ISSN 2229-5518

Character Localization From Natural Images Using Nearest Neighbours Approach

Shaila Chugh, Yogendra Kumar Jain

Abstract— Scene text contains significant and beneficial information. Extraction and localization of scene text is used in many applications. In this paper, we propose a connected component based method to extract text from natural images. The proposed method uses color space processing. Histogram analysis and geometrical properties are used for edge detection. Character recognition is done through OCR which accepts the input in form of text boxes, which are generated through text detection and localization stages. Proposed method is robust with respect to the font size, color, orientation, and style. Results of the proposed algorithm, by taking real scenes, including indoor and outdoor images, show that this method efficiently extract and localize the scene text.

Index Terms— Character Localization, Scene Text, Nearest Neighbours, Edge Detection, OCR, Histogram, Filters.

1 INTRODUCTION

TEXT detection and localization from natural scene is an active research area in computer vision field. Scene text appear as a part of scene, such as text in vehicle number plates, hoardings, books, CD covers, etc.

Various font sizes and styles, orientations, alignment, effects of uncontrolled illumination, reflections, shadows, the distortion due to perspective projection as well as the complexity of image backgrounds, makes automatic text localization and extraction scene a challenging problem. Localization of characters in images is used in many applications. Text detection can be used in the applications of page segmentation, document retrieving, address block location, etc. For extraction of text, different approaches have been suggested, based on the text characteristics.

The method proposed by Xiaoqing Liu et al. [2] is based on the fact that edges are a reliable feature of text, regardless of color/intensity, layout, orientations, etc. Edge strength, density and the orientation variance are three distinguishing characteristics of text embedded in images, which can be used as main features for detecting scene text. Their proposed method consists of three stages: target text area detection, text area localization and character extraction.

Wang et al. [3] proposed a connected-component based method which combines color clustering, a black adjacency graph (BAG), an aligning-and-merging-analysis scheme and a set of heuristic rules together to detect text in the application of sign recognition such as street indicators and billboards. Author has mentioned that uneven reflections have resulted in incomplete character segmentation that increased the false alarm rate. Kim et al. [4] implemented a hierarchical feature combination method to implement text detection in real scenes. However, authors admit that their proposed method could not handle large text very well due to the use of local features that represents only local variations of image blocks. Gao et al. [5] developed a three layer hierarchical adaptive text detection algorithm for natural scenes. It has been applied in prototype Chinese sign translation system which mostly has a horizontal and/or vertical alignment.

Cai et al. [7] have proposed a method that detects both low and high contrast texts without being affected by language and font-size. Their algorithm first converts the video image into an edge map using color edge detector [9] and uses a low global threshold to filter out definitely non-edge points. Then, a selective local thresholding is performed to simplify the complex background, then the edge-strength smoothing operator and an edge-clustering power operator highlights those areas with high edge strength or edge density, i.e. text candidates.

Garcia et al. [10] proposed a connected component based method in which potential areas of text are detected by enhancement and clustering processes, considering most of the constraints related to the texture of words. Then, classification and binarization of potential text areas are achieved in a single scheme performing color quantization and characters periodicity analysis. Lienhart et al. [11] and Agnihotri et al. [8] are also proposed connected component based approaches. Alain Trémeau et al. [14] have proposed a method for detection and segmentation of text layers in complex images, which uses a geodesic transform based on a morphological reconstruction technique to remove dark/light structures connected to the borders of the image and to emphasize on objects in center of the image and used a method based on difference of gamma functions approximated by the Generalized Extreme Value Distribution (GEVD) to find a correct threshold for binarization. Jianqiang Yan et al. [15] have used Gabor filters with scale and direction varied to describe the strokes of Chinese characters for target text area extraction and by establishing four sub-neural networks to learn the texture of text area, the learnt classifiers are used to detect target text areas.

Existing methods experience difficulties in handling texts with various contrasts or inserted in a complex background. In this paper, we propose a connected component based text extraction algorithm, a general-purpose method, which can quickly and effectively localize and extract text from both document and indoor/ outdoor scene image.

2 PROPOSED METHOD

In our proposed method we consider that text present in images is in the horizontal direction with uniform spacing be-

ISSN 2229-5518

tween words. The proposed method of text extraction and localization consists of four stages :

1. Color Processing
2. Detection of Edges
3. Text Area Localization
4. Gap Filling and Enhancement.

2.1 Color Processing

Pre-processing of image is done to facilitate easier detection of desired areas. The image is converted to the YUV color space and then, the luminance (Y) is used for further processing. The YUV color format describes a color by using the color components luminance and chrominance. The luminance component (Y) represents the brightness information of a color, the chrominance components (U and V) contain the color differences

o taking
ndwidth
enabling
ore effi-
a "direct"
le output
its color

Figure 1

2.2 Detection of Edges

Aim is to concentrate on regions where text is plausible. In this process, simply the gray-level image is converted into an edge image. The connected-component method is used to highlight possible text areas more as compared to the non-text areas.

Here all pixels are given gray-levels in accordance with their neighbors. The largest value between the pixel and its upper-right, upper and left direction neighbors is assigned as weight here.

In the algorithm, to generate the edge image (EdgeImagexy) from gray scale image (gray Image), each pixel (pixelxy) of grayscale image is processed, to assign the intensity, which is the largest difference among their left (pixelx-1), right (pixelx+1 , y) and upper right (pixelx+1 , y-1) pixels. The edge image initially included all black pixels. Figure 3 shows the edge image, generated through the gray scale image of figure 1.

Procedure 1
Input : gray image
Output : edge image
Step I : initialize all pixel value of edge image EdgeImagexy to zero.
Step II : repeat until all pixelxy € gray image are processed through Step III .
Step III : pixelxy € gray Image
EdgeImagexy : max[(diff(pixelx,y,pixelx-1,y), (diff(pixelx,y,pixelx+1,y), (diff(pixelx,y,pixelx+1,y-1))]
Step IV : improve contrast by using appropriate convolution filter on edge image.
End.

d, to en-
he edges
filter on
age.

Figure 4

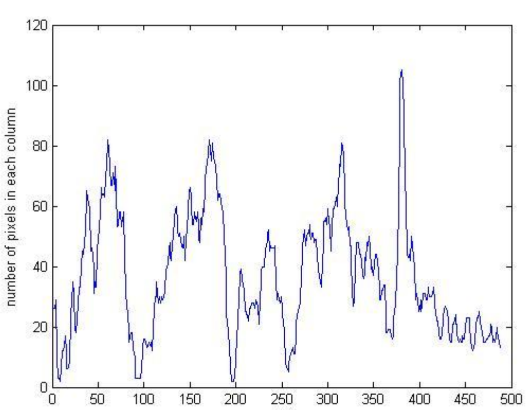
2.3 Text Area Localization

In this stage, height-wise and width-wise analysis is done through the horizontal and vertical histogram for the target text areas. For developing histogram with white text area against dark background, the sharpened image is taken as the input intensity image.

Procedure 2
Input : edge image
Output : generate an array of four coordinates (x0, y0, x1, y1) of text areas.
Step1: generate graph H1 for height wise analysis, and H2 for width wise analysis.
Step2: calculate minimum threshold min (T1) in H1 and maximum threshold max (T2) in H2.
Step 3: pixel value comes under these two threshold value are considered for text area and suppress all pixel those values is less than min(H1) and greater than max(H2)
Step 4: calculate x0, y0, x1, y1 coordinates which belongs to the text area.
End.

Figure 5. Procedure for localization of text area

The width-wise histogram represents aggregate of pixels in all columns and height-wise histogram represents aggregate of ...ns where ...resent in ...wo thre- ...in(T1) = ...10. Areas ...t. Figure ...d height



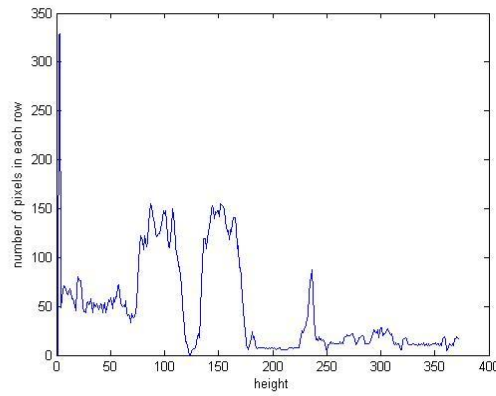


Figure 7

2.4 Refining And Enhancement

To remove the probable areas which are not text, the geometric proportion between length and breadth of the text is considered. This geometric proportion is taken after experimenting on various kind of images to get mean values. Here areas with vertical to horizontal axis proportion lower than ten are taken as target text areas. A gap image will be formed. It will be used as a basis to filter the localization of distinguished text areas. All pixels of the binary processed image, which are enclosed by background pixels in diagonal x and y directions, are replaced by the background intensity.

3 Experimental Results and Discussion

In order to evaluate the performance of the proposed method, we used 25 test images with variable font sizes, perspective and alignment under different lighting conditions. Figure 8 shows some of the results, from which we can see that our proposed method can extract text with various font sizes, perspective, alignment, any number of characters in a text string under different lighting conditions. Although there is no generally accepted method that can be used to evaluate the performance of the text localization and extraction method, we use several metrics used by others to facilitate a meaningful comparison. In our proposed number of correctly located characters, which are regarded as ground-truth. precision rate and recall rate, quantified to evaluate the performance are given by (1) and (2). The area in the image which is not text but is recognized as text, by mistake, by the proposed method, is defined as false positive. The area in the image which is text but could not be recognized by the method is defined as false

n of our
proposed
hod.

$$\text{Precision Rate} = \frac{\textit{Correctly}\ \text{Located}}{\textit{Correctly}\ \text{Located} + \text{False Positive}} \times 100\% \quad (1)$$

$$\text{Recall Rate} = \frac{\textit{Correctly}\ \text{Located}}{\textit{Correctly}\ \text{Located} + \text{False Negative}} \times 100\% \quad (2)$$

TABLE 1
PERFORMANCE COMPARISON

Image source no	Recall rate in edge based (%)	Recall rate in proposed method (%)	Precision rate in edge based (%)	Precision rate in proposed method (%)
1	90	100	50	75.5
2	100	100	47.6	50
3	16	93.3	50	93.3
4	13.2	50	12	25
5	84.6	86	20	30
6	100	100	45	60
7	80	82	80	82

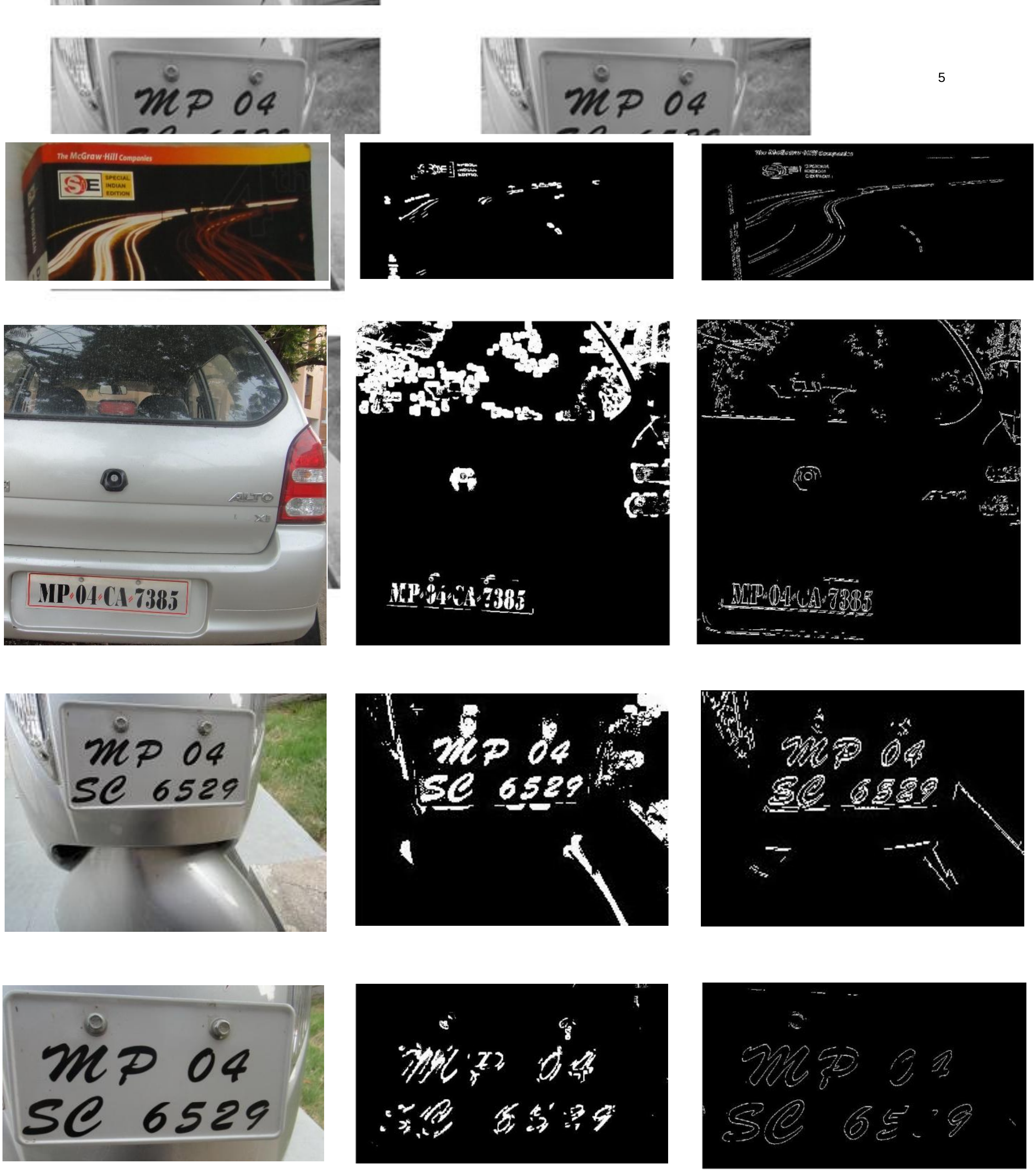


Figure 8. Images with variable font sizes, angles, perspective distortion, colors, scaling and resolutions (a) Original images (b) Extracted text by edge based text detection method [2] (c) Extracted text by proposed method.

ISSN 2229-5518

3 Conclusion

In this paper, nearest-neighbors based approach for localizing text from images is proposed. Results shows that the proposed method is effective in localizing the text areas from natural scenes. Proposed algorithm is robust with respect to aspect to font sizes and styles, orientations, alignment, uneven illumination, and reflection effects. Binary output can be directly be used as an input to an existing OCR engine for character recognition without any further processing. This method distinguishes text areas from texture-like areas, such as window frames, wall patterns, etc., by using the histogram thresholds. Here we used MATLAB for implementation purpose. To reduce the false positive rate, morphological operations can be used. This could increase precision rate also.

References

[1] Wonjun Kim and Changick Kim,"A New Approach for Overlay Text Detection and Extraction From Complex Video Scene," IEEE Transactions on Image Processing, vol.18, no. 2, pp. 401-411, February 2009.

[2] Xiaoqing Liu, Jagath Samarabandu, "Multiscale Edge-Based Text Extraction from Complex Images," ICME, pp.1721-1724, 2006 IEEE International Conference on Multimedia and Expo, 2006

[3] Kongqiao Wang and Jari A. Kangas, "Character location in scene images from digital camera," Pattern Recognition, vol. 36, no. 10, pp. 2287–2299, 2003.

[4] K. C. Kim, H. R. Byun, Y. J. Song, Y. M. Choi, S. Y.Chi, K. K. Kim, and Y. K. Chung, "Scene text extraction in natural scene images using hierarchical feature combining and verification," in Pattern Recognition, 2004, Aug. 2004, vol. 2 of ICPR 2004. Proceedings of the 17th International Conference on , pp. 679–682.

[5] Jiang Gao and Jie Yang, "an adaptive algorithm fot text detection from natural scenes," in Computer Vision and Pattern Recognition, 2001. CVPR 2001, 2001, Proceedings of the 2001 IEEE Computer Society Conference on, pp. II–84–II–89.

[6] X. Liu and J. Samarabandu, "An edge-based text re- gion extraction algorithm for indoor mobile robot navigation," in Proc. of the IEEE International Conference on Mechatronics and Automation (ICMA 2005), Niagara Falls, Canada, July 2005, pp. 701–706.

[7] M. Cai, J. Song and M. R. Lyu. A New Approach for Video Text Detection. I Proc. of International Conference On Image Processing, Rochester, New York,USA, pp. 117-120,2002.

[8] L. Agnihotri and N. Dimitrova, "Text detection for video analysis," in Content-Based Access of Image and Video Libraries, 1999. (CBAIVL '99), 1999, Proceedings. IEEE Workshop on, pp. 109-113.

[9] J. Fan, D.K.Y. Yau, A.K. Elmagarmid, and W.G. Aref, "Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing", IEEE Trans. on Image Processing, 10(10): 1454-1466, 2001.

[10] C. Garcia and X. Apostolidis. Text Detection and Segmentation in Complex Color Images. In Proc. Of international Conference on Acoustics, Speech and Signal Processing (ICASSP2OOO), Istanbul, Vol.4, pp. 2326-2330,2000.

[11] R. Lienhart and W. Effelsberg. Automatic Text Segmentation and Text Recognition for Video Indexing. Multimedia System, Vol. 8, pp. 69-81,2000.

[12] http://www.equasys.de/colorformat.html #YUVColorFormat

[13] http://en.wikipedia.org/wiki/YUV

[14] Alain Trémeau, Basura Fernando, Sezer Karaoglu and Damien Muselet, "Detecting text in natural scenes based on a reduction of photometric effects: problem of text detection", "IAPR Computational Color Imaging Lecture Notes in Computer Science", 2011, Volume 6626/2011, 230-244.

[15] Jianqiang Yan, Dacheng Tao, Chunna Tian, Xinbo Gao, Xuelong Li: Chinese text detection and location for images in multimedia messaging service. SMC 2010:3896-3901.