ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

A Novel Text Detection Technique Based On Corner Response

[1] Nishant Singh, [2] Charul Bhatnagar
[1][2] CEA Department GLA University Mathura (UP)

Abstract - **Data about the text incorporated in pictures and videos have a cardinal role in semantic assessments. In this paper, Novel text Detection and Localization (NTDL) algorithm is presented for text detection and localization in the background that incorporates noise in it. This algorithm is constituted by corner response. In contrast to the portions that do not contain text, there are some edges that are dense and corners in portions having text. So, some related strong reactions from regions of text and minimal reactions from portions that do not have text. These reactions furnish some cues that are highly useful for text detection and localization of pictures. By employing a basic schema constituted on the threshold, we obtain regions of candidates for text. These portions are evaluated by interlinking several characteristics like size and color of linked devices. Lastly, the text line is identified exactly by the projection of response from corners. The outcomes from illustrations present exactness speed and recalling for suggested methodology and we have obtained the recall of 93.25%, accuracy 97.96% and speed of 98.14% that greatly enhanced the performance of the system.**

Keywords: Edge detection; corner detection; connected component; textual approach.

I. INTRODUCTION

Text detection in pictures and videos has gained attention of various explorers from past few years. Texts are the source to furnish cardinal and instinctive data and they have a close relation to videos. Thus, it is very simple and easy to evaluate the logics of videos related to the data accumulated in texts. But this can be achieved if detection of text is done efficiently and inaccurate manner.

According to the survey of digital media it has been found that the extracted information can be used in a wide range of applications like Content based Image Retrieval System and Reading Foreign Language Text etc. Basically, the images are divided into three main categories: Scene images, document images, and born-digital images [1]. Scene images contain the text, such as advertising boards, banners, which is captured naturally when the scene images are taken by the camera, therefore scene text is embedded in the background as a part of the scene. Document images are the image format of the document. Born-digital images are generated by computer software and are saved as digital images. Compared with document images and scene images, there are more defects in born-digital images, such as more complex background, low resolution, compression loss. So, during the text extraction from born-digital images, it is very difficult to differentiate text from the background. Now if we talk about the text that appears on an image, then it is classified into two parts- Scene Text and Artificial Text [2]. Scene Text is a part of an image while Artificial Text is produced separately and laid over the image. Both the categories have their own methods of detecting text [3].

This paper is organized as follows: In section II, theoretical aspects of the existing work are described. The proposed work is explained in section III. Section IV discuss and compare the various experimental results and conclusion and future work are discussed in section V.

II. THEORETICAL ASPECTS

A. Existing Work

In the previous work text detection is divided into 3 sub-parts: First, technique constituted on interlinked devices estimate some confinements on sections containing text like consistent colors, defined sizes and alignments in space are taken into account. By linking the characteristics of size and color, Yu and Jain et al [4] detects the color as interlinked devices in frames of videos. The main issue we deal with this situation is that it is not generally incurred for every kind of picture. Assize, shape and color in a text may vary from pictures; Secondly, methodologies constituted on texture or edges [5] assume that there is smoothness in backgrounds than in portions of texts. So, one is able to distinguish between portions that have text and the portions that do not have as per the intensity of texture and edges. But the deduction of noise from complicated backgrounds is an issue which we still face.

A technique is suggested by Lyu et al [6] to identify texts that are in multiple language and resolution. Maps on the edges of Sobel are taken as characteristics and a defined particular threshold is opted to discover the text's candidate area. Techniques constituted on main moments of blocks of pictures are suggested by Li et al [7] they revealed that text

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

detection on various scales is possible by employing these attributes. Lastly, techniques constituted for learning by machines incur attributes which are pulled out from portion containing the text and the regions that do not have text to assist machine of supporting vector or a neutralized network and text detection become an issue that needs attention. An SVM based algorithm by invading reactions from stroke filter is suggested. The SVM based paradigm is suggested by Hu et al [8] constituted on difference on extreme gradients and other linked devices attributes that can attain fake relative rates.Conflicts that are coming across on the methodology of learning by machines are that it requires much quantity of samples of various types.

In this paper, a new technique for text detection and localization with SVM constituted on corner response [9]-[11] is suggested. CR is obtained as the outcome of a unique filter that can be pulled up the gray value of corners in pictures. The maximum of local in CR is said to be as Harris CR. Even though CR doesn't accumulate particular positioning data of the corner points, it reveals the assumption of a pixel that can be at corner point. It is seen as appropriate attribute for detection of text. It works fine even under several resolutions and thus texts in several sizes can be identified. Accumulated with several characteristics in consistent colors and interlinked device's size have given good outcomes.

B. Related Concepts

1) Text Region Detection and Localization

This portion gives explanation about techniques to identify portions of text in pictures constituted on the reactions of corner points. It is comprised of 3 levels: First, calculating reaction of corner points in the area of multi scaled and putting in a threshold [12] to attain a candidate portion of the text; Second, evaluating the portion of candidate by linking characteristics of size 7 color attributes; Lastly, identifying lines of text by making use of bounding box.

2) Computing Corner Response at Multiple Scale

Corner is defined as a unique feature in a 2-D plane that attains greater curvature in the boundary of portions. It can be identified by obtaining maxima of local in CR. At corner points, frame of video, that is deployed to produce components that are interlinked. Instead, they invade quantity of corner points, but not CRs to distinguish regions that have text and those that do not have text. Superiorities to deploy CR in respect to the quantity of CRs are explained in 2 aspects. At first, we do not need to acquire the location of corner points as we are concerned with the portion that can possess CRs. CR is an attribute that reveals the possibility of occurrence of corner points. Secondly, CR can obtain a consistent value for every pixel that can be handled easily for some procedures. In this, we explained the computation of CR in short. Provided a picture I(x, y), the standard form of CR am presented in equation (1).

$$\Re(x,y) = \sum u,v\ W(u,v)[I(x+u,y+v)] \quad (1)$$

In this, window function is determined as W(u, v). It is explained that CR can be calculated by approximations by the following formula.

$CR(x,y) = A(x,y)B(x,y)-(C(x,y))^2-weight*(A(x,y)+B(x,y))^2$ (2)

In this, A(x, y), B(x, y) and C(x, y) are calculated as:

(3) $$A(x,y) = W(u,v) * (\nabla_x I(x,y))^2$$

(4) $$B(x,y) = W(u,v) * (\nabla_y I(x,y))^2$$

(5) $$C(x,y) = W(u,v) * (\nabla_x I(x,y))^2 (\nabla_y I(x,y))^2$$

As per the above provided equation, ∇x I(x, y) and ∇y I(x, y) are amplitudes at edges in the direction of x and y by which we can attain operator of Sobel. W (u, v) is a template for Gaussian to smoothen it.

(6) $$W(u,v) = \exp - (u^2 + v^2/2\sigma)$$

And we can make selection of value of σ and size of the template. Various pictures and their corresponding CR are presented in Figure 1. We can attain the portions of text that comes out from background in a few pictures as in Figure 1 (a). Regions from complicated background can be identified as in Figure 1(b), but its elimination can be performed in the underneath steps. It is observed in Figure 1 (c) that CR is not much solid as a font of texts is very big. But CR can be made strong by down sampling the picture as it is presented in Figure 1(d). To identify the pictures in greater fonts, pyramid formed by scaling down the pictures is formulated at first and that calculations on CR [13], [14] are performed at every level of the pyramid. Font of the text and resolution of the picture are the 2 factors that determine the factor of scaling and level of the pyramid.

3) Text Candidate Region Generation

Region of candidate of text is produced that is constituted on CR. It is segregated in 2 blocks of small size. In this experiment, we select blocks of size of 8x8. Value of aggregated intensity for every block in CR Mblk is computed. Threshold Tblk is designated for Mblk. If the equation provided below gets satisfied, this present block is assumed to be one of them in the regions of candidate of text. Value of

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

threshold as deployed here is minimal as value of several pixels is 0 in CR picture.

$$M_{blk} > T_{blk} \quad (7)$$

$$\text{)lk} = \frac{1}{H * W} \sum_{x=0,y=0}^{H,W} CR(x,y) \quad (8)$$

4) Text Verification Using Color Constraints

The text contained in the picture has a uniform color and it is different from that of the background. This divergence in gray value of the character of text is minimal in contrast to the background and there is a much difference in gray value in background and text. Superiority is gained for these characteristics to banish blocks containing noise.
On every block, a threshold TCR for each pixel in CR and gathering of points Rb and Rt in every block.

$$\text{R(x,y)} \geq T_{CR}, (x,y) \in Rt \quad (9)$$

$$\text{R(x,y)} < T_{CR}, (x,y) \in Rb \quad (10)$$

Then the values of Dis and Dev are computed. The Gray value of g (x, y) is of the pixel.

$$\text{ev} = \sqrt{\frac{1}{Nt} \sum_{(x,y) in\ Rt} (g(x,y) - Mt)^2} \quad (11)$$

$$\text{is} = |Mt - Mb| \quad (12)$$

Mb and Mt aggregate gray value in Rb and Rt. At last, it is evaluated if the condition as given gets satisfied.

$$\text{s} > \text{TdisDev} \quad < \text{Tdev} \quad (13)$$

5) Text Line Localization

The text regions are already obtained after evaluation. But there is irregularity in shape of the portion and there is required for refinement to a rectangle that is aligned. As the portions of text are aligned horizontally or vertically in video, technique of projection is explained to attain the exact coordinates of lines of text. CR determines the projections that are preceded by the following steps. For every portion that has components interlinked, extension is done by 4 pixels along the border line. A bounding box is deployed for identification of portions. For every column or row in the bounding box, summation for the intensity of CR is computed and a curve is obtained in Figure 2(b). As a space lies there in characters, smoothening of the curve should be performed. A filter of Gauss [15], [16] is opted for purpose of smoothening. Lastly, a threshold is employed to identify the coordinates of a line of text. An illustration is presented in projections that are vertical and horizontal directions. It is observed that if the value of the threshold is put up as 30% of that of peak, outcomes are better as in figure 1(b).

III. PROPOSED WORK

In the design as presented, NDTL algorithm helps to identify the text. There is a scope for improvisation in recalling, speed and exactness. As per the past work, figures for recalling, speed and exactness are 91.63%, 70.20% and 95.86%. Improvisation is done with these attributes for text detection. We interlink NTDL which is abbreviated form of novel text detection and localization and SVM [17]-[19]. After linking these methodologies, speed and accuracy gets enhanced.

A.Algorithm of Proposed Methodology

Step 1 Actual image is transformed to down sampled image (if necessary).
Step 2 Calculating response from corner in multi scale spacing and threshold to obtain portion of text.
Step 3 Evaluating portion of candidate by interlinking size and color characteristics.
Step 4 Applying of SVM for improvisation in performance.
Step 5 Accessing line of text by making use of bounding box.
B.Detailed Description of Methodology

This portion gives detailed explanation about suggested methodology to obtain portions containing text in pictures that are constituted on response from corners. Initially a printed document is taken as an input image and if the font is very large then we down sampled the image because corner response technique is not as efficient for large fonts. Corner points are computed which is a intersection of edges present in an image with the help of corner response detector. Then threshold value is computed for the categorization of textual and non textual region. Color based verification is applied as text is of uniform color so it gives better identification of textual data. Then SVM is applied for the classification of identified text into different classes. Finally we have applied the projection based verification as the text can be of different orientation.

IV. EXPERIMENTAL RESULTS

We have performed our work on the Windows 7 operating system and Image Processing tool of MATLAB 2010 is used.

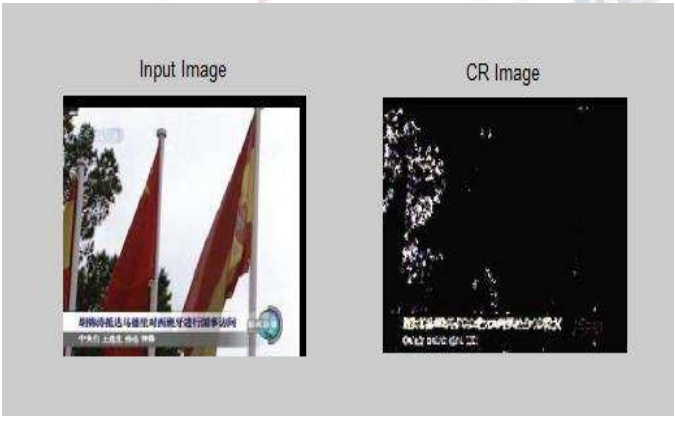
ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

The suggested text detection and localization with SVM algorithm is evaluated on several images based on actual life that is comprised of news of television and movies. Language of texts in the videos is either Chinese or English.

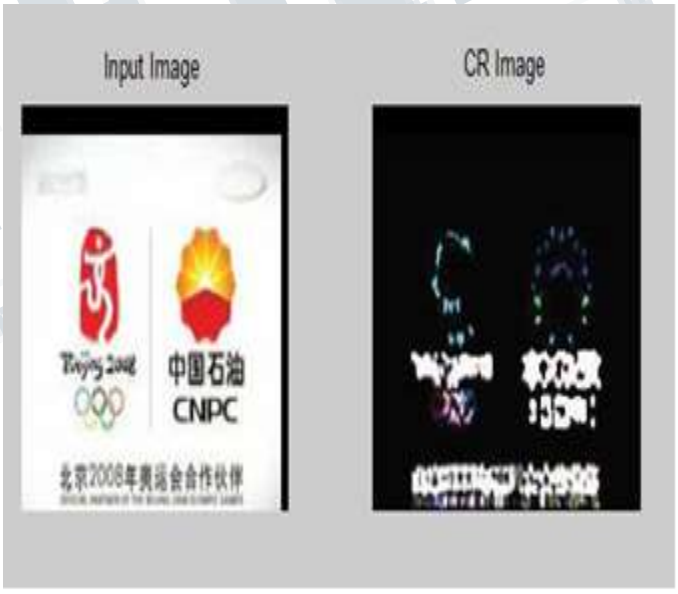
In contrast to the previous work, this methodology is signified on different criteria: First, it has more robustness than other edges and techniques constituted on textures. As the impact of CR is more and it has deduced noise at the level of extraction of characteristics. So, even if the backgrounds have some complicacies, we still will be able to identify the text from it. Second, it is able to identify text of bigger fonts as the features that are deployed are able to work in coarse and fine resolutions. Last, in contrast to other methodologies, this technique is most efficient. As computation of CR is much simpler and there is no need to determine coordinates of corner points.

(a)



(b)

(c)



(d)

Fig. 1. Corner reactions for few pictures: (a). CR for English texts; (b). CR for Chinese texts in noisy background; (c). CR for large texts; (d). CR for large text after down-sampling.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

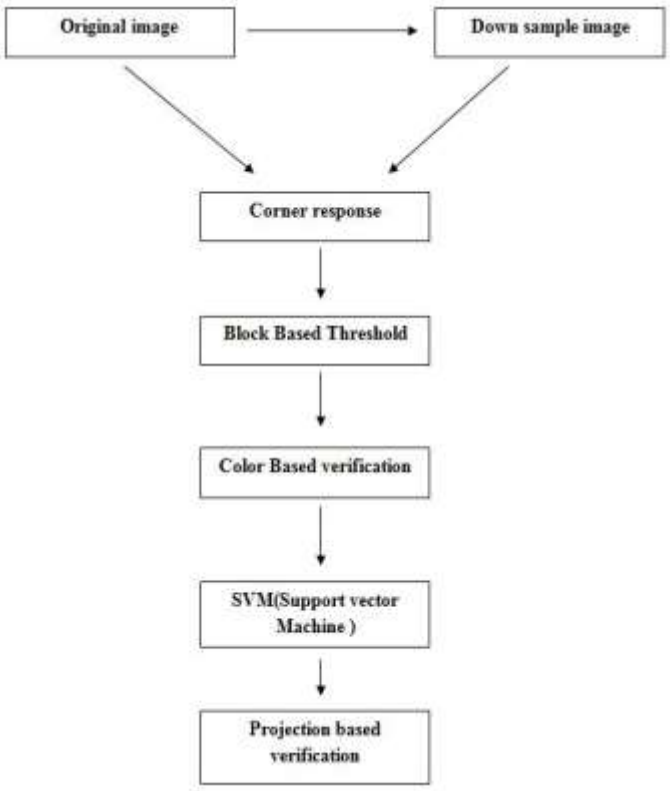


Fig. 2. Proposed methodology flow

Figure 3 shows the experimental result in which two images are considered. One with noisy background and another with large English text. Candidate regions are computed of these images and then verified with the help of SVM is shown.
We have also compared our approach with 4 existing approaches and considered 3 parameters for performance evaluation, namely recall, precision and speed shown in Table 1. Recall is defined as the total number of correctly detected text lines divided by the total number of all text lines, whereas precision is defined as the total number of correctly detected text lines divided by the total number of all detected text lines including the false alarm. In our approach we have greatly enhanced the performance of the system which helps in effectively and efficiently retrieve the information from the text.

V. CONCLUSION AND FUTURE WORK

The document suggests text detection & localization with the SVM method that is constituted on CR of a picture.

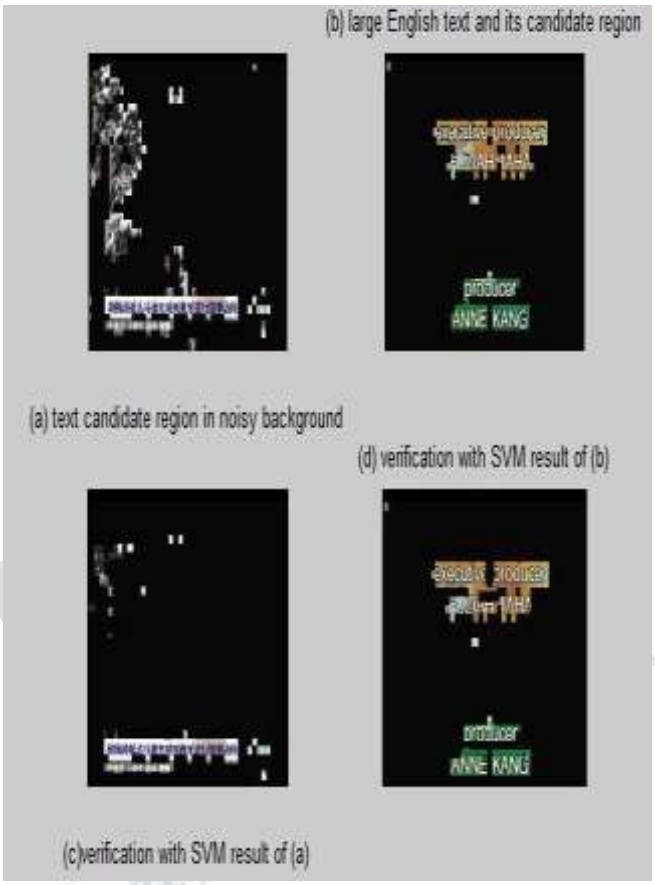


Fig. 3. (a) Text candidate region in noisy background; (b) Large English text ; (c) Verification with SVM results of (a) ; (d) Verification with SVM results of (b)

TABLE 1. PERFORMANCE COMPARISON

Methods	Recall %	Precision %	Speed(ms) %
Lyu in [6]	90.69	90.77	96.10
Li in [7]	91.32	92.47	77.50
Harris in [9]	91.84	94.58	105.40
Sun in [13]	91.63	95.86	70.20
Our Approach	93.25	97.96	98.14

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

There are 5 general steps in our suggested methodology. It is comprised of 5 levels: First, the actual image is transformed in down sampled pictures. Second, calculating the response from corner in multi special & threshold to attain a portion of text. Third, evaluating the region of the candidate by interlinking the size and color. Fourth, SVM implemented for improvisation of performance. Fifth, identifying the text line by making use of bounding box. Outcomes from illustrations reveal that correctness of the suggested technique on text detection & localization method. Results can be improved by improvement for SVM in the future.

REFERENCES

[1] R. C Gonzalez, (2009), Digital Image Processing, Pearson Education India.

[2] Jiawei Han, & Miche Line Kamber, (2006) Data Mining: Concepts and Techniques, 2nd ed., Elsevier.

[3] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. Image Processing, vol. 9, no. 1, pp. 147– 156, 2000.

[4] B. Yu and A. Jain, "A generic system for form dropout," IEEE Trans. Pattern Analysis And Machine Intelligence, vol. 18, pp. 1127–1134, 1996.

[5] A.K. Jain and B. Yu, "Automatic text location in images and video frames," Pattern Recognition, vol. 31, no. 12, pp. 2055–2076, 1998.

[6] M.R. Lyu and J.-Q. Song, "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE Trans. Circuits and System for Video Technology, vol. 15, no. 2, pp. 243–255, 2005.

[7] Xiaojun Li,WeiqiangWang, Shuqiang Jiang, Qingming Huang, and Wen Gao, "Fast and effective text detection," in Proc. of the IEEE International Conference on Image Processing (ICIP), 2008. iyan Hu and Minya Chen, "Adaptive fre/spl acute/chet kernel based support vector machine for text detection," in Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2005.

[8] Xian sheng Hua, Xiang rong Chen, Liu Wenyin, and Hong jiang Zhang, "Automatic location of text in video frames," in Proceeding of ACM Multimedia 2001 Workshops: Multimedia Information Retrieval (MIR2001), 2005.

[9] C.G. Harris and M.J. Stephens, "A combined corner and edge detector," in Proceeding of the 4th Alvey Vision Conference, 1988, pp. 147–152.

[10] Rainer Lienhart and Axel Wernicke, "Localizing and segmenting text in images and videos," IEEE Trans. Circuits and System for Video
Technology, vol. 12, pp. 256–267, 2002.

[11] J Gllavata, R Ewerth and B Freisleben, ―Finding Text in Images via Local Thresholding‖, IEEE, vol 1, pp. 539-542,Dec 2003.

[12] Li Sun, Guizhong Liu, Xueming Qian, Danping Guo," A Novel Text Detection And Localization Method Based On Corner Response " ©2009 IEEE

[13] L.Sun, G.Liu, X.Qian and D.Guo, ―A Novel Text Detection and Localization Method based on Corner Response‖, IEEE, vol 2, pp. 390-393, Jun-Jul 2009.

[14] X.Zahu, K.Lin, Y.Fu, Y.Hu, Y.Liu and T.Huang, ―Text From Corners: A Novel Approach to Detect Text and Caption in Videos‖ IEEE, vol 20, pp. 790-799, Mar 2011.

[15] Xueming Qian, Guizhong Liu, Huan Wang, and Rui Su, "Text detection, localization, and tracking in compressed video," Signal Processing: Image Communication, vol. 22, no. 9, pp. 752–768, 2007.

[16] Puri and S Kaushik, An Enhanced Fuzzy Similarity Based Concept Mining Model Using Feature Clustering‖, IEEE, pp. 1-6, March 2012

[17] S Puri and S Kaushik, A Technical Study and Analysis on Fuzzy Similarity Based Models for Text Classification‖, IJDKP, vol. 2, No. 2, pp. 1-15 Mar 2012.

[18] S. Puri and S. P. Singh, "Sentence Detection and Extraction in Machine Printed Imaged Document using Matching Technique," RAECS, unpublished, 2015.