International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

# Tourist's Translator based on Digital Image Processing and Hybrid Translation

**Rijwan Khan, Aryan Kaushal, Ayush Agarwal, Avdhesh Kumar**

***Abstract:*** ***Being a tourist in a foreign country is not easy when it comes to getting familiar with the local language. From reading signboards to getting unfairly charged while shopping, booking cabs and hotels for their stay, roaming around for sightseeing, communicating with the locals, everything requires an understanding of the local language. Nowadays, everybody has a smartphone, which proves to be the most helpful tool for travelling purposes. We aim to build up an Android application that will be capable of translating text, voice, and also textual information written on signboards from native language to the desired language, users have to use their smartphone camera for signboard translation.***

***We are using a hybrid machine translation approach for language translation. Text recognition from the captured images of signboards is done with the help of digital image processing. The purpose of this work is to reduce the failure of any single machine translation approach and to reduce miscommunication between tourists and the local people.***

***Keywords: Digital Image Processing, Tesseract Optical Character Recognition, Hybrid Machine Translation.***

## I. INTRODUCTION

We are living in a community where the language is our greatest mediator that allows us to relate and understand each other through various media. According to BBC - UK report1, around 75% of the world's population does not speak in English. This generates a worldwide communication gap. This language barrier prevents a person from seeking necessary information, especially for travelers. In the natural scene image, a massive amount of data contains in the form of alphanumeric and signs. Signs are all around in our surroundings, which tell us about the nearness of an actuality. Signs are used in the safety advisory, no parking, direction purpose, and so on. Traveling in non-English speaking countries, one can feel like he or she is inside a bizarre bubble, cut off from the rest of society, by this inability to communicate. To avoid this barrier, an interpreter or translator addresses such issues of linguistic barriers very well
by converting spoken or sign language statements from one language to another that indicates notices or risks. The comprehension stage consists of listening to, interpreting, and recall the information present in the source or input language and then reproduces the output in the desired language. In this system, our main aim is to recognize text from signs and to implement hybrid machine translation approach for text and voice translation using Yandex Server [1, 2, 3, 4].

The study is organized as follows. The methodology has been presented in section 2. The proposed method is discussed in section 3. Section 4 gives experimentation results with the comparison of the existing method. Finally, section 5 summarizes the work performed.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
**Rijwan Khan***, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, Affiliated to AKTU Lucknow, India. E-mail: rijwankhan786@gmail.com
**Aryan Kaushal**, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, Affiliated to AKTU Lucknow, India. E-mail: 8427ark@gmail.com
**Ayush Agarwal**, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, Affiliated to AKTU Lucknow, India. E-mail: agarwal0710@gmail.com
**Avdhesh Kumar**, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, Affiliated to AKTU Lucknow, India. E-mail: adrajabesit@gmail.com



## II. METHODOLOGY

### A. Digital Image Processing

*1. Image Capturing Module*

In-Camera capture module, tourists are privileged to capture the only selected textual image from signboard or natural scene by adjusting the camera capture box by touching the edges of the capture box on the screen. Throughout the session, the camera is in autofocusing mode automatically. The image is passed to the Tesseract OCR engine module after capturing the desired image.

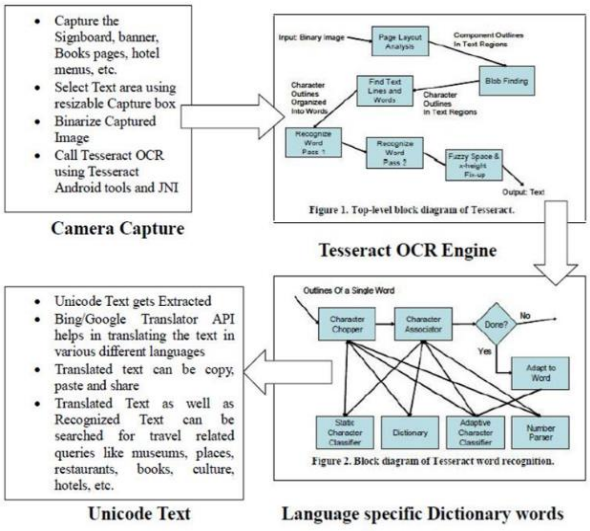


**Fig.1. Methodology Used**

*2. Tesseract OCR Engine Module*

The binarization method for the captured image is completed in this section and is evaluated after the text form.

ords and lines have been found after blobs detection. The term or string is passed to various phases. Every letter or character is extracted from the string in these phases, which analyzes the need to attach the extracted character or to disrupt related characters. At last, extracted characters are identified with embedded fuzzy features coordinated with a specific Unicode character language training data. Then after each character is joined to make a word in the same order as extracted and matched with the words of the particular language dictionary.

*3. Dictionary words Matching Module*

Each group of a word sequence is sent to Dictionary software installed on the Yandex server, where it will help in recognizing the term more precise and accurate than just providing the outcome of a meaningless word. Now, the identified text will be sent to the next module that is Unicode text post processing [5, 6].

*4. Unicode Text Postprocessing Module*

In the final module, Unicode characters will be displayed after the recognized text is passed to this module, and the user has the privilege to convert the identified text into a particular known language by selecting from the drop-down list under the settings [7, 8].

### B. Hybrid Translation Module

Hybrid machine translation is an approach in which more than one machine translation strategy is used. The main idea for developing a hybrid model is to reduce the failure of any single machine translation approach. In the proposed system, we are integrating our system with the YANDEX TRANSLATION server, which uses the CatBoost algorithm. The CatBoost algorithm is a collection of statistical machine translation and neural machine translation [9, 10].

- **Statistical Machine Translation**

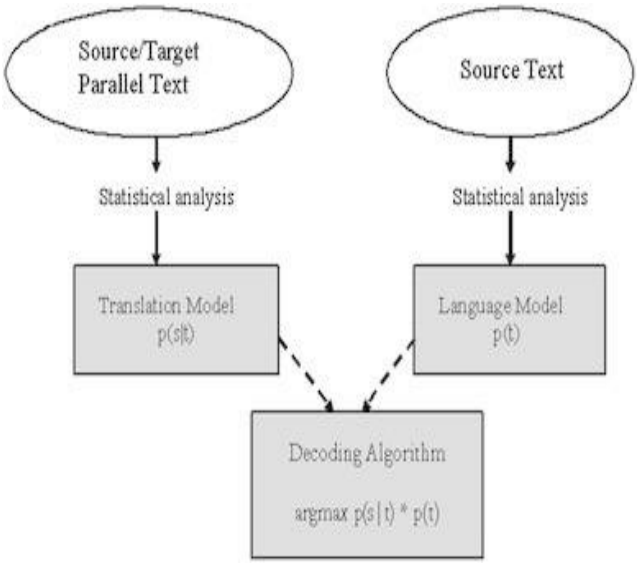


**Fig.2. Block diagram of Statistical Translation**

- **Neural Machine Translation**

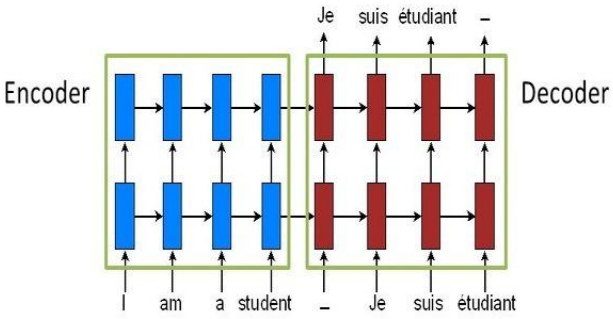


**Fig.3. Neural Translation**

## III. PROPOSED METHOD

The system captures images of the signboard via a smartphone camera. Captured image forwarded to the main server called the Yandex server. Then it is pre-processed and recognized after extraction of the text. After that, the identified text is translated into the target language and appropriately conveyed for user convenience. This process is reflected in Figure 1. The above process contains (i) Acquisition of signboard (ii) Text pulling and recognition, (iii) Text/Voice Translation.

### C. Acquisition of Signboard image

For capturing the signboard image, we use a handheld camera. To obtain good quality images, the gap between the mobile and the signboard should be up to 10 meters. The photos are taken with a smartphone camera; they are often noisy and contain complicated environments.

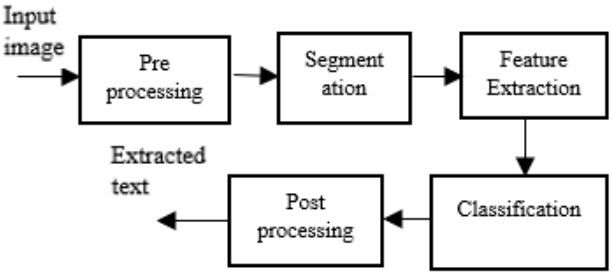


**Fig.4. Optical Character Recognition**

### D. Text Pulling and Recognition

Captured image text pulling and identification are termed as the Optical Character Recognition (OCR). We use a Tesseract OCR engine in this article to get the text from the picture you have acquired [11, 13]. The process of optical character recognition (OCR) is broken down into the various steps: Pre-processing, Segmentation, Feature Extraction, and classification, Postprocessing.

*1. Pre-processing of Image*

Pre-processing consists of the enhancement of image contrast [12], removal of noise, binarization, and smoothing. For better contrast, the images are equalized according to the histogram. Adaptive histogram equalization increases contrast of image and reduces image imperfections.

In order to remove noise content, improved images are additionally processed with a central filter. A local threshold process is used to produce a binary image as a result. Binarization transforms a gray or color image into a binary image, where the background is preferably represented by one color and the text by another. Smoothing finally that implies both filling and thinning. The standardization is applied to get the uniform size, slant, and rotation characters.

*2. Segmentation*

Image segmentation is a common technique used in the processing and analysis of digital images to divide images into multiple parts or regions, often based on the image pixels. Image segmentation could be based on color-or shape similarities between the foreground and the background or cluster regions.

*3. Feature Extraction and Classification*

The characteristics that are important to classify characters at the stage of recognition are extracted at this stage. Classification is the method of defining every character and assigning the appropriate class of character to it. The approaches to classification are in two types. 1) Decision-theoretic methods: Statistical methods, Artificial neural networks and Minimum distance classifiers are the principal approaches to decision-theoretical recognition. 2) Structural Methods: Similarity actions can be formulated using grammar concepts based on the relationship of the structural component. Assume we have two distinct character classes that are produced by S1 and S2 grammars. Given an unnamed character, we say that if S1 generates it, but not S2, it is more like the first class.

*4. Postprocessing of Image*

This stage performs symbol clustering and error handling. The way symbols are grouped into strings is commonly called grouping. It depends on where the symbol is found in the text. Symbols that consider themselves pretty similar are grouped.

### E. Text/Voice Translation

The Yandex translator (CatBoost algorithm) is used for text and voice translation, and we have used Google speech recognition to identify words or characters spoken by the user.

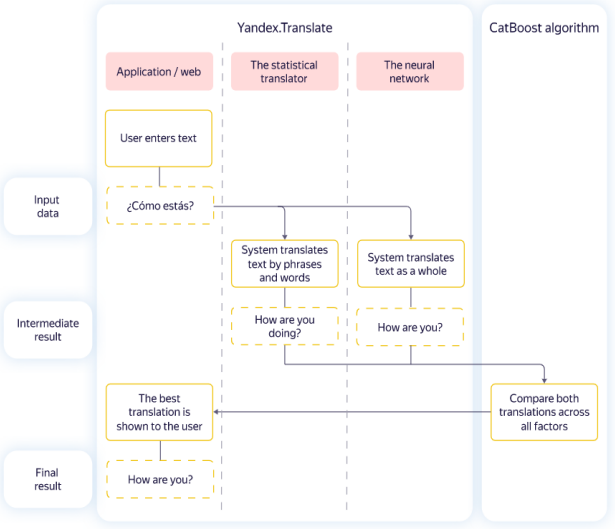


**Fig.5. Shows how this translator works**

You have to click on the text translator option next; you have to choose the desired language in which you want that text, you have to type text or speak from your microphone and click on the go. Take your data as an input by the app interface, and it will send it to the Yandex API Server where your data will be processed, and then it will fetch by the app interface, and then it will display the result in the selected language. This application supports about 50+ languages across the world and able to recognize nearly all possible standards symbols.

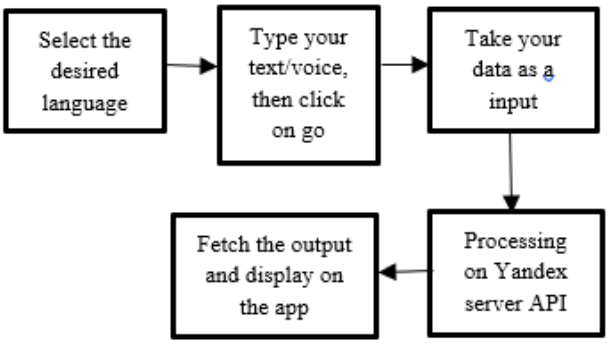


**Fig.6. Text/Voice Recognition**

## IV. EXPERIMENT RESULTS

This specific portion explains the results of the work we have done. Two criteria are used to calculate the performance of the proposed system, namely Precision and Accuracy.

- **Precision**

This can also be interpreted as identification levels for characters. Term precision is defined as the number of characters recognized correctly to the total number of characters tested where tc is the number of characters recognized correctly, and Tc tells the total number of characters

$$\text{Precision} = tc/Tc$$

- **Accuracy**

It is defined as the number of correctly recognized images (signboard) to the total number of test images (signboard). Here tw is the number of properly recognize characters, and Tw is the number of total characters.

$$\text{Accuracy} = tw/Tw$$

Up to 80 signboard images have experimented with our approach, and near about 72 instances are accurately converted to the language you want. That is, the machine reaches 90 percent accuracy. Figure 7 illustrates the image and the results achieved from the sample sign. We have experimented with this method with different fonts for the text, and the device recognizes and correctly translates them. Apart from that, we have checked the voice translation feature over more than 40 voice samples of different-different languages and found this application 94.4% accurate.

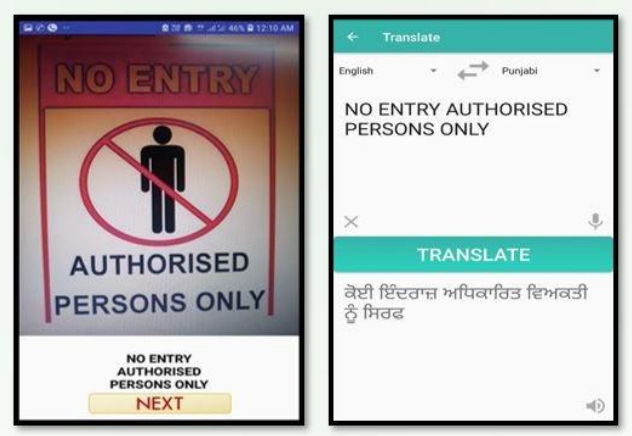


**Fig.7. (a) Captured image of Signboard. (b) Extracted text is in English and converted into Punjabi.**

## V. CONCLUSION

The proposed system successfully translated 50+ languages practiced and spoken across the globe, and results meet our expectations. We have used a hybrid machine (CatBoost algorithm) translation instead of single machine translation to make better use of technology and found that hybrid machine translation is better than single machine translation in terms of accuracy and efficiency. At last, we can say that this application will be more helpful to minimize miscommunication among people due to the language barrier.

## REFERENCES


1. N. Mishra and C. Patvardhan, "Atma: Android travel mate application," International Journal of Computer Applications, vol. 50, no. 16, 2012.
2. Azmi Can Özgen, Mandana Fasounaki, Hazim Kemal Ekenel, "Text detection in natural and computer-generated images," 2017.
3. Akhilesh A. Panchal, Shrugal Varde, M.S. Panse, "Character detection and recognition system for visually impaired people," IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016.
4. Pratik Madhukar Manwatkar, Dr. Kavita R. Singh, "A technical review on text recognition from images," IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO), 2015.
5. Milyaev, Sergey; Barinova, Olga; Novikova, Tatiana; Kohli, Pushmeet; Lempitsky, Victor (2013). "Image binarisation for end-to-end text understanding in natural images" (PDF). Document Analysis and Recognition (ICDAR) 2013. 12th International Conference on. Retrieved May 2, 2015.
6. Kumuda T, Basavaraj L. Detection and localization of text from natural scene images using texture features. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai: 2015. p. 1–4. CrossRef.
7. E. D. Haritaoglu and I. Haritaoglu, "Real-time image enhancement and segmentation for sign/text detection," Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 2003, pp. III-993.DOI: 10.1109/ICIP.2003.1247414
8. S. L. Fong, A. O. Elfaki, M. G. bin Md Johar and K. L. T. Aik, "Mobile language translator," 2011 Malaysian Conference in Software Engineering, Johor Bahru, 2011, pp. 495-500. DOI: 10.1109/MySEC.2011.6140723
9. 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.
10. A. Canedo-Rodriguez, S. Kim, J. H. Kim, and Y. Blanco-Fernandez, "English to Spanish translation of signboard images from a mobile phone camera," IEEE Southeastcon 2009, Atlanta, GA, 2009, pp. 356-361.DOI: 10.1109/SECON.2009.5174105
11. R. Smith, "An overview of the tesseract OCR engine," in Proc. Ninth Int. Conference on Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.
12. 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.
13. Simple Android Photo Capture. Available at: http://labs.makemachine.net /2010/03/simple-android photo-capture/


## AUTHORS PROFILE

**Dr. Rijwan Khan** is B.Tech (CSE), M.Tech (CSE) and Ph.D in Computer Engineering. He has total 13 years of teaching experience. Now working as a Head of Department CSE in ABES Institute of Technology. His are of research is software testing, soft computing, nature inspired algorithms. He published more than 25 research papers in different journals. He is reviewer of more than 6 different Scopus and SCI indexed journals like journals of Inderscience, Springer, and Elsevier etc. He is author of 3 books on C Programming, Data Structure using C, and Operating System and one book chapter in Springer book.

**Aryan Kaushal** is an undergraduate computer science & engineering student pursuing B. Tech at ABES Institute of Technology, Ghaziabad. His area of interest is AI, image processing and mobile application development. He has done training in Python and JavaScript.

**Ayush Agarwal** is an undergraduate computer science & engineering student pursuing B. Tech at ABES Institute of Technology, Ghaziabad. He has keen interest in Artificial Intelligence and Database Management System. He has done training in Database Management System and Python.

**Avdhesh Kumar** is an undergraduate computer science & engineering student pursuing B. Tech at ABES Institute of Technology, Ghaziabad. His Area of interest is android mobile application development using java. He has done some training in Java.