag{4}$$

where $\tau_{i,j}(t-1)$ is the intensity of pheromone in *jth*'s block next to *ith*'s in (t-1)th's time. $\tau_{i,j}(0)$ is initialized with 0.0001 here. $\eta_{i,j}$ is the heuristic information α, β are the relatively importance of pheromone and heuristic information. If β is high in value with respect to α, the algorithm will converge early. Here, we set $\alpha=10, \beta=0.1$. Ω_i is the set of the adjacent nodes of *i*. The definition of the heuristic guiding information will be given in section C.

3)Pheromone updating. In this paper, the pheromone will be updated in both locally and globally. In the local updating, the pheromone will be updated as formula (5) describes when the *kth*'s ant takes one step at *tth*'s time.

$$\tau_{i,j}(t)=\begin{cases}(1-\rho)\cdot\tau_{i,j}(t-1)+\rho\cdot\Delta\tau_{i,j}^{k} & \text{if } (i,j) \text{ visited by the kth ant}\\ \tau_{i,j}(t-1) & \text{else}\end{cases} \quad (5)$$

where ρ is an evaporation coefficient; $\Delta\tau_{i,j}^{k}$ has a relation with the heuristic guiding function, we here define $\Delta\tau_{i,j}^{k}=\eta_{i,j}$. After all ants finish their own travel, we update the global pheromone with (5).

$$\tau(t)=(1-\phi)\cdot\tau(t-1)+\phi.\Delta\tau \quad (6)$$

Where φ is the attenuation coefficient of pheromone.

4) Text Location. When all ants finish the iterations of travel, each block of image has a value of pheromone. If the value exceeds the threshold *T*, it will be taken as a text block. Otherwise, as the non-text block. The *T* is obtained by the method proposed in section D.

C. The Definition of Heuristic Guiding Function.

It can be seen from the above analysis, the function of heuristic information is to guide ants to choose the text blocks, it can be decided with 3 features talked above in section A. 3 windows are selected to compute the value of features. Their size are $16\times48, 48\times16, 16\times16$, where the experiment has shown that 16×48 is suitable to locate the horizontal characters, 48×16 is suitable to locate the vertical characters, while 16×16 is suitable to locate both directions, but the result is less effective than the formers. This paper is mainly focus on news video, where the most text areas are horizontal. Hence, the window-size is 16×48, step-size is 8, that means each sub-block of 8×8 size in the image has a different feature value. We respectively normalize the data of the 3 features. The statistical result has shown that the three

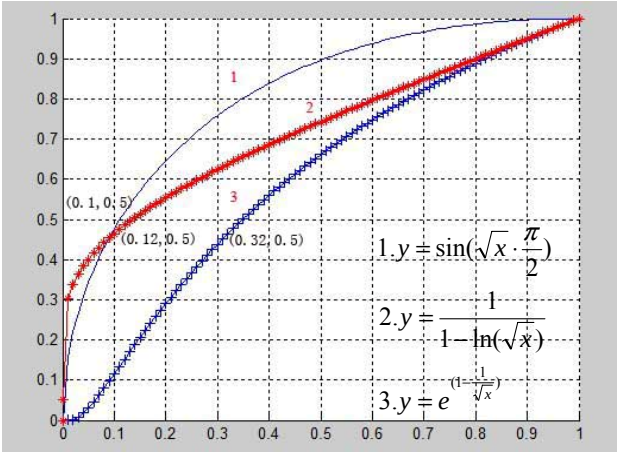


Fig 2 Coordinate graph of 3 conversion function

features have the different classification lines to split the text and non-text, which is respectively about 0.1, 0.12 and 0.32. If these data were directly used, the variance, whose value in (0.1, 1) can be taken as text, would have a greater impact on the result than corners with text recognition interval in (0.32, 1). To balance the weight of each feature, 3 conversion functions are made up to change their classification lines all to about 0.5.

It can be seen from the Fig 2 that the expected result can be achieved, through converting variance with function 1, converting wavelet coefficient with function 2, and converting corners with function 3. Therefore the heuristic guiding function is defined by (7):

$$\eta_{i,j}=\sin(\frac{\pi}{2}\cdot\sqrt{SD_{i,j}})\cdot\frac{1}{1-\ln(\sqrt{VarCoef_{i,j}})}\cdot e^{(1-\frac{1}{\sqrt{Corner_{i,j}}})} \quad (7)$$

where *SD* is the variance of normalization, *VarCoef* is the wavelet coefficient of normalization; *Corner* is the number of corners of normalization.

D. The New Binarization Method

There are many classic binarization algorithm like Bersen[16], LEVBB[17], OTSU[15]. The OTSU is a more popular one. It is a method called maximum variance between clusters which can automatic find a maximum variance between background and foreground to divide images into two parts. The otsu method considers the whole parts value, while isolated points would affect the final result. Here we proposed a new method with iterations, some isolated points can be removed in the first few iterations. The specific process is as follows:

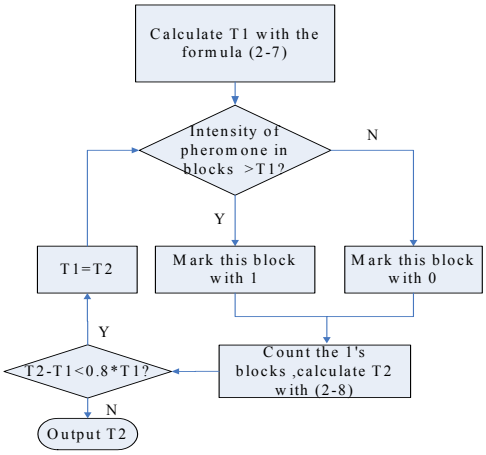


Fig 3 Process of the proposed binarization method

$$T1 = mean + 0.8 * std \ . \quad (8)$$

where *mean* and *std* is the mean and standard deviation of all blocks.

$$T2 = \theta \cdot (mean2 + std2) \ . \quad (9)$$

where *mean2* and *std2* is the mean and standard deviation of the blocks marked with 1. $\theta \in [0,1]$.

The algorithm can converge after only 2-3 iterations, so its speed is faster than the otsu algorithm. At end, the pheromone matrix is binarized with T2, and then close operation is applied using 80×2 structuring element.

E. Text Area Verifing Using Edge Intensity

After all these process, the text areas are initially located. But the false detection rate will be high when the background of the image contains leaves, grass, line-upped team which has the similar intensive edge as characters. Like the case shows in fig 4. To solve this problem, a method about calculating the edge intensity ratio of different directions is proposed. As usual, the Chinese character has similar edge intensity in different directions, but the edge intensity of non-text in different

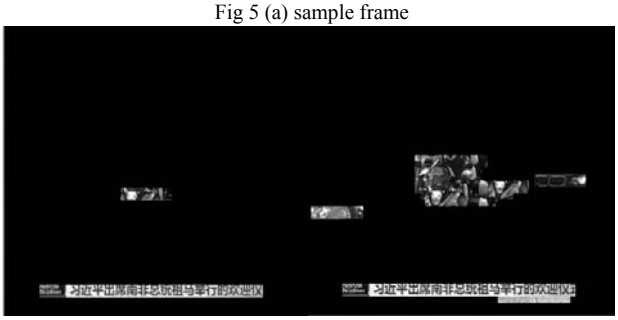
Fig 4. False detection in text location

directions is unpredictable. Firstly, use canny operator to detect the text area's edge that was found above. Then do with the edge graph using Transverse differential, Longitudinal differential, and 45 degree and -45 degree differential (Transverse differential is that move the image to the left with 1 pixel, subtracting it from original image. The others are in the same way). Now we have the horizontal, vertical, 45 degree and -45 degree information of edge. We compute these 4 ratios: horizontal to 45 degree, horizontal to -45 degree, vertical to 45 degree, vertical to -45 degree. Take the maximum of them to be the final edge intensity. The result of edge intensity of fig 3, from up to down, is 2.69, 10, 1.54. Obviously, an appropriate threshold can be to remove the non-text area.

III. Experimental and Comparative Results

The proposed ant colony algorithm is evaluated with 500 news video frames. Fig 4 shows a sample of original frames; fig 6 shows the result of ant colony algorithm text location, with a contrast between the binarization algorithm of otsu method and our method. The *recall* and

Fig 5 (a) sample frame



(b). Result using our method (c). Result using otsu method

Fig 6

precision are stated as follows:

$$recall = \frac{Net}{Tn} \quad (10)$$

$$precision = \frac{Net}{Td} \ . \quad (11)$$

where, *Net* is the exactly detected number of text lines. *Tn* is total number of text lines in test images. *Td* is the total number of text lines that the method detected. Text

line is the one that contains text blocks with the height of 8 pixels.

To prove the effectiveness of the verification of edge intensity, the initial text location and the location with verification are compared. The result is showed in table I.

It can be seen above that the precision is higher after using the verification while recall is a little lower. It is because the method we proposed can remove the most

TABLE I.
Recognition rate of text

Method	*Precision*	*Recall*
Location	65%	91%
Location+verification	80%	89%

false detections as well as some real text blocks under complex background.

The comparison of the proposed method by this paper with the ones proposed in reference [1]: extract the edge feature then using SVM to recognize. And reference [4]: use the wavelet transform to recognize (for comparison, here the window-size is also 16×48, step-size is 8) is listed in table II.

Table II has shown that the proposed method is better than the ones in reference [1] and [4]. The reason is that ref[1] uses SVM while the result of recognition is depended on the train sample. It is difficult to find a good

TABLE II.

THE DIFFERENT ALGORITHM COMPARISON

Method	*Precision*	*Recall*
Ref[1]	75%	83%
Ref[4]	71%	80%
The Proposed method	80%	89%

sample to train when the style of text and non-text area are not unique. Ref[4] uses only one feature, the wavelet coefficient, while our method uses three. Above all, the method proposed can handle more complex cases.

IV. CONCLUSION

In this paper, a text location method based on ant colony algorithm is proposed. The reason why use the ant colony algorithm is that text detection can be seen as a binary-class problem to differentiate text blocks and non-text blocks. Firstly, split the image into blocks, then extract 3 features of the sub-blocks to use ant algorithm for unsupervised classification. Finally, verify it using the edge intensity in different directions to reduce the false detection rate. This paper also compares it with other methods to prove its efficiency.

ACKNOWLEDGMENT

This work is supported by the National High Technology Development 863 Program of China (No.2011AA01A107) and the Zhejiang Provincial Technical Plan Project (No. 2011C13008).

REFERENCES

[1] CHEN D, BOURLARD H, THIRAN J P. Text identification in complex background using SVM[C]. Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2001, 2: 621-626.

[2] PALAIAHNAKOTE S, HUANG W, LIM T. An efficient edge based technique for text detection in video frames [C].The Eighth IAPR Workshop on Document Analysis Systems. Washington, DC: IEEE Computer Society, 2008: 307-314.

[3] WU V, MANMATHA R, RISEMAN E M. Text finder: An automatic system to detect and recognize text in images [J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(11):1224-1229.

[4] GLLAVATA J, EWERTH R, FREISLEBEN B. Text detection in images based on unsupervised classification of high-frequency wavelet coefficients[C]. Proceedings of 17th International Conference on Pattern Recognition. Washington, DC: 2004:425-428.

[5] SRIVASTAVI A, KUMAR J. Text detection in scene images using stroke width and nearest-neighbor constraints[C]. Proceedings of IEEE TENCON 2008. Washington, DC:IEEE Computer Society,2008: 1-5.

[6] KIM P-K. Automatic text location in complex color images using local color quantization[C]. TENCON 99: Proceedings of the IEEE Region 10 Conference. Washington, DC: IEEE Computer Society, 1999: 629-632.

[7] Huiping Li, David Doermann, Omid Kia. Automatic text detection and tracking in digital video[C]. IEEE Transactions on Image Processing. 2000, 9(1):147-156.

[8] Haralick R, Shanmugam K, Dinstein I, Textural features for image classification[C], IEEE Transactions on Systems, Man and Cybernetics, 1973, 3(6):610-621.

[9] Michael R.Lyu, Jiqiang Song, Min Cai. A comprehensive method for multilingual video text detection, location and extraction[C]. IEEE Transactions on Systems, 2005,15(2):243-256.

[10] J. Ohya, A. Shio. Recognition characters in scene images. IEEE Trans. Pattern Analysis and Machine Intelligence, 16(2):1994,214-220.

[11] Sara Saatchi and Chih Cheng Hung. Hybridization of the ant colony optimization with the K-means algorithm for clustering [J] .Springer-Verlag Lecture Notes in Computer Science, 2005:511-520.

[12] Haralick R, Shanmugam K, Dinstein I, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, 1973, 3(6):610-621.

[13] Harris C G, Stephens M J. A Combined Corner and Edge Detector. Proceedings of the 4th Alvey Vision Conference. 1988.

[14] LEE M, KIM S, CHO W, et al. Segmentation of brain MR images using an ant colony optimization algorithm [C]. Processing of the 9th IEEE International Conference on Image Processing. 2006, 1:985-988.

[15] Ostu N. A threshold selection method from gray-level histoSystems Man Cybernetic, 1978 (8): 62- 65IEEE Trans.

[16] Bernsen J. Dynamic Thresholding of Gray-level Images. Proc of 8thlntel Conf.on Patt. Recog. Paris, France: IEEE Computer SocietyPress, 1986: l251-l255.

[17] Ridler T W, Calvard S. Picture thresholding using an iterative selection methodIEEE-SNC.1978, 81:630-632.

[18] Zeng F, Zhang G, Jiang J. Text Image with Complex Background Filtering Method Based on Harris Corner-point Detection[J]. Journal of Software, 2013, 8(8): 1827-1834.
[19] Luo Q, Gao Y, Luo J, et al. Automatic Identification of Diatoms with Circular Shape using Texture Analysis[J]. Journal of Software, 2011, 6(3): 428-435.
[20] Wong F, Chao S, Chan W K. Cyclops-Snapshot Translation System Based on Mobile Device[J]. Journal of Software, 2011, 6(9): 1664-1671.

Ming Jiang He received the B.S. degree and M.S. degree in science in 1996 and 2001 respectively, and Ph.D. degree in Computer Science in 2004, all from Zhejiang University, China. He is currently a Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, Internet QoS provisioning, and network multimedia processing.

Taotao Zha He received his BSc in Software Engineering from Hangzhou Dianzi University in 2012. Currently he is a master student in this university. His primary research area focuses on image and video processing, text recognition .

Xingqi Wang He received his Bachelor and Master degree from Harbin Institute of Technology in 1997 and 1999, respectively, and Ph.D degree from Zhejiang University in 2002. He is an associate professor in college of Computer Science, Hangzhou Dianzi University, China. All his major are Computer Science. As a researcher, he visited CERCIA, University of Birmingham, UK from 2005 to 2006. His research interests include machine learning, data mining and multimedia content analysis.

Jingfan Tang He received the Ph.D. degree in Computer Science in 2005 from Zhejiang University, China. He is currently an Associate Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, quality assurance, process improvement and legacy system re-engineering..

Chunming Wu He received the B.S. degree, M.S. degree and Ph.D. degree in Computer Science from Zhejiang University, China, in 1989, 1992 and 1995 respectively. He is currently a Professor in college of Computer Science, Zhejiang University, and the director of NGNT laboratory. His research fields include Network Multimedia processing, reconfigurable network technology, network virtualization and artificial intelligence