Signboard Text Translator: A Guide to Tourist

Ilaiah Kavati, G Kiran Kumar, Sarika Kesagani and K Srinivasa Rao
Dept. of CSE, MLR Institute of Technology, Hyderabad-43, India

Article Info

Corresponding Author:
Ilaiah Kavati
MLR Institute of Technology
Dundigal, Hyderabad - 43
+91 - 9848916272
kavati089@gmail.com

ABSTRACT

The travelers face troubles in understanding the signboards which are written in local language. The travelers can rely on smart phone for traveling purposes. Smart phones become most popular in recent years in terms of market value and the number of useful applications to the users. This work intends to build up a web application that can recognize the English content present on signboard pictures captured using a smart phone, translate the content from English to Telugu, and display the translated Telugu text back onto the screen of the phone. Experiments have been conducted on various signboard pictures and the outcomes demonstrate the viability of the proposed approach.

1. INTRODUCTION

We live in a society where we speak with individuals and data frameworks through differing media. Large volumes of data is represented in natural scenes. Signs are all around in our lives. A sign is an article that proposes the nearness of an actuality. It can be a shown structure bearing letters or symbols, used to recognize something or publicize a business. It can also be a posted notification bearing a warning, safety advisory, or command, etc. Signs are great examples in regular habitats that have high data content. They make our lives simpler when we can read and follow them, but they posture issues or much threat when we are most certainly not. For instance, a traveler won't have the capacity to understand a sign in local language that indicates notices or risks [1]. In this work, we concentrate on recognizing text on signs. Sign board text translator framework uses a smart phone to capture the signboard image containing signs, recognizes it, and translates it into user specified language.

Automatic detection and recognition of content from common scenes are requirements for a signboard text translator. The main challenge lies in the assortment of text: it can vary in text style, size, font, and orientation. There may also blur in the text and can be blocked by other objects in the signboard. As signs exist in three-dimensional space, content on signs can be misshaped by inclination, tilt, and shape of objects on which they are found. A numerous OCR frameworks function very well on high quality images; however, they perform poor for signboard images because of low quality nature of them. The proposed approach uses a Tesseract OCR engine to extract the text from the acquired image [2]. We have successfully applied the proposed approach to a English-Telugu sign text translation framework, which can recognize English signs caught from a camera, and translate the recognized text into Telugu. To our knowledge, English to Telugu signboard translator has not been investigated previously.

Rest of the paper is organized as follows: The present advancements on text translators will be explored in Section II. Section III explores the proposed approach. The outcomes are discussed in Section IV. Finally, Section V concludes our work.

2. RELATED WORK

This section explores about current signboard text translator applications developed for different languages using Optical Character Recognition (OCR). The work is accomplished for English to Chinese, Japanese to English,

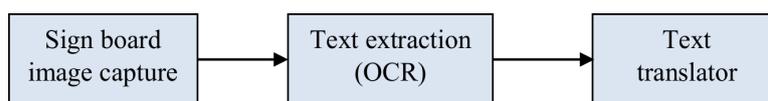| Sign board image capture	→	Text extraction (OCR)	→	Text translator

Figure 1. Block Diagram of the proposed approach

Chinese to English, Hindi to Tamil, English to Spanish, Malay to English/Arabic and Hindi to English on various versatile stages. Following are some articles exist in the literature for text translation in different languages.

Ma et al. built up an Android-based text translation application that can perceive the content caught by a cellular telephone camera, translate the content, and show the interpretation result back onto the screen of the cell telephone. This application empowers the clients to get text translation as simplicity as a button click. The camera catches the content and returns the interpreted result continuously. This framework contains automatic text detection, optical character recognition, text correction, and text translation [3].

Watanabe et al. report an application which interprets Japanese writings in a scene into English. The application is intended to keep running on a cell phone with a camera. It recognizes Japanese characters detected by an installed camera of a cellular phone and translate them into English [4]. In another approach, Chen et al. presents an approach to deal with automatic recognition of signs from natural scenes, and its application to a signboard translator. The proposed approach uses multiresolution and multiscale edge detection, color analysis in a hierarchy for sign identification. The strategy can essentially enhance text detection rate and OCR exactness. Rather than utilizing binary data for OCR, they retrieve features from the picture straightforwardly. They use a local intensity normalization technique to adequately handle lighting variations, trailed by a Gabor transform to get local features and lastly a Linear Discriminate Analysis (LDA) strategy for feature selection. They applied this methodology in developing a Chinese sign translation framework, which can consequently recognize Chinese signs taken from a camera, and translate the text into English [5].

In another approach, Authors proposed a design for English to Spanish translation of the signs present on JPEG pictures taken with a cellular phone camera. They distinguish the text using the frequency data of the DCT coefficients, binarize it utilizing a clustering algorithm, recognize it by the utilization of an OCR algorithm and phonetically translate it from English to Spanish. The outcome is a helpful, basic, moderate and robust framework that can deal with issues like lighting variations, focusing, low resolution and etc, in a short time [1]. Muthalib et al. explores a framework for text translation. It uses mobile technology and the framework is examined in which it included 30 global undergraduates of Malaysia. Quickly this study found that clients accept the Mobile-Translator well. The mobile translator particularly for use in Malaysia, in which the local language is Malay that can be translated into either English or Arabic [6].

Mishra and Patvardhan presents a greatly on-interest, quick and easy to understand Android Application ATMA. ATMA stands for Android Travel Mate Application. This application is helpful for local Tourists and Travelers who have Android Smart telephones. It empowers Travelers and Tourists to effectively catch the local nation language Books pages, signboards, flags and lodging menus and so on. The implicit OCR changes over the text translator in the caught image into Unicode content arrangement. It likewise gives interpretation office with the goal that Tourists can translate Native Language Unicode content into their own particular nation language. This Application has advanced features such as copy paste, share and search for travel related inquiries like museums, spots, hotels, books, restaurants and so forth [7].

3. PROPOSED METHOD

The proposed system capture the signboard images through a Camera phone. The captured image is then submitted to a central server. The image is preprocessed to extract the text and is reccognized. The recognized text is then converted to user specified language and delivered appropriately. The proposed approach is shown in Figure 1. The proposed approach follows these steps: (i) Sign board image acquisition, (ii) Extraction and recognition of text from the captured image, (iii) Translation of the extracted text into user specified language.

3.1. Sign board image acquisition

We use a mobile camera to capture the signboard image. We restrict the distance of the camera from the signboard should be a maximum of 10 meters to get good quality images. As the images are captured using a mobile camera, they may be generally noisy and contains complex backgrounds. Hence, the textual image should be pre-processed to make it clear so that the system can easily extract the text and recognizes it.
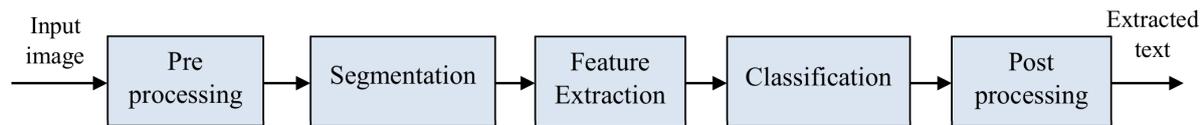
Input image → Pre processing → Segmentation → Feature Extraction → Classification → Post processing → Extracted text

Figure 2. Optical Character Recognition

3.2. Text Extraction and Recognition

Extraction and recognition of text from the captured image is called Optical Character Recognition (OCR). In this paper, we use a Tesseract OCR engine to extract the text from the acquired image [2]. The Tesseract is an open-source OCR engine that was developed at HP between 1984 and 1994. The Optical Character Recognition (OCR) process is divided into the following phases: Preprocessing [8], Segmentation, Feature Extraction and classification. Figure 2 shows the phases of OCR.

3.2.1. Preprocessing

The pre-processing involves: Image contrast enhancement [9], Noise removal, Binarization and smoothing. To enhance the contrast, the images are subjected to adaptive histogram equalization. The adaptive histogram equalization increases the contrast of the image and reduces the possible imperfections in the image. The enhanced images are further processed with median filter to suppress the noise content. The resulting image is subjected to locally adaptive threshold method to obtain a binary image. Finally smoothing which implies both filling and thinning. Filling eliminates small breaks, gaps and holes in the digitized characters. Thinning reduces the width of the line. The normalization is applied to obtain characters of uniform size, slant and rotation.

3.2.2. Segmentation

Segmentation is a process which helps to divide each character from a word present in a given image/page. The objective of the segmentation is to extract each character from the text present in the thinned image [10]. After performing segmentation, the characters of the string will be separated and it will be used for further processing. A horizontal projection profile technique [11] is used to segment the text area from the thinned image. This technique scans the imput image horizontally to find the first and last black pixels in a line. Once these pixels are found, the area in between these pixels represents the line that may contain one or more character.

3.2.3. Feature Extraction and Classification

In this stage, the features of the characters that are crucial for classifying them at recognition stage are extracted. The classification is the process of identifying each character and assigning to it the correct character class. The classification approaches are two types. 1) Decision-theoretic methods: The principal approaches to decision-theoretic recognition are minimum distance classifiers, statistical classifiers and neural networks. 2) Structural Methods: Measures of similarity based on relationships between structural components may be formulated by using grammatical concepts. Suppose that we have two different character classes which can be generated by the two grammars G1 and G2, respectively. Given an unknown character, we say that it is more similar to the first class if it may be generated by the grammar G1, but not by G2.

3.2.4. Postprocessing

This step performs grouping of symbols and errors handling. The process of performing the association of symbols into strings is commonly referred to as grouping. It is based on the symbols location in the document. Symbols that are found to be sufficiently close are grouped together.

3.3. Text Translator

This section explains the process of translating the extracted text from the signboard image into user specified language. To do this, we propose a translator module that maintain a dictionary of warnings or hazards in English (such as Poison, Do Not Drink, No Smoking, Danger, No Entry, Flammable Area, etc. which are generally appear on signboards) along with their meaning in user specified language i.e., Telugu. Whenever the system recognizes a text

<div align="center">(a)   (b)</div>
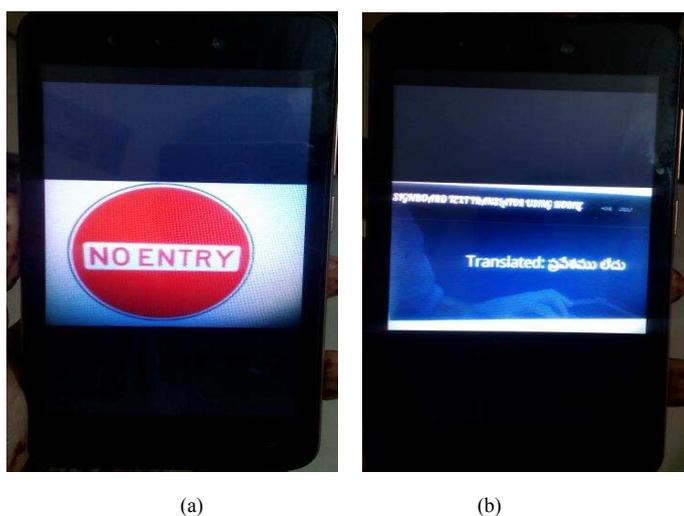
<div align="center">Figure 3. (a) Sample signboard image in English, (b) Translated text from English to Telugu</div>

in English language, it will verify the dictionary and the corresponding Telugu text is retrieved which is displayed over the cell phone display. This approach can be easily adopted to other launguages by feeding the required words and their meaning in the required language.

4. RESULTS

This section explain the results of our work. To measure the performance of the proposed system, we used two measures namely: Precision and Accuracy.

- **Precision:** Precision is also called character recognition rate. Precision is the number of characters correctly recognized to the total number of characters tested. This is given in Equation 1, where n_c is the number of correctly recognized characters and N_c is total number of characters.

$$Precision = \frac{n_c}{N_c} \tag{1}$$

- **Accuracy**: Accuracy is the number of signboard images correctly recognized to the total number of signboard images tested. This is given in Equation 2, where n_w is the number of correctly recognized characters and N_w is total number of characters.

$$Accuracy = \frac{n_w}{N_w} \tag{2}$$

We experimented our approach on 20 signboard images. Samples of some signboard images are shown in Figure 3, Figure 4 and Figure 5. The proposed system achieves a precision of 93.6 %. In other words, our system is able to extract the characters from the signboards and recognizes them effectively. Further, the accuracy of the system depends on the character recognition rate $i.e.$, Precision. As the system shows an acceptable level of character recognition, our system achieves enhanced accuracy. We experimented our approach on 20 signboard images and for 18 instances they are correctly translated into the specified Telugu language. In other words, the system achieves an accuracy of 90%. The sample signboard images taken and the obtained results are shown in Figure 3, Figure 4 and Figure 5. We also experimented this approach for the text with different fonts and the system recognizes/ translated them correctly.

5. CONCLUSION

This system is developed to make it easier for tourists from the states of Telangana and Andhrapradesh to translate the signboards written in English which could be diffficult for them to understand during their trips. We proposed a system to translate signboard images captured using a mobile phone from English to Telugu. The system is able to translate the text in different color, light, text on dark background, image with blurring, etc. The system

Figure 4. (a) Sample signboard image in English, (b) Translated text from English to Telugu
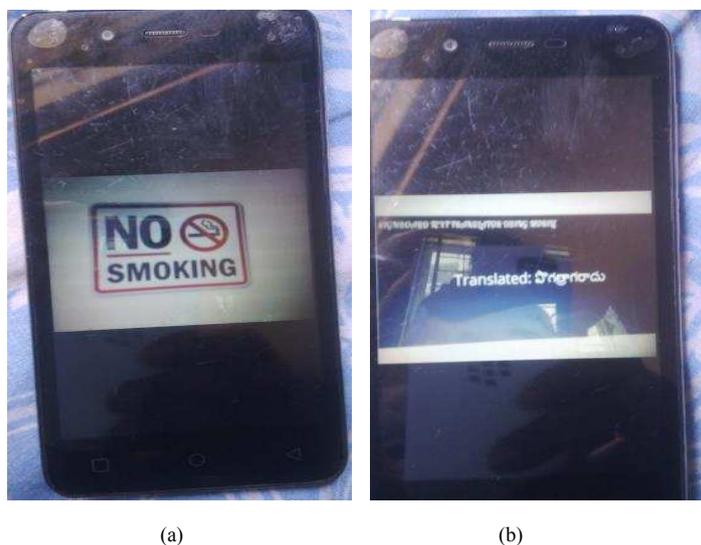

Figure 5. (a) Sample signboard image in English, (b) Translated text from English to Telugu

achieved a Precision $i.e.$, character recognition rate of 93.6% and translation accuracy of 90%. This shows the efficiency of the Tessaract OCR engine for text extraction and recognition. Our system shows some characteristics that make it interesting and deserve for further research. Future work involves: 1. automatic recognition and translation of handwritten English signboards into Telugu that involves language processors; 2. Usage of more accurate OCR for increased performance; 3. Developing a Mobile application.

REFERENCES

[1] A. Canedo-Rodríguez, S. Kim, J. H. Kim, and Y. Blanco-Fernández, "English to spanish translation of signboard images from mobile phone camera," in *IEEE Southeastcon 2009*. IEEE, 2009, pp. 356–361.

[2] R. Smith, "An overview of the tesseract ocr engine," in *Proc. Ninth Int. Conference on Document Analysis and Recognition (ICDAR)*, 2007, pp. 629–633.

[3] D. Ma, Q. Lin, and T. Zhang, "Mobile camera based text detection and translation," *Stanford University, Nov*, 2000.

[4] Y. Watanabe, K. Sono, K. Yokomizo, and Y. Okada, "Translation camera on mobile phone." in *icme*, vol. 3, 2003, pp. 177–180.

[5] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," *IEEE Transactions on image processing*, vol. 13, no. 1, pp. 87–99, 2004.

[6] A. A. Muthalib, A. Abdelsatar, M. Salameh, and J. Dalle, "Making learning ubiquitous with mobile translator using optical character recognition (ocr)," in *International Conference on Advanced Computer Science and Information System (ICACSIS)*.   IEEE, 2011, pp. 233–236.

[7] N. Mishra and C. Patvardhan, "Atma: Android travel mate application," *International Journal of Computer Applications*, vol. 50, no. 16, 2012.

[8] X. Tian, B. Yu, and J. Sun, "A preprocessing and analyzing method of images in pdf documents for mathematical expression retrieval," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 6, pp. 4579–4588, 2014.

[9] S. Sahu, "Comparative analysis of image enhancement techniques for ultrasound liver image," *International Journal of Electrical and Computer Engineering*, vol. 2, no. 6, p. 792, 2012.

[10] D. Aghlmandi and K. Faez, "Automatic segmentation of glottal space from video images based on mathematical morphology and the hough transform," *International Journal of Electrical and Computer Engineering*, vol. 2, no. 4, p. 463, 2012.

[11] H. Almohri, J. S. Gray, and H. Alnajjar, *A real-time DSP-based optical character recognition system for isolated arabic characters using the TI TMS320C6416T*.   University of Hartford, 2007.

BIOGRAPHIES OF AUTHORS

Dr Ilaiah Kavati is a Professor of Computer Science and Engineering Department at MLR Institute of Technology, Hyderabad, India. He obtained his Ph.D. in Computer Science from the University of Hyderabad (Central University), India in 2016. He recieved his B.Tech and M.Tech degrees from Jawaharlal Nehru Technological University, Hyderabad in 2004 and 2010 respectively. He is having 13 years of teaching and research experience. His current researches are in the fields of image processing, biometrics, security, internet of things and big data analytics. He is a life member of ISTE professsional society. He published more than twenty publications in international conferences and journals.

Dr G Kiran Kumar is a Professor of Computer Science and Engineering Department at MLR Institute of Technology, Hyderabad, India. He obtained his Ph.D. in Computer Science & Engineering from Nagarjuna University, Andhra Pradesh, India in 2015. He recieved his B.Tech and M.Tech degrees from Osmania University, Hyderabad. He is having more than 15 years of teaching and research experience. He is a life member of ISTE. His current researches are in the fields of Spatial Data Mining, image processing, big data and cloud computing.

Sarika Kesagani is an Assistant Professor of Computer Science and Engineering Department at MLR Institute of Technology, Hyderabad, India. She recieved her B.Tech and M.Tech degrees from Jawaharlal Nehru Technological University, Hyderabad in 2012 and 2016 respectively. Her current researches are in the fields of image processing and Internet of Things.

Dr K Srinivas Rao is the Professor of Computer Science and Engineering Department at MLR Institute of Technology, Hyderabad, India. He obtained his Ph.D. in Computer Science & Engineering from Anna University, Tamilnadu, India. He recieved his B.Tech and M.Tech degrees from Osmania University, Hyderabad. He is having more than 20 years of teaching and research experience. His current researches are in the fields of Data Mining, image processing and big data analytics.