Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **2**, March **2016**

A Joint Approach of Harris Corners Detection and Baseline Searching for Localization of Uyghur Text Lines in Image Sequences

Dilmurat Tursun, Li Kai, Gulxax Halelbek
Institute of Information Science and Engineering
Xinjiang University of China
No.666 Shengli Road, Urumqi, 830046, China
tursundilmurat@gmail.com

Askar Hamdulla*
School of Software
Xinjiang University of China
No.666 Shengli Road, Urumqi, 830046, China
*Corresponding Author: askarhamdulla@gmail.com

Received February, 2015; revised October, 2015

Abstract. *This paper proposed an approach for detection of Uyghur text lines in video frames according to the writing styles and characteristics of Uyghur characters. Firstly, according to the comparative values of color sub-components between the text lines located area and background, gets the corner maps of the corresponding grayscale images by using the algorithm of Harris corner detection. Then the relatively isolated corners are removed in the corner maps by filtering, and then the corner images are fused into text regions. And finally, the edges of the grayscale images are detected by using Sobel operators, and the detected text regions are checked whether those are the true Uyghur text lines located areas or not according to the heuristic features of a true target area and the baseline feature of Uyghur text lines. The experimental results show that the proposed method achieves good effects and robustness in localization of Uyghur text lines in images sequences.*
Keywords: Uyghur, Text detection, Harris Corner, Edge detection, Text baseline.

1. **Introduction.** In recent years, the multimedia information such as images, audio and videos are rapidly becoming the mainstream of the exchange of information and services. Traditional database search of retrieval based on keywords cannot meet the needs of the people. Content based multimedia retrieval is becoming more and more popular now. Therefore, text detection and extraction from image sequences are very important issue in this field.

The text in the images and videos can be roughly divided into two categories: scene text and artificial text. Scene text often appears in the scene such as the text in the billboards and road signs, the number in the athletics and so on. Meanwhile, the artificial text as the complement of visual content is often added to the movies and TV series subtitles, news headlines, TV subtitles and so forth.

Most existing approaches can be generally classified into three categories: namely, edge based methods [1, 2], connected component based methods [3], and texture based methods [4]. The text should be detected precisely if we want to extract the text from the images. Although so far the location and extraction technologies for Chinese, English has been matured, but the proposed algorithms for those languages could not be directly applied to the images contain Uyghur texts very well because of the special characteristics of the Uyghur fonts. The method proposed by [5] couldn't achieve good effects in the videos which contain of Uyghur text lines, because he/she does not take the unique baseline characteristic of Uyghur text into account. Based on the combination of the Harris corner and special baseline characteristics of Uyghur font streams, this paper proposes a method to detect the Uyghur text areas in videos images.

2. **Image preprocessing.** In order to detect the Uyghur text regions in the video images, the algorithm flow chart of this paper is shown in Figure 1. Firstly, the color image is converted into gray scale image, and then the Harris corner detection procedure [6] is employed on the gray image and the corners outside the text regions were filtered out. The morphological dilation operator is used to merge the corners into the candidate text region. Finally the baseline characteristic of Uyghur text lines and the heuristic rule of the text area are used to verify whether the candidate text region is the true text region or not.

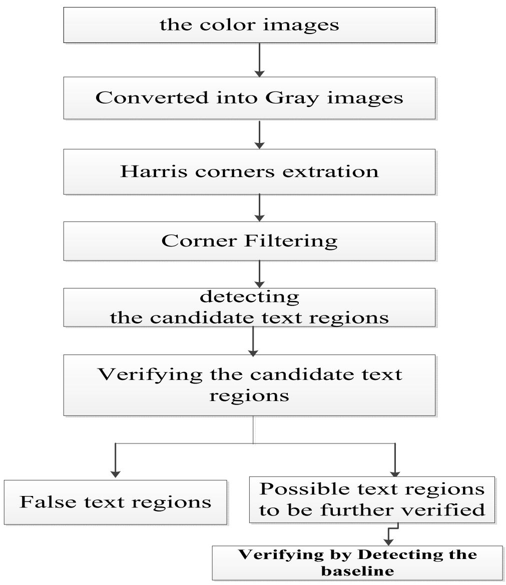


Figure 1. Overall Scheme of our Algorithm

2.1. **Converting the color image into gray scale image.** The preprocessing of image makes very much convenience to do further operations on image and for the experiments in research. The image preprocessing contains image enhancement, color into gray-scale transformation, image binarization, edge detection etc. As the Harris corner detection is usually applied on gray image, the color images are converted into gray image in advance. According to the photosensitive characteristics of the color from human psychology research [7], the gray value of each pixel can be calculated by

$$Y = 0.229 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

We can use (2) instead of (1) to avoid the floating-point computations,

$$Y = (229 \times R + 587 \times G + 114 \times B + 500)/1000 \quad (2)$$

Taking the integer number division and rounding into account, (2) can be simplified as (3),

$$Y = (23 \times R + 59 \times G + 11 \times B + 50)/100 \tag{3}$$

This allows the operation time can be further reduced. So in this paper we calculate the gray value in (4),

$$Y = (77 \times R + 151 \times G + 28 \times B + 128)/256 \tag{4}$$

Because 256 is a power of 2, the compiler optimized it for the shift operation, so that the running time can be saved.

3. **The localization of possible text regions.**

3.1. **The selection of the corner operator.** In this paper, we select the corners as a essential characteristics of the text, because corners have two major advantages as below:

(1) Corners, the frequent and essential features of text regions, have strong stability and robustness than other low-level features. Therefore, the background noise can be eliminated to a large extent.

(2) The distribution of corners in the text regions is more regular than the non-text regions; therefore, the corners of irregular non-text can be filtered out according to the designing specific rules like in this paper.

As the text in images and videos contain rich corners, this paper proposes the combination of the Harris corner detection algorithm and mathematical morphological operations to achieve the possible text located regions in the images and videos.

Harris corner has better stability, better noise immunity, and better time complexity than Susan corner [8]. Harris operator is point feature extraction operator from a signal and it is based on the C. Harris and M. J. Stephenss Moravec algorithm. This operator inspired by the autocorrelation function of the signal processing, produces matrix M associated with the autocorrelation function. The Eigen values of the matrix is first-order curvature of the autocorrelation function, if both Eigen values are large, it means that the point is a corner. The corner detection principle of the image is:

If the offsets of a point in any directions can cause a large changes in gray scale on the image plane, we can conclude that the point as a corner [9].

Harris corner detection is defined as follows:

$$R = det(M)/tr(M) \tag{5}$$

$$M(x,y) = \begin{bmatrix} I_u^2(x,y) & I_{uv}(x,y) \\ I_{uv}(x,y) & I_v^2(x,y) \end{bmatrix} \tag{6}$$

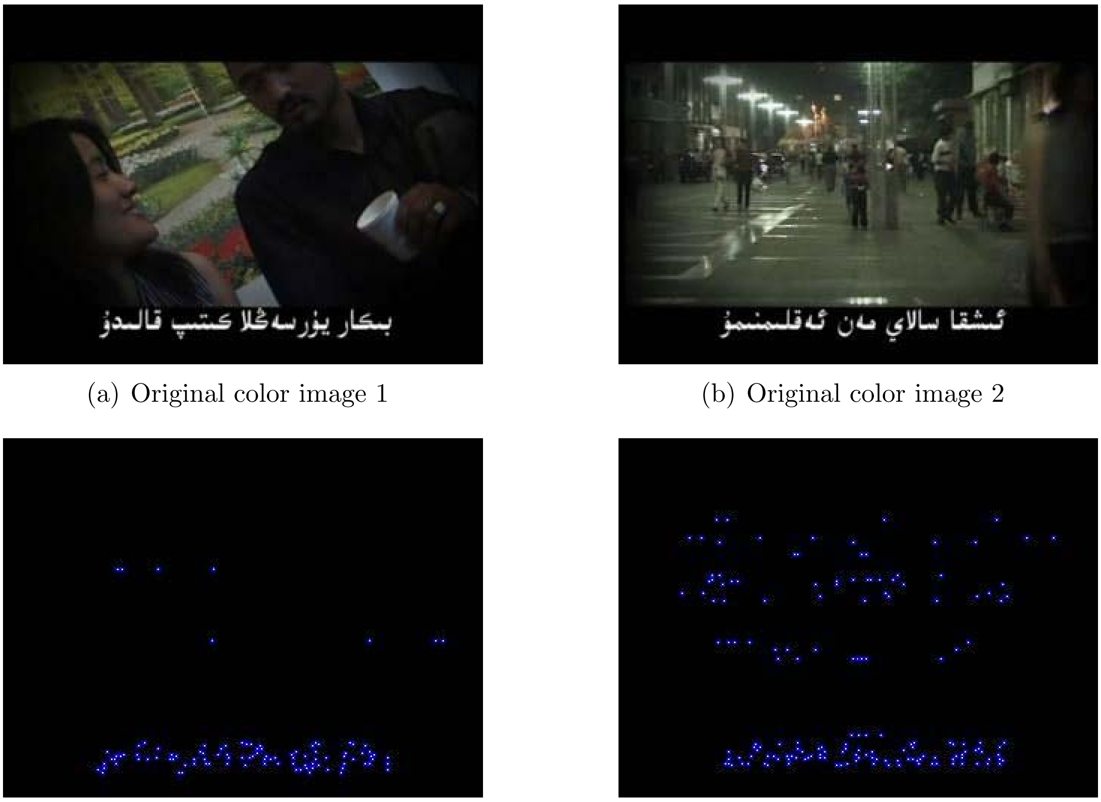
$$I_u^2(x,y) = X^2 \bigotimes h(x,y) \tag{7}$$

$$I_v^2(x,y) = Y^2 \bigotimes h(x,y) \tag{8}$$

$$I_{uv}(x,y) = XY \bigotimes h(x,y) \tag{9}$$

$$h(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \tag{10}$$

$I_u(x,y)$ and $I_v(x,y)$ is the partial derivatives at point (x,y) in u and v directions. $I_{uv}(x,y)$ is mixed derivatives in u and v directions. $h(x,y)$ is a Gaussian function. The main purpose of the Gaussian function is reducing the effect of noisebecause the first-order partial differential is sensitive to noise. If R exceeds the specified threshold value, it is considered that this point is a corner point. The samples of the original video images and corner points detected are shown in Figure 2.



(a) Original color image 1

(b) Original color image 2

(c) Detected Corners from image1

(d) Detected Corners from image 2

FIGURE 2. Original color images and corresponding detected corners

3.2. **Filtering the corners.** As a lot of corners exist outside the text regions, the corners in images are seen as noise. In order to filter the corners of noise, the specific steps are conducted as follows:

(1) Assuming the height of image is H, the width of the image is W. When,

$$Corner(i,j) = \begin{cases} 0 & \text{if corner exist at } (i,j) \\ 1 & \text{if corner not exist at } (i,j) \end{cases} \quad (11)$$

$Corner(i,j)$ is the flag at the point of (i,j) representing whether contain corner or not. Here we assume $corner(i,j) = 1$.

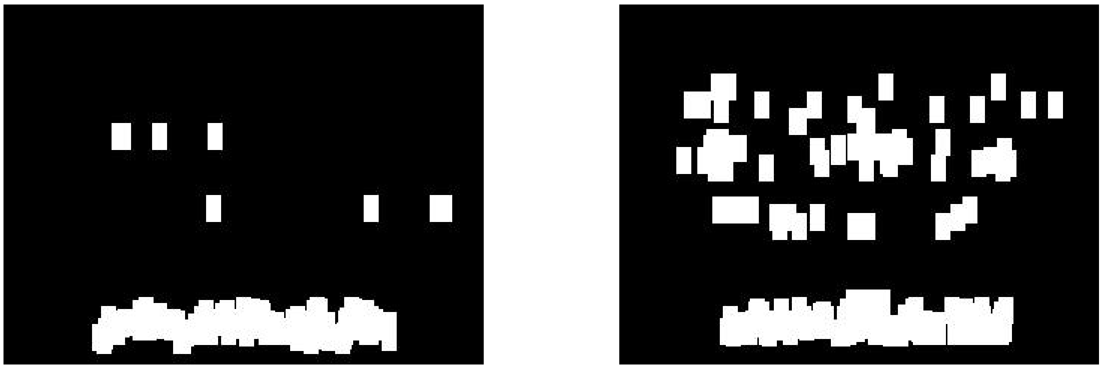
According to the characteristics of the text area in the images, we assume the $Count[i]$ is the sum of all pixels in the row i. when $Count[i] < 4$, all the corners in the row i are removed.

(2) As the distribution of corners is relatively dense, when a corner exists in coordinate (i,j), some other corners also exist within a certain area around point of (i,j) . We assume $Corner(i,j) = 1$, $m \times m$ is the matrix which is centered by (i,j),

$$RoundCount = \sum_{p=i-5}^{i+5} \sum_{q=j-5}^{j+5} Corner(i,j) \quad (12)$$

Where, $5 \leq i \leq H-5, 5 \leq j \leq W-5$. Many experiments show that in the case of $m = 11$ gives the better results. When *RoundCount* < 3 corners with the scope of matrix which is centered by (i,j) are all filtered out.

3.3. **The generation of candidate text areas.** We can see from the corner images in Figure 2, the corner points are located relatively near to each other. This paper puts the corner of the same region together to generate the candidate text area through mathematical morphology dilation [10]. This paper chooses the rectangular elements with the length of 18 and the height of 10 to execute the expansion operation. The area obtained by corner connection operation are shown in Figure 3.



(a) Text regions detected from image1 (b) Text regions detected from image2

FIGURE 3. the candidate text regions detected

3.4. **The validation of the candidate text regions.** The candidate text areas obtained by corner connection (CC for short) described in section 3.3 may still include some non-text regions. As text regions usually have similar characteristics such as the area of the connected region (Ra), the height of the connected region (Rh), the ratio of the connected regions width to height (Ras), and the area occupancy rate (Rs).

$$Ra = sum(CC) \tag{13}$$

$$Rh = Height(CC) \tag{14}$$

$$Ras = \frac{Width(CC)}{Height(CC)} \tag{15}$$

$$Rs = \frac{Area(CC)}{Area(BoundingBox(CC))} \tag{16}$$

As the text regions in video images often have certain area and height, the connected regions with too small area or too small height can be removed. According to the main difference between the text region and non-text region, this paper developed the following typical rules to remove the typical non-text area shown in Table 1.

TABLE 1. The feature differences between text area and non-text area

features / regions	Ra	Rh	Ras	Rs	$Ro(inDegree)$
$Text$	> 50	> 5	> 2	$[1/3, 1]$	$[-10°, 10°]$
$Non-Text$	$[1, 50]\bigcup[20000, +\infty]$	< 5	< 2	$[0.1, 0.25]$	$[-90°, 20°]\bigcup[20°, 90°]$

The parameters of Table 1 are also described as follows:
1: Retain those connected regions with area larger than 50.
2: Remove those connected regions with height less than 5.
3: The non-text regions with ratio of width to height less than 2.

4: Retain those text regions with area occupation more than 1/3.

According to the filtering rules above, the text regions detection results of Figure 2(a) and 2(b) are shown in Figure 4(a) and 4(b) separately.

(a) Text regions detected from image1 (b) Text regions detected from image2

FIGURE 4. the candidate text regions detected

However, when there are leaves and grasses in the background of images and have similar corner density with the text area, the false detection rate will be increased if we use only the heuristic rules described above. For example, the corners map generated from the original image from 5(a) is shown in figure 5(b), and the corresponding image including candidate text regions is shown in figure5(c).

3.5. **The verification of the candidate text regions.** As for the false text regions emerged in figure 5(c), this paper proposes a novel method to verify the candidate text regions, which used the baseline characteristic of the Uyghur text lines to verify whether the candidate text regions are true text regions or not. The Uyghur text baseline characteristic is described as follows: In Uyghur baseline sample shown in Figure 6, the two lines are the Uyghur baseline, the difference of Uyghur letters against to Chinese, English scriptures is the presence of the baselines. The baseline is the horizontal line which Uyghur letter is connected along and the most pixels are located in. Uyghur baseline is generally below the midline, text width ratio (the distance from the baseline to the top is divided by the distance from baseline to the bottom) is generally slightly greater than 1.

The contoured image of the text regions is shown in Figure 7.

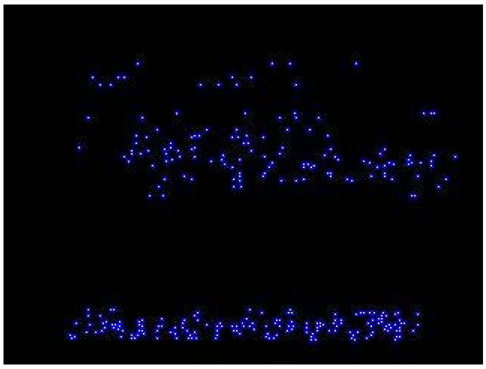
This paper uses the horizontal projection method to detect the baselines of those possible text areas, and the horizontal projection scheme is shown in Figure 8.

In Figure 8, although the area on the left appears small peak, but obviously peak ratio of the upper and lower boundary is much larger than 1. So we can remove the upper regions in figure 5(c), the final text area detected is shown in figure 5(d).

In this paper, the text area validation procedure are conducted as follows: Firstly, horizontal projection process is performed, the formula of the horizontal projection is

$$HP(i) = \sum_{j=1}^{h} E(i,j) \tag{17}$$

(a) Original color image 3



(b) Filtered Corners from image 3

(c) False text area also detected

(d) False text area eliminated

FIGURE 5. The not successfully localized text areas and correction results

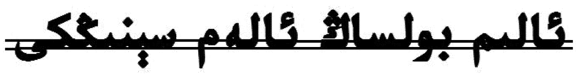


FIGURE 6. the baseline of the Uyghur (the line in the middle)

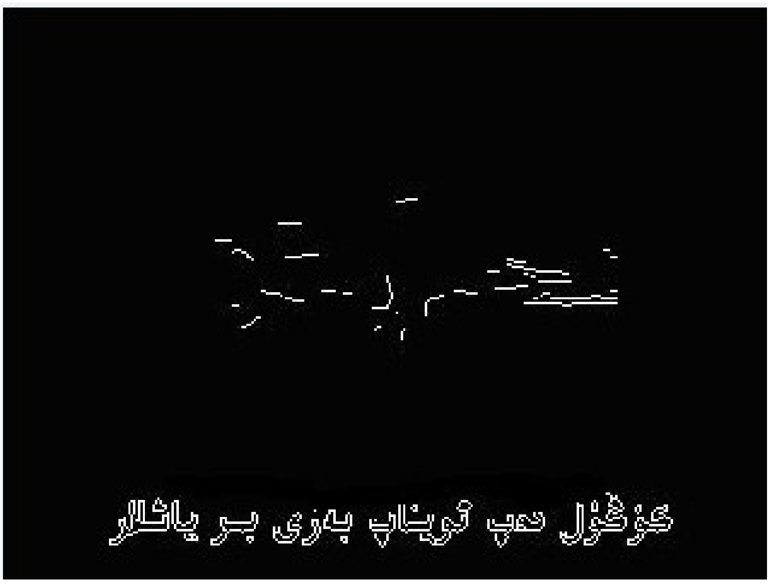


FIGURE 7. the contours of the text regions

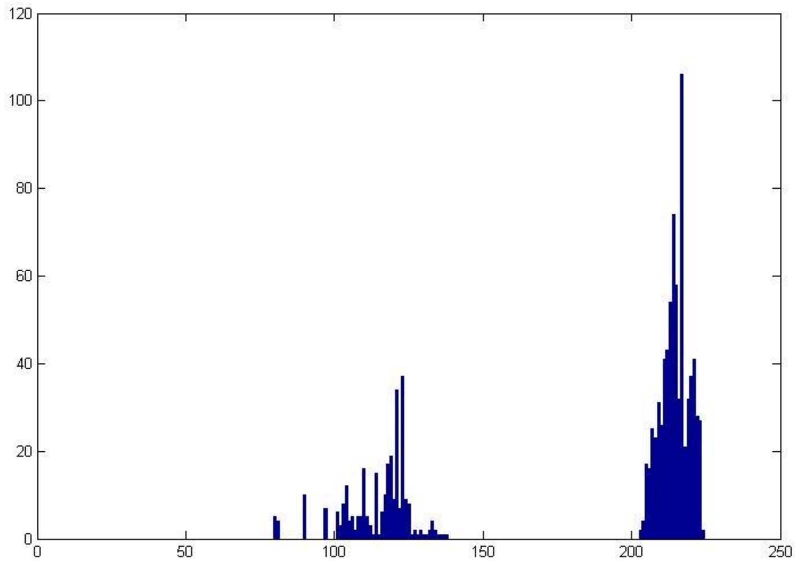


Figure 8. Horizontal projection

Where, E is the edge map of the text regions, $HP(i)$ is the horizontal projection of E. And Finally, candidate text regions are easily verified according to the characteristic of Uyghur baseline.

4. **Experiments and Remarks.** In order to evaluate the proposed text detection algorithm, this paper used 570 Uyghur video frames size of 320*240 in which 475 frames have simple backgrounds and 95 frames have complex backgrounds. The experiments is conducted on a computer with Pentium(R) Dual-Core CPU T4400, 2.20G Hz clock speed and 1G memory using un-optimized Matlab code on the Microsoft Windows XP SP3 platform. This paper uses Recall, Precision and Rate to evaluate the performance. Table 2 and Table 3 compares the results by different approaches that the approach proposed in this paper with no baseline features (named as baseline method), the approach proposed in reference paper [11, 12] and the approach proposed in this paper with baseline features, respectively on different image sets of simple and complex backgrounds.

Table 2. Performance Comparisons (on simple background images)

Methods	Precision (in %)	Recall (in%)	Rate (in s/frame)
Baseline Method	82.4	71.5	4.1
Method of Paper [11]	93.2	75.3	22.0
Method of Paper [12]	96.4	75.0	24.8
Our Method	95.8	78.0	5.1

We can see from Table 2 and Table 3 that the proposed approach holds higher precision comparing with the approach in which the baseline features are not considered, so it is necessary to verify the candidate text region by the characteristic of baseline features. Although the recalls are almost at same level, but the rate of proposed approach is much faster than of reference [11,12]. The proposed algorithm requires at least 5 characters occurred in the candidate text regions, otherwise the text areas would be mistakenly undetected, and it is mainly because of the counts of the corners detected are less due to

Table 3. Performance Comparisons (on complex background images)

Methods	Precision (in %)	Recall (in%)	Rate (in s/frame)
Baseline Method	65.3	68.3	4.8
Method of Paper [11]	62.2	64.4	22.8
Method of Paper [12]	75.6	73.5	25.6
Our Method	77.8	74.0	5.6

the counts of available characters are a few, and the true corners are mistakenly filtered out by row filtering. From Table 3, we can see that considering of the precision, recall and rate, the approach proposed in this paper has a greater advantage and better robustness than the other two methods especially on the images of complex backgrounds.

5. **Conclusions.** Based on the principle of the Harris corner detection and mathematical morphology, this paper proposes a color image text region detection algorithm. As Harris corner has better stability, noise immunity, time complexity than the Susan corner, this paper uses Harris corner and mathematical morphological methods to generate candidate text region. Some typical non-text regions are removed according to the heuristic rules. Finally we use the baseline feature of Uyghur text to verify candidate text regions. The result of experiment shows that the proposed approach can achieve good results in complex background as well as simple background.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China under grant of 61461049. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] M. Long, C. H. Wang, B. H. Xiao, Text detection in natural images based on multi-scale edge detection and classification, *3rd International Congress on Image and Signal Processing(CISP)*, Yantai, China, vol.4, pp.1961–1965, 16-18 October, 2010.

[2] M. H. Li, M. Bai, A mixed edge based text detection method by applying image complexity analysis,*10th World Congress on Intelligent Control and Automation(WCICA)*, Peking, China, pp.4809–4814, 6-8 July, 2012.

[3] Q. D. Zhu, L. Jing and R. S. Bi, Exploration and Improvement of OSTU Threshold Segmentation Algorithm, *8th World Congress on Intelligent Control and Automation (WCICA)*, Jinan, China, pp.6183–6188, 7-9 July, 2010.

[4] K. I. kim, K. Jung and J. H. Kim, Texture-Based Approach for Text Detection Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm, *IEEE Transactions on Pattern Analysis And Machine Intelligence*, vol.25, no.12, pp.1631-1639, 2003.

[5] A. Halidan, Y. Yilihamu, M. Kuerban, Segmentation algorithm of Uygur characters under complex background, *Computer Engineering and Applications*, vol.43, no.20, pp.163–165, 2007.

[6] B. Fabio, T. Domenico, V. Cesare, Improving Harris corner selection strategy, *Computer Vision*, vol.5, no.2 ,pp.87–96, 2011.

[7] F. Yuan, Windows Graphics Programming, *Publishing House of Mechanical Industrial*, Translated by Yingyu Studio, Peking, pp.269–271, 2002.

[8] Z. H. Li, J. Chen, J. Zhang, D. H. Li, The Comparisons of Two Corner Detection Algorithms, *Second International Symposium on Intelligent Information Technology Application*,Shanghai, China, no.2, pp.211–215, 21-22 December, 2008.

[9] R. P. Wurtz, T. Lourens, Corner detection in color images by multi-scale combination of end-stopped cortical cells, in *International Conference on Artificial Neural Networks(ICANN)*, W. Gerstner, A. Germond, M. Hasler, and J. Daniel Nicoud (eds.), Lecture Notes in Computer Science vol.1327, Springer Verlag, pp.901–906, 1997.

[10] D. F. Zhang, Lecture Notes on MATLAB based Digital Image Processing, *Publishing House of Electronics Industry*, Peking, no.7, pp.254–255, 2010.

[11] Q. Ye, W. Gao, Weiqiang Wang, Wei Zeng, A Robust Text Detection Algorithm in images and Video Frame, *Fourth International Conference on Information, Communications and Signal Processing and Fourth IEEE Pacific-Rim Conference On Multimedia*, Singapore, vol.2, no.3, pp.802–806, 15-18 December 2003.

[12] K. Li, A. Hamdulla, Edge and baseline detection algorithm for localization of Uyghur text in images, *Computer Engineering and Applications*, vol.50, no.10, pp.203-207,2014.

[13] D. L. Zhang, J. Q. Qiao, J. B. Li, L. Y. Qiao, Shu-Chuan Chu and John F. Roddick, Optimizing Matrix Mapping with Data Dependent Kernel for Image Classification,Journal of Information Hiding and Multimedia Signal Processing, vol. 5, no. 1, pp. 72-79, January 2014.

[14] G. L. Zhang and C.-L. Yang, Anomaly Detection for Hypaerspectral Imagery Using Analytical Fusion and RX,Journal of Information Hiding and Multimedia Signal Processing, vol. 5, no. 2, pp. 179-186, April 2014.

[15] C. Qin, C. C. Chang and S.-T. Wang, A Novel Lossless Steganographic Scheme for Data Hiding in Traditional Chinese Text Files, Journal of Information Hiding and Multimedia Signal Processing, vol. 5, no. 3, pp. 534-545, July 2014.