Journal of Science and Technology (JST)
Volume 2, Issue 4, July - August 2017, PP 33-41
www.jst.org.in *ISSN: 2456 - 5660*

Smart Multilingual Sign Boards

Pravin A. Dhulekar[1], Niharika Prajapati[2], Tejal A. Tribhuvan[3], Karishma S. Godse[4]

[1](Asst Professor, Department of Electronics and Telecommunication Engineering, Sandip Institute of Technology and Research Center, Sandip Foundation, Nashik, India)

[2,3,4](Student, Department of Electronics and Telecommunication Engineering, Sandip Institute of Technology and Research Center, Sandip Foundation, Nashik, India)

Abstract : *The Project is based on design & implementation of smart hybrid system for street sign boards recognition, text and speech conversions through character extraction and symbol matching. The default language use to pronounce signs on the street boards is English. Here we are proposing a novel method to convert identified character or symbol into multiple languages like Hindi, Marathi, Urdu, etc. This Project is helpful to all starting from the visually impaired, the tourists, the illiterates and all the people who travel. The system is accomplished with the speech pronunciation in different languages and to display on screen. This Project has a multidisciplinary approach as it belongs to the domains like computer vision, speech processing, & Google cloud platform. Computer vision is used for character and symbol extraction from sign boards. Speech processing is used for text to speech conversion. GCP is used for multiple language conversion of original extracted text. Further programming is done for real time pronunciation and displaying desired output.*

Keywords - *Street Sign Boards Recognition, Character Extraction, GCP, Symbol Matching, Computer Vision*

I. INTRODUCTION

Communication is a very important tool in a human life. It is an essential requirement in this world to survive. We can look back in the times of ignorance when no language was developed even than communication existed in form of sign language and other forms. To overcome the difficulties faced in communication; this paper has been implemented to achieve a real time system with feature/character/symbol extraction, text to speech conversion and then text conversion into different languages. It starts with capturing any image may be color, gray scale or monochrome via any video acquisition device and that image is stored in various file formats (such as bmp, jpeg, tiff, etc.). After it is stored, image is read by the software and feature extraction is performed. After character recognition is done, it is displayed and speech signal is generated which can be heard via speakers. GCP here is used as real time language translation platform. This platform help us to get the conversion of the extracted character in different languages , which are decided according to the users for e.g. English, Hindi, Marathi, Urdu, etc., as we use the Google cloud platform there is none of the boundary for the language which is displayed. This complete system is very helpful for visually impaired, illiterate, tourists and everyone who travels.

To make the cities of India a smart city, we must focus on their needs and the greatest opportunities to improve lives, increase security on roads, accident avoidance and safe driving.

II. LITERATURE REVIEW

A recent survey says that 285 million people are estimated to be visually impaired worldwide, 39 million are blind and 246 have low vision. Out of every 100, 82 are living with blindness are aged 50 and above 20% of all such impairment‟s still cannot be prevented or cured. In today‟s scenario, the overall literacy rate for people aged 15 and above is 86.3% which means there are still 13.7% people illiterate around the world. Starting from the most illiterate country like South Sudan, Afghanistan to the most literate countries like Norway, Cuba, etc., everywhere there is a requirement of local language for the illiterate or the rural people to communicate.

Over the past years, tourism has been growing rapidly strong and an underlying contributor to the economic recovery. Around 1.1 billion people traveled abroad in 2014 and Europe is the most visited area in the world. In 2015, 244 million people, or 3.3% of the world's population, lived outside their country‟s origin. It is predicted that immigration rates will continue to increase over time and every time they travel, have to face with different languages at different places.

Hence, for all such people across the globe a common solution, a speech generation system using language translation software is implemented which will be useful to everyone. In [3] presented a work on converting the multilingual text into speech for visually impaired. Basic working principle used here is u law companding. This system used full for the blind and mute people. In [4] presented a paper based on the nature

prosody generation in the text to voice for English using the phonetic integration .This system is very complicated to record and accumulate the word but is used the less memory space. In [5] worked on a system of converting the text in the form of speech. This paper includes the recognized text and converted proper sound signal. OCR techniques are implemented to get the output of the system. In [6] worked on the text to voice conversion for the application on android environment. The basic objective of this paper is extending the offline OCR technologies to the android platform. They proposed there system to detect the text which modified to the multi-resolution and converted into different languages using language translator. In [7] presented the basic idea of interfacing of speech to computer. The automatic speech recognition (ASR) played the key role of taking the speech technology to the people. The system is classified further as feature extraction & feature recognition. The aim was to study and to explore how neural network can be implemented to recognize a particular word speech as an alternative to the traditional methods.

In [8] presented the voice recognition system automatically by machine means an easy mode of communication between human & machine. The area of voice processing has implied to application like text to speech processing, speech to text processing, telephonic communication call routing etc. Here various features extraction, technique are discussed such as relative spectral processing (RASTA), linear predictive coding (LPC) analysis and many more. In [9] presented the idea of an innovative and real time low cost technique that makes user comfortable to hear the contents of the images or any kind of abstracted document instead of reading through them. It has used a combination of OCR "Optical Character Recognition and TTS text to speech synthesizer ".This system helps the visually impaired person to interact with the machine like the computer easily through the vocal interface.

In [10] proposed a system which is mainly designed for the blind user , which have interest in the field of education , this people study with the help of audio recording provided by NGO"s or the Braille Books. This will provide them to have an audio material of their own choice of any printed material of object. The whole system mainly includes OCR "Optical Character Recognition " which will perform/include the various operations like thresholding, filtering, segmentation and many more.

Figure 1: Street Sign Boards

III. PROPOSED TECHNIQUE

To make a hybrid system, the methodology used here is Optical Character Recognition and language translation. The images will be captured via web camera, then extracting text from image to convert that text into speech. After that, there is further process on extracted text of language translation using Google translation software on cloud. Then the whole system will be helpful to all starting from the visually impaired, the tourists, the illiterates and all the people who travel.

IV. ALGORITHM

Text reading system has two main parts: image to text conversion and text to voice conversion. Image into text and then that text into speech conversion is achieved using programming on MATLAB platform.

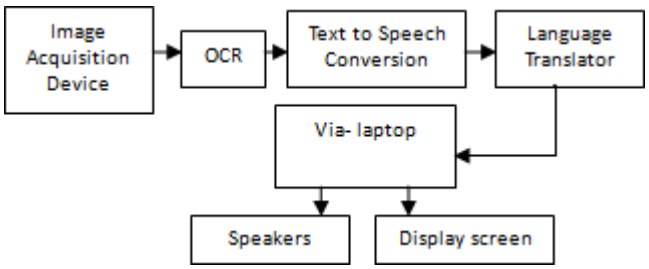

Figure 2: Flowchart

A. Work Flow

1) Image Acquisition Device: A computer vision system processes images acquired from an electronic camera, which is like the human vision system where the brain processes images derived from the eyes[8].The camera or web-cam is used as the acquisition device, which captures the image symbol or sign which will ultimately help us to get the sample image for the further process and to get the symbol for recognition. Also we get the video from this device through which we can extract the required area of the interest and can be processed further for various operations to get the final results.

a) Web Camera:

Figure 3: HP Laptop Web Camera

Technical Specification:

- Screen Size: 15.6 inches
- Screen Resolution: 1366 x 768
- Frame rate up to 30 frames per second

*2) Optical Character Recognition:*This block is used for character extraction and symbol matching which is included in algorithm and explained in detailed, this is the area where all the input images or videos captured by camera or webcam are processed and then the feature extraction is done by various techniques. Here we also perform the template matching ,if the character or the symbol is to be matched, then the operations are performed as character/symbol to text conversion[5].

Syntax:

```
txt = ocr(I)
txt = ocr(I, roi)
[_] = ocr(_,Name,Value)
```

Example:

```
businessCard = imread(„businessCard .png“);
ocrResults = ocr(businessCard)
recognisedText = ocrResults.Text;
figure;
imshow(businessCard);
text(600, 150, recognizedText, „BackgroundColor“, [1 1 1]);
```

Steps:

Step 1: Image acquisition: Different types of images like monochrome, gray scale and color, any image can be used as input to any video acquisition like web camera whose primary operation is to sense and capture.

Step 2: Preprocessing: Not mandatory but preprocessing is done when the input image is not clear and requires deblurring, denoising, resizing or reshaping.

Step 3: Detect MSER Regions: Maximally Stable Extremal Regions (MSER) is used as a method of blob detection in images. Mathematically,

Let $Q_1,\ldots,Q_{i-1},Q_i,\ldots$ be a series of nested extremal regions ($Q_i \subset Q_{i+1}$). Extremal region Q_{i*} is maximally stable if and only if $q(i)=|Q_{i+\Delta} \setminus Q_{i-\Delta}|/|Q_i|$ has a local minimum at i*. (Here | | denotes cardinality). $\Delta \in S$ is a parameter of the method.

The equation checks for regions that remain stable over a certain number of thresholds. If a region $Q_{i+\Delta}$ is not significantly larger than a region $Q_{i-\Delta}$, region Q_i is taken as a maximally stable region.

3) Text to Speech Conversion: Basically called as "speech synthesis". In this block the output obtained from OCR i.e. the matched image, sign or symbol is in the form of text which we have converted by symbol to text conversion technique here it is converted into speech by using text to voice converter technique, this speech is processed further for various languages ,which will help the user to get the information about the total area which is of the required interest

4)Language Translation: For this, we are using Google Cloud Platform which allows us to use its application program interface at a minimal price. We chose Google translation API to connect with the MATLAB by creating our own console on GCP and generating an API key that we have used to send data to and get data from the cloud itself.

5)Via-laptop: We have implement the complete system by using laptop. We use the in-build camera of the laptop for capturing the images, and by doing preprocessing on the captured image, we got the required results which are displayed on the screen and at the same time the result is spelled out by the speaker of laptop.

6)Speakers and Display Device: It is the laptop speakers and the screen which is used for getting the voice in the different languages like Hindi, English, Marathi, Urdu etc. which will help the user to get the complete info in his own language it will be able to pronounce the different language to make the user easy to get the basic info. This info will help the visually impaired person which is ultimately the goal of the system. The display is used for displaying the different types of language outputs, it gives better results and which is the basic need of the system.

B. Optical Character Recognition

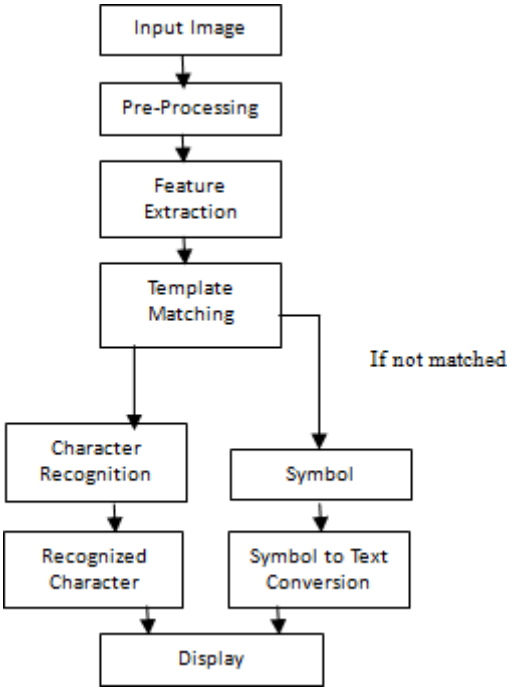

Figure 4: Algorithm of OCR

1) Input Image: Capturing the image may be on streets while travelling, in classroom while teaching, for information reading while browsing is all done by any video acquisition device such as a web camera. Then the image is read with the help of „imread" command. „imread" reads image from graphics file.

2) Pre-Processing: Pre-processing is not a mandatory step. It is used only when the image is corrupted. It consists of a number of steps to make the raw data usable for recognizer. It is mainly used for removal of noise, deblurring, resizing and reshaping of the image. MSER technique is also used for blob detection.

3) Feature Extraction: The objective of feature extraction is to capture the essential characteristics of the symbols. The most precise way of obtaining a character is by its raster image. Other way is to uproot certain features that still characterize symbols, but leaves out the unimportant attributes [5].

4) Template Matching/Symbol Matching: A simpler way of template matching uses a convolution mask (also called as template) structured to a related feature of the search image, which we want to detect. This technique can be easily performed on grey images or edge images. We will match the extracted images with the collected data or the data base. Here the information is already stored with the use of programming and hence, created the data base.

5) Character/Symbol Recognition: Character recognition deals with cropping of each line and letter, performing correlation and writing to a file. Before performing correlation we have to load the already stored model templates so that we can match the letters with the templates. After we got the character by character segmentation we store the character image in a structure. For pattern recognition, SURF (Speeded Up Robust Feature) technique is used to match scene image with the box image with a max ratio of 0.65 and a threshold of minimum 25 interest points.

6) Character/Symbol to Speech Conversion : First, it transforms text including symbols like numbers or abbreviations into its equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization[9].

7) Display : Display are used to display the text when we capture image by the camera or any type of the scanner processing it and convert that symbol and character into the text and display on the screen.

C. Pattern Recognition (Algorithm)

SURF feature extraction technique is used for pattern based recognition. In this template images are already stored in the database called as the „box image" and the image which is being captured through camera is known as the „scene image". First interest points are obtained from the scene image and then matching is done with the box image with a minimum threshold of 25 points. Also the max ratio is set to 0.65 based to get the most accurate results. Also the matched result is shown by the matching lines clearly in the figure placed below. The red and the green color points show the interest points. Algorithm of SURF technique is as follows:

1. Covert images to gray-scale.
2. Extract SURF interest points.
3. Obtain features.
4. Match features.
5. Get 50 strongest features.
6. Match the number of strongest features with each image.
7. Take the ratio of- number of features matched/ number of strongest features (which is 25)

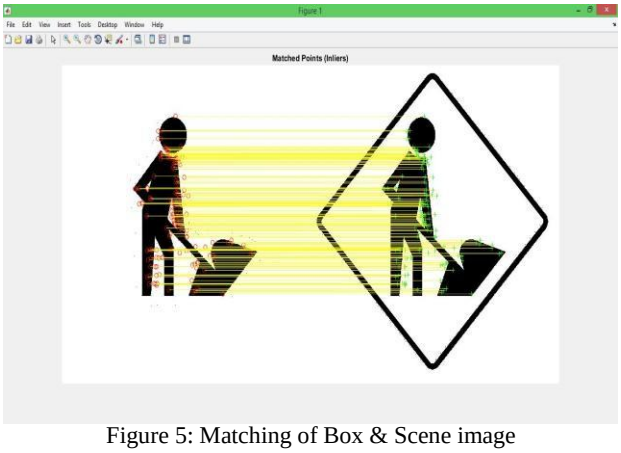

Figure 5: Matching of Box & Scene image

V. MATHEMATICAL ANALYSIS

1. Standard Data Set

This are the standard data base images which shows the100% accuracy and the following table shows the response time for the image to give the output, which is very less.

Table 1: Standard Based

Sr. No.	Name on Sign Board	Accuracy	Response Time
1.	Horns Prohibited	100%	0.217834 seconds
2.	No Entry	100%	0.220310 seconds
3.	No Parking	100%	0.198523 seconds
4.	Overtaking Prohibited	100%	0.234994 seconds

2.Pattern Based Data Set

The pattern based recognition shows the following Accuracy and the response time. Hence it shows the good accuracy of the pattern or symbol based sign boards on which the processing is done.

Table 2: Pattern Based

Sr. No.	Name on Sign Board	Accuracy	Response Time
1.	Men At Work	99%	0.298642 seconds
2.	School Ahead Go Slow	97%	0.383578 seconds
3.	No U Turn	96%	0.379566 seconds
4.	Turn Left	98%	0.286479 seconds

3.Real Time Data Set

This is the results of the real time data images which are captured by the user and shows 93% to 95% accuracy. Also the response time is mentioned.

Table 3: Real Time Based

Sr. No.	Name on Sign Board	Accuracy	Response Time
1.	U Turn Prohibited	93%	0.400865 seconds
2.	Straight Prohibited	95%	0.389267 seconds

3.	Do Not Honk Your Horn	91%	0.305811 seconds
4.	Overtaking Prohibited	95%	0.477447 seconds

VI. RESULTS AND DISSCUSSIONS

Algorithm is piloted and tested and hence the results got generated as shown below:

This is the GUI of the complete project run on MATLAB which includes standard sign board recognition based on both character and pattern recognition. Also real time sign board recognition is performed

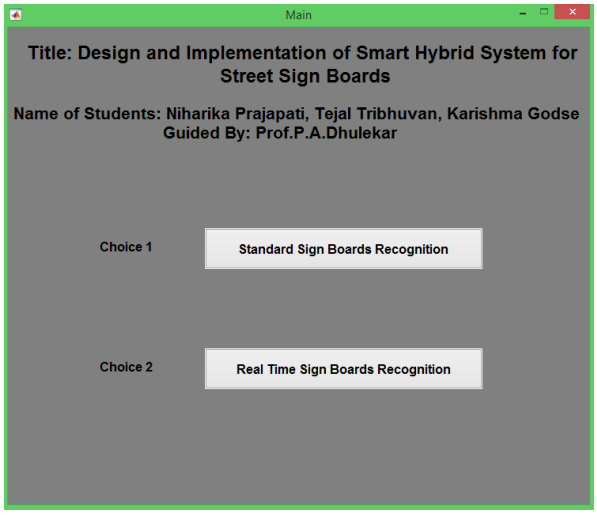

Figure 6: Result Image I

Case 1(a): Result shows that each and every character from the image is extracted successfully irrespective of their fonts using OCR & also voice is generated and displayed in three different languages..

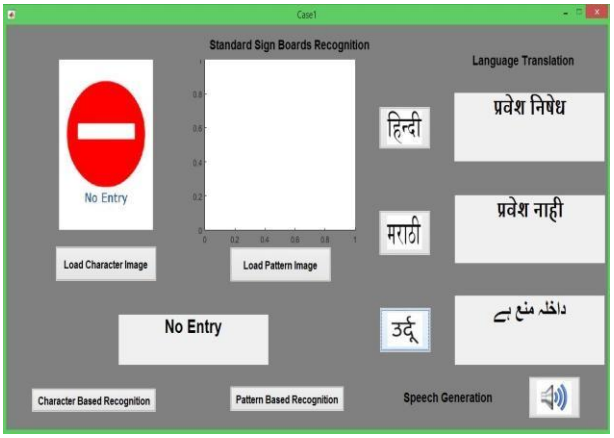

Figure 7: Result Image II Case I(a)

Case 1(b): Result shows that how pattern of the captured image is matched with the template image & also voice is generated and displayed in three different languages.

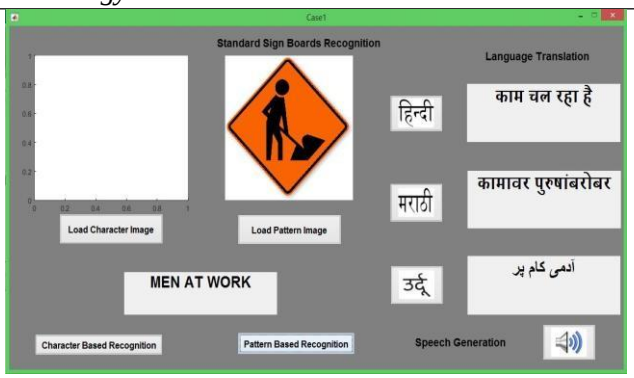

Figure 8: Result Image III (Case I(b)

Case 2: Result shows that each and every character from the image is extracted successfully in real time using OCR & also voice is generated in three different languages and displayed for the same on real time platform.

Figure 9: Result Image IV (Case II)

VII. CONCLUSION

This paper is an effort to implement an innovative robust approach for character extraction and text to voice conversion of different images using optical character recognition and text to speech synthesis technology. A user friendly, cost effective, reliable to all and applicable in the real time system is achieved. Using this methodology, we can read text from a document, street sign boards, web page or also an e-Book and can generate synthesized speech through any system i.e. computer's speakers or phone‟s speaker. In developed software, the use of computer vision has set all policies of the characters corresponding to each and every alphabet extraction, matching, its pronunciation, and the way it is used in grammar and dictionary. Speech processing has given robust output for various images. Other application of this system includes such as making information browsing for people who do not have the ability to read or write. This approach can be used in part as well. If requirement is only for text conversion then it is possible or else text to speech conversion is also done easily. People with vision impairment or visual dyslexia or complete blindness can use this approach for reading the documents, books and also while travelling. People with vocal impairment or complete dumb person can utilize this approach to turn typed words into vocalization. Tourists having language barrier can also use this approach for understanding different languages. People travelling in cars and buses can save their time using this feature. Experiments have been performed to test the text and speech generation system and good results have been achieved. The work is done very effectively for the symbol extraction. We got required results as provided in the above concept, the symbol extraction is done by using the SURF algorithm. Hence got the suitable outputs and can overcome the problem of such a sign boards which are only in symbol format. We got the outputs of the all the 3 methods as for the standard data base 100%, for the pattern base we acquired 96% to 98% and for the real time data set 91% to 95% accuracy with less response time . The developed system is now able to deal with any sign boards of the RTO which is the basic need of the society, In this way we also contribute to the concept of SMART CITY by contributing towards the SMARTTRANSPORTATION.

REFERENCES

[1] Preeti.Kale, S. T. Gandhe, G. M. Phade, Pravin A. Dhulekar, "Enhancement of old images and documents by Digital Image Processing Techniques" Presented in IEEE International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India. 16-17 January 2015.

[2] Harshada H. Chitte, Pravin A. Dhulekar, "Human Action Recognition for Video Surveillance" Presented in International Conference on engineering confluence, Equinox, Mumbai, Oct 2014.

[3] Suraj Mallik, Rajesh Mehra,"Speech To Text Conversion For Visually Impaired Person Using M Law Companding " Publish In Iosr Journal Of Electronics And Communication Engineering (Iosr-Jece) E-Issn: 2278-2834,P- Issn: 2278-8735.Volume 10, Issue 6, Ver. Ii (Nov - Dec .2015).

[4] Mrs. S. D. Suryawanshi, Mrs. R. R. Itkarkar, Mr. D. T. Mane ," High Quality Text To Speech Synthesizer Using Phonetic Integration "Publish In International Journal Of Advanced Research In Electronics And Communication Engineering (Ijarece) Volume 3, Issue 2, February 2014.

[5] Pooja Chandran, Aravind S, Jisha Gopinath And Saranya S S ," Design And Implementation Of Speech Generation System Using Matlab " Publish In International Journal Of Engineering And Innovative Technology (Ijeit) Volume 4, Issue 6, December 2014.

[6] Devika Sharma ,Ranju Kanwar ," Text To Speech Conversion With Language Translator Under Android Environment " Publish In International Journal Of Emerging Research In Management &Technology Issn: 2278-9359(Volume-4,Issue-6).

[7] Siddhant C. Joshi, Dr. A.N.Cheeran "Matlab Based Back-Propagation Neural Network For Automatic Speech Recognition", IJAREEIE(An Iso 3297: 2007 Certified Organization)Vol. 3, Issue 7, July 2014.

[8] Siddhant C. Joshi, Dr. A.N.Cheeran, "A Comparative Study of Feature Extraction Techniques for Speech Recognition System" ISSN: 2319-8753, IJIRSET ,Vol. 3, Issue 12, December 2014.

[9] K.Kalaivani, R.Praveena,V.Anjalipriya,R.Srimeena, "Real Time Implementation Of Image Recognition And Text To Speech Conversion",IJAERT,Volume 2 Issue 6, September 2014, Issn No.: 2348 – 8190.

[10] Kalyani Mangale, Hemangi Mhaske, Priyanka Wankhade, Vivek Niwane, "Printed Text To Audio Converter Using OCR", International Journal Of Computer Applications (0975 – 8887) (NCETACt-2015).

[11] Pravin A.Dhuleka,Niharika Prajapati, Tejal A Tribhuvan, Karishma S. Godse." Automatic Voice Generation System After Street Board Identification For Visually Impaired" ICSPICC-2016,Bambori,Jalgoan,

Maharashtra.22-24 December 2016.

[12] Mark Nixon, Alberto Aguado, Feature Extraction and Image Processing, 2nd Edition. Sunil Jadhav, P. A. Dhulekar, Dr.G.M.Phade, "Hybrid Optical Character Recognition Method for Recognition of Text in Images", IEEE FIFTH International Conference on Computing of Power ,Energy & Communication" ICCPEIC -2016, April 20th and 21th – 2016