Proceedings of the Language Technologies for All (LT4All) , pages 56–60
Paris, UNESCO Headquarters, 5-6 December, 2019.

A Real Time Instruction Extractor from Traffic Signal for Translation

Professor Dr. Mohammad Nurul Huda, Sabbir Arif Siddique
United International University, eGeneration Limited
mnh@cse.uiu.ac.bd, sabb.a.sidd@gmail.com

Abstract

This paper has developed and demonstrated a system to build traffic instruction detection and translation tools that can extract and convert Bangla text from natural images containing traffic instruction. In the process of developing the system, we have applied various techniques to extract and convert information from natural images. These techniques involve Image Processing, Machine Learning, Optical Character Recognition and Machine Translation. The proposed system consists of three steps, which are Text extraction from image, Post Processing by Language Model and Machine Translation.

Keywords: Optical Character Recognition, Image Processing, Machine Translation, Language Model

Résumé

এই গবেষণাটি দ্বারা আমরা একটি মেশিন লারনিং সমৃদ্ধ ট্রাফিক সিগন্যাল অনুবাদক টুল তৈরি করেছি। এই টুলটি বাংলা অপ্টিকাল ক্যারেক্টার রেকগনিশন ও মেশিন ট্রান্সলেশন এর সম্মিলনে তৈরি হয়েছে। এই টুলটি ব্যাবহার করে বিদেশী ব্যাক্তি বাংলা ট্রাফিক সিগন্যাল ইংরেজি ভাষায় অনুবাদ করে পরতে ও বুঝতে পারবে।

1. Paper Submission

The problem of understanding traffic signs in Bangla has been identified as a major problem for the foreigners. As these traffic signs contain both images of visual traffic signal along with Bangla text, it is nearly impossible to acknowledge the signs for a foreign citizen. Figure 1 is an illustration of a few existing traffic signs found on the roads of Dhaka.

Figure 1: Traffic Signs in Bangladesh.

Moreover, placement of traffic signs does not follow any international standard. Therefore, it may be rather difficult for non-local residents to find the signs without much effort. In our paper, we have proposed a state-of-the-art solution to address the mentioned problems. In this study we have used image processing mechanism and machine translation in this purpose.

The main goal of our image processing part of this research is to analyze a captured image, find, and segment the Bangla letters from there. In addition, we have also incorporated an efficient machine translator to translate the extracted Bangla text into English and other major languages.

The paper [1] proposed a novel system for the automatic detection and recognition of text in traffic signs. The authors have proposed a system in their work, which is capable of defining search area within the image. The paper [2] has recommended a system that can detect and recognize instruction from traffic signals. The authors of this paper have proposed a system to integrate in the Advanced Driver Assistance System (ADAS). We have recognized and implemented techniques expressed and illustrated in this paper. Moreover, we have incorporated additional techniques to improve the outcome of the Bangla OCR. These techniques include Edge Detection using Canny method [3], Gaussian filter [4], Edge Tracking by Hysteresis [5], B/W labeling, Character Segmentation [6], Character Recognition through Back Propagation Neural Network [7] to process the text extracted from the image and Example Based Machine Translation [8] algorithm.

Our proposed method for Bangla detection and translation from traffic signs is comprised of three stages. The first stage detects the traffic signs from natural images. In consequence, the second stage extracts Bangla text from the natural image. In the final stage, the text is translated into English. This paper represents the first endeavor in developing a traffic sign detection and translation system for Bangla language. Although Google and Bing have similar products, they however do not have support for Bangla yet.

2. Previous Study

There are many works for Bangla OCR from documents like Bangla OCR by UIU and first commercial OCR “Puthi OCR” by Team Engine. Most prominently there are two notable thesis work for Bangla OCR from image. The first one is from Khulna University by Zahid et.al and other one is from Computer Vision & Pattern Recognition Unit,

Indian Stat. Inst., Kolkata, India. In this research, we have incorporated techniques analyzed from the above-mentioned sources and combined them into a single system application.

3. Proposed System

The proposed system processes the captured images and converts them into English instructions. Distinct modules of the system execute in sequence to acquire the targeted goal from the input. Each of these modules employs diversified tools and contemporary algorithms. These modules are explained with demonstration and relevant diagrams in the following section. The proposed system is illustrated in the system diagram in Figure 2.

3.1 Image Processing

Captured image that contains Bangla traffic instruction is processed through a sequence of techniques, which are clarified by demonstration in the following sub sections.

3.1.1 Pre-Processing

After the natural image is captured, the preprocessing mechanism is conducted on the image. The input and output of the process is illustrated in Figure 3.

Preprocessing resizes and adjusts the RGB value of the captured images. The outcome of this stage is the B/W image with the corrected proportion.

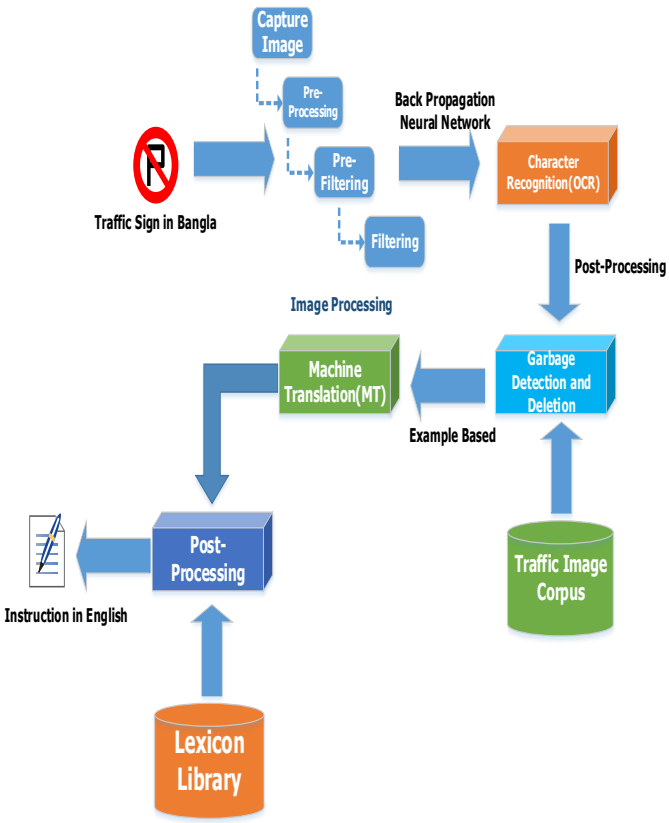

Figure 2: Proposed System Diagram.

Figure 3: Capturing and Preprocessing of Natural Image.

3.1.2 Pre-Filtering

On completion of the preprocessing pre-filtering is applied

on the processed image. The pre-filtering of the image is conducted by employing Edge Detection by Canny Edge Detection Method. The Canny Method is less likely than other methods to be fooled by noise. The general criteria for edge detection include the following steps.

I. Detection of edge with low error rate, which means that the detection should accurately catch as many edges shown in the image as possible.

II. The edge point detected from the operator should accurately localize on the center of the edge.

III. A given edge in the image should only be marked once, and where possible, image noise should not create false edges.

After edge detection process is conducted, Gaussian filter [4] is applied on the output to further fine-tune the detected edges. The equation for a Gaussian filter kernel with the size of (2k+1) * (2k+1) is shown as following:

$$H_{ij} = \frac{1}{2\pi\sigma^2} * \exp\left(-\frac{(i-k-1)^2 + (j-k-1)^2}{2\sigma^2}\right)$$

Here is an example of a 5x5 Gaussian filter, used to create the image to the right, with = 1.4. Here the asterisk denotes a convolution operation.

$$\mathbf{B} = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} * \mathbf{A}.$$

After applying the filter, the intensity gradient of the image is established. The edge detection operator (Roberts, Prewitt, and Sobel for example) returns a value for the first derivative in the horizontal direction (Gx) and the vertical direction (Gy). From this, the edge gradient and direction can be determined by the following equations.

$$\mathbf{G} = \sqrt{\mathbf{G}_x{}^2 + \mathbf{G}_y{}^2}$$

$$\mathbf{\Theta} = \mathrm{atan2}(\mathbf{G}_y, \mathbf{G}_x)$$

In consequence, edge-thinning technique termed Non-maximum suppression is enforced on the produced output. After application of non-maximum compression, the edge pixels are quite accurate to present the real edge. However, there are still some edge pixels at this point caused by noise and color variation. In order to get rid of the spurious

responses from these bothering factors, it is essential to filter out the edge pixel with the weak gradient value and preserve the edge with the high gradient value. Thus, two threshold values are set to clarify the different types of edge pixels, one is called high threshold value and the other is called the low threshold value. On resolving the double threshold value, edge tracking is conducted by Hysteresis. Afterwards structural elements of the image are extracted and then dilated. On the completion of the above-mentioned processes, the cropped images are acquired. The cropped elements along with some garbage is illustrated in

Figure 4.

Figure 4: Cropped Elements after Pre-Filtering.

3.1.3 Filtering

Filtering techniques is further applied on the pre-filtered output. These techniques include range estimation of the pre-filtered output. In the mentioned process, the garbage elements are removed and actual Bangla texts from the image is revealed. The flow diagram of the filtering process is illustrated Figure 5.

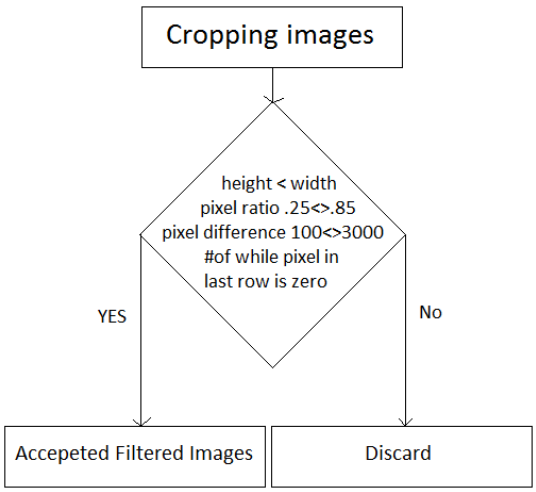

Figure 5: Flow Diagram of Filtering.

The output of this stage is illustrated in Figure 6.

Figure 6: Filtered Image.

3.1.4 Character Segmentation

The character segmentation process segments the characters in two categories. The first category is Characters without KAR (Bangla −□□□).

Figure 7: Characters without KAR.

The second category is characters with KAR.

Figure 8: Characters with KAR.

Therefore, the final output will be like the illustration in Figure 9.

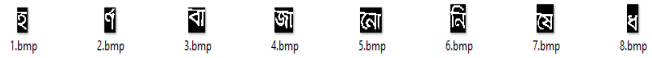

Figure 9: Final Output of Character Segmentation.

As we train our NN with black letters which have white background so after segmenting those letters we simply reverse the black pixel with white pixels of each letters and the output is given in Figure 10.

Figure 10: Output after Pixel Reversing.

In addition, for better image processing we reshape the image into a constant height and width. We use 45 × 45 (= 2025) constant shapes for each letter. This output is illustrated in Figure 11.

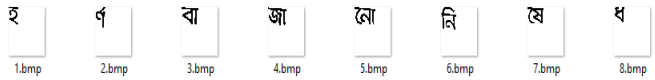

Figure 11: Characters after Reshaping.

The flow diagram of the character segmentation is illustrated below in Figure 12.

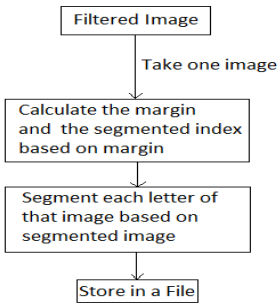

Figure 12: Flow Diagram of Character Segmentation.

3.2 Character Recognition and Post-Processing

After segmentation process, the output must be converted into machine-readable text. Neural network is employed to generate that conversion. However, the output of the neural net may contain a few garbage, which must be eliminated

to extract clean text. The processes in Figure 12 are detailed in the following sub-sections.

3.2.1 Character Recognition Using BP ANN

Backpropagation Artificial Neural Network (BP ANN) is employed in the proposed system to convert the segmented characters into electronic text. The text is retrieved in Unicode font. Backpropagation (BP) artificial neural network is the most commonly used algorithm in OCR, as it is highly effective in the given context. A typical BP ANN is illustrated in Figure 13.

BP ANN employs the following technique to extract the electronic character from the character segmentation output, which is depicted in Figure 14.

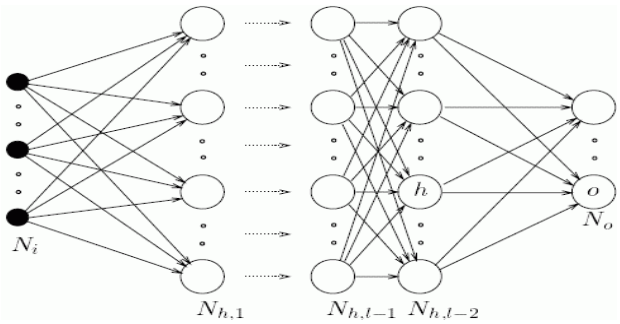

Figure 13: Back Propagation Artificial Neural Network.

3.2.2 Garbage Detection and Deletion

After character segmentation, post processing is conducted. Post processing is primarily consisted of garbage detection and deletion. To detect garbage from multi characters we will perform a partial string matching. Partial string matching is an approach to identify garbage value and useful to predict words from a partially correct word. Therefore, here is our algorithm.

- Split the result string.
- Iterate through all words.
- if(word.length > 1)
  Perform partial matching for each of the Bangla words in dictionary. Find the best matched Bangla words and return.

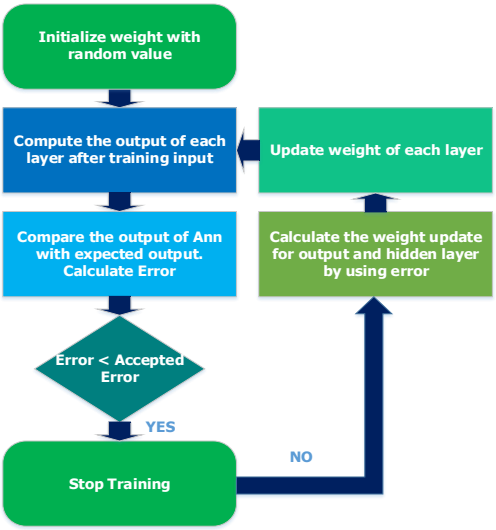

Figure 14: Flow Diagram of Character Recognition

Afterwards Levenshtein's Distance is employed to acquire best matching strings from the string dictionary. In figure 15 there is an illustration of an extracted charaqcter string with garbage values.

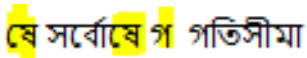

Figure 15: Sample Output with Garbage.

Now the first ষে in Figure 15 will be removed as that will not be partially matched with any word. গ in Figure 15 will also be removed, as it is a single character.

Instead of detecting সর্বোচ্চ our BP ANN returns সর্বোষে but it's partially matched so that will be replaced with correct one! Partially matched with সর্বোচ্চ because it will need three moves to transform one to another, which is minimum among other words in dictionary, and similarity between two words is 62%. Hence after post processing of the sample output we acquire the clean and authentic string as illustrated in Figure 16.

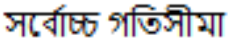

Figure 16: Output after Post Processing.

3.3 Machine Translation [8]

Now we have successfully extracted the authentic and clean text from the natural image. The next step is to convert the Bangla text into English. Machine Translation is a process of translating one word/sentence to another language's corresponding word/sentence. Machine translation is a complex problem because there are thousands of things that are needed to be considered. In basic level, we can just replace the words in a sentence with corresponding word in target language. That is not able to produce a good translation as the sentence structures are different and the recognition of whole phrases with their closest counterparts in the target language is needed. The approach [8] that we have taken in this paper is illustrated in Figure 17.

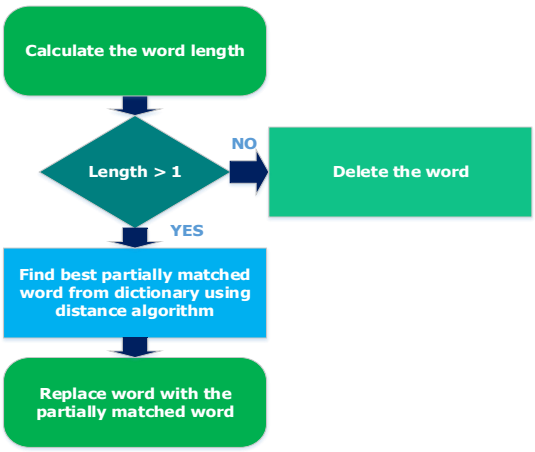

Figure 17: Flow Diagram of Garbage Deletion Process.

This approach can successfully translate most of the common traffic instructions. However, the English meanings that are constructed using multiple Bangla words is not considered here. The process of our approach of machine translation is illustrated in the Figure 18. The translated output of the extracted Bangla text is shown in Figure 19.

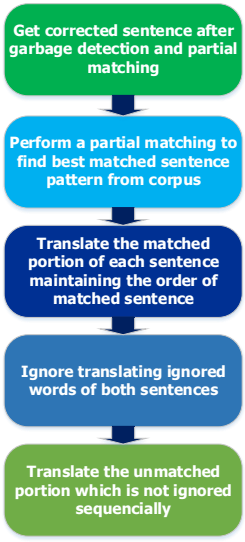

Figure 18: Flow Diagram of Machine Translation.

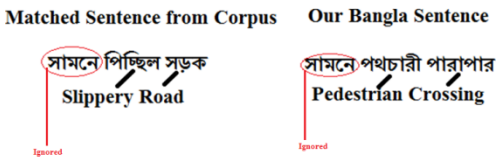

Figure 19: Machine Translation of Extracted Bangla Text.

4. Experimental Result and Analysis

The image of traffic instructions is very rare in internet. In fact, the traffic instructions are hard to find. Therefore, we really did not manage to get a plenty number of images for training and testing. That is one of the biggest difficulties that we have faced. Therefore, we have to test our system for limited training and testing data.Training Image Corpus:22 and test corpus:6. Backpropagation Artificial Neural Network is used where input features: 45×45 and output size : 18. Number of words in Traffic Instruction Database: 45. A demo corpus of pattern matching is shown in Table 1 and experimental result is shown in Table 2.

Table 1: Demo Corpus for Pattern Matching.

Bangla Sentence	English Sentence
পার্কিং নিষেধ	No parking
পথচারী চলাচল নিষেধ	No Pedestrian
সামনে টি-জাংশন আছে	T-junction
সামনে পথচারী পারাপার	Pedestrian Crossing

5. Conclusion and Future Works

In this research work, state of the art algorithms to translate Bangla Traffic sign into English for Foreigners were implemented. Because Canny edge detection method is applied in the pre-filtering process to detect edges from the captured image, it will be less prone to get deceived by noise. Consequently, the system is able to analyze signs manipulated with rain, leaves and dirt and produce output that is quite accurate.

In the process of conducting the research work, we have identified a number of constraints and area of improvements. The most notable of them are listed as following.

- Limited Size of the training corpus
- Limitation of OCR for angled photos
- Image adjustment is not dynamic
- Overfitting of data from the Neural Network
- Machine translation needs optimization

Moreover, the authors would like to incorporate driver movement detection techniques through accelerometer, gyroscope and compass sensor data to align and compare that with the instruction from the traffic sign. Therefore, if the driver's movement data is conceived as illegitimate according to the traffic signs, the system will generate a warning sound.

Table 2: Demo Experimental Result.

Input Sentence Bangla	Output Sentence English
সামনে সরু সেতু আছে	Narrow Bridge
সামনে স্কুল	School
হাসপাতাল	Hospital
সামনে ওয়াই-জাংশন আছে	Y-Junction
থামানো নিষেধ	No Stopping
হর্ণ বাজানো নিষেধ	No Horn Honking
সর্বোচ্চ গতিসীমা	Highest Speedlimit
বিপদজনক খাঁদ	Dangerous Dip
বনভোজন এলাকা	Picnic Site

References

[1] Jack Greenhalgh, Majid Mirmehdi, "Recognizing Text-Based Traffic Signs," IEEE Transactions on Intelligent Transportation Systems, Volume 16 Issue 3, June 2015

[2] Swati M, K.V. Suresh, "Automatic traffic sign detection and recognition - A Review," 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)

[3] J. Canny, "A Computational Approach to Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986.

[4] K. Ito, "Gaussian filter for nonlinear filtering problems," Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187), Sydney, NSW, 2000, pp. 1218-1223 vol.2.

[5] Yan Zhang and Lee Makowski, "Auto-thresholding Edge Detector for bio-image processing," 2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC), Troy, NY, 2015, pp. 1-2.

[6] A. Rehman, "Offline touched cursive script segmentation based on pixel intensity analysis: Character segmentation based on pixel intensity analysis," 2017 Twelfth International Conference on Digital Information Management (ICDIM), Fukuoka, 2017, pp. 324-327.

[7] Y. Li, Y. Fu, H. Li and S. Zhang, "The Improved Training Algorithm of Back Propagation Neural Network with Self-adaptive Learning Rate," 2009 International Conference on Computational Intelligence and Natural Computing, Wuhan, 2009, pp. 73-76.

[8] Linsen Yu, Yongmei Liu and Tianwen Zhang, "Using Example-Based Machine Translation Method For Automatic Image Annotation," 2006 6th World Congress on Intelligent Control and Automation, Dalian, 2006, pp. 9809-9812.