App on online Text to Speech of Name board in Public Transport

[1]Jagadeeshwaran.R, [2]Hemalatha.R, [3]Javid Mohamed Z, [4]B.S Charulatha

[1]Student, [2]Student, [3]Student,[4]Professor
[1]Computer Science and Engineering,
[1]Rajalakshmi Engineering College, Chennai, India

Abstract: The concept of cosmopolitan community is fast developing in India. People from one region migrate to other for education, job, business and a variety of reasons. All official communication will be done using a common language. But for the day-to-day life the common language is not to support anymore. The passengers during the transit to other place, the language falls as a barrier. The identification of place displayed on the name board of public transport is gruelling to the passengers. In the public transport, usually the name of the places will be displayed in local language or occasionally in English. Technology comes in handy to the cosmopolitan society. The use of technology can be devised as an app, which will be accessed by the users. This paper concludes with the design of app that reads out the destined place when focused on name board of public transport. The proposed system is interfaced with image processing technology, OCR and Text-to-Speech. Thus, the role of a language translator can be reduced thereby people can commute in public transport without the aid of local people. The work is confined with Telugu Language and fixed letters associated with the destination name displayed on the name board.

Index Terms - **Image Processing, Optical character recognition, Text to Speech**

I. INTRODUCTION

In day to day life of cutting-edge technologies, goes forward and beyond thinking developers faces huge challenging aspects towards them. Acquisition of immediate information has become possible because of the advances taking place in the internet. In this way Image processing is one the emerging outcome of the world technological knowledge. Intrigue in image processing methods grows from two principal application areas mainly for the improvement of pictorial information for the human interpretation and processing of image data for various applications.

1. IMAGE PROCESSING:

Image processing technique algorithm takes an input, an image or a sequence of images and produces an output, which may be a modified image or a description of the input image contents. Image Processing take out information from images and amalgamate it for several applications. Image processing is the way of processing the images using some logical and mathematical operations by using any form of signal processing for which the input is an image or a series of images, such as pictures captured by a camera. In general, Image processing deals with the images that are in two-dimensional entities like satellite images, scanned images, etc. captured electronically. In depth image processing is mainly classified into two types, they are analog and digital image processing.

The analog image processing is the one which applied on analog type of signals and then it processes two-dimensional signals only. The images are operated by the electrical signals. In analog image processing, analog signals can be either periodic signals or non-periodic signals.
The Digital Image processing is the one which is applied to digital images. For operating the images, there are number of software and algorithms that are applied to perform manipulation. Digital image processing is the fastest growing industry which impacts everyone's life.

As stated in [3] in real world we define an image as a function of two real variables we say a (x, y) where a is defined to be the amplitude of the image and the real coordinate positions of the image. This concept produces the fact that images contain collections of objects each of which can be the basis for a region. In image processing system, it should be able to apply certain image processing operations for the regions that are selected.

Image processing has its applications in pattern recognition, text recognition, image sharpening and restoring, Robot/Machine vision, Transmission and encoding, Video processing, medical field, etc. Merits of image processing is cost-effective and faster. One needs less time for processing the images as well as less film and photographing equipment's.

It also removes noises, provide images with accurate density and contrast. The only limitation of Image processing is that it is impossible to process the images if the camera has very less resolution, battery consumption, memory card problems.

1.1 STEPS INVOLVED:

The steps involved with image processing as taken from [3] are as the following Image acquisition, Image Enhancement, Image Restoration, color Image Processing, Multiresolution Processing, Compression, Morphological Processing, Segmentation, Representation and Description, Feature Matching and Recognition. The figure 1.1 describes the block diagram of steps involved in Image processing methodology.
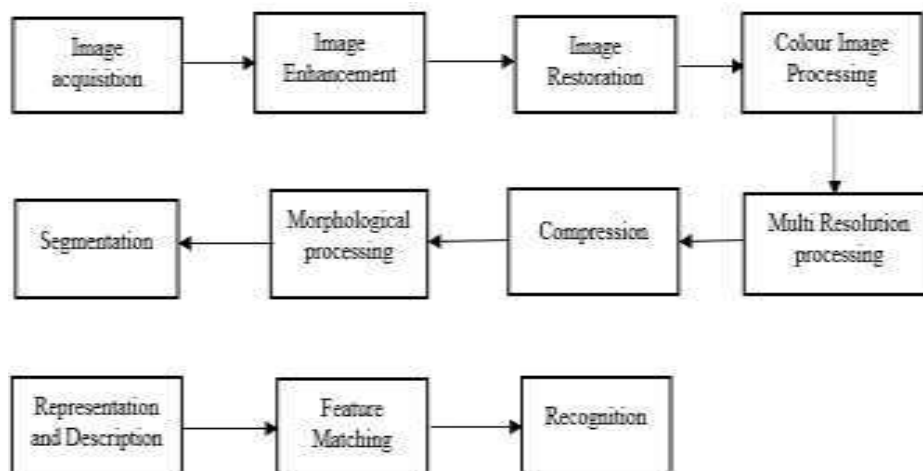
Fig 1.1 Steps involved in Image Processing

The first stage of image processing basically involves the capture of images and then involves pre-processing. Image Enhancement is done next where filtering takes place in order to improve the quality of images.

Image Restoration is the process of improving the appearance of the images by reducing the blurring of an image by mathematical or probabilistic models. Color Image Processing take place when there is the use of digital images. Multiresolution processing involves representation of various degree of resolution.

Compression is done next where the size of the images is reduced with minimum deterioration. Morphological processing is the process of extracting image components that are useful in the representing the images and description of the shapes. Segmentation is the way of partitioning the images into multiple segments. Representation involves boundary representation and then regional representation.

In boundary representation it focuses on the external shape characteristics such as corners and then inflections. In regional representation it mainly focuses on the internal properties such as textual and skeletal shape.

The recognition is the process of assigning labels to an object based on its description. It uses machine vision technologies along with artificial intelligence and trained algorithms to recognize the images through a camera.

2. OPTICAL CHARACTER RECOGNITION:

Optical Character Recognition (OCR) or Text Recognition is the technology used to distinguish the hand written or the printed characters inside the analog or digital images in the physical documents or the image captured in the desired format (.jpg, .jpeg, .png). OCR systems are shaped up with both hardware and software used to convert the content into desirable machine-readable text. Software takes the advantage of Artificial Intelligence to implement more advanced methods of intelligent character recognition in the way of identifying languages or style of handwriting.

2.1. PHASES OF OCR:

The text recognition phases as taken from [7] are as follows: pre-processing, character segmentation, feature extraction, classification, and post-processing.

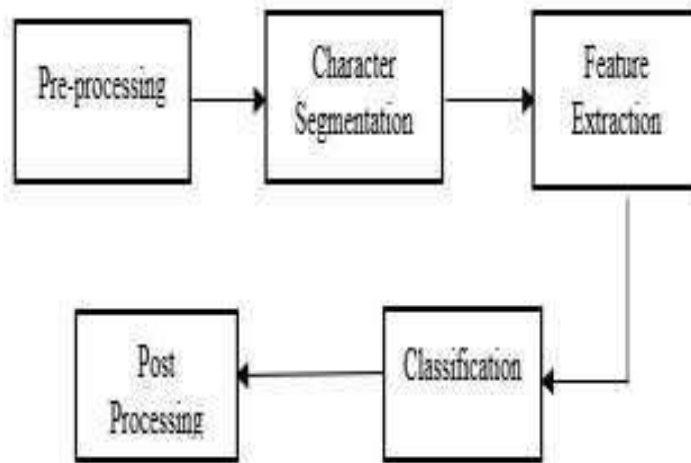*Figure 1.2 Steps involved in Optical Character Recognition*

Fig 1.2 Steps involved in Optical Character Recognition

In the pre-processing phase once the image has been captured, non-identical pre-processing steps can be performed for the improvement of the quality of the images. In these different pre- processing techniques, there could be removal of noise, thresholding and extraction of image base line etc. In Character segmentation the characters in the captured images are separated such that they can be passed to the recognition engine.

In complex situations, where the characters are overlapping or broken in the image advanced recognition techniques can be used. In feature extraction, the characters segmented are then passed to extract non-identical features. Based on the features the characters are been recognized. The extracted features should be efficient for computation, able to minimize intra- class variations and maximize inter-class variations. Classification involves the mapping of features segmented in previous stages with different categories.

The classification of characters involves various types. Structural classification techniques are based on features that are extracted from the structure of images and it uses different decision rules to classify characters. Statistical pattern classification methods are based on probabilistic models and other statistical methods are also involved to classify the characters. In post processing, the results are not accurate for 100% especially for complex languages. It is performed to improve the accuracy of OCR systems. These techniques utilize natural language processing, geometric and linguistic context to correct errors in OCR results. Post processor explained in [1] employs a spell checker and dictionary, probabilistic models like Markov chains and n- grams in order to improve the accuracy. The time and space complexity of a post processor should not be very higher and the application of a post-processor should not cause new errors.

3. TEXT-TO-SPEECH:

Text-to-speech is one of the assistive technologies, that reads the text aloud. It is sometimes called "read aloud" technology. A system which is used for this purpose is speech synthesizer and can be implemented in software and hardware products. A Text-to-Speech (TTS) system converts the normal language into speech, other systems deliver the linguistic representation like phonetic transcriptions into speech. The main phases in the text to speech conversion as stated in [2] are Text processing and Speech synthesis. In Text processing, the input text is analyzed, normalized and transcribed into phonetic or linguistic representation. Speech synthesis is the artificial creation in form the human speech. Speech synthesis are used in variety of applications. Communication aids have developed from low quality talking calculators towards the modern three- dimensional applications, such as talking heads. The implementation method depends on applications that are being used in day to day life. In some cases, such as announcement or warning systems, vocabulary that are unrestricted is not necessary and the result is certainly achieved with some simple form of messaging system. The application field of synthesis of speech is proliferating fast whilst the quality of Text to Speech systems is also increasing steadily. Speech synthesis systems are also becoming more affordable as common for customers, which makes these systems more suitable for daily use. The better availability of TTS systems may increase employment possibilities for people with communication difficulties. Text to speech mainly include applications for most of communication aids. It has its applications for the blind people, applications for the deafened and vocally handicapped [6], educational applications, Applications for Telecommunications and Multimedia and other applications for the future.

II. LITERATURE SURVEY

Noman Islam, Zeeshan Islam, & Nazia Noor, "A survey on optical character recognition system" 2016 [10] – the paper describes that the authors their mentioned their approach of summarizing the results that have been done so far in the field of OCR. The paper provides an overview of similar aspects of Optical Character Recognition and also explains corresponding proposals aimed for solving issues of OCR. The problems addressed in the paper is by improving accuracy and speed of OCR for diverse style documents printed/ written in unconstrained environments. The methodology used include Acquisition Pre-processing, Segmentation, Feature Extraction, Classification, Postprocessing.

Victor Wiley & Thomas Lucas, "Computer Vision and Image processing" 2018 [12] – the authors contribute to recent development on reviews related to computer vision, image processing, and their related studies. The authors categorized the computer vision mainstream into four group, e.g., image processing, object recognition, and machine learning. They also provide brief explanation on the up-to-date information about the techniques and their performance.

Ray Smith, Daria Antonova & Dar-Shuang Lee, "Adapting the Tesseract Open Source OCR Engine for Multilingual OCR" 2009 [11] – the authors describe efforts to adapt the Tesseract open source OCR engine for multiple scripts and languages. The authors explained about concentration on authorization of generic multi-lingual operations which has reduced the customization for producing new language to provide corpus of text and more. Beyond this change were made on certain modules which includes layout analysis, post processing in linguistic manner, changes were not required in the character classifier by changing for fewer limits. The Tesseract classifier adapted easily to Simplify the Chinese language. Testing results on English, a mixture of European languages, and Russian languages, taken from a sample of books, show a reasonably and consistent word error rate between 3.72% and 5.78%, Also results that the, Chinese language results in a character error rate for only 3.77%.

X. Huang, A. Acero, H.-W. Hon, Spoken Language Processing, Prentice Hall PTR, 2001 [15]. The methodology involved in the book is that the steps involved in the TTS.

Konkimalla Chandra Prakash, Y. M. Srikar, Gayam Trishal, Souraj Mandal & Sumohana S. Channappayya, "Optical Character Recognition (OCR) for Telugu: Database, Algorithm and Application" 2017 [9] – the authors used the methodology of data set, pre-processing techniques and classification. The algorithms used are skew correction, MSER variation, character segmentation. Initially they trained the small subset of the existing data and found that the performance of the classifier was not up to the mark. Accuracy for character CNN and vattu CNN were 95.8 and 91.1 %. They used their end to end OCR to find the good results on the data.

Chaw Su Thu & Theingi Zin, "Implementation of text to speech recognition" 2014 [2] – the authors address that the main aim of this paper is the study on Optical Character Recognition along with speech synthesis technology in order to develop a cost-efficient user-friendly image to speech conversion system using MATLAB. In this work, the OCR system is implemented for the recognition of English character in capitals from A to Z and numbers from 0 to 9. Each character is recognized at once. Here the recognized character is saved as text in a notepad file. In this work a text-to-speech conversion system that can get the text through image and directly input in the computer then speech through that text using MATLAB. The methodology used are Template file Creation, Creating the Neural Network and Character Recognition. The limitation of the system is only one character can be converted into text at once.

Karolina Kuligowska, Paweł Kisielewicz & Aleksandra Wlodarz, "Speech synthesis systems" 2018 [8] – The authors explained in detail about the current classification, construction and functioning of speech synthesis systems which gives the idea of the implemented synthesizers. They also addressed the limitations and drawbacks of the system has Spontaneous speech Naturalness of Sparsely spoken languages. The methodology used in the approach is that Concatenative speech synthesis, Statistical parametric speech synthesis system, Hybrid speech synthesis.

C. V. Jawahar, M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran "A Bilingual OCR for Hindi-Telugu Documents and its Applications" 2003 [6] – The authors idea is to use Support Vector Machine (SVM) and Principle Component Analysis (PCA) based OCR for two Indian languages. The accuracy is up to 96.7% and based on it applications are been sketched.

Itunuoluwa Isewon, Jelili Oyelade & Olufunke Oladipupo," Design and implementation of Text to Speech conversion for Visually Impaired People" 2014 [5] – The authors designed an application of Text to speech by using Natural Language Processing (NLP) and Digital Signal Processing (DSP) and then converted the text into speech which is synthesized that represents the text. They developed the application by using Java programming language and used the free open source API called has SWT and DJ Native Swing.

Jisha Gopinath, Aravind S, Pooja Chandran & Saranya SS, "Text to speech Conversion System Using OCR" 2015 [7] – The main idea implemented by the authors is that to recognize the text character and then convert that into speech signal. The method used for this implementation is by using MATLAB by pre-processing the image then by segmenting the characters from each other. Segmentation is followed by the extraction of letters and then to resize them.

Arun K pujar, C dhanujaya, BC Jinaga "An Adaptive Character Recognizer for Telugu Scripts using Multiresolution Analysis and Associative Memory" [1] The authors designed a robust OCR for Telugu text. They used wavelet multi-resolution analysis methodology for extracting features and recognition of letters. The authors used the wavelet analysis for extracting the features of Telugu scripts.

III. PROPOSED SYSTEM

A. Problem Statement

For the people in the remote areas who are not able to understand the international language (English) in the name board of public transport. To remove this language barrier an android application is designed to recognize the language. The application consists of modules which can be accessed by the user with the help of the internet connection. The application is implemented with OCR by Image processing methodology and also Text-to-Speech interface from which the output can be rendered in the form of voice to the user. In the proposed system, the image from the name board is captured from the user. Once they are captured the user have one main advantage of enlarging the image for the pixel clarity for the pre-processing of image. The methodology used is to train the set of data that is set of words in order to obtain with greater accuracy. In our system the main idea is to scan the source and the destination of the bus.
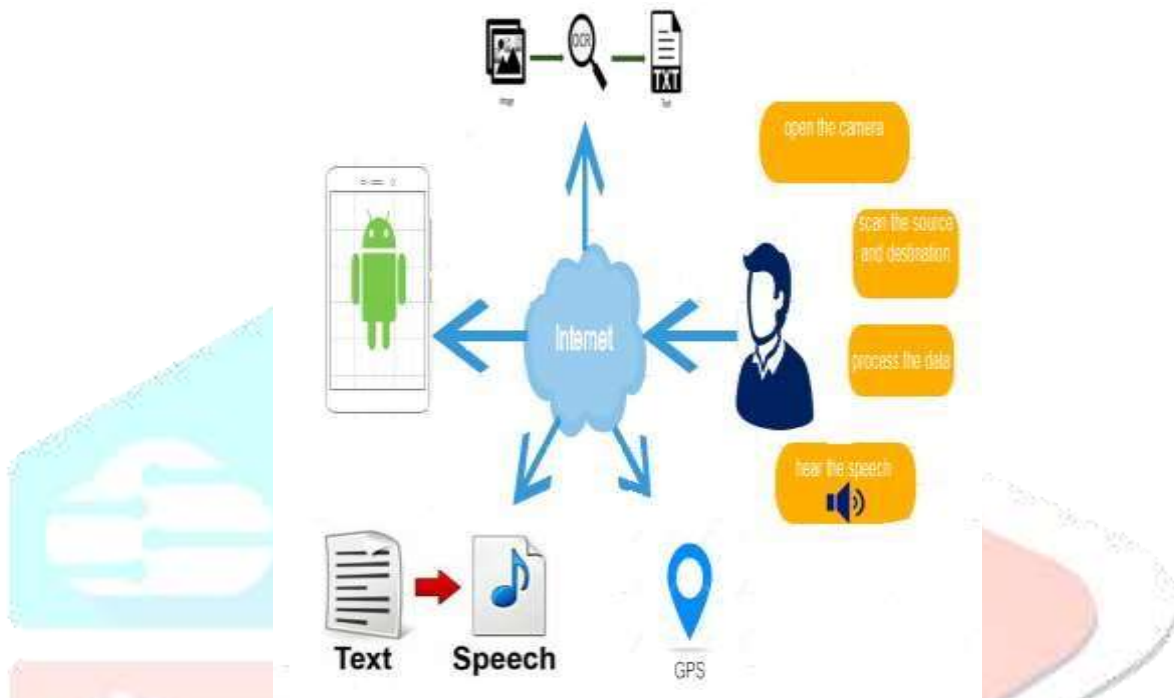

Fig 1.3 System Architecture

B. Proposed system with explained with a scenario

Let us take a place Chennai as source and the Chittoor as destination which is displayed as analog image in the nameboard of the bus. Here the place Chennai is captured from the android device and it is processed by applying the steps involved in the optical character recognition. Then the image is being compared with the trained set of words and if matched with the Telugu word Chennai the corresponding Telugu word is rendered as voice to the user by using the Text to speech API. The application also provides with the GPS for the user to know the route from the source and the destination. The GPS will be implemented by using the Google Cloud API.
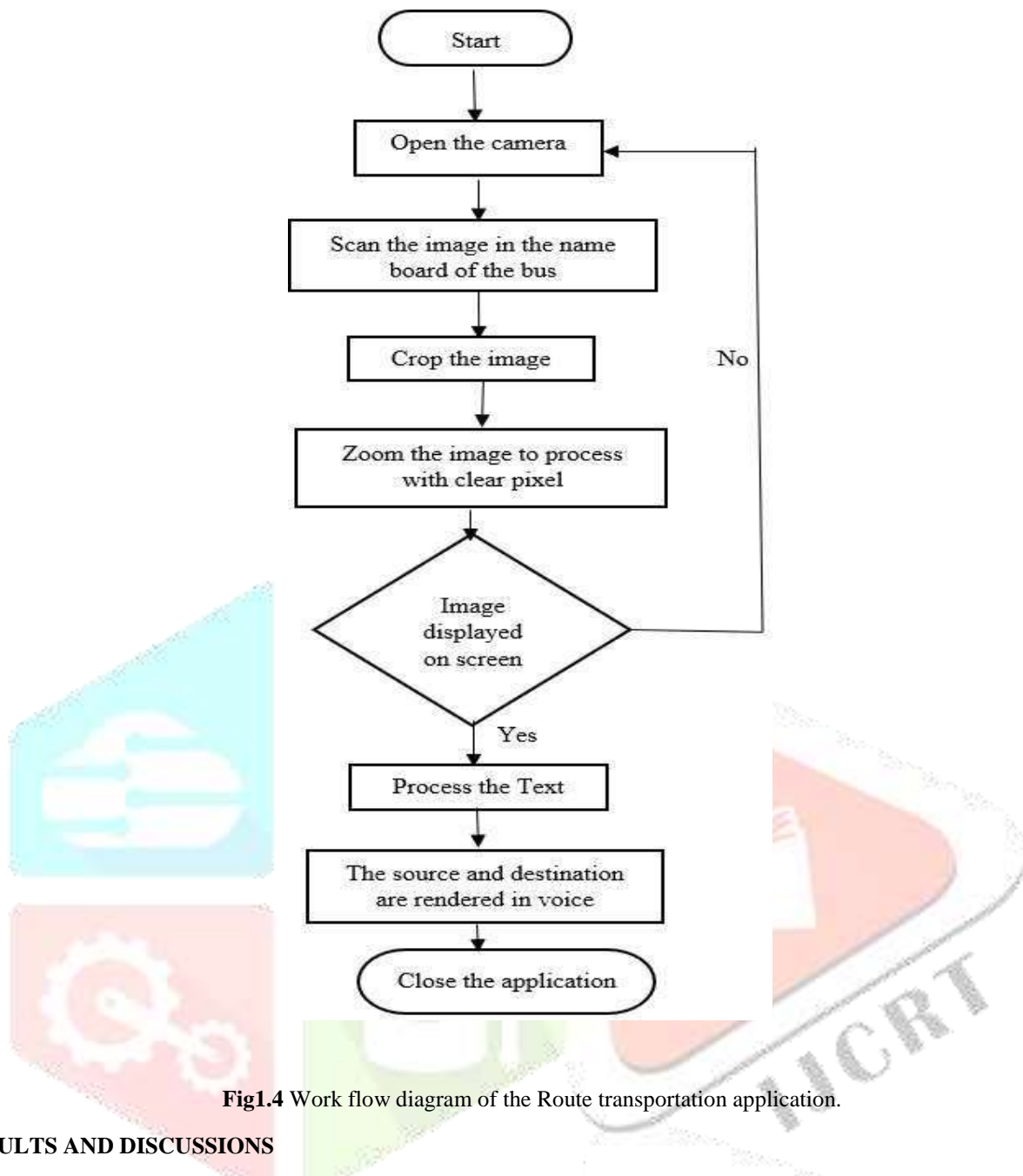
Fig1.4 Work flow diagram of the Route transportation application.

IV. RESULTS AND DISCUSSIONS

As the usage of mobile phones and apps are exploding in day to day life, we have finally developed an app that can be used by the people who are barrier to other languages when they move from one place to another in a public transport. The app is developed by the OCR methodology, in which the words in other languages can be recognized. Initially, we have trained only few sets of words for comparing. When the camera is opened the can be able to capture the image from the name board of public transport. Then the user will be able to crop the image for obtaining the word from the image. One of the advantages of cropping is the image gets automatically zoomed so that the user can process the word faster. When the image is zoomed the accuracy for getting the word from the image is up to 80%. Once the word is obtained from the image it can be processed and then can be viewed on the screen. After that the text is processed further in order to hear the text in international language.

Fig1.5 OCR Scanning of Telugu text

Fig 1.6 Final results in voice format

V. CONCLUSION AND FUTURE WORK

The proposed work contains modules which can be implemented in Java Programming language. It can be concluded that the proposed system based on natural language processing which involves the translation of one language to another by recognition of characters is typically used to help the people in the remote areas in order to avoid the language barrier of international language (English) without the aid of others.

Although it contains certain limitations, the application requires higher-grade cameras, in order to improve the accuracy of scanning and faster performance. If the public transport waits at the bus stop for 10-15 seconds the user will not be able to scan the places in the name board. Since the android version gets updated to more faster versions the libraries and packages used needs to be imported and also SDK should be installed.

The future scope of the work would be updated for scanning the intermediate stops from the source to destination online that would include bilingual languages.

REFERENCES

[1] Arun K pujar, C dhanujaya, BC Jinaga "An Adaptive Character Recognizer for Telugu Scripts using Multiresolution Analysis and Associative Memory" unpublished

[2] Chaw Su Thu & Theingi Zin, "Implementation of text to speech recognition" 2014 International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 3, March – 2014.

[3] https://www.tutorialspoint.com/dip/image_processing_ introduction.html.

[4] https://sisu.ut.ee/imageprocessing/book/1

[5] Itunuoluwa Isewon, Jelili Oyelade & Olufunke Oladipupo, "Design and implementation of Text to Speech conversion for Visually Impaired People" 2014 International Journal of Applied Information Systems (IJAIS) – ISSN 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 7– No. 2, April 2014 – www.ijais.org.

[6] C. V. Jawahar, M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran "A Bilingual OCR for Hindi-Telugu Documents and its Applications" 2003 Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03) 2003 IEEE.

[7] Jisha Gopinath, Aravind S, Pooja Chandran & Saranya SS, "Text to speech Conversion System Using OCR" 2015 International Journal of Emerging Technology and Advanced Engineering, www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 1, January 2015).

[8] Karolina Kuligowska, Paweł Kisielewicz & Aleksandra Włodarz, "Speech synthesis systems" 2018 International journal of Engineering and Technology 7(2.28) 234-238.

[9] Konkimalla Chandra Prakash, Y. M. Srikar, Gayam Trishal, Souraj Mandal & Sumohana S. Channappayya, "Optical Character Recognition (OCR) for Telugu: Database, Algorithm and Application" 20 Nov 2017.

[10] Ray Smith, Daria Antonova & Dar-Shuang Lee, "Adapting the Tesseract Open Source OCR Engine for Multilingual OCR" July 25,2009.

[11] Victor Wiley & Thomas Lucas, "Computer Vision and Image processing" International Journal of Artificial Intelligence Research ISSN: 2579-7298 Vol 2, No 1, June 2018, pp. 28-36.

[12] www.ques10.com/p/33595 what-is-image-processing- explain-fundamental-steps/.

[13] www.tutorialspoint.com/image_processing_introduction. htm.

[14] www.tutorialspoint.com/dip/applications_and_usage.htm.

[15] X. Huang, A. Acero, H.-W. Hon, Spoken Language Processing, Prentice Hall PTR, 2001.