Multi-oriented text detection in scene images

Article *in* International Journal of Pattern Recognition and Artificial Intelligence · February 2013
DOI: 10.1142/S0218001412550105

M. Basavanna
Davangere University

Palaiahnakote Shivakumara
National University of Singapore

International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 26, No. 7 (2012) 1255010 (19 pages)

DOI: 10.1142/S0218001412550105

World Scientific
www.worldscientific.com

MULTI-ORIENTED TEXT DETECTION IN SCENE IMAGES

M. BASAVANNA

School of Computing Science, VELS University
Chennai, Tamil Nadu, India
Basavanna_m@yahoo.com

P. SHIVAKUMARA*

School of Computing
National University of Singapore, Singapore
shiva@comp.nus.edu.sg

S. K. SRIVATSA

St. Joseph College of Engineering
Chennai, Tamil Nadu, India
profsks@rediffmail.com

G. HEMANTHA KUMAR

Department of Studies in Computer Science
University of Mysore, Mysore, Karnataka, India
ghk.2007@yahoo.com

Received 21 May 2012
Accepted 25 September 2012
Published 21 December 2012

We present a new run-length based method for multi-oriented text detection in scene images. We consider one ideal Sobel edge image of the horizontal text image to compute run-lengths for multi-oriented text images. Then the method proposes a Max−Min clustering to find ideal run-lengths that represents text pixel from an array of run-lengths of ideal image. The run-lengths computed for the input multi-oriented and horizontal text images are matched with the ideal run-lengths given by the Max−Min clustering to find potential run-lengths. The boundary growing method is introduced to traverse multi-oriented text lines given by the potential run-lengths and then the method eliminates false positives to clear the background using angle-proximity features of the text blocks. The non-horizontal text image is rotated to horizontal direction based on angle of the text lines to ease the implementation. The method explore new idea based on zero-crossing to separate text lines from the touching text lines given by the boundary growing method. The proposed method is tested on our own multi-oriented scene data captured by high resolution camera and mobile camera, and the benchmark database (ICDAR 2003 competition scene images) to evaluate the performance of the proposed method. The

*Corresponding author.

results are compared with the existing methods to show that the proposed method outperforms the existing methods in terms of measures.

Keywords: Multi-oriented text; run-lengths for multi-oriented text; boundary growing; zero crossing; scene text detection.

1. Introduction

With the increasing use of digital image capturing devices, such as digital cameras, mobile phones, and PDAs, content based image analysis techniques are receiving intensive attention in recent years. Among all the content in images, text information has inspired great interests because it bridges the gap between high and low level features and finds wide applications in license plate reading, sign detection, translation, mobile text reading, content-based web image search and so on. Jung *et al.*[13] have proposed text extraction system with four stages: text detection, text localization, text extraction and enhancement, and recognition. Among these stages, text detection is critical to improve the overall method performance. In the last decade, many methods have been proposed to address image and video text detection and localization problems, and some of them have achieved impressive results for specific applications.[7,9,12,21,25,27] However, accurate text detection in natural scene images is still challenging due to the variations of text font, font size, color and alignment orientation and it is often affected by complex background, illumination changes, image distortion and degrading.[21]

There are methods in document analysis for detecting and extracting text in the scanned images, however, these methods expect not only high resolution but also plain background images to segment and extract the text lines accurately. Since these methods use connected component analysis and projection profiles, it cannot be used directly to detect text lines in natural scene images where complex background, multi-color, multi-oriented text are present.[12] Therefore, the methods proposed in document analysis are not suitable for this application.

2. Literature Review

The existing methods of text detection and localization can be roughly categorized into two groups: region based and connected component based. Region based methods use texture analysis as features and classifiers to classify text and non-text region because text regions have distinct textural properties from non-text ones, these methods can detect accurately even when images are noisy. On the other hand connected component based methods directly segment candidate text components by edge detection or color clustering. The non-text components are then pruned with heuristic rules or classifiers. Since the number of segmented candidate components is relatively small, connected component based methods have lower computation cost and the located text components can be directly used for recognition. Although the existing methods have reported promising text detection performance, there still remain several problems to solve. For region based methods, the speed is relatively slow and the performance is sensitive to font size and font. On the other hand, connected component based methods cannot segment text components accurately when complex background is present in the images.[21]

Lienhart and Wernicke[14] have proposed a method for segmenting text in images and videos based on multi-layer feed forward network and temporal information. Automatic detection and recognition of signs from natural scenes is proposed in Ref. 8 which works based on multiresolution, multiscale edge detection, adaptive searching, color analysis and affine rectification in a hierarchical framework. Method for Devanagari and Bangla text extraction from natural scene images is proposed in Ref. 4 based on connected component analysis and mathematical morphology operation to achieve good accuracy. However this method works well for those two scripts since the features are language dependent. Pan *et al.*[20] have proposed a robust method to detect text in natural scene images based on cascade AdaBoost classifier. Non-text components are filtered out using Markov random fields (MRF). Chen and Yuille[6] have presented a method for text detection in natural scene images based on statistical analysis and AdaBoost classifier. Recently, Epshtein *et al.*[10] have proposed a method based on stroke width transform for text detection in scene images. This method works well as long as stroke width is constant in the character shape. Hybrid approach for text detection in natural scene images is proposed in Ref. 21 to overcome the problems of existing methods. This paper introduces text region detector to estimate the text existing confidence and scale information in image pyramid, which help segment candidate text components by local binarization. Text localization and recognition in complex scene images using local features is proposed by Zheng *et al.*[29] In this work, they explored SIFT features, automatic template images for matching and multiple-size sliding window operation is performed to identify the text candidates. Neumann and Matas[17] have proposed a general method for text localization and recognition in real-world images. It exploits maximally stable external regions (MSER) which provides robustness to geometric and illumination conditions. Text detection in natural scene images by stroke Gabor words is proposed by Yi and Tian.[28] They compute a set of Gabor filters that can describe principle stroke components of text by their parameters. Then k-means algorithm is applied to cluster the descriptive filters. Neumann and Matas[19] have proposed a method for text localization in real-world images using efficiently pruned exhaustive search. The method exploits higher order properties of text such as word text lines. The method demonstrates that grouping the stage plays a key role in text localization performance and that a robust and precise grouping stage is able to compensate errors of the character detector. Guo *et al.*[11] have proposed a method for localization and recognition of the scoreboard in sports video based on SIFT point matching. In this work, they try to utilize the temporal frames for text localization. Neumann and Matas[18] have proposed a method for real-time scene text localization and recognition. The external region of the character is defined which is robust to blur, illumination, color, texture variation and handles low contrast text. However, above methods have focused on horizontal scene text especially on ICDAR-2003 competition dataset. Since their focus is on horizontal text, the methods take advantage of the direction of the line to extract features and to improve the performance. But recently, the method for arbitrary orientation in natural scene images is proposed by Yao *et al.*[26] In this method, they adopt stroke width transform (SWT) and also design various features that are intrinsic to texts and robust to variations. Based on connected component analysis, candidate linking and chain analysis, the method solves the problem of multi-oriented text detection. It is also seen that few methods have addressed the issue of multi-oriented text detection in video[22,23,30] and the few methods are experimented on natural horizontal scene text data to show that the methods can be used for natural scene text detection.[5,15,24]

From the above literature review on text detection and localization of natural scene images, it is observed that most of the methods have used connected component analysis (foreground information) to extract different feature along with training and classifiers to improve text detection and localization accuracy. In addition, few methods have addressed the multi-oriented text in natural scene images. However, none of the methods exploit the space inside the characters, inter characters and word to detect multi-oriented text in natural scene images as it is proposed in the present work which uses run-lengths of space inside the characters, inter characters and words to identify the text candidates. But recently, we have proposed a method[2] based on run-lengths and heuristics to detect horizontal text in natural scene images. The scope of the work is limited to horizontal but not non-horizontal text detection. We have also tested our run-length concept on non-horizontal text detection in Ref. 1 where in we use horizontal run-lengths to classify text pixel from non-text pixels. This work does not solve the problem of multi-oriented text detection problem since text extraction depends on projection profiles. Therefore, this method fails to address the touching lines problem which is common in multi-oriented text detection in video and scene text images. Hence, in this work, we use the same run-length as basis in different way to detect multi-oriented text in natural scene images. On top of that, we present boundary growing for traversing multi-oriented text lines and new zero crossing based method for separating touching text lines. In this way, the proposed method is different from the existing method and the literature.

3. Proposed Methodology

The proposed method works on the basis that space inside the character, inter character and words have regular spacing. This fact motivated us to propose a new method that computes run-length of the space inside the characters, inter character and words as it is inspired by our work in Ref. 2 where it is shown that run-lengths information helps in finding accurate text candidates detection. The method uses Sobel of the input image for computing the run-lengths because we believe that Sobel edge operator is good for high contrast text pixels and text pixels usually have high contrast compared to its background. Therefore, we prefer Sobel edge operator rather than Canny and other operators for this application. To solve the multi-oriented text detection problem, we propose new idea of selecting potential run-lengths with the help of run-lengths computed for the Sobel edge image of the ideal horizontal text image. The boundary growing is proposed for fixing bounding box for the multi-oriented text. Since input image contains complex background, boundary growing sometimes group adjacent text lines if there is a less space between text lines. Therefore, we propose new method for separating touching text lines by finding zero crossing in the output of boundary growing. These are the main contributions of the proposed work. The proposed method is divided into three sub-sections. Section 3.1 describes how to select potential run-lengths for multi-oriented text using horizontal ideal run-lengths. Boundary growing is presented in Sec. 3.2. Zero crossing for separating touching text lines is discussed in Sec. 3.3.

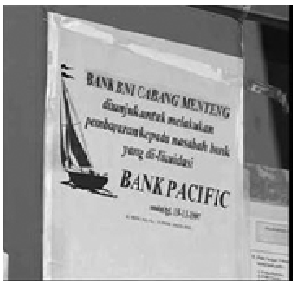
3.1. *Run-lengths for multi-oriented text detection*

To identify the potential run-lengths for text candidates from the multi-oriented text input image, we choose one ideal image as shown in Fig. 1(a) for which we get Sobel edge image as shown in Fig. 1(b). Then we compute horizontal run-lengths for the Sobel edge image by counting number of consecutive black pixels between white pixels in the Sobel edge image and we propose Max−Min clustering to classify the frequencies of run-lengths which represents text pixels as shown in Fig. 1(c) where one can see text patches for the text regions. Here Max−Min clustering algorithm chooses maximum and minimum frequency from the array of run-lengths and then the frequencies in the array are compared with the maximum and the minimum to choose the frequency that is close to maximum frequency. The reason for proposing this simple Max−Min clustering is that we believe that the frequency array containing frequency of run-lengths give high values for text and low values for non-text since the run-lengths are computed based on regular spacing between the intra and inter characters. This results in ideal run-lengths for the multi-oriented text detection as shown in Fig. 1(c) where ideal run-lengths are displayed as white patches. This is true because high frequencies of run-lengths usually occur when there is a text due to transition from background to text and text to background. For multi-oriented text input image shown in Fig. 1(d), we apply the same horizontal run-length criteria on the Sobel edge image (Fig. 1(e)) of the input image Fig. 1(d) to find ideal run-lengths with the help of Max−Min clustering algorithm as shown in Fig. 1(f) where the ideal run-lengths are displayed as white patches for the multi-oriented text. It is noticed from Figs. 1(c) and 1(f) that just ideal run-length does not give better results for multi-oriented text image. Therefore, we find new way to obtain potential run-lengths from ideal run-lengths by comparing the ideal run-lengths of the multi-oriented text with the ideal run-lengths of horizontal text obtained from the horizontal input image shown in Fig. 1(c), which we call potential run-lengths as shown in Fig. 1(g) where clear text patches can be seen compared to the text patches in Fig. 1(f). In this way, the proposed method uses the horizontal ideal run-lengths to

(a) Horizontal input

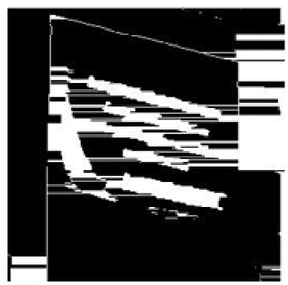
(b) Sobel edge

(c) Ideal run-lengths

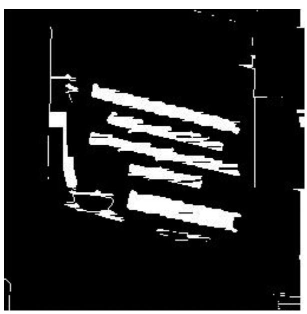


(d) Multi-oriented text

(e) Sobel edge



(f) Run-lengths



(g) Potential run-lengths

Fig. 1. Run-lengths for multi-oriented text.

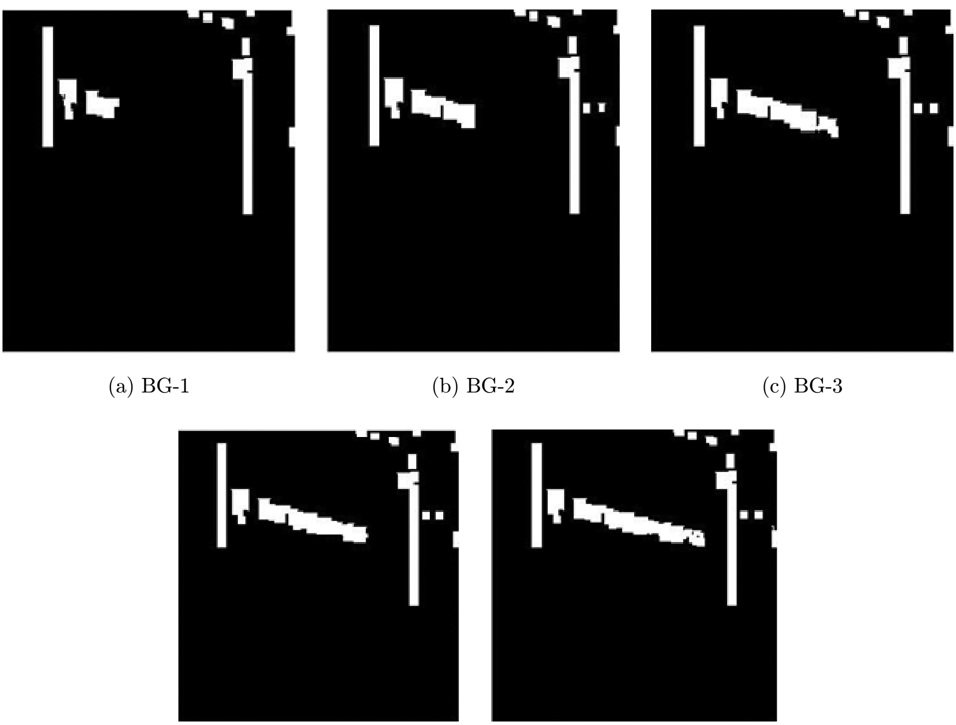
obtain potential run-lengths for all input text images in this work. The horizontal ideal run-lengths are computed only once for the image shown in Fig. 1(a). We choose this image as ideal image based on our experimental study on both horizontal and non-horizontal images.

3.2. *Boundary growing method for traversing*

For text patches obtained from the previous section shown in Fig. 1(g), we present a boundary growing (BG) method to traverse multi-oriented text lines along text direction to extract text lines in the image. Since projection profiled based method do not work for multi-oriented text, we propose BG method in this work. For each component in the text patches shown in Fig. 1(g), the BG method fixes bounding box first and then it allows boundary to grow by pixel by pixel until it reaches neighbor pixel of the component. This is valid because the space between the characters is less than the space between the words and lines. Therefore, BG method grows along text line direction and it reaches end of the text line. End of the line is decided based on experiments on space between characters and words. The process of BG method is illustrated in Figs. 2(a)–2(f) where BG-1–BG-5 show the number of iterations required to extract the first text line in Fig. 1(g), respectively in Figs. 2(a)–2(f). Figure 2(g) shows all text lines which are extracted by the BG method after eliminating false positives. To eliminate false positives, we use our existing method[3] which uses the proximity of the pixels in the text block and angle information of the text block. For more details, refer the method presented in Ref. 3.

3.3. *Zero crossing for separating text lines from touching*

Due to complex background of natural scene images, the BG method sometimes connects adjacent text lines while growing if the space between text lines is less as



(a) BG-1 (b) BG-2 (c) BG-3

(d) BG-4 (e) BG-5

Fig. 2. Boundary growing (BG) for multi-oriented text lines.

(f) BG-text line (g) BG-for all text lines in Fig. 1(d)

Fig. 2. (*Continued*)

shown in Fig. 2(g) where first, second, third and fourth lines are connected and form one component, and fifth and sixth lines are not connected. As a result, fixing bounding box for each text line has become hard. Therefore, we propose new idea that the method fills white pixels for the results obtained by the BG as shown in Fig. 3(a) where binary multi-oriented text patches can be seen. For each component in the

(a) Multi-oriented text with touching (b) Text line segments (c) Column-maximum zero crossing

(d) Text line separation (e) Text detection

Fig. 3. Zero crossing method for rotated multi-oriented touching text line segmentation.

image shown in Fig. 3(a), we compute angle using PCA by passing coordinates of each component. The average of angles of all the components is considered as actual angle of the whole image. The oriented image is rotated to horizontal direction with the help of average angle as shown in Fig. 3(b) where text lines are in horizontal direction. To overcome the problem of touching for fixing bounding box, we propose a zero crossing method which does not require complete spacing between the text lines, to fix the boundary for such text lines. Zero crossing means transition from 0 to 1 and 1 to 0. The method counts the number of transitions from 0 to 1 and 1 to 0 in each column from top to bottom for the image shown in Fig. 3(b). Next it chooses the column which gives the maximum number of transitions to be the boundary for the text lines as shown in Fig. 3(c) where the vertical line shows the column which has highest number of transitions. Here we ignore transition if the distance between two transitions is too small. With the help of the number of transitions, the method draws horizontal boundaries for the text lines as shown in Fig. 3(d) where horizontal boundary for each text lines can be seen. Further, the method looks for spacing between the text components within two horizontal boundaries to draw the vertical boundary for the text lines as shown in Fig. 3(e) where all text lines are separated with their bounding boxes.

4. Experimental Results and Comparative Study

We consider five datasets to show that the proposed method is capable of handling different situations and diversified dataset. The proposed method is tested on 418 High Resolution Multi-oriented Camera Images (HMCI), 394 Low Resolution Multi-Oriented Mobile Camera Images (LMMI), 523 High Resolution horizontal text Camera Images (HHCI), 485 Low Resolution Mobile Camera Images (LMCI) and the 230 standard dataset ICDAR-2003 competition data[16] to evaluate the performance of the proposed method in terms of recall, precision, f-measure and misdetection rate. In total, the proposed method is tested on 2050 images to show that the proposed method is superior to existing methods. The measures are defined as follows. Recall is the number of text blocks detected by the proposed (M) method divided by the number of actual text blocks (N). Precision is the number of text blocks detected by the proposed method (M) divided by the total number of text blocks (T). Misdetection rate is the number of misdetection divided by the number of text blocks detected by the proposed method (M) and f-measure is (2 $\times$ Recall $\times$ Precision)/(Recall + Precision). Here the actual number of blocks are counted manually and considered them as ground truth. Text block is the block which contains text information and misdetection is the text block that if it overlaps less than 20% area with the ground truth then it is considered as a misdetection. In other words, if the area of detected text block overlaps with 80% area with the ground truth then it is considered as true text block. More details about measures and definitions can be found in Ref. 23 as we follow the same instructions given in Ref. 23.

According to discussion on evaluation in Ref. 23, we can see two kinds of evaluation such as one in the line of video text detection and another one is in the line of scene text detection in natural scene images as in ICDAR-2003 competition. The methods that use different dataset along with ICDAR-2003 competition data usually follow the former criteria (in the line of video text detection) and the methods that use only ICDAR-2003 competition for experimentation follow latter criteria (given in ICDAR-2003 competition). This video text detection criteria allows more tolerance than ICDAR-2003 competition measures and it counts at text line level but not at word level as in the ICDAR-2003 competition. Since our method experiment on different dataset, we use the measure as in the video text line instead of ICDAR-2003 competition measure because the current work is not focusing on recognition. In this work, we consider both multi-oriented text and horizontal text images for experimentation to show that the proposed method works irrespective of rotations.

4.1. *Experiments on high resolution multi-oriented camera images (HMCI)*

Figure 4 shows sample results of the proposed method for multi-oriented text detection on different images captured by high resolution camera (six mega pixels). First column in Fig. 4 shows sample images of different background, fonts, font size

Fig. 4. Sample results for HMCI.

Table 1. Performance of the proposed method for multi-oriented text images.

Dataset	Recall	Precision	f-measure	MDR
HMCI	0.94	0.86	0.89	0.04
LMMI	0.95	0.87	0.90	0.05

and orientation and the second column in Fig. 4 shows the result of the proposed methods for the images in the first column. It is observed from Fig. 4 that the proposed method detects multi-oriented text in the HMCI well. The quantitative results are reported in Table 1 where recall, precision, f-measure and misdetection rate are promising and encouraging for multi-oriented text detection in HMCI.

4.2. *Experiments on low resolution multi-oriented mobile camera images (LMMI)*

Figure 5 shows sample results of the proposed method for multi-oriented text detection on different images captured by low resolution mobile camera (two mega

Fig. 5. Sample results for LMMI.

pixels). First column in Fig. 5 shows sample images of low resolution as we can see dim text in the images. The second column in Fig. 5 shows the result of the proposed methods for the images in the first column in Fig. 5. It is observed from Fig. 5 that the proposed method detects multi-oriented text in the LMMI well. The quantitative results are reported in Table 1 where recall, precision, f-measure and misdetection rate are promising and encouraging for multi-oriented text detection in LMMI. It is noticed from Table 1 that the proposed method gives good results for low resolution images compared to the result of high resolution. This is because the complex background of high resolution camera images sometime overlaps with the text blocks while the background in low resolution does not. However, misdetection rate for low resolution is higher than the high resolution images.

4.3. *Experiments on high resolution horizontal camera images (HHCI)*

The objective of this experiment is to show that the proposed method works well for horizontal text detection images when it works well for multi-oriented text images. Figure 6 shows sample results of the proposed horizontal text detection on different

Fig. 6. Sample results for HHCI.

Table 2. Performance of the proposed and existing method for HHCI.

Methods	Recall	Precision	f-measure	MDR
Proposed	0.97	0.86	0.91	0.03
Basavanna *et al.*[2]	0.86	0.41	0.55	0.03

images (First column in Fig. 6). The second column in Fig. 6 shows the result of the proposed method for the images in the first column. It is observed from Fig. 6 that the proposed method detects all the text lines in the images properly. The quantitative results of the proposed method are reported in Table 2 where recall, precision, f-measure and misdetection rate are promising and encouraging. The results are compared with our previous method[2] since the method considers horizontal text images for experimentation. According to Table 2, the existing method is not good compared to the proposed method because the existing method produces more false positives for this dataset.

4.4. *Experiments on low resolution horizontal mobile camera images (LHMI)*

The objective of this experiment is to show that the proposed method works well for low resolution horizontal text images also when it works for high resolution images. Figure 7 shows sample results of the proposed method on different images. First column in Fig. 7 show sample images and the second column in Fig. 7 show the result of the proposed method for the images in first column. It is observed from Fig. 7 that the proposed method detects all text lines in the images properly. The quantitative results reported in Table 3 show that recall, precision, f-measure and misdetection rate are promising and encouraging for low resolution images. The results are compared with our previous method[2] and are reported in Table 2. According to Table 2, the existing method does not give good results compared to the proposed method since existing method is not good for eliminating false positives. On the other hand, the proposed method is good in terms of measures because of advantage of the false positive elimination and zero crossing.

4.5. *Experiments on ICDAR-2003 competition data*

This dataset is benchmark data for scene text detection available publicly. Our method is tested on this dataset to show that the proposed method is suitable for this dataset also as this dataset is challenging due to complex background, non-uniform illumination and unfavorable characteristics of scene text. Figure 8 shows sample results of the proposed method for text detection. It is observed from Fig. 8 that for the images in first column, the proposed method detects text properly as shown in second column in Fig. 8. The quantitative results reported in Table 4 show that the proposed method is good for scene text detection in terms of measures. It is noted

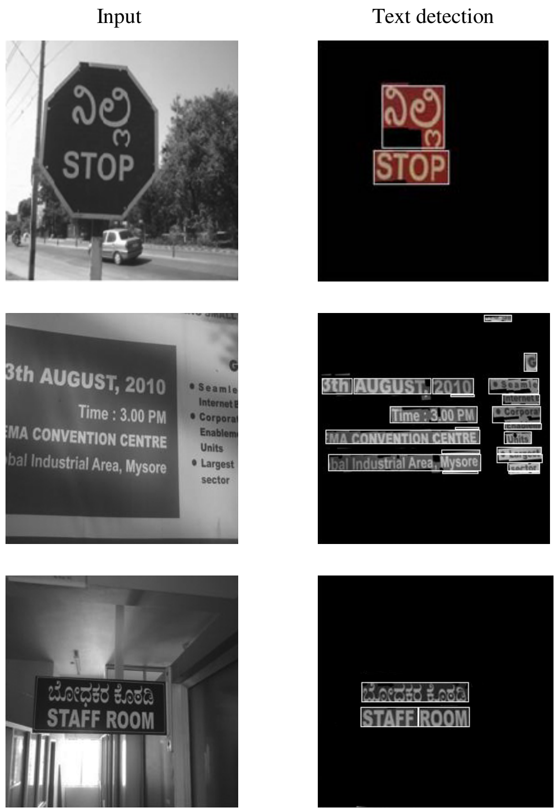


Fig. 7. Sample results for LHMI.

Table 3. Performance of the proposed and existing method for LHMI.

Methods	Recall	Precision	f-measure	MDR
Proposed	0.96	0.87	0.91	0.02
Basavanna *et al.*[2]	0.92	0.45	0.60	0.02

that the method in Ref. 22 is compared with the existing methods on the same dataset with the same evaluation criteria. Therefore, our method is compared with the methods as listed in Table 4. According to Table 4, accuracy of the proposed method is higher than the existing methods. The reason for poor accuracy for the existing methods is that the main focus of the existing methods[5,15,22–24,30] is to detect text lines in the video images but not in the scene text images and mobile camera captured images while the proposed method's focus is to detect scene text images of both high and low resolution images.

Fig. 8. Sample results for ICDAR-2003 competition data.

Table 4. Performance of the proposed and existing methods for ICDAR-2003 competition data.

Methods	Recall	Precision	f-measure	MDR
Proposed	**0.94**	**0.85**	**0.89**	**0.02**
Basavanna *et al.*[2]	0.81	0.64	0.71	0.06
Shivakumara *et al.*[22]	0.87	0.72	0.78	0.14
Shivakumara *et al.*[23]	0.86	0.76	0.81	0.13
Zhou *et al.*[30]	0.66	0.83	0.73	0.26
Liu *et al.*[15]	0.53	0.61	0.57	0.24
Wong and Chen[24]	0.52	0.83	0.64	0.08
Cai *et al.*[5]	0.67	0.33	0.44	0.43

It is observed from Tables 1–4 that the proposed method gives consistent results for different datasets. This is because of the way the proposed method computes run-lengths and used the Sobel edge image for detecting text detection. Besides, the pattern of text and spacing between the character and words may not change much when resolution and background changes. Therefore, the proposed method gives consistent results for different datasets.

5. Conclusions and Future Work

We present a new run-length based method, boundary growing method and the methods based on zero crossing for multi-oriented text detection in both high and low resolution images. The simple horizontal run-lengths are used in novel way to identify the potential run-lengths for multi-oriented text detection. The touching between text lines is solved by proposing new methods on angle information of the text given by the boundary growing and zero crossing between text lines. The experimental results on different dataset and comparative study reveal that the proposed method outperforms the existing methods in terms of f-measures. However, the limitation of the work is that for the curve shaped text lines and single image containing different oriented text lines, the proposed method does not give results since zero crossing method requires horizontal text images and single orientation text images but the potential run-lengths and boundary growing methods give results for arbitrary text lines in the images. Therefore, we are investigating a new method to replace zero crossing to expand the scope of this work.

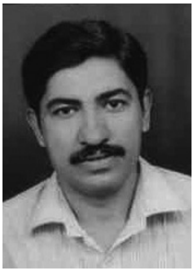
Acknowledgment

This research work is supported by University Grants Commission (UGC), New Delhi, India (F.No.UGC/MRP(s)/800/10-11/KAMY022). Authors would like to thank Prof. D. Krishne Gowda Principal, Government College (Autonomous), Mandya, Karnataka, India for providing good infrastructure to carry out research work.

References

1. M. Basavanna, G. Shivanand, P. Shivakumara and S. K. Srivatsa, Automatic non-horizontal scene text recognition from document image, *Int. J. Mach. Intell. (IJMI)* **3** (2011) 164–167.
2. M. Basavanna, P. Shivakumara, S. K. Srivatsa and G. Hemantha Kumar, A new run-length based method for scene text detection, in *Proc. IICAI* (2011), pp. 1730–1736.
3. M. Basavanna, P. Shivakumara, S. K. Srivtasa and G. Hemantha Kumar, Angle-proximity based text block verification method for text detection in scene images, *Int. J. Mach. Intell. (IJMI)* **3** (2011) 245–250.
4. U. Bhattacharya, S. K. Parui and S. Mondal, Devanagri and bangla text extraction from natural scene images, in *Proc. ICDAR* (2009), pp. 171–175.
5. M. Cai, J. Song and M. R. Lyu, A new approach for video text detection, in *Proc. ICIP* (2002), pp. 117–120.
6. X. Chen and A. Yuille, Detecting and reading text in natural scenes, in *Proc. CVPR* (2004), pp. 366–373.
7. D. Chen, J. M. Odobez and H. Bourlard, Text detection and recognition in images and video frames, *Pattern Recogn.* **37** (2004) 595–608.

8. X. Chen, J. Yang, J. Zhang and A. Waibel, Automatic detection and recognition of signs from natural scenes, *IEEE Trans. Image Process.* (2004), pp. 87–99.
9. D. Doermann, J. Liang and H. Li, Progress in camera-based document image analysis, in *Proc. ICDAR* (2003), pp. 606–616.
10. B. Epshtein, E. Ofek and Y. Wexler, Detecting text in natural scenes with stroke width transform, in *Proc. CVPR* (2010), pp. 2963–2970.
11. J. Guo, C. Gurrin, S. Lao, C. Foley and A. F. Smeaton, Localization and recognition of the scoreboard in sports video based on SIFT point matching, in *Proc. MMM* (2011), pp. 337–347.
12. K. Jung, Neural network-based text location in color images, *Pattern Recogn. Lett.* **22** (2001) 1503–1515.
13. K. Jung, K. I. Kim and A. K. Jain, Text information extraction in images and video: A survey, *Pattern Recogn.* **37** (2004) 977–997.
14. R. Lienhart and A. Wernicke, Localizing and segmenting text in images and videos, *IEEE Trans. CSVT* **12** (2002) 256–268.
15. C. Lu, C. Wang and R. Dai, Text detection in images based on unsupervised classification of edge-based features, in *Proc. ICDAR* (2005), pp. 610–614.
16. S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, ICDAR 2003 robust reading competitions, in *Proc. ICDAR* (2003), pp. 1–6
17. L. Neumann and J. Matas, A method for text localization and recognition in real-world images, in *Proc. ACCV* (2010), pp. 770–783.
18. L. Neumann and J. Matas, Real-time scene text localization and recognition, in *Proc. CVPR* (2012), pp. 3538–3545.
19. L. Neumann and J. Matas, Text localization in real-world images using efficiency pruned exhaustive search, in *Proc. ICDAR* (2011), pp. 687–691.
20. Y. F. Pan, X. Hou and C. L. Liu, A robust system to detect and localize text in natural scene images, in *Proc. DAS* (2008), pp. 35–42.
21. Y. F. Pan, X. Hou and C. L. Liu, A hybrid approach to detect and localize text in natural scene images, *IEEE Trans. Image Process.* (2011), pp. 800–813.
22. P. Shivakumara, R. P. Sreedhar, T. Q. Phan, S. Lu and C. L. Tan, Multi-oriented video scene text detection through Bayesian classification and boundary growing, *IEEE Trans. CSVT* **22** (2012) 1227–1235.
23. P. Shivakumara, T. Q. Phan and C. L. Tan, A Laplacian approach to multi-oriented text detection in video, *IEEE Trans. PAMI* **33** (2011) 412–419.
24. E. K. Wong and M. Chen, A new robust algorithm for video text extraction, *Pattern Recogn.* **36** (2003) 1397–1406.
25. V. Wu, R. Manmatha and E. M. Riseman, Text finder: An automatic system to detect and recognize text in images, *IEEE Trans. PAMI* **21** (1999) 1224–1229.
26. C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, Detecting texts of arbitrary orientations in natural images, in *Proc. CVPR* (2012), pp. 1083–1090.
27. Q. Ye, Q. Huang, W. Gao and D. Zhao, Fast and robust text detection in images and videos frames, *Image and Vision Computing* **33** (2005) 565–576.
28. C. Yi and Y. Tian, Text detection in natural scene images by stoke Gabor words, in *Proc. ICDAR* (2011), pp. 177–181.
29. Q. Zheng, K. Chen, Y. Zhou, C. Gu and H. Guan, Text localization and recognition in complex scene images using local features, in *Proc. ACCV* (2010), pp. 121–132.
30. J. Zhou, L. Xu, B. Xiao and R. Dai, A robust system for text extraction in video, in *Proc. ICMV* (2007), pp. 119–124.



M. Basavanna received his B.Sc. degree in Computer Science from the University of Mysore, Mysore in 1996, M.Sc. degree in Computer Science from the University of Mysore, Mysore in 1999 and M.Phil. degree from the Alagappa University, Karaikudi, Tamil Nadu, India in 2007. Since 2009 he has been a researcher at the VELS University, Chennai, Tamil Nadu, and has been working for the project on "A New Robust System for Scene Text Recognition Application to Blind Persons". From 1999, he was a Lecturer in Computer Science, and became Assistant Professor in 2009 in the Department of Collegiate Education, Government of Karnataka, India.

He is the coordinator of the Carrier Oriented Program in Computer Network (COP) at Autonomous Government College, Mandya, Karnataka, India funded by UGC, New Delhi.

His research interests include digital image processing, pattern recognition and scene text analysis.

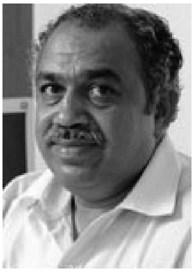
P. Shivakumara is a Research Fellow in the Department of Computer Science, School of Computing, National University of Singapore. He received his B.Sc., M.Sc., M.Sc. Technology by research and Ph.D. degrees in Computer Science respectively in 1995, 1999, 2001 and 2005 from University of Mysore, Mysore, Karnataka, India.

From 1999 to 2005, he was Project Associate in the Department of Studies in Computer Science, University of Mysore, where he conducted research on document image analysis, including document image mosaicing, character recognition, skew detection, face detection and face recognition. He worked as a Research Fellow in the field of image processing and multimedia in the Department of Computer Science, School of Computing, National University of Singapore, from 2005 to 2007. He also worked as a Research Consultant at Nanyang Technological University, Singapore for six months on image classification in 2007. He has published around 100 research papers in national, international conferences and journals. He has been a reviewer for several conferences and journals.

His research interests are in the area of image processing, pattern recognition and video text processing.

Dr. S. K. Srivatsa received his Bachelor of Electronics and Telecommunication Engineering degree from Jadavpur University, Calcutta, India. He obtained his Master's degree in Electrical Communication Engineering and Ph.D. from the Indian Institute of Science, Bangalore, India.

Dr. S. K. Srivatsa was a Professor of Electronics Engineering at Anna University, Chennai, India. He was a Research Associate at the Indian Institute of Science, Bangalore. Presently he is a Senior Professor in the Department of Instrumentation & Control Engineering at St. Joseph College of Engineering, Chennai, India. He has taught around twenty different courses at undergraduate and post graduate levels. He has reviewed five textbooks for Tata McGraw Hill and one textbook on Digital Electronics for ISTE. He has evaluated UGC minor research projects. His current research activities pertain to computer networks, design and analysis of algorithms, coding theory and artificial intelligence and robotics.



Dr. G. Hemantha Kumar received his B.Sc. degree in Computer Science from the University of Mysore, Mysore in 1980, and M.Sc. degree in Computer Science from the University of Mysore, Mysore in 1988, and Ph.D. from the University of Mysore, Mysore in 1998.

Dr. G. Hemantha Kumar is serving as Course-Coordinator of the Chinese B.Tech. Programme. Since 2007, he has been working as a Professor in the Department of Studies in Computer Science, Manasagangotri, University of Mysore and presently, he is a Professor and Chairman in the Department of Studies in Computer Science, University of Mysore, Manasagangotri. He has about 87 research papers on various reputed journals and conference proceedings. His current research activities pertain to image processing, pattern recognition, document image analysis and face recognition.