AUTOMATIC DETECTION AND TRANSLATION OF BENGALI TEXT ON ROAD SIGN FOR VISUALLY IMPAIRED

S. M. Anamul Haque[†], Shahida Arbi, Tabassum Tamanna and Sadia Mahsina Itu
[†]Department of Computer Science and Engineering, Northern University Bangladesh.
Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
Email: anam_ecs@yahoo.com, munni_222@yahoo.com, ovi_cse@yahoo.com, sadia_itu@yahoo.com

Abstract: *Large amount of information are embedded in natural scenes. Signs are good example of natural objects with high information content. Visually impaired individuals are unable to utilize the significant amount of information in signs. This paper presents a system for detecting and recognizing the signs in the environment specially written in Bengali and voice synthesizing their contents. In this paper we present an algorithm for detection and localization of text in road sign using edge detection approach. We use an adaptive thresholding method to binarize the text blocks and recognize the text using a neural network based OCR module. Finally, the recognized text is synthesized as voice output message convenient for visually impaired. The detection, recognition and speech synthesis modules each perform well on their respective task, and initial evaluation of the complete system is promising.*

Keywords: *road sign, text localization, OCR, speech synthesis, visually impaired.*

1. Introduction

Signs are everywhere in our lives. A sign is an object that suggests the presence of a fact. They make our lives easier when we are familiar with them. Unfortunately, the visually impaired are deprived from such information, which limits their mobility in unconstrained environments. Road signs are typically placed either by the roadside or above roads [1]. Road signs provide important information for guiding, warning in order to make movement safer and easier. For example, when a visually impaired person passes though the road in a city they find the guideline and information from road sign. This information will significantly improve the degree to which the visually impaired can interact with their environment that a sighted person does. Hence the research field of road signs text recognition and translated to suitable form receives a growing attention due to the great variety of potential application.

Automatic sign translation system utilizes a camera to capture the image with signs, detect the text on signs, recognize the text and translate the result of sign recognition in to target formats. Such a system relays on technologies of sign detector, OCR, and translation. Hence, our aim is to develop a system which is capable of capturing image from natural scene, automatically detecting and recognizing text from Bangladeshi road signs, and translating the text as voice output message. Our approach for detecting the text region is based on edge detection techniques together with several heuristic methods. The recognition method is neural network based conventional Optical Character Recognition (OCR) system. Microsoft Speech SDK (TTS) is used to translate the recognition text to voice stream. The work is related to existing research in text detection from general background or video image [2]-[7], Bangla Optical Character Recognition (OCR) system [8]-[10]. Some researchers published their efforts on texture-based [11]-[13] text detection. Some other researchers used the edge based method [14]-[16] to locate the text on image. Some have created system specifically to detect and read text from signs in natural images. Jing Zhang et al. [17] proposed a PDA-based sign translator that can capture sign image, auto segment and recognize the Chinese sign, and translate it into English. Silspschote et al. [18] demonstrated a system for detecting and

recognizing signs in the environment and voice synthesizing their contents.

This paper is organized as follows: section 2 provides an overview of the overall system. Section 3 describes text detection challenges and the approach we have followed. Section 4 deals with recognition of characters based on neural network. The speech synthesis technique is explained in section 5. Section 6 discusses experimental result and the paper concludes with section 7.

2. Structure of the Proposed System

The system proposed in this paper is designed for automatic detection and recognition of Bengali text on road signs and translate into voice stream. The system consists of three modules: Text detection and extraction, Optical Character Recognition (OCR) and speech synthesis. First module includes: (i) Data acquisition (ii) Pre-processing (iii) Text detection and localization. The OCR module includes: (i) Character segmentation (ii) Feature extraction (iii) Character Classification using Artificial Neural Network. The speech synthesis module converts the recognized text into voice streams as the system output. Figure-1 shows the diagram of the system architecture.

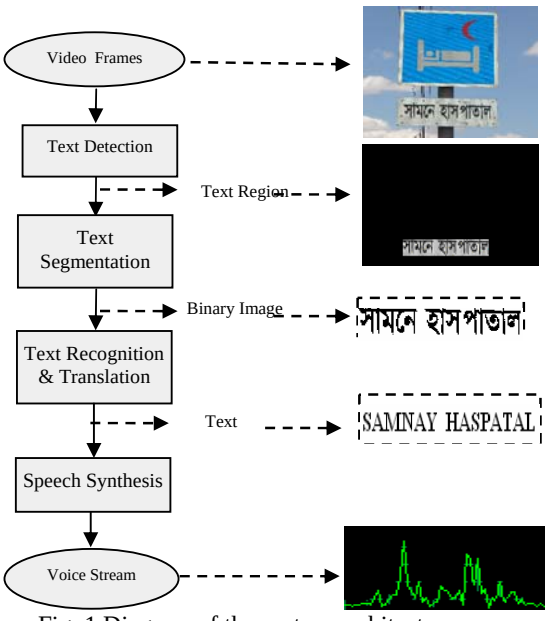

Fig. 1 Diagram of the system architecture

3. Text Detection and Extraction

Automatic detection of text from natural scene is a challenging task because they usually embedded in the environment. The task is related to text detection and extraction from video image. Text extraction from video consists of three major steps. The first one is to find text region in original images. Then the text needs to be separated from background. And finally a binary image has to be produced.

Since the gradient of intensity (edge) is more stable than intensity itself in lighting changes; the edge based method is applied for text detection in our system. The main idea behind this approach is that text contrasts a lot with background. After the edges will be computed, the number of edges in x and y direction will be calculated and if it is higher than a certain threshold then it will be considered as a text area. Then each text area will be binarized using the luminance. So in a final result, the text will be in white and the background in black (or the inverse). Images taken from camera are histogram equalized for better contrast and then the following algorithms are sequentially applied to locate and extract the text from the image.

First, the captured images are converted into Gray scale image and applied a 3-by-3 median filter to blur the background. The characters are slightly blurred, but the background areas are more blurred in the image as shown in Figure-2(b). The edge of the image is computed using Sobel vertical edge emphasis filter as shown in Figure 2(c). The text area is dilated by applying 6-by-6 maximum filter and compute its horizontal projection which depicted in Figure 2(d). The horizontal projection histogram is scanned and excluded the line having a projection value less than a threshold value. The outcome of this stage is shown in Fig. 2(e). Then, we have applied Sobel horizontal edge emphasizing filter for each possible text area and computed their histogram to exclude them properly and obtained the segmented text area (Fig. 2 (g)).

Finally we have converted our segmented image to binary image by using a luminance threshold and final result is shown in Fig. 2(h). Finally skew correction is performed on

the extracted text and send to the OCR module for recognition.

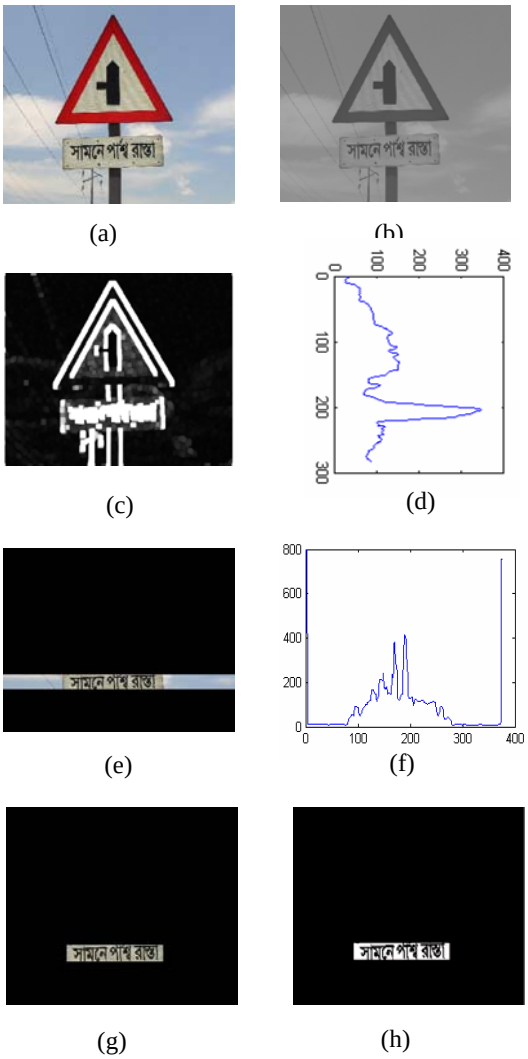

Fig. 2 (a) Original Image (b) Blurred Background (c) Sobel vertical edge emphasize Result (d) Horizontal histogram (e)Vertical Exclusion (f) Vertical histogram (g) Extracted text area (h) Binary image

4. Optical Character Recognition (OCR) Module

We used a neural network based optical character recognition method to recognize the extracted binary text. The character recognition module that we have developed consists of five major components which are illustrated in Figure-3. The input of this module is the binary text image and the output is the recognized ASCII text. The whole process is discussed below.

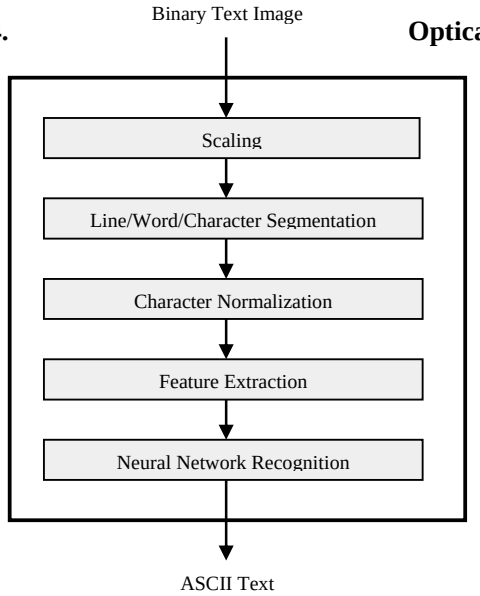

Fig. 3 Diagram of the OCR module

4.1. Scaling

In order to segment and recognize size, independent printed Bangla text, scaling techniques have been used. The proposed system uses 18 pt fonts to segment the printed Bangla text. If a document containing characters of size more than 18 pt font then the system will be scaled down all of the characters to 18 pt font and if less than 18 pt font then the system will be scaled up all of the characters to 18 pt font.

4.2. Text Segmentation

Segmentation of Bengali text is a process by which the text is partitioned into its coherent parts. The text image contains a number of text lines. Each line again contains a number of words. Each word may contain a number of characters [8]. Our text Segmentation process includes: a) Text Line Detection b) Word Segmentation c) Zone finding d) Character Segmentation

4.2.1. Text Line Detection

Text line detection has been performed by scanning the input image horizontally. Frequency of black pixels in each row is counted in order to construct the row histogram .The position between two consecutive lines, where the number of pixels in a row is zero denotes a boundary between the lines [10].

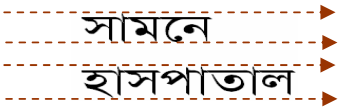

Fig. 4 Text Line Segmentation

4.2.2. Word Segmentation

After a line has been detected, it is scanned vertically. In order to find the column histogram, number of black pixels in each column is calculated. If there exists *n* consecutive scan that find no black pixel we denote it to be a marker between two words. The value of n is taken experimentally. Figure-5 shows the word segmentation process.

Fig. 5 Word Segmentation

4.2.3. Zone Finding

A word may be partitioned into three zones. The upper zone denotes the portion above the headline, the middle zone covers the portion of basic (and compound) characters below headline and the lower zone is the portion where some of the modifiers can reside. The imaginary line separating middle and lower zone is called the base line [8].

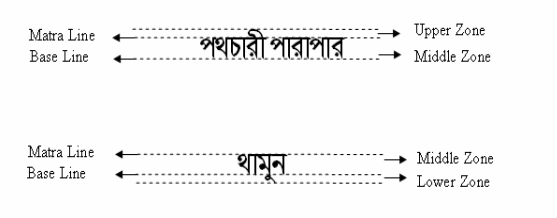

Fig. 6 Zone Finding

The *Matra* line is that line where the upper zone and the middle zone are separated and in the Matra line the number of black pixels will be maximum (see Figure-6). A row wise histogram of black pixels is performed to find out the *Matra* line of a word. In an 18 pt font, the average height of characters is constant. We use this average height to calculate the base line.

4.2.4. Character Segmentation

To detect the character boundary, a vertical scan is initiated from the row that is just under the Matra line to the base line of a word. The starting boundary of a character is the first column where the first black pixel is found. After finding the starting boundary of a character, it continues scanning until a column without any black pixels is found, which is the ending boundary of the character being processed. After getting the character boundary, we check the upper zone, lower zone and calculate the width. Then fed it to the trained neural network to recognize the character. If the neural network does not recognize the character then we collect the upper zone window of that character. If the upper zone window contain any symbol like 'ে' or 'ৈ' then we replace it by the modifier 'ি' or 'ী' respectively. We also collect the lower zone window of that character. If the lower zone window contain any symbol like 'ু' then we replace it by the modifier 'ু'. If the width of the character is greater then the average width then we perform the piecewise linear scanning [8] to separate the character and separate the characters which is shown in Figure-7.

(a) (b)

Fig. 7 Character Segmentation Process (a) Linear Scanning (b) Piecewise Linear scanning

4.3. Feature Extraction

Feature extraction provides an important role in character recognition systems. This effectively reduces the number of computations and hence faster the recognition process. In our feature extraction algorithm, the image is resized into 200 X 200 matrix containing '0s' and '1s', where 0 represent the presence of the character and 1 represent the absence of the character. Our algorithm uses 10X10 window and count the number of black pixel within a window. If the number of black pixels is more than or equal to the 60% of the total number of pixels in the window, the cell is considered as 0 (i.e. black pixel) otherwise the cell value is 1 (i.e. white). This process is applied in the whole image. After completing this process finally 200X200 images is reduced into 20X20 image. The algorithm for this conversion is presented in Algorithm-1.

```
1       X =1
2       For I=1 to M increased by 10
3          Y=1
4          For J=1 to N increased by 10
5          BEGIN
6             BlackPixelCount=0
7             For K=I to I+9
```

```
8            For L=J to J+9
9            BEGIN
10             IF Image (K, L) =BlackPixel then
11           BlackPixelCount=BlackPixelCount+1
12             END
13           IF BlackPixelCount > =60 then
14             FeatureMatrix (X, Y) = 0
15           Else
16             FeatureMatrix (X, Y) = 1
17         Y=Y+1
18         X=X+1
19       END
```

4.4. Neural Network

We use a neural network to recognize the road sign characters. At first, a multilayer neural network is trained using back propagation learning algorithm [19] with adequate training character data set. Three layers neural networks have been used for improving the classification capability of the neural network with minimum error tolerance rate. The overall neural network architecture is shown in Figure-8. The network has 400 inputs according to the feature matrix. The numbers of hidden neurons are 50 and 70 which are selected by trial and error method and the number of output neurons is 20.

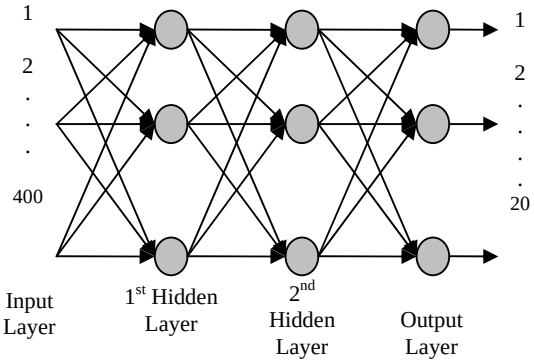

Fig. 8 Multi Layer feed-forward (MLP) network Model

5. Speech Synthesis

Speech Synthesis or text-to-speech (TTS) is the process of converting text into spoken language. It generates sounds similar to those created by human vocal cords.

Even now, Bengali Text-to-Speech or speech synthesis is not available in commercially. But, a number of English Text-to-speech or speech synthesis engine is available. In our system, we have used Microsoft Speech SDK (Version 5.1) to generate Bangla speech. Since, this engine is able to generate speech only from English text; we represent our recognized Bengali word in *US Roman letters* as shown in Table-1.

Table 1: Translation Bengali text into US Roman letters format

Text in Bengali letters	Text in US Roman letters
সামনে	SAMNAY
হাসপাতাল	HASPATAL
বাস ষ্টপ	BUS STOP

The speech synthesis engine converts the Bengali Roman letter words into phonetic and prosodic symbol, generate digital audio stream. Then, Sound card converts the audio stream to acquisition signal and amplifies through speaker. Finally, the recognized text on road sign is hard from the speaker.

6. Experimental Result

Experiments have been performed to test the proposed system and to measure the accuracy of the system. We have simulated our software using MATLAB 7.1, Visual Basic 6.0 as programming language, Microsoft Access 2000 as database, and Microsoft voice engine tools for speech synthesis system. A number of video sequences were collected for experiments with one sign included in each sequence. We have divided our system in three main phases: Text Extraction, Segmentation and Recognition. So the overall performance of the system directly depends on the performance of the three individual phases. The experiment results are summarized in the tables in the next page.

Table 2: Result of Text Extraction

Sequence	No. of Frames	No. of Text Extraction	Accuracy
Seq 1	1050	900	85.7 %
Seq 2	2000	1696	83.8 %

Table 3: Result of segmentation

No. of Extracted text	Line segmentation rate	Word segmentation rate	Character segmentation rate
900	97.1 %	96.3 %	92 .08 %
1696	95.4 %	93.9 %	88.02 %

Our training set comprises of characters of 8 Bangla fonts (of size 18) namely SutonnyMJ, Ruposhe, Daleshari, PorashMJ , KarnaphuliMJ, SushreeMJ, DhanshirhiMJ and ShurmaMJ which are usually used to write text in the road sign. We have tested the neural network with characters of various size and fonts. Our experimental result of recognition is shown in table-4.

Table 4: Character Recognition Result

Total Segmented Character	Correct Recognition	Wrong Recognition
1000	95.7 %	4.3 %

Table 5: Test Result of whole system

Modules	Correct	Error
Text Detection	84.7%	15.3%
Segmentation	93.9%	6.1%
Character Recognition	95.7%	4.3%
Whole system	91.4%	8.6%

The successful text detection percentage of 84.7% is reasonably good but would probably not satisfactory compared to the segmentation and OCR module. The 15.3% of signs that the system failed to locate the text correctly due complex background or low intensity signs or multiple sign boards in the same frame as shown in Figure-9.

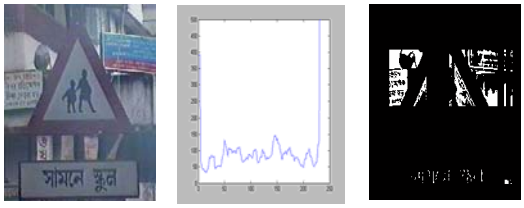

Fig. 9 Difficulties of sign detection

7. Conclusion

In this paper, we described a system which processes the text from Bangladeshi Road Sign for the visually impaired to access textual information in the environment as voice stream. This integrated system uses video, image processing, Optical Character Recognition (OCR) and text-to-speech (TTS) engine. Our current system is focused on automatic text detection and translation only from road sign. Our future goal is to develop a system that can detect and translate a wide variety of signs, including traffic, government, public, and commercial sign and operate on wearable device like PDA (Personal Digital Assistant).

8. References

[1] Marwan A. Mattar, Allen R. Hanson Erik, G. Learned-Mille. “Sign Classification for the Visually Impaired,” Retrieved April 11, http://www.cs.umass.edu/~elm/papers/techrep05_14.pdf.

[2] N. Ezaki, M. Bulacu and L. Schomaker, “Text detection from natural scene images:Towards a system for visually impaired persons”, In Proceedings of the International Conference on Pattern Recognition, 2004, pp. 683-686.

[3] T. Yamaguchi, Y. Nakano, M. Maruyama, H. Miyao, and T. Hananoi, “Digit classification on signboards for telephone number recognition”, In Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003), volume I, Edinburgh, Scotland, 3-6 August 2003, pp. 359–363.

[4] K. Matsuo, K. Ueda, and U. Michio, “Extraction of character string from scene image by binarizing local target area”, Transaction of The Institute of Electrical Engineers of Japan, 122-C(2), February 2002, pp. 232–241.

[5] T. Yamaguchi and M. Maruyama, “Character Extraction from Natural Scene Images by Hierarchical Classifiers,” In Proc. of the Int’l Conf. on Pattern Recognition, 2004, pp. 687-690.

[6] J. Yang, J. Gao, Y. Zang, X. Chen, and A. Waibel, “An automatic sign recognition and translation system”,In Proceedings of the Workshop on Perceptive User Interfaces, Nov. 2001, pp. 1-8.

[7] H. Li, D. Doermann, and O. Kia, “Automatic Text Detection and Tracking in Digital Video,” IEEE Trans. Image Processing, Vol. 9, No. 1, Jan 2000.

[8] B.B.Chaudhuri and U.Pal, “A Complete Printed Bangla OCR System,” Pattern Recognition Vol-31, 531-549, 1997.

[9] 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’2002, 27-28 December, East West University, Dhaka, Bangladesh.

[10] Jalal Uddin Mahmud, Mohammed Feroz Raihan, Chowdhury Mofizur Rahman. “A Complete OCR System for Continuous Bengali Characters”.

[11] H. Li, D. Doermann, and O. Kia, “Automatic Text Detection and Tracking in Digital Video,” IEEE Trans. Image Processing, Vol. 9, No. 1, Jan 2000.

[12] Y. Zhong, H. Zhang, and A.K. Jain, “Automatic Caption Extraction of Digital Videos,” Proc. ICIP’99, Kobe, 1999.

[13] T. Sato and T. Kanade, “Video OCR: Indexing digital news libraries by recognition of superimposed caption,” ICCV Workshop on Image and Video retrieval, 1998.

[14] W. Qi, et. al. "Integrating Visual, Audio and Text Analysis for News Video". 7th IEEE International Conference on Image Processing (ICIP2000), Vancouver, Canada, 10-13 September 2000.
[15] A. Wernicke, R. Lienhart. "On the Segmentation of Text in Videos". IEEE Int. Conference on Multimedia and Expo. pp. 1511-1514, July 2000.
[16] T. Sato, T. Kanade, E. Hughes, and M. Smith. "Video OCR for Digital News Archives," IEEE Int'l Workshop on Content-Based Access of Image and Video Databases, pp. 52 - 60, Jan 1998.
[17] Jing Zhang, Xilin Chen, Jie Yang, Alex Waibel, "A PDA-based sign translator," http://www.is.cs.cmu.edu/papers/speech/icmi02/icmi02_xilin.pdf
[18] Piyanuch Silapachote et al., " Automatic Sign Detection and Recognition in natural Scene," http://www.cs.umass.edu/~weinman/pubs/silapachote05automatic.pdf
[19] Simon Haykin, "Neural Networks: A Comprehensive Foundation," Second edition, Pearson Education Asia, 2001.

S. M. Anamul Haque was born in Dhaka, Bangladesh. He received the B. Sc. degree in Electronics and Computer Science from Jahangirnagar University, Dhaka, Bangladesh. He is currently working toward the M. Sc. degree at department of computer Science and Engineering in the same university. Since December 2006, he has been a senior lecturer at the Department of Computer Science and Engineering, Northern University Bangladesh. He has authored and co-authored more than eight publications in journals and international conference proceedings. His research interest includes automation, neural network, pattern recognition, image processing, computer vision and Bioinformatics.

Shahida Arbi received the B. Sc (Engg.) degree in computer Science and Engineering from International Islamic University Chittagong, Bangladesh. Her area of research interest includes image processing, pattern recognition, Computer vision, neural network and fuzzy logic.

Tabassum Tamanna received the B. Sc (Engg.) degree in computer Science and Engineering from the International Islamic University Chittagong, Bangladesh. Her research interests are in the areas of computer vision, pattern recognition, neural networks and system programming.

Sadia Mahsina Itu received her B. Sc (Engg.) degree in computer Science and Engineering from International Islamic University Chittagong, Bangladesh. Her research interests include image processing, computer vision, MIS and E-commerce.