2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies

Rule Based Filtering Approach for Detection and Localization of Bangla Text from Scene Images

Rashedul Islam
Computer Science and Engineering Discipline
Khulna University
Khulna-9208, Bangladesh
rashedcse98@yahoo.com

Md. Rafiqul Islam
Computer Science and Engineering Discipline
Khulna University
Khulna-9208, Bangladesh
dmri1978@gmail.com

Kamrul HasanTalukder
Computer Science and Engineering Discipline
Khulna University
Khulna-9208, Bangladesh
k.h.t@alumni.nus.edu.sg

Abstract— ***Detection, and localization of Bangla text from natural scene images are important prerequisites for developing Bangla OCR as well as many content-based image analyses. But there is no standard Bangla OCR to be used in the daily work. Due to the presence of some unique features, detection and localization of Bangla text have become more challenging than English text. In this paper, we have proposed MSER based method along with rule-based filtering for efficiently detect and localize Bangla texts from scene images. As the MSER based method is the winning method of the benchmark data, such as ICDAR 2011, this algorithm has been applied to get a better result than related existing methods. By using MSER, candidate text regions are detected. False positives are present into the detected regions. To remove the false positives, rule-based filtering technique has been applied. In this process, geometric properties of text like aspect ratio, eccentricity, Euler number, extent, and solidity have been used to filter out non-text regions. As there is no publicly available database containing scene images of Bangla text, we have developed such database to perform the experiment. The proposed method has been evaluated on 50 sample images of our present dataset containing Bangla and also evaluated on 50 images of ICDAR 2013 benchmark dataset and we have got better results in terms of precision, recall, and f-measure in both cases. A comparison has been made among existing related method and the proposed method and found that the proposed method is better.***

Keywords— ***aspect ratio; Euler number; filtering; ICDAR; MSER;***

I. Introduction

Text in scene images provides important information about semantic of the images [1] that help people to understand the meaning of the images. In some cases, text may be the main component of scene images. Automatic detection and extraction of text from scene images have drawn the attention of the researchers due to its wide application areas like content-based image retrieval, text-based image indexing, automatic annotation of the image, robotics, document analysis, keyword-based image search, etc. It is also useful for those who have language barriers like visually impaired persons and foreigners [2]. It will assist in establishing a text editable method from a scanned document. The efficiency of text extraction depends on proper detection. So at first, we have to design an algorithm that will be capable of detecting and localizing text from the input image with high accuracy. The existing text detection method can be divided into two groups; texture based method and connected component based method. The complexity of texture based method is higher though it is robust. But connected component based (CC-based) methods are simple but they are not robust in all the cases [3].

In Bangla alphabet, there are 39 consonants, 11 vowels, and 10 numerals. Modifiers are used for writing texts in Bangla language. Two types of modifiers are used here; vowel and consonant modifiers and they are used only with consonants. A modifier can be used on any side of a Bangla character. The presence of headline is a unique property of Bangla text. It is a horizontal line always located at the upper portion of a character. Among the 50 basic characters, 10 do not have headlines, 8 have half headlines and the rest have full headlines. There are some characters that contain curved line above them. Such characters are ই,ঈ,উ,ঊ,ঐ,ঔ,ট,ঠ. As per the structure of Bangla texts, they can be partitioned into three zones. The upper zone denotes the portion above the headline, the middle zone covers the portion of basic characters or compound characters below the head-line and lower zone is the portion where some of the modifiers can reside. The middle and lower zones are separated from each other by an imaginary line called the base line. Fig. 1 shows the zones of Bangla script.

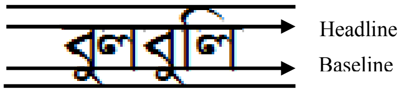

Fig. 1. Different zones of Bangla text.

The concept of uppercase and lowercase is absent in Bangla script and writing style of Bangla is from left to right in a horizontal manner. Due to the presence of some unique features as mentioned above, Detection and recognition of Bangla text have become more challenging than English text. By using the property of having matra or headline, U. Bhattacharya et al. [2] described the process of extraction of Devanagari and Bangla text from natural scene images. They have used morphological operation to perform the task. Rangit

978-1-5090-6266-9/17 $31.00 © 2017 IEEE
DOI 10.1109/ICRTEECT.2017.16

Ghoshal et al. [4] proposed a scheme for extraction and recognition of Bangla text from natural scene image through perspective correction. For text extraction, the authors have detected headline using morphological opening operation. Separation of the components that are closed to or attached with the headline takes place first. Additionally, the authors have distinguished between text and non-text by using certain shape and position based conditions. Text components have very low variation in stroke thickness in comparison with non-text counterparts and presence of a headline along with a few vertical downward strokes originating from this headline are two major characteristics of Bangla and Devanagari text [3]. On the basis of the above features, Aruni Roy Chowdhury et al. [3] present a robust scheme for detection of Devanagari and Bangla text from scene images. In this paper, we propose a robust text detection and extraction method. To achieve the goal, we have applied Maximally Stable Extremal Region (MSER) algorithm [5] along rule-based filtering technique to get better accuracy than existing method [6]. The MSER algorithm is applicable to an image with low quality. This algorithm does not depend on a language while detecting text from scene images [1]. The complexity of the algorithm proposed by J. Matas et al. is O (nlog(log(n))) [5]. Where n is the number of pixels in the image. MSER algorithm produces a large number of false positives. To eliminate these false positives we have applied rule-based filtering technique. After that, all the individual text characters have been merged to form single rectangular bounding box around individual words.

The organization of remainder of the paper is as follows. Related works are described in Section 2, Section 3 gives methodology, Experimental results are described in Section 4, and Section 5 provides conclusions

II. Related Work

Scene image may contain useful information that would be helpful for different categories of people. From this point of view, many researchers are working in this field. Earlier works in this field were confined in scanned documents only where the texts were presented in black color under white background [7]. U. Bhattacharya et al. [2] proposed an efficient method for extraction of Bangla and Devanagari text from scene images by analyzing connected components obtained from the binary image. They did it by using the common feature of this two script i.e. presence of headline. The authors have calculated height, mean, and standard deviation of every candidate headline components and used morphological operations to detect true headlines. Here, morphological operation (opening an object A by linear structuring element B) helps to identify headlines. They performed the experiment on their set of 100 images and got precision=68.8% and recall=71.2%. Their algorithm fails if there is small sized curved text present in the image. Another morphological approach was proposed by Ghoshal et al. [8]. In their proposed method they at first detected unattached text components and then segmented connected components from an image containing Bangla/Devanagari characters. The restriction of the proposed approach is that it can capture only highlighted texts. Uniform stroke thickness and presence of headline are the two major characteristics of Bangla and Devanagari scripts. Aruni Roy Chowdhury et al. [3] described the method of text detection from scene images using these two major characteristics of Bangla text. In the preprocessing stage, they extracted connected components by using morphological closing operation along with canny edge detector. Then they selected candidate text regions by computing stroke thickness. The authors have tested the algorithm on 100 sample images of their database and obtained precision=72% and recall = 74%.

An image, when captured by a digital camera may have perspective distortion. The above algorithms cannot handle the images having perspective distortion. Rangit Ghoshal et al. have corrected the distortion by using homographic transformation [4]. After correcting the perspective distortion, the authors have detected headlines by applying morphological operation. Then the authors have separated the components those are attached with the headline. The components selected by the above procedure may include texts as well as non-texts. The authors have separated headline attached text components by applying some characteristics of text like elongation, holes, aspect ratio, object to background pixel ratio. There are some text symbols that are not attached with headlines. For separation of these texts, the authors have taken the measure of increasing the area of the bounding box enough so that the text components lie inside it and thus separated. Then they removed the headlines and separated the text components. Another approach for scene text detection is the use of stroke width feature. Epshtein et al. estimated this stroke width by dense calculation of “stroke width transform” in a bottom-up approach from pixel level. They proposed a novel CC-based text detection algorithm [9].

Md Zahidul Islam et al. [15] used the texture based approach in which, an image is decomposed and pre-processed to extract features. Then they have used these features to train two classifiers that are based on ANN and SVM classifiers. They have used 56 features from which 8 are calculated from the mean, second order and third order central moment and the other 48 are calculated from Wavelet Histogram Energy (WHE). The overall technique includes an iterative scan through the input by a fixed block size and defines whether it is text or non-text. The authors used a 16×16 block size and a scan interval of 4 pixels. Their system is divided into the following phases:

- Image decomposition
- Feature extraction and selection
- Text detection

Prakriti Banik et al. [14] proposed a method of segmentation of characters or their parts from Bangla texts extracted from scene images. The proposed algorithm can detect background and texts by combining unsupervised learning k-means clustering and Otsu’s threshold selection. In order to select optimal K value for K-means clustering, the authors have proposed criteria to select. The segmentation is based on region growing and extraction of both headline and baseline of Bangla texts. The algorithm proposed by the authors work in the following steps as shown in *Fig. 2*.

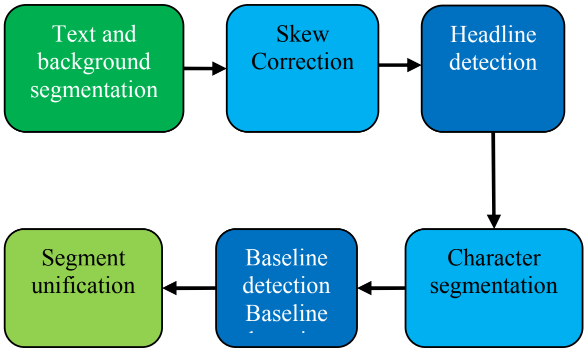

Fig. 2. Different zones of Bangla text [14].

H. Chen et al. [10] proposed an algorithm that uses a combination of MSER and Canny edge detector for detecting text candidates. Image blur is efficiently handled by this combination. Because close symbols are distinguished by the canny detector. This is achieved by removing the MSER pixels outside the boundary formed by the canny edges. Different types of filtering techniques like size, aspect ratio, the number of holes, stroke width, have been used by the authors to filter out the non-text regions. For the construction of text candidates, the authors have used the single-link clustering algorithm. Main parameters used for this task are spatial distance, width, height, and aspect ratio. There is an additional check after text candidates are built. The probability of line being a text line if it contains three or more text objects. A text line is rejected if a significant portion of the objects is repetitive. The text lines are then split into individual words using Otsu's method [11]. In order to detect text from scene images, prior knowledge is used by most of the text detection methods. The datasets those contain scenes of only English language is used by such algorithms [1]. In [1], the authors have described the algorithms where the language of the text is known and investigate them as an algorithm that does not depends on language. They have worked on the methods proposed by Chen et al. [10], Yin et al. [12], and Gomez et al. [13]. Different approaches were used by these algorithms and produced good results on a benchmark dataset of ICDAR. MSER algorithm was used by all of these methods for extraction of character candidates [1].

On the basis of common properties of text, different approaches have been proposed for the extraction of text from images. Victor Wu et al. stated that text has some common distinctive characteristics in terms of frequency, orientation information, and also spatial cohesion [16]. Spatial cohesion indicates that position of every text character under a string is near to each other. Their orientation, height, and spacing are nearly same. The spatial cohesion of characters is determined by edge-based algorithm [17] and connected component (CC) based algorithm [18] of text characters. Rashedul et. al. [6] proposed an approach to extract text regions from scene image, which is based on hybrid method formed by combining edge based and connected component (CC) based methods. They compared their approach to Edge based and CC based methods and showed better results than these two existing methods. They tested the methods on only a few images (8 images). In order to compare the performance of our proposed method, we have implemented the hybrid method [6] and tested on 50 images from our own database and 50 images from ICDAR 2013 database and the result shows that our method is better in respect of precision, recall, and f-measure.

III. Proposed Approach

Proper detection is the key issue of text detection and localization algorithm. The efficiency of this algorithm much depends on text detection approach. So our first target is to design an algorithm to detect text regions from the input scene image with a better result than existing methods. As filtering or removing non-text is the key challenge of every text detection algorithm, we tried to use better filtering technique to achieve our goal. In our proposed approach, Maximally Stable External Region (MSER) based method that was proposed by Matas et al. [20] have been used for detecting candidate text regions. To achieve higher precision and recall rates, rule-based filtering technique has been introduced in this proposed approach. The overall organization of the proposed method is shown in *Fig. 3*.

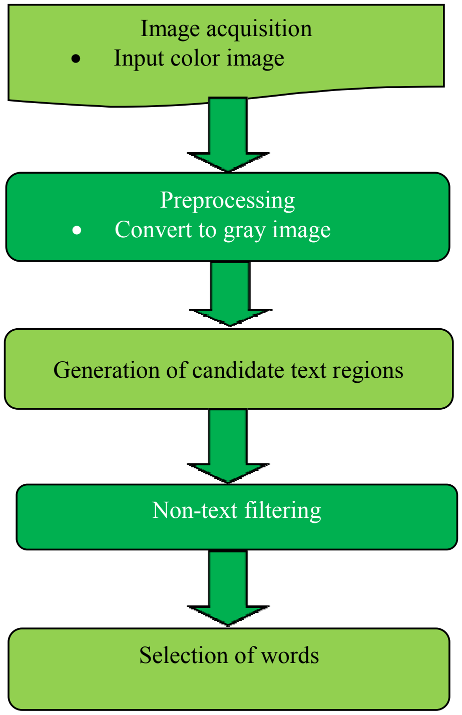

Fig. 3. Proposed method of Bangla text localization.

A. Image Acquisition

The image has captured by using a digital still camera.

B. Image Preprocessing

In this step, the input color image is converted to gray scale image (0 to 255-pixel values).

C. Generation of candidate text regions

Candidate text regions can be found by the Maximally Stable Extremal Regions (MSER) algorithm. Texts in scene images have consistency in color and high contrast that leads to stable intensity profiles. For this reason, MSER algorithm is used to select candidate text regions from the obtained gray scale image. The MSER algorithm is composed of the following major parts.

1) Preprocessing: As we know, the intensity of a gray scale image varies in a different location of the image. According to intensity order, pixels are sorted and the number of pixel in each intensity is determined for further processing.

2) Clustering: In this step, a representation of all regions at each intensity level is created.

3) MSER detection: On the basis of variations of intensity, MSER regions are marked on the gray scale image by ellipses

4) MSER display: Candidate text regions are surrounded by ellipses as shown in *Fig. 4*.

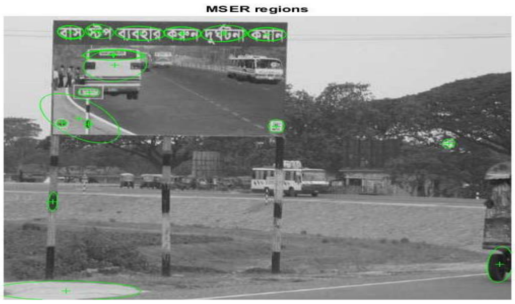

Fig. 4. MSER regions of an image.

D. Non-text filtering

In order to filter out non-text regions from selected MSER's following geometric properties of text can be used.

- Aspect ratio
- Eccentricity
- Extent
- Euler number
- Solidity

Brief explanations of the above rules mentioned in [19] are stated below

- *Aspect ratio:* Ratio of width and height of a candidate region is known as the aspect ratio. If it is more than 3, the region is considered as a non-text region and thus eliminated. It is shown in (1).

$$AR = \frac{\text{witdh of a region}}{\text{height of the region}} \qquad (1)$$

- *Eccentricity:* Ratio of the distance between the foci of an ellipse and its major axis length is termed as eccentricity. The range of the value is in between 0 and 1. An Ellipse having the value of eccentricity =0 represents a circle, while an ellipse with eccentricity = 1 is a line segment. Eccentricity (E) is shown in (2).

$$E = \frac{D}{L} \qquad (2)$$

where D is the distance between the foci of the ellipse and L is the major axis length of the ellipse.

- *Euler number:* The numeric value obtained by subtracting the number of holes in the objects of a region from the number of objects in that region is known as Euler number. Euler number (EN) of a region is shown in (3).

$$ENL = num\_\_obj - num\_\_hole \qquad (3)$$

- *Extent:* Ratio of the area of a region to the area of the bounding box is called extent. Extent (ET) is shown in (4).

$$ET = \frac{\text{area}}{\text{area of bbox}} \qquad (4)$$

- *Solidity:* Solidity of a region can be defined as the proportion of the pixels in the convex hull that is also in the region. Solidity (S) is shown in (5).

$$S = \frac{\text{area}}{\text{convex\_\_area}} \qquad (5)$$

We have extracted all of the above mentioned features of the MSER regions from the input image. By using the feature values, non-text regions have been filtered out from the candidate MSERs.

E. Identification of words

At first, every single character or region has bounded by small rectangular boxes. In order to get individual words as a single unit, these small bounding boxes have merged by following the procedures as stated below.

- Expand the area of each small bounding box by 2%
- Find out neighboring regions.
- A series of coinciding bounding boxes will be produced.
- Computing overlap ratio between all the bounding box pairs.
- Merge these boxes to get a single rectangular box around individual words.

Finally, an output image has obtained by applying the above procedures on an image. *Fig. 5* shows it where red colored rectangular box indicates the single region of texts.

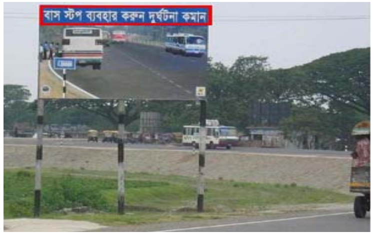

Fig. 5. Finally detected texts.

IV. Experimental results

The proposed method has been implemented on Windows platform using MATLAB r2016a. As there is no publicly available standard database of scene images containing Bangla text, we have developed one such database where images are captured by OLYMPUS VH-210 (14 MP) Digital still camera from different places and downloaded from google image. We have an intention to make this database available to the researchers. 250 sample images have been taken from the said database to collect simulation result of the proposed method. A few of these sample images with detected texts are shown in *Fig. 6*.

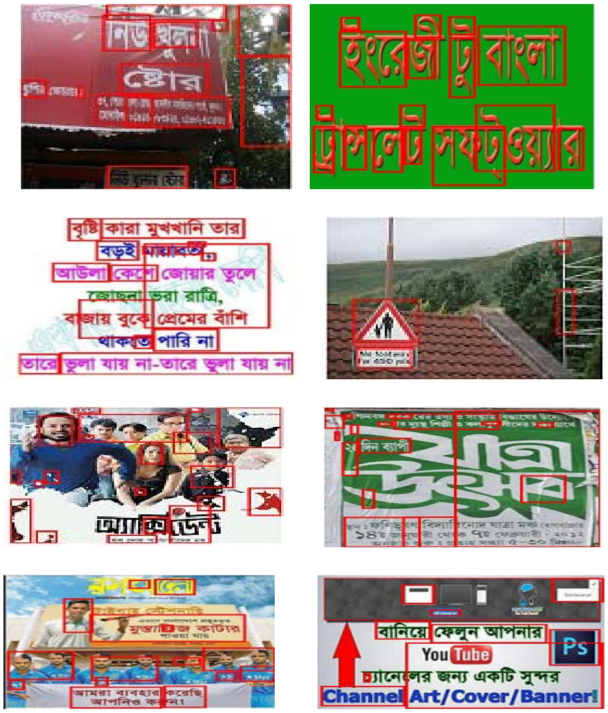

Fig. 6. A few sample images from our database. Detected texts are marked by red color rectangular area.

The simulation results have been summarized by manually counting the number of Correctly Detected Characters (CDC), False Positives (FP), and False Negatives (FN). This process of counting is known as ground truth [17]. We have calculated precision, recall, and f-measure defined in (6), (7) and (8) respectively.

$$P=\frac{\text{CDC}}{\text{CDC+FP}}\times 100\% \tag{6}$$

$$\text{R}=\frac{\text{CDC}}{\text{CDC+FN}}\times 100\% \tag{7}$$

$$f-measure = \frac{2\times P\times R}{(P+R)}\times 100\ \% \tag{8}$$

where P is the precision and R is the recall

The proposed method provides precision=73.47%, recall= 88.18%, and f-measure=77.55%. To observe the results of the proposed method, 50 sample images from ICDAR 2013 database were taken and got precision=76.26%, recall= 80.68% and f-measure=75.21%. Fig. 7 shows a few samples of the images of ICDAR 2013 with detected texts marked by red colored rectangular boundaries.

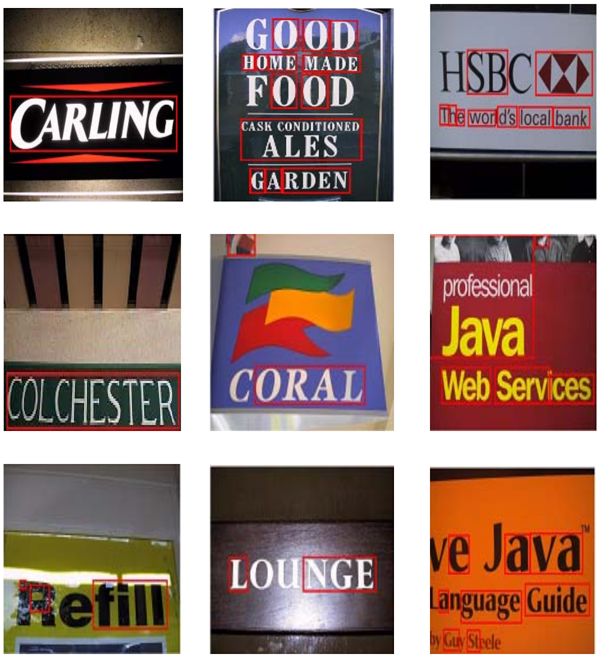

Fig. 7. : A few samples from ICDAR 2013 database of scene images where detected texts are marked by red colored bounded boxes by proposed method

Table I and Table II show the comparison results of proposed method with existing hybrid method and Connected Component (CC) based methods on 50 sample images of our database and ICDAR 2013 benchmark database. Graphical representation of the comparison of the three methods according to Table I is shown in *Fig. 8*.

TABLE I. PERFORMANCE MEASURE OF 50 SAMPLE IMAGES FROM OUR DATABASE

Method	Precision (%)	Recall (%)	f-measure (%)
Proposed	73.47	88.18	77.55
Hybrid [6]	67.28	81.80	70.70
CC [18]	64.78	76.52	68.59

TABLE II. PERFORMANCE MEASURE OF 50 SAMPLE IMAGES FROM ICDAR 2013 DATABASE

Method	Precision (%)	Recall (%)	f-measure (%)
Proposed	76.26	80.68	75.21
Hybrid [6]	68.57	70.59	65.39
CC [18]	65.47	69.72	62.07

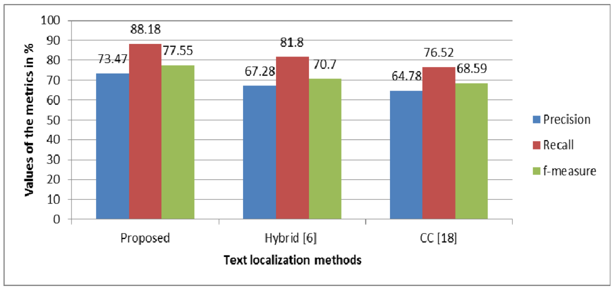

Fig. 8. A comparison among proposed method, a hybrid method, and CC based method on the basis of precision, recall, and f-measure on 50 sample images from our database contains Bangla scene images.

5. Conclusions

Detection and localization of Bangla text from the scene images is more challenging to achieve satisfactory performance in all the applications due to large variations in character font, size, orientation, color etc. In order to increase the performance of the proposed algorithm, rule-based filtering technique has been used. The advantage of MSER algorithm is that it is invariant to language. Though it produces many false positives, the average precision, recall, and f-measure that were achieved was remarkable. The results show that our proposed method performs well than the existing method [6]. From our experiment, it has observed that in some cases the proposed method generates many false positives. So our future plan is to use machine learning approach to enhance the task of filtering technique that will improve the performance of the proposed method. We also plan to make our database (that contains scene images of Bangla text) publicly available for academic researchers.

REFERENCES

[1] Mikhail Zarechensky, Text detection in natural scenes with multilingual text. In the Proceedings of the Tenth Spring Researcher's Colloquium on Database and Information Systems, Veliky Novgorod, Russia,2014.

[2] Bhattacharya, U., Parui, S. K., Mondal, S.: Devanagari and Bangla Text Extraction from Natural Scene Images. In Proceedings of the 10th International Conference on Document Analysis and Recognition, pp. 171–175, Washington, DC, USA 2009.

[3] Aruni Roy Chowdhury, U. Bhattacharya, S. K. Parui,Text Detection of Two Major Indian Scripts in Natural Scene Images, CBDAR 2011, LNCS 7139, pp. 42–57, 2012.

[4] Ranjit Ghoshal, Anandarup Roy, Swapan K. Parui, Recognition of Bangla text from Scene Images through Perspective Correction, In proceedings of International Conference on Image Information Processing (ICIIP), IEEE conference paper, pp.1-6, 03 Nov - 05 Nov 2011, Shimla, Himachal Pradesh, India.

[5] J. Matas, O. Chum, M. Urban, and T. Pajdl. Robust wide baseline stereo from maximally stable extremal regions. In British Machine Vision Conference, volume 1, pages 384–393, 2002

[6] Rashedul Islam, Md. Rafiqul Islam, Kamrul HasanTalukder, An Approach to Extract Text Regions from Scene Image, 2016 International Conference on Computing, Analytics and Security Trends (CAST), College of Engineering Pune, India. Dec 19-21, 2016

[7] A.O.M Asaduzzaman, Md. Khademul Islam Molla and M. Ganjer Ali, Printed Bangla Text Recognition using Artificial Neural Network with Heuristic Method, Proc. ICCIT'2001, 28-29 December, East West University, Dhaka, Bangladesh

[8] Ranjit Ghoshal, Anandarup Roy, Tapan Kumar Bhowmik and Swapan K. Parui, Headline based Text Extraction from Outdoor Images, Pattern Recognition, and Machine Intelligence, Lecture Notes in Computer Science, 2011, Volume 6744/2011, 446-451, DOI: 10.1007/978-3-642-21786-9_72.

[9] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. Proceeding of CVPR, pp. 2963-2970, 2010

[10] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, V. Chandrasekhar, G. Takacs, R. Vedantham, R. Grzeszczuk, and B. Girod. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. IEEE International Conference on Image Processing, Sep 2011

[11] N. Otsu. A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics, vol.9, no. 1., pp. 62-66, 1979.

[12] Xu-C. Yin, X. Yin, K. Huang, H. Hao, "Robust text detection in natural scene images, " in IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 36, no.5, pp. 970-983, 2014.

[13] L. Gomez and D. Karatzas, Multi-script text extraction from natural scenes, ICDAR, 2013.

[14] P. Banik, U. Bhattacharya, S. K. Parul, Segmentation of Bangla Words in Scene Images,Proceedings of the 8th Indian conference on computer vision, Graphics and Image processing, Article no. 70, ICVGIP'12, December 16-19, 2012, Mumbai, India.

[15] Md Zahidul Islam_ and Amit Kumar Mondal, Towards a Standard Bangla PhotoOCR: Text Detection and Localization, 17th International Conference on Computer and Information Technology, pp. 198-203, 22-23 December 2014, Dhaka, Bangladesh.

[16] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, Text-finder: An automatic system to Detect and Recognize Text in Images, IEEE Transaction on Pattern analysis and Machine intelligence, vol. 21, no. 11, November 1999

[17] Xiaoquing Liu and Jagath Samarabandu, An Edge-based text region extraction algorithm for indoor mobile robot navigation, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol. 1, No. 7, pp. 2043-2050, 2007.

[18] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, A robust algorithm for text detection in images, Proceedings of the 3rd International symposium on Image and signal Processing and analysis, 2003.

[19] http://www.mathworks.com/help/images/ref/regionprops.html [Accessed on September 2016] .

[20] Matas, O. Chum, M. Urban, and T. Pajdla. , Proceeding of British Machine Vision Conference, pages 384-396, 2002.

[21] Li, Yao, and Huchuan Lu, Scene text detection via stroke width, Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012