Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):2577-2582

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Automatic text detection based on multi-resolution medical image fusion

Xinghui Zhu

College of Information Science & Technology, Hunan Agricultural University, Changsha, China

---

ABSTRACT

As to the problem of automatic text detection in medical video, this paper proposed a efficient algorithm for text location and medical video searching. In order to overcome the challenge of the different size of text in medical video frames, an algorithm which based on multi-resolution image fusion and text block 's feature was presented. The method include two steps: firstly, the wavelet feature and LBP feature of positive samples and negative samples were extracted which can be trained by support vector machine (SVM). And then, the test video for text detection should be ergoticly detected by multi-resolution method. Finally, the result image of text detection can be gained by multi-resolution image fusion. The experimental results show that this method has the superiority of accuracy rating compared with the traditional method based on edge detection, even so that the video frames are transformed quickly.

Keywords: wavelet feature; eLBP; text detection; multi-resolution fusion

---

INTRODUCTION

News video image text contains a lot of useful information, as the video plays more and more important role in the modern daily life, and realize automatic analysis of news video based on text contents and management requirements increasingly strong [1]. The traditional video retrieval is done through artificial means, efficiency is low and this way. With the development of image analysis technology, the use of image processing, automatic analysis technology to realize video retrieval has become a current an important trend[2]. Now the OCR technology has obtained certain achievement in character recognition, but due to a low resolution of news video, text is also embedded in the complex background[3], in the character recognition must first to test the video frame of the text area location, to ensure the accuracy of the OCR text recognition. Therefore, how to accurately detect text area has important practical significance.

Aiming at this requirement, many researchers in the text area location field did a lot of work, the representative of the method is based on edge detection and text line detection characteristics of subtitles [4], which can be more rapid detection of text, but the detection error rate is high, can't solve the problem of text in video sizes, and the constraint parameter setting is more complex.

To overcome this problem, this paper proposes a multi-scale image fusion based on news video text area detection localization algorithm. First of all the artificial collection of positive and negative samples wavelet high-frequency characteristics and local binary pattern such as feature extraction, the two features can reflect the similarities and differences of background of video and text image block; Then use support vector machine (SVM) training samples, obtained classifier; Last video frame of the test for multi-scale image sub-block traversal classification, finally the fusion detection results, test results in the end to the image. Compared with the method based on edge detection, because of video frames for the multi-scale detection, this algorithm can overcome the difficult problem of detecting text sizes, high positioning accuracy.

TEXT AREA FEATURE EXTRACTION

Select and extract text image features is an important foundation for accurate detection[5], text area analysis found that text area have commanded the edge and texture information. In this paper, considering the characteristic of text images, in order to distinguish it from video background image and considering the advantage of wavelet transform in image analysis and text image texture characteristic, put forward using wavelet high-frequency coefficients and local binary pattern characteristics to build the sample feature vector.

2.1. ***High frequency wavelet coefficient characteristics***

Also known as wavelet, wavelet is a finite oscillation waveform, the mean is zero. Wavelet analysis and Fourier analysis is very similar[6], the basic mathematical thought comes from classical harmonic analysis. Compared with the Fourier analysis and wavelet transform is LAN transform time and frequency, can more effectively local signal analysis. At the same time have the characteristics of multiresolution wavelet analysis, the signal can be divided into coarse part and detail part, namely the signal low frequency part and high frequency part.

Analysis text regions can be found in the video images, text area focus on performance in detail, so wavelet analysis are used to get the high frequency subband, and analyses characteristics of high frequency subband, can get the details of the text area. Sample image for Haar wavelet decomposition, respectively for horizontal, vertical and diagonal direction of the three high frequency subband, HH, HL HD, then extract the three high frequency subband average M, second order central moment μ_2 and third-order central moment μ_3, its calculation formula is:

$$M(I) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} I(i,j) \tag{1}$$

$$\mu_2(I) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\Big(I(i,j) - M(I)\Big)^2 \tag{2}$$

$$\mu_3(I) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\Big(I(i,j) - M(I)\Big)^3 \tag{3}$$

Among $I(i,j)$ is the value of the high frequency sub-bands of wavelet high frequency sub-bands of each form three characteristic value, you get 9 dimensional feature vector.

2.2. ***Local binary pattern characteristics***

Local binary pattern (LBP) characteristics of the operator by Ojala et al. [4], first proposed as a no-argument operator is used to measure local contrast to texture classification effectively. LBP operator is a very good texture description operator, its calculation is simple and can represent the image of local characteristics, and has a gray monotone and rotation invariance. LBP operator is commonly defined in 33 window, based on grey value of the window center threshold, do other pixel binarization processing within the window, when the center point neighborhood pixel gray value greater than or equal to center, its corresponding location assignment 1, or assign a value of 0. Depending on the pixel position weighted summation, the window of LBP values, values of LBP is a range of integers, 256 values according to 256 kinds of unique texture pattern. LBP operator computation formula is as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{4}$$

Including i_c for window center pixel gray value of(x_c , y_c), i_n ($n \in$ [0,7]) represent eight adjacent pixel gray value, the function of $s(x)$ is defined as follows:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

This is the standard, the basic idea of LBP operator but for standard LBP operator was flawed characters. Here USES a local boundary can describe the image texture of the boundary characteristics of local binary patterns (eLBP). ELBP operator on window pixel binarization, if neighborhood window center pixel gray value close to the center pixel gray value is 0, its corresponding location assignment or assign a value of 1.

According to the eLBP operator, with 256 units of the histogram, in order to decrease the number of histogram unit considering the shape characteristics of the characters at the same time, the neighborhood pixels in the calculation of eLBP operator can be divided into two categories: horizontal, vertical and diagonal direction horizontal vertical direction local binary pattern of operator and local binary pattern diagonal direction operator calculation formula is:

$$eLBP_{v-h}(x_c, y_c) = \sum_{n=0}^{3} s_e(i_{2*n+1} - i_c)2^n \tag{6}$$

$$eLBP_{diag}(x_c, y_c) = \sum_{n=0}^{3} s_e(i_{2*n} - i_c)2^n \tag{7}$$

Function of $s_e(x)$ defined as follows:

$$s_e(x) = \begin{cases} 1, & |x| \geq e \\ 0, & |x| < e \end{cases} \tag{8}$$

Calculate according to the operator $eLBP_{v-h}$and $eLBP_{diag}$ operator, is respectively two histograms with $2^4 = 16$ units, which makes the texture feature extraction computation have greatly reduced the characteristic dimension of.

Images of 16×16 fixed size based on local binary pattern 3×3 level of vertical direction operator local binary pattern $eLBP_{v-h}$ and diagonal direction operator $eLBP_{diag}$, get will be the two histograms with 16 units, respectively, to extract histogram of each unit value, can get 32 d characteristics.

Integrated wavelet high-frequency coefficients features and local binary pattern, can to each sample image feature extraction, to get a 41 dimension feature vector, the feature extraction of the sample.

TEXT AREA LOCATION

Upon completion of the sample, on the basis of feature extraction, training need to sample characteristics, get the text and the text classifier, used to classify test image. Considering the relatively small size of the training sample set, USES the SVM for training to sample characteristics. SVM has good generalization ability, small sample can to a certain extent, the disadvantages of fewer samples and SVM is a two classifiers, which just meets our requirements of text and non-text classification.

3.1. *The SVM training*

Support vector machine (SVM) is initially proposed to deal with linear separable problem [6, 7], is developed on the basis of statistical learning theory, a new generation of learning algorithms in text classification, image analysis and other fields obtained better application, compared with easy to excessive fitting of the neural network, support vector machine (SVM) for test sample not seen has better generalization ability.

Through theoretical derivation, the SVM in the end, four dual quadratic optimization problem to maximize the objective function:

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(\psi(\overrightarrow{x_i}) \cdot \psi(\overrightarrow{x_j})\right) \tag{9}$$

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \psi(\overrightarrow{x_i})$$

There was .

Input the SVM classifier to the characteristics of the training sample, optimize the training, can obtain a second classifier, to text classification.

3.2. *Text area location*

Due to the size of the training sample, so the news video frame text area detection location and need to use the small piece of video frames for a certain step traverse, for each little piece of identification of a text area classification. Text is difficult to determine the size of the video, and at the same time there are different height of the font line, an important innovation point of this article is integration of multi-scale's test results. Frame to be detected at 1, 0.5, 0.25, three scales to text area detection, and combined the three results, and the fusion results are connected domain analysis, and to shape the region constraint, remove noise region, improve the detection accuracy, text location steps of the specific as follows:

Initialize the horizontal and vertical direction scanning step size dx, dy; The input video frame to be detected, according to the scanning step length, with the size of the window of 16×16 scan, for each window to extract a 41 d characteristic vector and end up with a characteristic matrix; Using the trained SVM classifier detection classification, the classification result is zero, the window all pixel values are set to 0, classification result is 1, the window all pixel values are set to 1. Binary mark image I_0; The video frame to be detected and scaling to 0.5, 0.25, respectively, repeat steps 2, 3, and amplify the detection result to the original scale, get binary marker image. I_1, I_2;Integration of I_1, I_2, I_3 and the result of the fusion analysis of connected domain constraints, only keep area is greater than 25×25, regional height is greater than 20 and text area of the aspect ratio greater than 1.5.

3.3. Overall the implementation process of algorithm

According to the above description, can be based on the multi-scale image fusion of news video text region detection algorithm is the implementation of the process, as shown in fig.1.

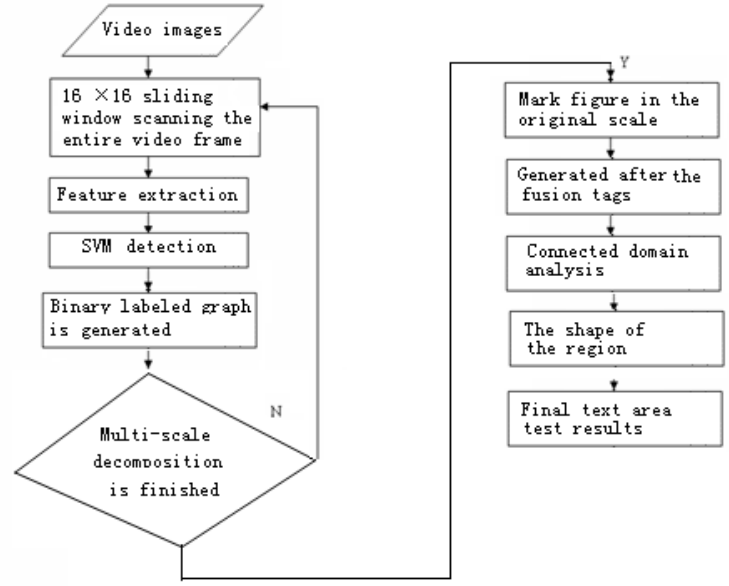


fig.1. Algorithm flow chart as a whole

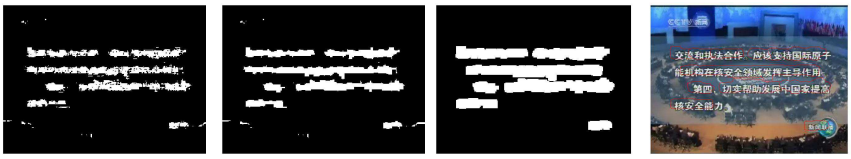
THE EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the effectiveness of the algorithm in the matlab platform, launched a video frame the text area location simulation experiment. First by artificial respectively collected 200 size is 16 * 16 for text image samples and erbal image samples, positive and negative samples, respectively, as shown in fig. 2.

fig. 2. (left) of positive and negative training samples sample (right)

Characteristics of positive and negative samples are extracted respectively, obtained a 400 * 400 characteristic matrix, and then use SVM to characteristics of the training, get a classifier C. Use classifier C, according to the process of figure 1, some frames in news video text detection results as shown in fig. 3.

（a）Video frame （b）Test results I_0 （c）Test results I_1 （d）Test results I_2



（e）fusion results I'（f）Connected domain（g）result of constraint（h）final location map analysis results
fig. 3. news video a frame detection results

According to test results in the figure 3, the algorithm can effectively detect the resolution is poor and complex background, the text area in news video, and make an accurate positioning, is this correct subsequent video content based retrieval can automatic accurate completion of important premise.

Define the recall to assess effectiveness of the algorithm, namely:

Recall = correct detection of text blocks/all text block

Based on 30 frames the news video detection location and statistical analysis, this article with text detection algorithm based on edge detection algorithm of text detection recall results as shown in table 1.

table 1 Text area results

Detection algorithm	recall ratio
Based on edge detection	80.1%
this paper algorithm	95.3%

Experiments show that the algorithm is compared with the previous proposed based on edge detection method, has the obvious superiority, on the positioning accuracy is improved greatly, word recall ratio increased by 15.2%, while in the process found that the present algorithm can also overcome text, as a result of the rapid transformation between video frames jumping too big problems, has certain practical significance.

CONCLUSION

To overcome the words in the news video size and frame between rapid transformation leads to the problem of the large text jumps, this paper proposes a multi-scale image fusion based on news video text area detection localization algorithm. First of all the artificial collection of positive and negative samples wavelet high-frequency characteristics and local binary pattern such as feature extraction, the two features can reflect the similarities and differences of background of video and text image block; Then use support vector machine (SVM) training samples, obtained classifier; Last video frame of the test for multi-scale image sub-block traversal classification, the final test results to fusion, and connected domain analysis and regional constraints, test results in the end to image. Compared with the method based on edge detection, because of video frames for the multi-scale detection, this paper algorithm in addition to text size problem can be overcome, and raised the recall text detection, has certain practical significance.

Acknowledgments
This project was supported in part by the National Key Technology R&D Program of China (2012BAD35B07).

REFERENCES

[1] YANG Gelan, Yue WU, and Huixia JIN. *Journal of Computational Information Systems*. **2012**, 8(10) 4315-4322.

[2] Xuezhan Liang, Xiaolin Liu. *The Computer Simulation*. **2012**, 29(3-9): 223-226.
[3] Jinlin Guo. *Computer application research*, **2011**,28(8):3148-3151.
[4] JAIN K, YU B. *Multimedia System*, **2000**, 8(8)69-81.
[5] Yueting Zhuang, LiuJun Wei,and Fei Wu. *Journal of computer-aided design and graphics*, **2002**,14(8):750-753.
[6] Gelan Yang, Huixia Jin, and Na Bai. *Mathematical Problems in Engineering*, Volume **2013**, Article ID 272567,10 pages.