International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.4 (2014), pp.409-420
http://dx.doi.org/10.14257/ijsip.2014.7.4.39

Caption Region Detection in Video Images by Improved Corner Detector

Xin Liu[1], Jin Dai[1], Yuanyuan Jia[2] and Rubin Liu[1]

[1]School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[2]College of computer science, Chongqing University, Chongqing, 400044, China
liuxin@cqupt.edu.cn

Abstract

Caption region detection is significant for text segmentation and recognition in video images. In this paper, caption region detection includes the following three steps: text corner detection, candidate text region generation, non-text region elimination. Firstly, an improved Harris corner detection algorithm based on second difference, called sHarris, is proposed to detect text corners. sHarris detector is effective to detect text corners existing in video image. Secondly, morphologic operations are utilized to generate candidate text regions after text corner points are detected. Finally, non-text regions are eliminated from candidate text regions by some validation rules. In the end of this paper, experiments are used to analyze and evaluate the detection results of caption regions, which indicate the effectiveness of the proposed method.

Keywords: *corner point, second difference, text region detection*

1. Introduction

Texts in videos and images can help us to obtain more information, and comprehend these videos and images much better. So, the detection and recognition of them have drawn more and more researchers' attention recent years [1-25]. Text region detection is a key step of text detection and recognition. The following features usually used to detect text region: (1) Color clustering feature. For example, Hase, *et al.,* divided a full color image into several representative binary color images to nominate character strings by using multi-stage relation [4]. This feature is commonly applied to the case that color of texts in images is similar or distinct with other background. If texts have various colors, the color features would be invalid. The feature also becomes ineffective when the color of texts is similar to the surrounding region, or other background region. (2) Edge feature. Because there are different contrasts between background and texts, edge is a widely-used feature for text region detection. Shivakumara et al. proposed an efficient edge based technique for text detection in video frames [9, 18]. However, edge feature is always combined with some other methods to generate and identify candidate text region. Many researchers applied edges to construct connected component for candidate text region generation [6,19, 20, 25, 26]. In addition, Mita and Hori [14, 24] used And Operation between stable edge images and low variance image for candidate text region detection. Lyu, *et al.,* [7] proposed a coarse-to-fine localization scheme based on edge detecion to identify text region. The shortcoming of edge feature is that text region detection is easily influenced by complicated edge information of the background in images or videos. (3) Connected component (CC) feature. As previously stated, to get candidate region, we may need to construct CC after edge detected. Agnihotri, *et*

ISSN: 2005-4254 IJSIP

al., [26] and Yangxing, *et al.*, [27] used various 8-neighborhood-connectivity algorithms to construct connected component. Different from them, Chen, *et al.*, [25], Jung, *et al.*, [19] and Cheolkon, *et al.*, [28] utilized morphologic operation to obtain connected component. Otherwise, Zhu, *et al.*, [5] proposed a non-linear Niblack method to form CCs, and trained these CCs in a cascade of classifiers by Adaboost algorithm to get the detection results. (4) Texture feature. The text and non-text region contain different texture information through the transformation of DCT, FT, WT, *etc.*, so, texture feature can be used to detect the text region. Kumar, *et al.*, putted forward two-class Fisher classifier based on matched wavelet filters to extract the textual areas of an image [3]. Zhao, *et al.*, adopted Neural network to locate the text region based on DWT and the feature of Kurtosis [10]. Texture analysis was performed by Crandall, *et al.*, in 8×8 block-wise DCT for text detection and localization [23]. (5) Corner feature. Strokes of texts contain abundant corners which can be used to detect text region. There are two typical corner detection algorithms, Susan and Harris corner detection algorithms. Guo, *et al.*, [15] designed an algorithm of video text localization based on Susan corner point detection and color clustering. A novel detection and localization method for video texts is proposed based on Harris Corner Response [16]. Zhao, *et al.*, also used Harris corners detection to detect text and caption in videos [17]. Zou, *et al.*, [29] made a comparison between the two methods, and concluded that Harris corner detection algorithm is superior to SUSAN corner detection algorithm on the whole.

In this paper, detecting caption text region in videos is discussed such as News, Films, TV series, TV talk shows. Caption texts often have different contrasts, sizes, colors or languages, and it makes the detection work more difficult. Considering the characters of caption texts itself and the complexity of background in video images, Corner feature is introduced to detect caption text region. The detection work includes the following three steps: text corner detection, candidate text region generation, non-text region elimination. In the first step, an improved Harris corner detection algorithm based on Second difference, called Harris, is proposed to detect text corners. Next, morphologic operations are used to generate candidate text regions after text corner points are detected. Finally, non-text regions are eliminated from candidate text regions by some validation rules.

2. Harris Corner Detector

The Harris operator, developed by Chris Harris and Mike Stephens in 1988 [30], is based on the local auto-correlation function of a signal, which measures the local changes when the signal shifted by a small amount in different directions [30, 31]. Let (x, y) be a pixel point in a digital image, $(\Delta x, \Delta y)$ be the magnitude of shift at (x, y), and $f(x, y)$ be the image function. The auto-correlation function is defined as [30, 31],

$$\begin{aligned} V(x,y) &\approx \sum_{w}\left[f(x_i,y_i)-f(x_i+\Delta x,y_i+\Delta y)\right]^2 \\ &=\begin{bmatrix}\Delta x & \Delta y\end{bmatrix}\begin{bmatrix}\sum_{w}(f_x(x_i,y_i))^2 & \sum_{w}f_x(x_i,y_i)f_y(x_i,y_i) \\ \sum_{w}f_x(x_i,y_i)f_y(x_i,y_i) & \sum_{w}(f_y(x_i,y_i))^2\end{bmatrix}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} \end{aligned} \quad (1)$$

Where, (x_i, y_i) are the points in the Gaussian window w centered on (x, y). The Gaussian template $w(x, y) = \exp -(x^2+ y^2)/2\sigma$, the value σ and size of the Gaussian template need to be chosen. $fx(\cdot, \cdot)$ and $fy(\cdot, \cdot)$ are the partial derivatives in x and y directions, respectively.

Harris corner detector is defined as follows:

$$R(x, y) = A(x, y)\, B(x, y) - [C(x, y)]^2 - k\, [A(x, y) + B(x, y)]^2 \quad (2)$$

Where, $A(x,\ y) = \sum_{w}(f_x(x_i, y_i))^2$, $B(x,\ y) = \sum_{w}(f_y(x_i, y_i))^2$, $C(x,\ y) = \sum_{w} f_x(x_i, y_i) f_y(x_i, y_i)$,
When set a threshold CT for $R(x, y)$ to find local maxima, then, we can get the corners. k is a constant. $k = 0.04$ to 0.06 give the best results according to empirical tests [16, 32].

3. Text region Detection

3.1. Text Corner Detection by Improved Harris Corner Detection Algorithm

Given that the image signal is a sequence of samples from some function f, then $f'(x_i)=\Delta f(x)/\Delta x$. Set the sample spacing is $\Delta x = 1$, derivative of $f(x)$ can be approximated by

$$f'(x_i) = f(x_i) - f(x_{i-1}) \tag{3}$$

The masking template of first difference could be represented as $M' = [-1, 0, 1]$. Then, the 3×3 horizontal and vertical masking templates of first difference are shown in Figure 1.

−1	−1	−1
0	0	0
1	1	1

−1	0	1
−1	0	1
−1	0	1

Figure 1. 3×3 Masking Templates of First Difference

−1	−1	−1
2	2	2
−1	−1	−1

−1	2	−1
−1	2	−1
−1	2	−1

Figure 2. 3×3 Masking Templates of Second Difference

And, the second difference of $f(x,y)$ about x and y can be approximated by

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y), \quad \frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \tag{4}$$

Then, the 3×3 masking templates of second difference can be denoted as Figure 2.

The masking templates shown as Figure 1 and Figure 2 can both be used to detect edges. For example the image in Figure 3, the results of edge detection with first difference and second difference are showed in Figure 4 and Figure 5. The left images are results of vertical edges, and the right ones are results of horizontal edges. In Figure 4, the white lines mark the edges of non-texts detected by first difference masks, which disappear in Figure 5 detected by second difference masks. Given a signal S, if S is upward step edge, downward step edge, bright impulse or line, second derivative masks produce zero-crossings at points of high contrast. But, first derivative masks just produce high absolute values at points of high contrast. So, the second difference is introduced into Harris in this paper.

Figure 3. Original Image

Figure 4. The Vertical and Horizontal Edges by First Difference Masks

Figure 5. The Vertical and Horizontal Edges by Second Difference Masks

Let $f(x, y)$ be the image function, $(\Delta x, \Delta y)$ be the magnitude of shift at (x, y), $(f(x_i+\Delta x, y_i+\Delta y)-f(x_i, y_i))$ and $(f(x_i, y_i)-f(x_i-\Delta x, y_i-\Delta y))$ are shift magnitudes in two opposite directions. The auto-correlation function about the difference of the two shift magnitudes is defined as

$$V(x,y)=\sum_w\left[(f(x_i+\Delta x, y_i+\Delta y)+f(x_i-\Delta x, y_i-\Delta y)-2f(x_i,y_i)\right]^2$$

$$=\sum_w[((f(x_i,y_i)+(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y})f+\frac{1}{2!}(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y})^2 f)-f(x_i,y_i))$$

$$-(f(x_i,y_i)-(f(x_i,y_i)-(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y})f+\frac{1}{2!}(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y})^2 f))]^2$$

$$=\sum_w\left[\begin{bmatrix}\Delta x & \Delta y\end{bmatrix}\begin{bmatrix}\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x\partial y}\\ \frac{\partial^2 f}{\partial x\partial y} & \frac{\partial^2 f}{\partial y^2}\end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}\right]^2$$

$$=\begin{bmatrix}\Delta x & \Delta y\end{bmatrix}(MI)\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}$$

$I=\begin{bmatrix}\Delta x^2 & \Delta x\Delta y\\ \Delta x\Delta y & \Delta y^2\end{bmatrix}$, Let I be an identity matrix. $M=\begin{bmatrix}A & C\\ C & B\end{bmatrix}$,

Where, $A=\sum_w(\frac{\partial^2 f}{\partial x^2})^2+\sum_w(\frac{\partial^2 f}{\partial x\partial y})^2$, $B=\sum_w(\frac{\partial^2 f}{\partial x\partial y})^2+\sum_w(\frac{\partial^2 f}{\partial y^2})^2$,

$C=\sum_w(\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial x\partial y})+\sum_w(\frac{\partial^2 f}{\partial x\partial y}\frac{\partial^2 f}{\partial y^2})$,

$$c(x,y)=\det(M)-k(\mathrm{trace}(M)); \quad (5)$$

Where, $\det(M)=\lambda_1\lambda_2=AB-C^2$; $\mathrm{trace}(M)=\lambda_1+\lambda_2=A+B$; When set a threshold CT for $c(x,y)$ to find local maxima, the corners can be obtained. In Eq. (5), k is a weight value which is also set to be the same as the Harris corner detector.

This paper named the improved Harris corner detector as sHarris.

3.2. Candidate Text Region Generation

To obtain candidate text regions in a video image, the corners which are detected by sHarris detector must be further processed. Input video images, the output of sHarris detector are their corresponding corner images. Some morphologic operations are applied to generate connected components here. Find the position of these corners in original video images, and repeated dilating operation will be implemented. About twelve calculations with dilating operator are necessary. Then, some connected components would be generated at the center of corners which are candidate text regions. The template of the operator is a 3×3 square matrix. Figure 6(b) is the corners detected by sHarris from Figure 6(a). In Figure 6(c), two candidate regions are obtained after twelve dilating operations. One of candidate regions contains all the text information in the video image.

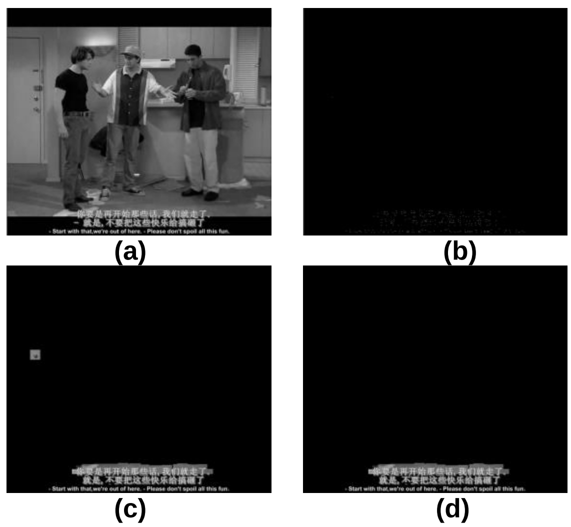


Figure 6. Detected Example: (a) Original Image; (b) Corners Detected by sHarris; (c) Candidate Regions After Dilating Operation; (d) Final Text Region

3.3. Non-text Region Elimination

Label all the candidate regions after they are generated through morphologic operation. Calculate the horizontal diameter (h_d) and vertical diameter (v_d) of candidate regions. Then, the ratio of h_d and v_d (r_{hv}), the ratio of v_d and h_d (r_{vh}) are obtained. Given that the i-th candidate region is connected component c_i, $c_i(x)$ and $c_i(y)$ are the coordinate values of c_i in horizontal and vertical direction. A valid value of h_d, v_d, r_{hv}, r_{vh} could be used to eliminate the non-text region.

A text region can't be judged by only the size of connected component. Before a region is really eliminated, there is another important step: Choose the maximum between h_d and v_d of c_i, and pixel scanning is implemented at the location of 2/5 and 3/5 respectively in the maximum direction. In Figure 6, there are two candidate regions, the scan results of upper one is shown as Figure 7(a), and the lower one as Figure 7(b). The two scan lines will be resemble each other in general structure for a regular text region as shown in Figure 7(b).

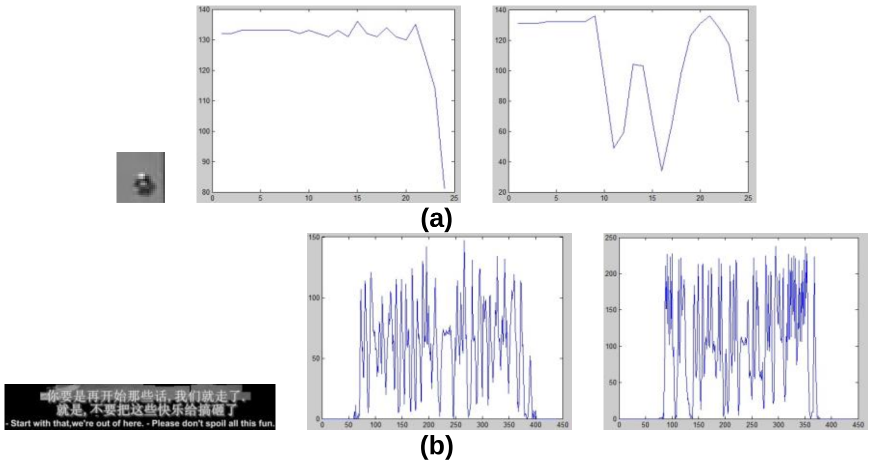


Figure 7. Scan Lines of the Candidate Regions in Figure 6

The whole workflow of the text region detection in this paper is shown in Figure 8.

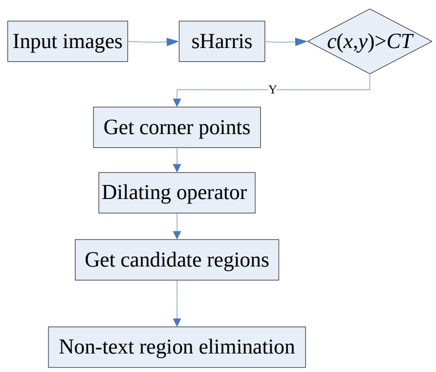


Figure 8. The Whole Workflow of Text Region Detection

4. Experimental Results

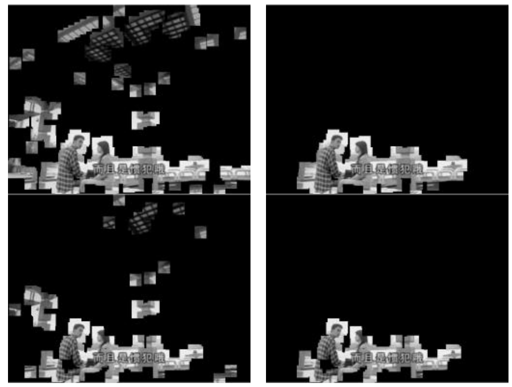
The testing images are a number of real-life video clips including television, movies and TV news, which come from TVB, CCTV, BBC, or some TV series of Europe, America, Japan and Korea. The languages of texts in these videos include Chinese, English, or other languages. Given that $\sigma = 2$ and Gaussian window typically has a radius of 3 times the standard deviation. The results of text region detection are compared and analyzed between Harris and sHarris according the detection workflow in Figure 8. Finally, some criterions are introduced to evaluate the detection results of these methods, including the method in [16] and the method in [7].

When the background and texts have high contrast, or the background is not very complex, detection results of Harris and sHarris are approximate. Conversely, if the background and texts have low contrast, or the background is complex, detection results of sHarris are better than Harris. When *CT* gets different values 10 and 20, text regions in Figure 9 detected by sHarris are explicit in Figure 10. However, detected regions by Harris conclude background in Figure 11.

Figure 9. Testing Image One

(a) (b)

Figure 10. Detection for Testing Image One by sHarris, CT=10 (Top Line), 20 (Bottom Line), (a) Text Candidate Regions, (b) Text Region

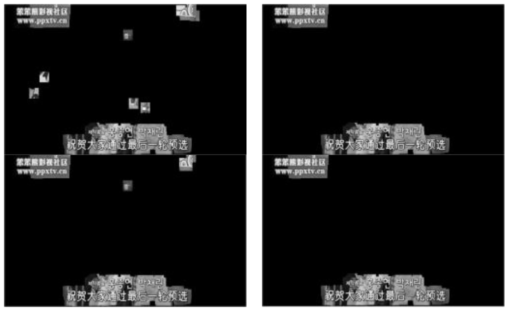


(a) (b)

Figure 11. Detection for Testing Image One by Harris, CT=10 (Top Line), 20 (Bottom Line), (a) Text Candidate Regions, (b) Text Region

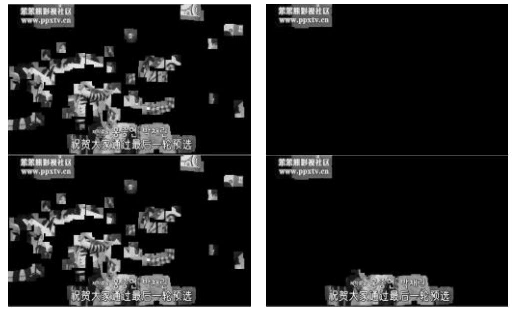
How the detection performance could be satisfactory, whether the bigger or smaller value of *CT*? Background around texts in Figure 12 is complex, and the corners including texts and background are detected by Harris. Dilating operation makes text region and background connected, then, the candidate will be removed as false detection, as shown in Figure 14 and *CT*=10. Quantities of experiments have showed that text regions are well detected by sHarris when *CT* is between 10 and 20.

Figure 12. Testing Image Two



(a) (b)

Figure 13. Detection for Testing Image Two by sHarris, CT=10 (Top Line), 20 (Bottom Line), (a) Text Candidate Regions, (b) Text Region



(a) (b)

Figure 14. Detection for Testing Image Two by Harris, CT=10 (Top Line), 20 (Bottom Line), (a) Text Candidate Regions, (b) Text Region

Finally, some criterions are introduced to compare the detection results. Hundreds of images are randomly selected from aforementioned video clips. Compared methods include sHarris and Harris detector based on the workflow in Figure 8, the method in [16] and the method in [7]. The comparison results are shown in Table 1, in which the criterions are as follows [9, 22, 23]:

(1) Correctly detected text region (CDR): a detected region that contains text.

(2) Falsely detected text region (FDR): a detected region that does not contains text or complete text information, or the background holds a large proportion in the detected region.

(3) Recall rate (RR): RR=CDR/total text region.

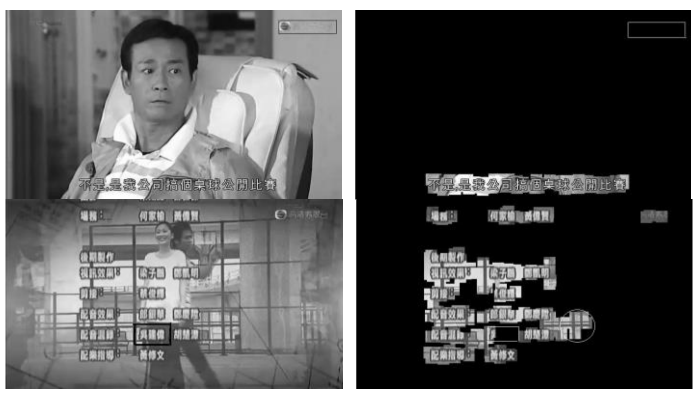
(4) Precision rate (PR): PR= CDR/ (CDR+FDR).

Table 1. Results of Text Detection Using the Following Methods

	Total text region	CDR	FDR	RR	PR
This workflow using sHarris	378	351	14	92.86%	96.16%
This workflow using Harris	378	319	71	84.4%	81.8%
Method in [16]	378	342	19	90.48%	94.74%
Method in [7]	378	337	26	89.15%	92.84%

As shown in Table 1, although this workflow using Harris and method in [16] both employ Harris detector, the results of the latter is better than the former. The reason is that the dilating operation in Figure 8 makes the text region and background to connect together, and the candidate regions contained texts are eliminated as false detected text region, as shown in Figure 14. Images were divided into small blocks in [16], and threshold which are set to be the mean corner response of these blocks are used to get candidate text region. Then, color feature and size range of connected component are used to verify candidate regions and obtain text regions. The detection results by Harris are distinct when thresholds *CT* are different. The improved detector sHarris can detect these points which denote mostly text existing. Therefore, there is only a global threshold *CT*, and the detection time will be shortened. As shown in Table 1, it is concluded that the proposed method has a better performance than other three methods on text region detection. There are some detection examples in Figure 15. Of course, there are some missed and false text detection results shown in Figure 16. The white lines in Figure 16(b) are marked the missed and false detection regions corresponding to the black lines in Figure 16(a). Polychrome caption texts do not be considered in this paper as shown in Figure 17. In addition, texts in book cover or other images can be detected, for example, Figure 18.

Figure 15. Examples of Text Region Detection



(a) Original image **(b) Text region**

Figure 16. Examples of Missed and False Detections

Figure 17. Polychrome Caption texts

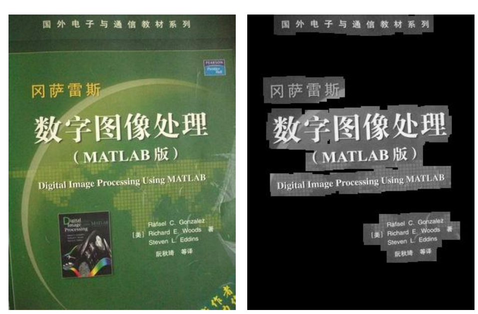


Figure 18. Text Detection for Book Cover

5. Conclusion

As mentioned above, there are different contrasts between caption texts and background in video image, and there are high-density corner points in text region. According to these features, a novel method is proposed based on improved Harris corner detector. Second difference is introduced into the improved Harris corner detector, called sHarris. Then, morphologic operations are performed to generate candidate text region, and the size ranges of these regions and the similarity of scan lines in candidate regions are used to validate text region and eliminate the non-text region. The proposed method is employed to detect caption text region, especially caption regions in some TV series or TV news. Finally, some experiments are used to show that the proposed method is robust to detect caption text region when caption texts have different contrasts, sizes, colors or languages.

Acknowledgement

This work is supported by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant no. KJ1400409, no. KJ1400412; and Natural Science Foundation Project of CQ CSTC under Grant no. cstc2013jcyjA40009; and the National Science Foundation of Chongqing University of Posts and Telecommunications under Grant no. A2014-82.

References

[1] D. Chen, J. Luettin and K. Shearer, "A survey of text detection and recognition in images and videos", Institut Dalle Molled'Intelligence Artificielle Perceptive (IDIAP) Research Report, IDIAP-RR 00-38, **(2000)**.
[2] K. Jung and K. I. Kim, "Text information extraction in images and video: a survey", Pattern recognition, vol. 37, **(2004)**, pp. 977-997.
[3] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and mrf model", IEEE Transactions on Image Processing, vol. 16, **(2007)**, pp. 2117-2128.
[4] H. Hase, T. Shinokawa and M. Yoneda, "Character string extraction from color documents", Pattern recognition, vol. 34, **(2001)**, pp. 1349-1365.
[5] K. Zhu, F. Qi, R. Jiang, L. Xu, M. Kimachi, Y. Wu and T. Aizawa, "Using adaboost to detect and segment characters from natural scenes", Proceedings of the International Workshop on Camera-based Document Analysis and Recognition, **(2005)** August, pp. 52-58.
[6] Y. Liu, S. Goto and T. Ikenaga, "A robust algorithm for text detection in color images", International Conference on Document Analysis and Recognition, vol. 1, **(2005)** August 29-September 1, pp. 399-403.
[7] M. R. Lyu, J. Song and M. Cai, "A comprehensive method for multilingual video text detection", localization, and extraction, IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, **(2005)**, pp. 243-255.
[8] Q. Ye, Q. Huang, W. Gao and D. Zhao, "Fast and robust text detection in images and video frames", Image and Vision Computing, vol. 23, **(2005)**, pp. 565-576.
[9] P. Shivakumara, W. Huang and C. L. Tan, "An efficient edge based technique for text detection in video frames", The Eighth IAPR International Workshop on Document Analysis Systems, **(2008)** September, pp. 307-314.
[10] T. Zhao, G. Sun, C. Zhang and D. Chen, "Study on video text processing", IEEE International Symposium on Industrial Electronics, **(2008)** June 30-July 2, pp. 1215-1218.
[11] W. Kim and C. Kim, "A new approach for overlay text detection and extraction from complex video scene", IEEE Transactions on Image Processing, vol. 18, **(2009)** February, pp. 401-411.
[12] G. Miao, G. Zhu, S. Jiang, Q. Huang, X. Changsheng and W. Gao, "A Real-Time Score Detection and Recognition Approach for Broadcast Basketball Video", IEEE International Conference on Multimedia and Expo, **(2007)** July, pp. 1691-1694.
[13] X. S. Hua, P. Yin and H. J. Zhang, "Efficient video text recognition using multiple frame integration", Citeseer, International Conference on Image Processing, vol. 2, **(2002)**, pp. 1522-4880.
[14] T. Mita and O. Hori, "Improvement of video text recognition by character selection", Sixth International Conference on Document Analysis and Recognition, **(2001)** September, pp. 1089-1093.

[15] G. Guo, J. Jin, X. Ping and T. Zhang, "Automatic video text localization and recognition", Fourth Internaltional Conference on Image and Graphics, **(2007)** August, pp. 484-489.
[16] L. Sun, G. Liu, X. Qian and D. Guo, "A novel text detection and localization method based on corner response", IEEE International Conference on Multimedia and Expo, **(2009)** June 28 – July 2, pp. 390-393.
[17] X. Zhao, K. Lin and Y. Fu, "Text From Corners: A Novel Approach to Detect Text and Caption in Videos", IEEE Transaction on Image Processing, vol. 20, **(2011)** March, pp. 790-799.
[18] P. Shivakumara, W. Huang, T. Q. Phan and C. L. Tan, "Accurate video text detection through classification of low and high contrast images", Pattern recognition, vol. 43, **(2010)**, pp. 2165-2185.
[19] C. Jung, Q. Liu and J. Kim, "Accurate text localization in images base d on SVM output scores", Image and Vision Computing, vol. 27, **(2009)**, pp. 1295-1301.
[20] Y. Song and W. Wang, "Text Localization and Detection for News Video", Second International Conference on Information and Computing Science, vol. 2, **(2009)** May, pp. 98-101.
[21] S. Chowdhury, S. Dhar, A. Das, B. Chanda and K. McMenemy, "Robust extraction of text from camera images", 10th International Conference on Document Analysis and Recognition, **(2009)** July, pp. 1280-1284.
[22] Y. Su, Z. Ji, X. Song and R. Hua, "Caption Text Location with Combined Features for News Videos", Internatioanl Workshop on Education Technology and Training and Geoscience and Remote Sensing, vol. 1, **(2008)** December, pp. 714-718.
[23] D. Crandall, S. Antani and R. Kasturi, "Extraction of special effects caption text events from digital video", International Journal on Document Analysis and Recognition, vol. 5, **(2003)**, pp. 138-157.
[24] O. Hori and T. Mita, "A robust video text extraction method for character recognition", Systems and Computers in Japan, vol. 36, **(2005)**, pp. 87-96.
[25] D. Chen, J. M. Odobez and H. Bourlard, "Text detection and recognition in images and video frames", Pattern recognition, vol. 37, **(2004)**, pp. 595-608.
[26] L. Agnihotri and N. Dimitrova, "Text detection for video analysis", IEEE Workshop on Content-Based Access of Image and Video Libraries, **(1999)**, pp. 109-113.
[27] L. Yangxing and G. Satoshi, "A Robust Algorithm for Text Detection in Color Images", Eighth International Conference on Document Analysis and Recognition, vol. 1, **(2005)** August, pp. 399-403.
[28] J. Cheolkon, L. Qifeng and K. Joongkyu, "Accurate text localization in images based on SVM output scores", Image and Vision Computing, vol. 27, **(2009)**, pp. 1295-1301.
[29] L. Zou, J. Chen, J. Zhang and L. Dou, "The Comparison of Two Typical Corner Detection Algorithms", Second International Symposium on Intellegent Information Technology Application, vol. 2, **(2008)** December, pp. 211-215.
[30] C. Harris and M. Stephens, "A combined corner and edge detector", Alvey Vision Conference, **(1988)**, pp. 147-152.
[31] K. G. Derpanis, "The Harris corner detector", York University, **(2004)**.
[32] D. Kerr, S. Coleman and B. Scotney, "A Space Variant Gradient Based Corner Detector for Sparse Omnidirectional Images", Journal of Mathematical Imaging and Vision, vol. 38, **(2010)**, pp. 119-131.

Authors

Xin Liu, she works at Chongqing University of posts and telecommunications, China. She graduated as Doctor in Computer Science from Chongqing University, China. Her research domain includes image processing, grey system and machine learning.

Jin Dai, he works at Chongqing University of posts and telecommunications, China. He graduated as Doctor in Computer Science from Chongqing University, China. His research domain includes intelligent information processing and natural language processing techniques.

Yuanyuan Jia, she is currently PhD student in College of Computer Science in Chongqing University, China. She received her B. S degree in Computer Science and Technology, Hebei Normal University, 2010. She is doing a combined Master-PhD program since 2010. Her major interests are in the field of image processing.

Rubin Liu, he is currently an undergraduate student in College of software in Chongqing University of posts and telecommunications, China. His major interest is application of software engineering.

