Journal of Electrical and Electronic Engineering
2017; 5(3): 86-91
http://www.sciencepublishinggroup.com/j/jeee
doi: 10.11648/j.jeee.20170503.11
ISSN: 2329-1613 (Print); ISSN: 2329-1605 (Online)

Fast Image Embedded Chinese Text Extracting by Homogeneous Space Mapping

Li Ying[1, *], Liu Lisha[1], Cui Yan-peng[2], Zhuang Huaiyu[3]

[1]School of Electronic Engineering, Xidian Univercity, Xi'an, China

[2]Institute for Internet Behavior, Xidian Univercity, Xi'an, China

[3]China Mobile Group Guangdong Co., Ltd. Guangzhou, China

Email address:

Liying6@chinamobile.com (Li Ying), liulisha413@126.com (Liu Lisha), 82302873@qq.com (Cui Yan-peng), zhuanghuaiyu@chinamobile.com (Zhuang Huaiyu)

[*]Corresponding author

To cite this article:

Li Ying, Liu Lisha, Cui Yan-peng, Zhuang Huaiyu. Fast Image Embedded Chinese Text Extracting by Homogeneous Space Mapping. *Journal of Electrical and Electronic Engineering*. Vol. 5, No. 3, 2017, pp. 86-91. doi: 10.11648/j.jeee.20170503.11

Received: April 17, 2017; **Accepted**: May 13, 2017; **Published**: May 16, 2017

Abstract: Text-embedded images are popular in the mobile Internet to spread malicious information. A fast text-embedded image Chinese text extracting algorithm based on homogeneous space mapping is proposed. Image enhancement functions are used to highlight edge and texture features of images. Sobel operator is used to extract the edge feature and wavelet packet is used to extract the 24-dimensional texture feature vectors in the enhanced images. The texture features and edge features are used to describe the homogeneity of an image, which construct the homogeneous feature map of the image. The differences between the non-text and the text region homogeneity are used to distinguish them and reduce non-text region further. Thus the text regions are highlighted. Then, homogeneous text samples are used to train the text region detector, which greatly reduces the computational complexity. Finally, the characters are segmented and recognized. Some experiments to verify the validity and practicability of the proposed algorithm have been conducted. The recognition rate achieves 86%, which is higher than that of other methods in industry. The algorithm is verified on the operator's malicious information monitoring system, which provides secure malicious filtering performance.

Keywords: Chinese Text Embedded Image, Homogeneous Mapping, Text Extraction, Information Security

1. Introduction

With the development of multimedia information retrieval, Internet and 4G network streaming technologies, multimedia and video have become the mainstream information carrier of exchanging multimedia information. With the emergence of the 4th generation (4G) digital communication network, information dissemination of instant communication channel has been more rapid, extensive, like WeChat. While the Internet brings the rich knowledge to people, the security of content emerges as an issue. Currently, filtering malicious information in plain text has been relatively mature, but how to effectively block the malicious text information in multimedia and video is still difficult. Therefore, the identification and interception of malicious information by analyzing the image and video content, have significant meaning in reducing the spread of malicious information and protecting the mental health of teenagers. How to detect and filter the malicious information of the embedded text in the image is the focus of the Internet information security. The methods based on connecting domain in Ref. [1] and [2] assume that the color of the characters in the image are consistent, then use the color region to determine the candidate text area. They use the heuristic rules to filter the text areas. The methods based on edge detection in Ref. [3] to [5] using the high contrast between text and background, first detect the edge, then use the morphological operator to connect the edge of the text area, and finally use heuristic rules for character filtering. It is difficult for the region-based methods to extract accurate connected domains when the background of images is complex or the quality of images is poor. In addition, the heuristic rules used in text filtering

depend on the prior knowledge, but the accurate prior knowledge is difficult to be obtained. Besides, there are many rigid thresholds are be determined, which lead to the poor robustness of the algorithms.

Texture-based methods in Ref. [6] and [7] regard the text area as a special texture, using the different texture characteristics of the text area and background area to detect, extract and identify text. They use sliding window to scan and detect the small area texture in the images, then use the trained classifier to judge whether the current small area is the text area, and finally merge all the text small area to form the candidate text area, which realize the extracting and identifying text are on the candidate text. In a complex context, the texture-based approaches are more robust than the methods based on the connected domain, and the versatility is better. Texture-based method has good versatility, but also has high computing cost, and is sensitive to the text size and font. Methods in Ref.[8] and [9] are aimed at extracting the text with rich corner information, like Chinese character. They use corner detection operators to get corner images. Then according to the distribution, density and other characteristics of text area corner, they filter and cluster the corner point to get the candidate text area. Corner detection-based methods are effect at detecting Chinese and other pictographic characters, but it is sensitive to the text size. Methods based on the video scene feature in Ref. [10] and [11] make full use of the color of the text area in the video scene as well as information redundancy in multi-frame images. Some new ideas for video text extraction have been provided in this paper.

This paper presents an extraction algorithm of embedded Chinese text in images based on homogeneous space mapping with the lowest cost, which is robust and fast to filter malicious information. It provides technical support in Ref. [12] and [13] for the malicious text interception in multimedia information transmission, and security for content security on mobile operators' Internet pipelines.

2. Homogeneous Mapping for Embedded Text Extraction in Images

The function scheme of extracting text from images include: image enhancement, feature extraction, homogeneity mapping, text area extraction, character extraction, recognition detector training and testing. Figure 1 shows the block diagram of the proposed embedded text extraction method. Each functional unit is described in the following Figure 1.

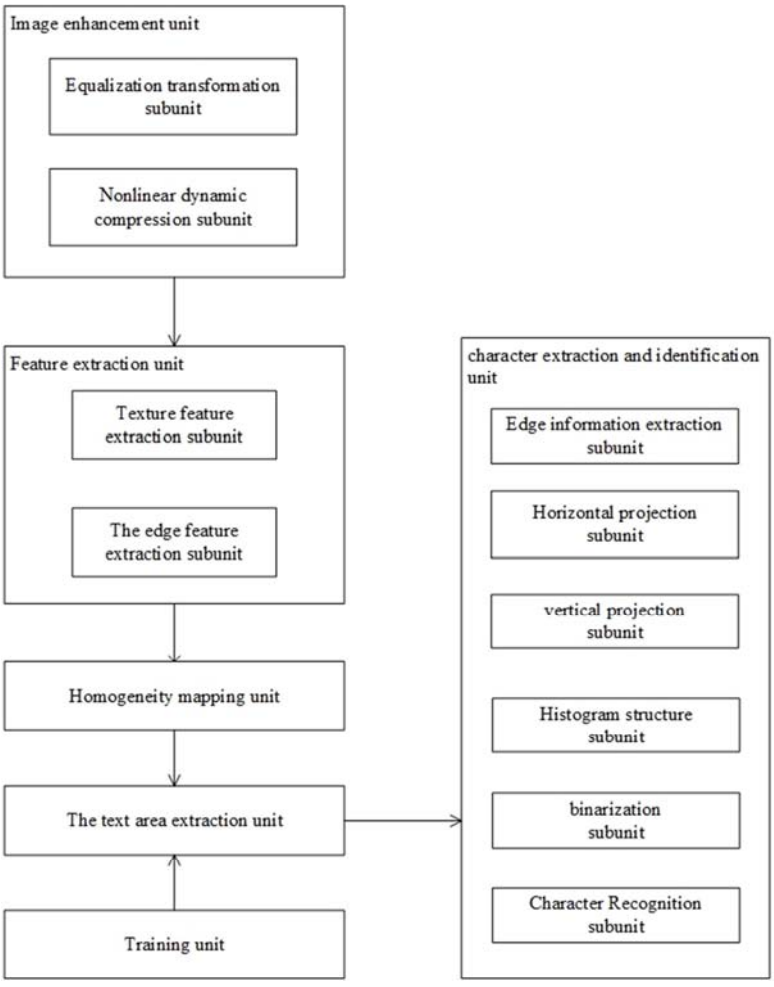

Figure 1. *The block diagram of embedded text extraction method in images.*

2.1. Image Enhancement

Nonlinear dynamic compression of the original image in gray level can fully highlight the edge, texture and other attributes of multimedia, inhibit the useless information, and provide more distinctive features for the subsequent text area extraction. The specific method is as follows:

(1) The equalization transformation function can regulate the gray histogram of original multimedia.

$$k = INT\left[(L-1)\cdot\sum_{j=0}^{k^*}P(j)+0.5\right] \tag{1}$$

Adjusting the gray range of the original multimedia can increase the multimedia clarity, where $INT[\ ]$ is rounded symbol, $P(j)$ is the gray color histogram of original multimedia, k^* is the gray level of original multimedia, $k^*=0,1,\cdots,L-1$. k is the gray level of the multimedia after equalization transformation.

The monotonically increasing S-type function as a nonlinear dynamic compression transformation function to process the multimedia is selected as following equalization (2) transforming:

$$S(k;\alpha,\beta,\gamma)=\begin{cases}0, & k\le\alpha\\ 2\left(\dfrac{k-\alpha}{\gamma-\alpha}\right)^2, & \alpha<k\le\beta\\ 1-2\left(\dfrac{k-\gamma}{\gamma-\alpha}\right)^2, & \beta<k<\gamma\\ 1, & k\ge\gamma\end{cases} \tag{2}$$

where α is the gray value corresponding to the first peak of the grayscale histogram after equalization. γ is the gray value corresponding to the last peak of the grayscale histogram after equalization. Obviously, $\alpha<\gamma$. The maximum information entropy criterion to determine β as following: assuming $P(k)$ is the gray scale histogram of original multimedia after equalization, from gray histogram definition $P(k)=n_k/n$, $k=0,1,\cdots,L-1$, $P(k)$ is the probability distribution of the multimedia gray level k, where n_k is the pixels number of the multimedia gray value k, and n is the total pixels number of the multimedia. A grayscale threshold v that divides the multimedia into two areas: one's gray value is 0 ~v, then it's entropy is:

$$H_1(v)=-\sum_{k=0}^{v}\frac{P(k)}{P(v)}ln\frac{P(k)}{P(v)} \tag{3}$$

Another region's gray value is $v+1\sim L-1$, then it's entropy is:

$$H_2(v)=-\sum_{k=v+1}^{L-1}\frac{P(k)}{1-P(v)}ln\frac{P(k)}{1-P(v)} \tag{4}$$

where $P(v)=\sum_{k=0}^{v}P(k)$. The total entropy is expressed as $H(v)=H_1(v)+H_2(v)$. Two types data can be separated when the entropy is the largest, so the best threshold v can be obtained when total entropy is the biggest, $v^*=arg_v max[H_1(v)+H_2(v)]$, then $\beta=v^*$.

The monotonically increasing S-type function as the nonlinear dynamic compression transform function to dynamically compress the multimedia is used in this paper. It not only fully highlight the edge of the multimedia, texture and other attributes, but also greatly improving the performance of multimedia enhancement and segmentation algorithms.

2.2. Feature Extraction

(1) A wavelet packet decomposition to extract the texture of enhanced multimedia information as following:

1) Function $W_0(x)$ is given to generate a set of orthogonal wavelet bases:

$$\begin{aligned}W_{2m}(x)&=\sqrt{2}\sum_k h(k)W_m(2x-k)\\ W_{2m+1}(x)&=\sqrt{2}\sum_k g(k)W_m(2x-k)\end{aligned} \tag{5}$$

where $W_{2m}(x)$ is the scale function, $W_{2m+1}(x)$ represents the wavelet function, $h(k)$ and $g(k)$ are the orthogonal wavelet basis of the filtering coefficients. The wavelet packet is $W_m(2^n x-k)$, where n is the scale parameter, k is the translation parameter, and m is the vibration parameter, $n,k\in Z, m\in N$.

2) Two-dimensional filter can be obtained by taking two one-dimensional wavelet packets direction as the inner product in the horizontal or vertical:

$$\begin{aligned}h_{LL}(k,n)&=h(k)\cdot h(n), h_{LH}(k,n)=h(k)\cdot g(n)\\ h_{HL}(k,n)&=g(k)\cdot h(n), h_{HH}(k,n)=g(k)\cdot g(n)\end{aligned} \tag{6}$$

where the first and second subscripts of the filter represent taking high-pass H or low-pass L filter in the x and y directions, respectively.

3) Take two-dimensional wavelet transform on the enhanced multimedia by a two-dimensional filter. That is, up-sampling the impact of the two-dimensional filter to get a plurality of sub-multimedia that have a translation invariance and containing intermediate frequency information (texture information), and selecting a predetermined number (e.g., eight) sub-multimedia with the largest variance.

4) The statistics-based first order gray distribution to describe the texture features, and the energy characteristics, entropy characteristics and mean deviation characteristics of

each sub-multimedia, which are defined as following:

$$F_1(x,y)=\frac{1}{255^2}\sum_{i=x-n}^{x+n}\sum_{j=y-n}^{y+n}f(i,j)^2 p[f(i,j)]$$
$$F_2(x,y)=\frac{-1}{log255}\sum_{i=x-n}^{x+n}\sum_{j=y-n}^{y+n}p[f(i,j)]log\{p[f(i,j)]\} \quad (7)$$
$$F_3(x,y)=\frac{1}{(2n+1)^2}\sum_{i=x-n}^{x+n}\sum_{j=y-n}^{y+n}\left|f(i,j)-\overline{f}(i,j)\right|$$

$p[f(i,j)]$ represents the probability of the gray scale of the point (i, j) in the subgraph. $\overline{f}(i,j)$ represents the gray scale of all the pixels in the operation window centered on the dot (i, j). For each selected feature subgraph, the above three features are calculated point by point. 24-dimensional texture feature vectors for every point in texture multimedia are provided, $F(x,y)=\left\{F_m^i(x,y)\right\}, i=1,2,\cdots,8, \quad m=1,2,3$

$F_m^i(x,y)$ represents the *m-th* feature extracted by the point (x, y) in the *i-th* feature sub-graph. The Sobel operator is used to extract the edge feature information of the enhanced multimedia, and the multimedia edge feature $E[I(x, y)]$ can be calculated.

2.3. Homogeneity Mapping

Homogeneity features are constructed as follows:

(1) Normalize the texture and edge features separately:

$$F_t^*(x,y)=\frac{F_t(x,y)-F_{tmin}(x,y)}{F_{tmax}(x,y)-F_{tmin}(x,y)},$$
$$E^*[I(x,y)]=\frac{E[I(x,y)]-E_{min}[I(x,y)]}{E_{max}[I(x,y)]-E_{min}[I(x,y)]} \quad (8)$$

(2) The homogeneity of the multimedia at (x, y) as following:

$$Y(x,y)=\{F_1^*(x,y),F_2^*(x,y),\cdots,$$
$$F_{24}^*(x,y),E^*[I(x,y)]\} \quad (9)$$

This step takes the multimedia texture information and edge information into account. It selects the texture information and edge information to form the multimedia homogeneity and gets the characteristics of multimedia through the mapping. This step makes full use of the different homogeneity between non-text area and text area in the multimedia to distinguish it from the text area, thus it suppress the information of the non-text area and highlighting the characteristics of the text area.

2.4. Text Area Extraction

The text area can be obtained by inputting the feature of the multimedia to be detected to the text area detector for detection. The text area detector is obtained by training the sample multimedia in a homogeneous space. Specific training steps are as follows:

(1) Construct a weak classifier for each feature in the feature vector;

(2) AdaBoost in Ref. [14] method is used to select the weak classifier with the best performance, and determine the weight τ_t of each weak classifier. It is inversely proportional to the error rate of the weak classifier c_t . T feature weak classifier is integrated into a strong classifier C according to the weight, to get the text area detector:

$$C(\mathrm{s})=\begin{cases}1, & \sum_{t=1}^{T}\tau_t c_t(\mathrm{s})\geq\frac{1}{2}\sum_{t=1}^{T}\tau_t\\ 0, & others\end{cases} \quad (10)$$

If the $C(s)$=1, the S is judged to be text, otherwise it is judged as non-text.

According to the characteristics of the text area and the AdaBoost classification training method, the text area detector with finer identification ability is obtained, which not only conforms to the feature description of the text, but also greatly reduces the computational cost and improves the detection efficiency.

2.5. Character Decomposition

In the extracted text area, because of the complexity of the text background, it is difficult to use a simple threshold method to separate the text from the background. And the background of the word is relatively uniform, so first decompose the text area into the character area, and then binarize the character multimedia. The specific method is as follows:

(1) Use the Canny operator to extract the edge information of the text area;

(2) Get the threshold of edge information after the horizontal projection to separate the text line of multimedia;

(3) Get the threshold of edge information after the vertical projection to separate the text line of multimedia;

(4) Constructs the gray histogram of the character multimedia;

(5) Using the Otus method to determine the optimal segmentation threshold of the gray histogram of the character multimedia;

(6) Binarize the character multimedia according to the optimal segmentation threshold, and remove the residual background based on the connected domain analysis method to get the character multimedia without background.

2.6. Character Recognition

(1) Use B-spline interpolation function to enlarge character multimedia after clearing the background to improve its resolution;

(2) Use Microsoft Office Document Imaging interface to the character recognition of the character recognition after amplification;

(3) According to the Chinese character code, filter out non-

Chinese characters in result.

(4) Match identification results with malicious information keyword list in multi-mode, find whether the keyword is included in the recognition result, and use the HASH table to create a keyword list, effectively improving the efficiency of the search;

(5) Generate recognition results file, and save the results in the database.

3. Experimental Analysis and Application

In order to verify the effectiveness and practicability of the algorithm, the 3000 embedded text images are selected for experiment, including Internet MMS, WeChat images, TV subtitles, and photos and so on. The average test results of the three experiments are shown in Table 1. Figure 2 shows the extraction of embedded text in image. The validity of the algorithm is further verified in the operator's bad information monitoring system. The embedded text in image and the monitor political sensitive information can be rapidly identified in real-time.

Table 1. *Recognition experiment of embedded text in image.*

Background type	Text recognition rate (%)	Text filtering rate (%)	Speed (frame per second)
Monochrome background	100	100	5
Simple background	99	99	3
General background	85	82	2
Complex background	78	71	1
overall	86	82	3.4

Like a spider tryin' to run from the rain

(a) Text area (b) Text

That you can't take
your pretty eyes away from me

(a) Text area (b) Text

Figure 2. *Illustration of image inline text extraction diagram.*

4. Conclusion

This paper presents a fast extracting algorithm for image embedded text based on homogeneous space mapping. Take the image with embedded text in the mobile Internet as the research object. Highlight the edge of the image and texture features through the image enhancement function, extract the texture feature vector of 24 dimensions by wavelet packet and the edge feature information of enhanced image by Sobel operator. Obtain the feature image by using texture features and edge features to construct and map homogeneity of image. Using the difference between the non-text and the homogeneity of the text area to distinguish the non-text area information, and highlight the text area characteristics. Then use the homogeneous space to train the sample image to get the text area detector. It's not only more in line with the description of the text, but also greatly reduces the amount of calculation. Finally, text is extracted by character segmentation and character recognition.

References

[1] Yao Cong, Bai Xiang, Liu Wenyu, et al. Detecting texts of arbitrary orientation in natural images [C], 2012 IEEE conference on computer Vision and Patten Recognition (CVPR). IEEE, 2012:1083-1090.

[2] Epshtein Boris, Ofek Eyal, Wexler Yonatan. Detecting text in natural scenes wstroke width transform [C], 2010 IEEE conference on Computer Vision a Pattern Recognition (CVPR). IEEE, 2010:2963-2970.

[3] Li Xueyan, GuoShuxu, GaoFengli. Text extraction in video based on wavelet modulus maximum [J], Computer Engineering, 2007, 33(5): 26-28.

[4] Zhao Ming Li Shutao, Kwok James. Text detetion in images using sparse representation with discriminative dictionaries [J]. Image and Vision Computing, 2010, 28(12):1590-1599.

[5] Park Jonghyun, Lee Gueesang, Kim Euichul,et al. Automatic detection and recognition of Korean text in outdoor signboard image [J], Pattern Recognition Letters. 2010, 31(12):1728-1239.

[6] Ikica, Andrej, Peer, Peter. An improved edge profile based method for text detection in images of natural scenes [C] EUROCON-Intenational Conference on Computer as a Tool (EUROCON). IEEE, 2011:1-4.

[7] Yan Jianqiang, Tao Dacheng, Tian Cunna, Gao Xinbo, Li Xulong. Chinese Text Detection and Location for Images in Multimedia Messaging Service[C], //Proc of Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, Istanbul Turkey: IEEE Conference Publications,2010: 3896-3901.

[8] Zhong Yu, Zhang Hongjiang, Anil K. Jain. Automatic Caption Localization in Compressed Video [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(4):385-392.

[9] Lienhart R, Wernicke A. Localizing and Segmenting Text in Images and Videos[J], IEEE Transactions on Circuits and Systems for Video Technology, 2002,12(4):256-268.

[10] Wonjun K, Changick K. A New Approach for Overlay Text Detection and Extraction from Complex Video Scene [J], IEEE Transactions on Image Processing, 2009, 18(2):401-411.

[11] Pratheeba T, Kavitha V, Rajeswari S R. Morphology based Text Detection and Extraction from Complex Video Scene [J], International Journal of Engineering and Technology, 2010, 2(3): 200-206.

[12] LI Ying, ZHUANG Huaiyu, LI Xiangwei, A New Technique on Text Region Location in Images, Journal of Xidian University,2013,40(6):187-192.

[13] LI Ying, CUI Yan-peng, GAO Xin-bo, A Novel Algorithm for Text Data Compression Based on Arithmetic Codec, Journal of University of Electronic Science and Technology of China, 2016, 45(6): 929-933

[14] Huang Jianhua, Cheng Hengda, Wu Rui. Text Detection Method Based on Fuzzy Homogeneity Mapping [J]. Journal of Electronics and Information Technology, 2008, 30(6): 1376-1380.