A Combined Edge and Connected Component Based Approach for Kannada Text Detection in Images

Shahzia Siddiqua
School of Computing & Information Technology
Reva University, Bengaluru, India
sshaz79@gmail.com

Naveena C
Dept. of Computer Science & Engineering.
SJBIT, Bengaluru, India
naveena.cse@gmail.com

Sunil Kumar Manvi
School of Computing & Information Technology
Reva University, Bengaluru, India
manvi.sunil@reva.edu.in

Abstract—**This paper presents an approach for Kannada text detection from scene images. The sturdiness of the system comes from three facets. Firstly, a morphological edge detection method is employed. In this a threshold level is determined resulting in an edged image. Secondly, a text feature filtering method is used which removes maximum non-letter components. Also, the method results in majority of letter components that are within the range of mean height of the residual bounding boxes. Finally, a text region binarization method is applied using a threshold from the residual bounding boxes. The performance of the proposed method was tested on self-captured images having kannada text and the results are satisfactory.**

Keywords—***Text detection; Kannada text; Edge detection; Feature filtering; Binarization***

I. Introduction

Texts are significant entities embedded in images. They often contain useful information such as dangerous warnings, transportation signals, describing the scene, billboards, etc. Automatic text recognition encourages a number of potential applications such as: helping a foreigner to figure out the contents on an information board (by interpreting the identified text into his native language), help blind people to walk freely and independently in a street (by translating the text to audio data) …etc. As a result research on text detection in images is becoming a hot topic. To detect text from an image various challenges are encountered like lighting and surface reflections, shaded or textured background, non-planar objects, blur or low contrast, complex images, variations in font size, style, color, orientation and alignments, etc.

Primarily there are three categories of approaches for text detection: *texture based, component based and hybrid approach.*

Texture based approach perceives texts as a form of texture and makes use of their textural attributes to separate text from non-text regions in the image.

Earlier works of Zhong *et al.* [1] suggests a technique using Horizontal spatial variance to localize text. Later, Li *et al.* [3] used the first-order and second-order instants and mean of wavelet coefficients to decompose an image. Kim *et al.* [6] used the pixel intensity to classify pixels, and formulated as SVM classifier. This method works well with simple backgrounds, but not the same for complex frames. To manage multilingual texts, Lyu *et al*. [8] suggested a search scheme, using properties of text like solid edge and extreme contrast to differentiate text from non-text areas. As the method involves many metrics it is tough to deal with texts of diverse styles.

Zhong *et al.* [4] planned a concept that smoothly spot text in Discrete Cosine Transform (DCT) area. The benefit of this approach is that the image is not decoded prior to detection. Yet, the detection exactness is limited. To increase the velocity of text detection techniques, Chen *et al.* [7] suggested a cascaded Adaboost classifier skilled for intensity variance, gradient histogram, mean strength, horizontal difference or vertical difference features. The detection method is considerably quicker but accuracy is lesser compared to other algorithms. Lately, Wang *et al.* [13] recommended a routine for locating particular words from images using sliding window. This algorithm can only detect words in the given set, which makes the possibility range quite narrow. Texture based procedures are computationally costly, for all locations and scales have to be tested. Also, these procedures generally handle straight texts and are subtle to orientation.

Components based approach first removes candidate components and then sieve non-text components. These methods are more proficient, as the count of features to be processed is comparatively less. Also, these methods are inconsiderate to font variation, orientation and scale change.

Using the attribute that letters have almost same stroke breadth, Epshtein *et al.* [14] recommended an operator as Stroke Width Transform (SWT). This operator retrieves character strokes from edge charts and extracts text components of different measurements and tracks. This method considers horizontal texts only and works on many rules and parameters. Neumann *et al.* [15] advised a method based on Maximally Stable Extremal Regions (MSER). Here the MSER regions are taken out with the help of a trained classifier that eliminates invalid candidates. Later, the remaining candidates are collected in text lines using a chain

978-1-5090-6701-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ICRAECT.2017.35

of linking rules. This algorithm works well for horizontal orientation only. Yi *et al.* [16] suggestions distinguishes slant texts in images. According to the scattering of pixels in color space, the image is splitted into several sections. These sections are merged into connected components. However, this method is not efficient with extensive complex data sets.

Shivakumara *et al.* [17] aimed logic for multi-angled text detection. Using Fourier-Laplace space for clustering, candidate regions are extracted. The elements works for text blocks but deny stoke characters. Focused on SWT [14], Yao *et al.* [19] recommended a procedure that can spot texts having random orientations in images. This procedure is armed with a two- stage classification outline and two groups of rotation and rotation-invariant structures aimed for seizing texts in natural scenes. To crack the mismatch glitch of edge points in the original SWT, Huang *et al.* [20] devised a novel operator termed Stroke Feature Transform (SFT). SFT presents color evenness and constricts relationships of resident edge points, generating improved element extraction results for horizontal texts only. In [21], Huang *et al.* integrated Maximally Stablextremal Regions (MSER) to mine text candidates and Convolutional Neural Networks (CNN) classifier to identify true text candidates in scene images. This process attained considerably greater performance above conventional methods.

Hybrid approaches make use of the advantage of texture based techniques and component based techniques. Liu *et al.* [11] suggested a technique via detailed edge identification policy edge pixels. Also candidate text sections are generated by verifying the gradient and geometrical properties of region contours. Contrasting [11], Pan *et al.* [18] suggested a hybrid method by using multi-scale probability maps. A classifier instructed on a set of texture structures (HOG structures [9]) processed with a cluster of predefined designs serves the probability maps. A Conditional Random Filed (CRF) template [10], which merging unary component factors and binary related connections, is made use of to distinguish text elements from non-text elements. These two methods detect horizontal texts only.

From the literature survey, it is observed that majority of work is done for foreign languages like English, Arabic, Chinese … etc. While very less work is done for Indian regional language like Kannada. The Kannada script is quite complex when compared to English and Chinese. In this work we propose text detection algorithm for Kannada in natural images.

This paper is organized as follows. In Section II, The methodology employed is discussed stage wise. In Section III, Experimental results with sample outputs is discussed. Finally, Conclusion and future enhancement are mentioned.

II. Methodolgy

Proposed method consists of the three facets namely *Edge detection, Feature filtering and Binarization.* To demonstrate the method flow in this paper, a sample input image is used as shown in Figure 1.

Figure 1.Input Image

A. Edge Detection

For the input true color image, its luminance intensity is determined using the CCIR (Case Centre for Imaging Research) 601 recommendations (1), where R, G and B are the red, green and blue components respectively of the digital image.

$$I = 0.299R + 0.587G + 0.114B \quad (1)$$

To reduce the amount of data processed, the image is compressed to a lower resolution without compromising its visual quality. This is done by down sampling the image to 1/3 size of I. The image is then blurred using combination of morphological open and close operation to suppress negative edges and excess segmentation. The edges of the image are mapped using a morphological gradient operation with a difference in structuring element characteristics. After this, an intensity level is estimated using the central difference edge detection filter. Next an orthogonal pair of filters are created and convolved with the image. The image is then thresholded to obtain a binary edged image as shown below in Figure 2.

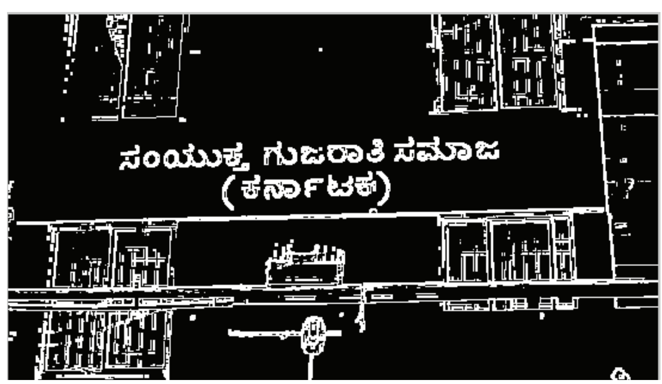


Figure 2. Image after Edge Detection

B. Feature Filtering

A region analysis is done resulting in a number of connected components and their bounding box properties. Individual component is tested against

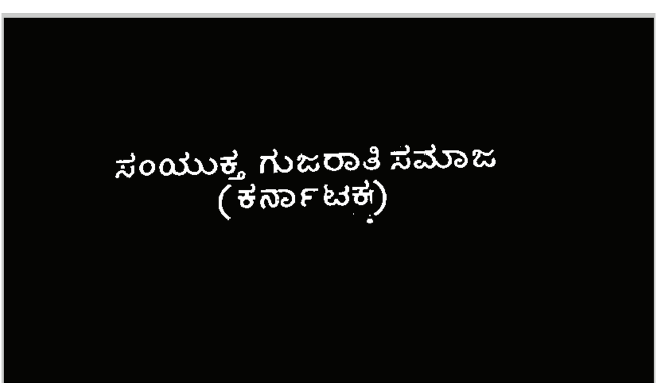
- its width and height dimensions > 10 pixels each,
- their aspect ratio always < 10,
- center width <= 0.2 and
- center height <= 0.5

Failing which majority of non-text components will be filtered out. A second round of filtering is done by computing the mean heights of the residual bounding boxes and extracting only those components whose height is close to mean. The result after filtering is reflected in the Figure 3.

Figure 3. Image after Feature Filtering

C. Binarization

Each residual bounding box is used as a mask to the original image. A cluster based thresholding is done to obtain the threshold of the masked image for binarization. Finally, the recognized text are assigned as the focus 1, and the left behind image as 0. The result is a set of single line binary text extracted images which will be used for the next phase of image processing i.e. Text Recognition. Figure 4(a) shows the input image after binarization, while Figure (b) and (c) show the two single line text resultant images after binarization.



(a) Input image after Binarization

(b) Result1 first single line text

(c) Result12 second single line text

Figure 4. Resultant images after Binarization

III. EXPERIMENTAL RESULTS

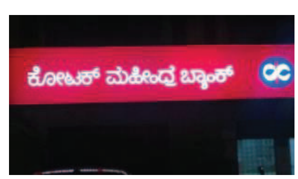
The framework is designed in MATLAB R2010a version 64bit system with 1.7GHz Intel Core i5 processor on Windows 7 OS. The self-captured dataset of 200 images highlighting Kannada text is used to evaluate the method. The details of the dataset is shown in Table1 below:

Table 1. Dataset Details

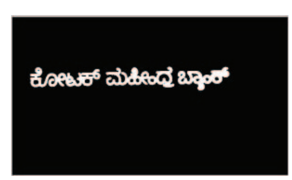
Description	***Details***
Camera Maker-Model	Canon – Canon EOS 500D Nokia – Lumia 520(5MP), 730(6.7 MP) Samsung – KENOX S760
Dimension min – max	85x41 – 1728x3072 pixels
Deep	24 bit deep
Focal Length	3mm – 40mm
Size	4KB – 16 MB

The dataset includes images focused w.r.t variety of situations like text in arbitrary-orientation, scene, graphics, day light, night, shadow, blur, multi lingual, ,illuminated, surface reflections, still video image, cd cover, book cover, different and mixed font style, size and color…etc. The resultant image/s for each input is a single text line which is an 8 bit deep jpg image. Its size varies from 1KB to 48 KB and dimension differs from a minimum of 39x54 pixels to a maximum of 1632x231 pixels.

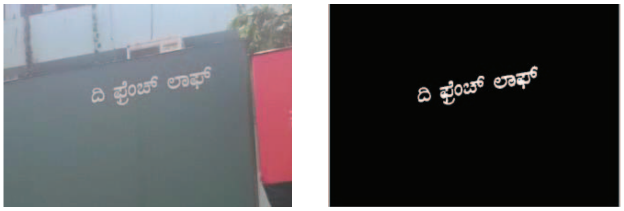
In Figure 5, the sample inputs and their binarized image of successful results of the proposed method are shown. It is clearly seen in this figure, that the method works efficiently for various situations like light text, blur and non-horizontal text. This method also works efficiently for Multi-lingual text images as reflected in Figure 6. However, the method was quite inefficient to detect vertical text and curved text as shown in Figure 7.



(a)Light Text Input



(b)Light Text Binarized



(c)Blur & Angled Text Input (d)Blur & Angled Text Binarized

Figure 5. Sample Results for Light, Blur & Angled Text in Images

(a) Multi-lingual Input (b) Multil-ingual Binarized

Figure 6. Sample Result for Multi-Lingual Scenario

(a)Circular Text Input (b)Circular Text Binarized

(c)Verticle Text Input (d)Verticle Text Binarized

Figure 7. Sample Result for Undetected Situations

IV. Conclusion

In this paper, a method is proposed to detect kannada text from images. The method uses morphological operations to implement edge detection, performs region analysis ignoring non-text areas and binarize using threshold of the remaining bounding boxes. The method was tested against self-created dataset. It is observed that the computational efficiency of the method on an average is 0.76 sec. This result emphasizes the usage of the method to real time applications. Experimental outcomes of the diversity of the dataset show that the proposed method performs well for text detection regardless of contrast, angle, background, text size and style. Yet, the proposed method needs to detect text accurately for smaller and denser font size. Further refinement of the algorithm is required to increase its performance.

References

[1] Y. Zhong, K. Karu, and A. K. Jain, "Locating Text in Complex Color Images," Pattern Recognition, vol. 28, issue 10, pp. 1523-1535, Oct. 1995.

[2] A. K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames," Pattern Recognition, vol. 31, issue 12, pp. 2055–2076, Dec. 1998.

[3] H. Li, D. Doermann, and O. Kia, "Automatic Text Detection and Tracking in Digital Video," IEEE Trans. on Image Processing, vol. 9, issue 1, pp.147-156, Jan. 2000.

[4] Y. Zhong, H. Zhang, and A. K. Jain, "Automatic Caption Localization in Compressed Video," IEEE Trans. PAMI, vol. 22, issue 4, pp. 385–392, Apr. 2000.

[5] Y. Y. Hasan and L. J. Karam, "Morphological Text Extraction from Images," IEEE Trans. on Image Processing vol.9 issue 11, pp. 1978-83, Nov. 2000.

[6] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. PAMI, vol. 25, issue 12, pp. 1631–1639, Dec. 2003.

[7] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," Proceeding of CVPR 2004, pg. 366-373, Jul. 2004.

[8] M. R. Lyu, J. Song, and M. Cai. "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction." IEEE Trans. CSVT, vol. 15, issue 2, pp. 243–255, Feb.2005.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proceeding of CVPR 2005, pg. 886-893, Jun 2005.

[10] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proceeding of ICML 2001, pg. 282-289.

[11] Y. Liu, S. Goto and T. Ikenaga, "A contour-based robust algorithm for text detection in color images." IEICE Transactions on Information and Systems, vol. E89-D issue 3, pp. 1221–1230, Mar. 2006.

[12] B. Epshtein, E. Ofek and Y. Wexler. "Detecting Text in Natural Scene with Stroke Width Transform," Proceeding of Computer Vision and Pattern Recognition, pp. 2963–2970, 2010.

[13] K.Wang and S. Belongie, "Word spotting in the wild," Proceeding of ECCV 2010, pg. 591-604, Sept. 2010.

[14] B. Epshtein, E. Ofek and Y. Wexler. “Detecting Text in Natural Scene with Stroke Width Transform,” Proceeding of Computer Vision and Pattern Recognition, pp. 2963–2970, 2010.

[15] L. Neumann and J. Matas, “A method for text localization and recognition in real-world images,” Proceeding of ACCV 2010, pg. 770-783, Nov. 2010.

[16] C. Yi and Y. Tian, “Text string detection from natural scenes by structure-based partition and grouping,” IEEE Trans. On Image Processing, vol. 20 issue 9, pp. 2594–2605, Sept. 2011.

[17] P. Shivakumara and T. Q. Phan and C. L. Tan, “A laplacian approach to multi-oriented text detection in video,” IEEE Trans. PAMI, vol. 33 issue 2, pp. 412–419, Feb. 2011.

[18] Y. Pan, X. Hou, and C. Liu, “A hybrid approach to detect and localize texts in natural scene images,” IEEE Trans. Image Processing, vol. 20 issue 3 pp. 800–813, Mar 2011.

[19] C. Yao, X. Bai, W. Liu, Y. Ma and Z. W. Tu, “Detecting texts of arbitrary orientations in natural images,” Proceedings of CVPR 2012, pg. 1083-1090, Jun 2012.

[20] W. Huang, Z. Lin, J. Yang, and J. Wang, “Text Localization in Natural Images using Stroke Feature Transform and Text Covariance Descriptors,” Proceeding of ICCV 2013, pp. 1241-1248.

[21] W. Huang, Y. Qiao, and X. Tang, “Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees,” Proceedings of ECCV 2014 pp. 497-511 Sept. 2014.