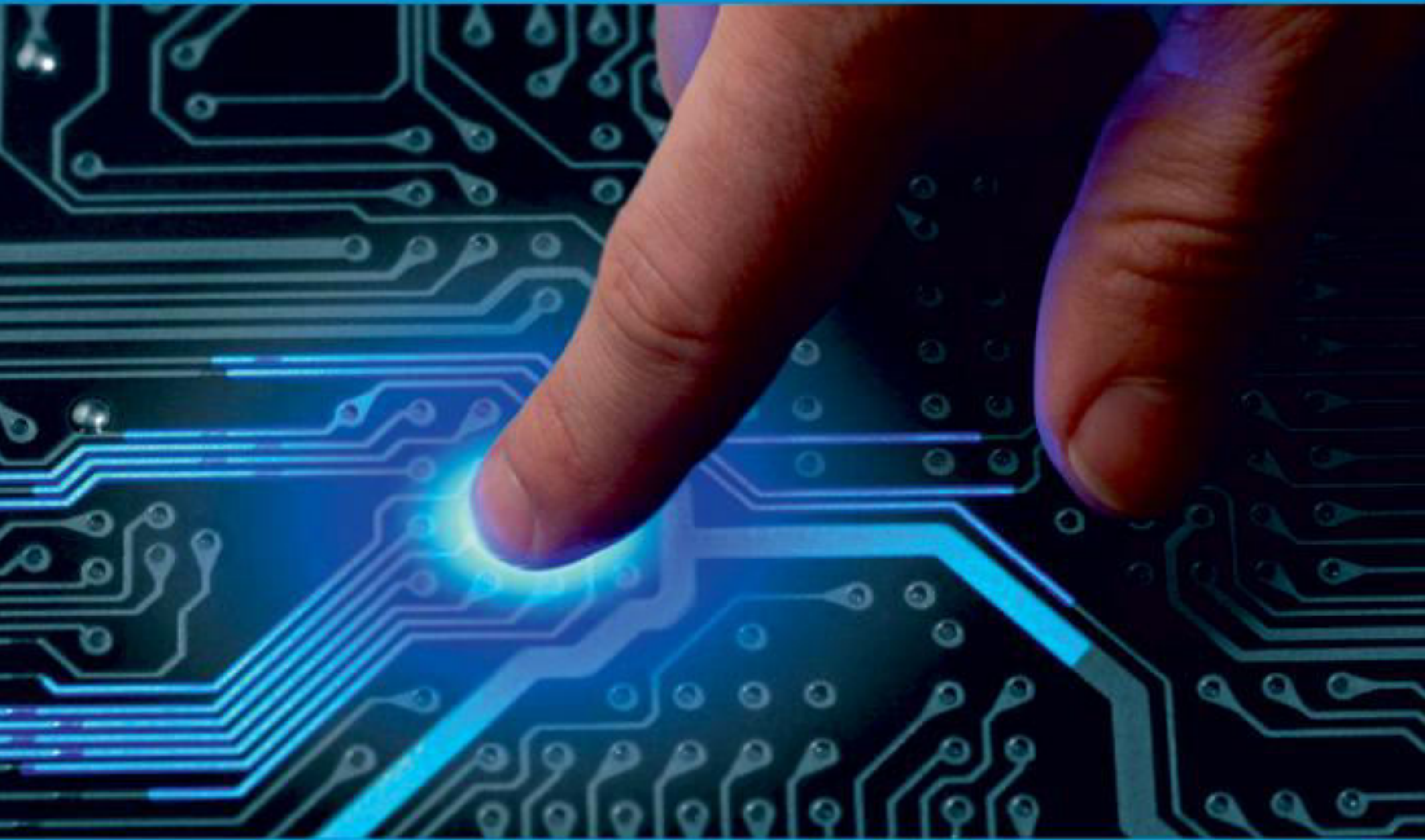
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

Text Extraction and Translation from an Image

Dr. Vijayalakshmi G V[1], Amit Anveri[2], Abdul Rehman[3], K Siddartha Shetty[4]

Associate Professor, Dept. of ECE, BMSIT&M, Bengaluru, Karnataka, India[1]

UG Student, Dept. of ECE, BMSIT&M, Bengaluru, Karnataka, India[234]

ABSTRACT: Communication is an essential part of human beings, and in the world today there are almost thousands of languages being used for communication, a person knowing and learning all the languages is not possible. Hence this causes the problem of languages, acting as barrier for communication and leads to various other problems. So, translation is required to convert one language to other languages, so that people can easily communicate each other across the globe. In the proposed system the problem is overcome by using software tools that convert the required image or document to any desired languages. The proposed method is an assistive text reading that helps to read text present on the text labels, printed notes and products in their own respective languages. It combines the concept of Optical Character Recognition (OCR), Translator and Text to Speech synthesis.

KEYWORDS: Optical Character Recognition, Text to speech, Text Extraction, Image preprocessing,Translator.

I. INTRODUCTION

The only means by which human beings' abstract reality is through language. Language is an efficient and effective medium of communication which explicitly represents ideas and expressions of human mind. Over 5000 languages exist in the world today and this gives evidence of linguistic diversity. It is difficult for an individual to know and understand all the languages of the world. Hence methodology of translation has been adopted to communicate messages from one language to another. The translation software comes into picture for efficient translation of one language to other languages. The written documents such as signboards, or any other papers to be translated needs to be recognized beforehand, so that the methodology of translation can be applied and the language can be translated. Optical character recognition (OCR) is used for recognition of characters in the language. It is a process run by a software that attempts to read the characters into recognizable full text. It is a quick process that enables automated conversion of images into full text files that can then be searched by word or character. Hence combination of the OCR and translation software can be used to convert any document or image into desired languages. The language acts as barrier many times, as one person cannot learn all the languages in the world and cannot be confined to one place as well, hence translation of the other languages to the known languages are needed. The tourism industry has always need of the software which could help them in communication with the tourists who are from various parts of the world, who just knew their native languages. Every state or country have their own native languages, which are used by people living there. And majority of signboards, addresses or written documents may be in their respective languages and people, who doesn't know these languages face problems.In India there are several languages spoken in several parts of the state, and the person visiting these places faces problem in communication as these regions use their languages for signposts, hoarding's, addresses, etc., and state governments also function in their own languages, like in Karnataka, Kannada is used as medium in various official documents to convey the messages by the government. The proposed system motivates us to solve the problem of language by translating the written or printed documents in the required language. The proposed system can also boost tourism as many tourists face these problems every year, in locating locations, etc.

II. LITERATURE SURVEY

Literature Survey is an important phase in the system development life cycle as we collect and acquire the necessary information to handle or develop a project during this phase. A literature review is a description of the literature relevant to particular field or topic. It gives an overview of what has been said, who the key writers are, what are the prevailing theories and hypothesis and what methods and what methodologies are appropriate and useful. G. Vamvakas et al [3] proposed the large number of documents in today's world is being digitalized; this creates a need for accurate and fast processing system. Historical documents are of more importance as they reflect our cultural heritage. During the last decades a lot of research has been done in order to develop systems and software that can be used to preserve and digitalize documents. One such system for processing documents is the "Optical character recognition system" also

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

known as OCR. Many commercial products have been designed that can convert digitized documents into word documents usually in ASCII format. These products can only read machine printed documents but when it comes to handwritten document's they are not able to produce satisfactory results. Till date recognition of historical, poor quality and handwritten documents is an evolving field in the study of OCR.Noman Islam et al [1] has found that character recognition is not a new problem but its roots can be traced back to systems before the inventions of computers. The earliest OCR systems were not computers but mechanical devices that were able to recognize characters, but very slow speed and low accuracy. In 1951, M. Sheppard invented a reading and robot GISMO that can be considered as the earliest work on modern OCR. GISMO was able to read and recognize musical characters and as well as words. But it was limited to only 23 characters. J. Rainbow, in 1954, devised a machine that can read uppercase typewritten English characters, one per minute. The early OCR systems were criticized for not being accurate and fast. Hence it has been an evolving field since the 1950's. Because of the complexities associated with recognition, it was felt that three should be standardized OCR fonts for easing the task of recognition for OCR. Hence, in 1970 OCRA and OCRB were developed by ANSI EMCA, that provided comparatively acceptablerecognition rates. Othman Saleh Mahdy et al [2] developed a system to process documents is Computer assisted translation also known as "CAT" or machine translation (MT). Translation plays a very important role for interaction between two different communities that speak and understand different languages. Computer-assisted translation (CAT) is the developments of computer technology that have created new opportunities for translators that cannot be found in traditional ways(Hutchins 1992) defines the term Machine Translation (MT) as the traditional and standard name for computerized systems responsible to produce translations from one language into another, with or without human assistance.Ankush Garg et al [4], Warren Weaver proposed using computers to solve the task of machine translation. Earlier research focused on rule-based systems, which gave way to example-based systems in the 1980s. Statistical machine translation gained prominence starting late 1980s, and different word-based and phrase-based techniques requiring little to no linguistic information were introduced. With the development in deep neural networks in 2012 application of these neural networks has become a major area of research. Various fields in the area of natural language processing (NLP) have been boosted by the rediscovery of neural networks. Early attempts used feed forward neural language models for the target language to re rank translation lattices. The first neural models which also took the source language into account extended this idea by using the same model with bilingual tuples instead of target language words, scoring phrase pairs directly with a feed forward network, or adding a source context window to the neural language model. The advent of Neural machine translation (NMT) certainly marks one of the major milestones in the history of MT, and has led to a radical and sudden departure of mainstream research from many previous research lines. NMT has already been widely adopted in the industry and is deployed in production systems by Google, Microsoft, Facebook, Amazon, SDL, Yandex, and many more.

III. PROPOSED METHODOLOGY

This project presents a prototype system for extraction of text present in the image and then translates it into desired language. As illustrated in the block diagram the system framework consists of three functional components: Optical Character Recognition (OCR), Translator and Text to speech conversion. In the first step the recognized text present in the image is extracted using OCR engines. In this project we use Tesseract OCR engine which helps to extract the recognized text. The aim of Optical Character Recognition (OCR) is to classify optical patterns (often contained in a digital image) corresponding to alphanumeric or other characters. The process of OCR involves several stepsincluding segmentation, feature extraction, and classification. In principle, any standard OCR software can now be used to recognize the text in the segmented frames. However, a hard look at the properties of the candidate character regions in the segmented frames or image reveals that most OCR software packages will have significant difficulty to recognize the text. Document images are different from natural images because they contain mainly text with a few graphics and images. Due to the very low-resolution of images of those captured using handheld devices, it is hard to extract the complete layout structure (logical or physical) of the documents and even worse to apply standard OCR systems. For this reason, a shallow representation of the low-resolution captured document images is proposed. In case of original electronic documents in the repository, the extraction of the same signature is straightforward; the PDF or PowerPoint form of the original electronic documents is converted into a relatively high-resolution image (TIFF, JPEG, etc.) on which the signature is compute. Finally, the captured document's signature is compared to with all the original electronic documents' signatures in order to find a match.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

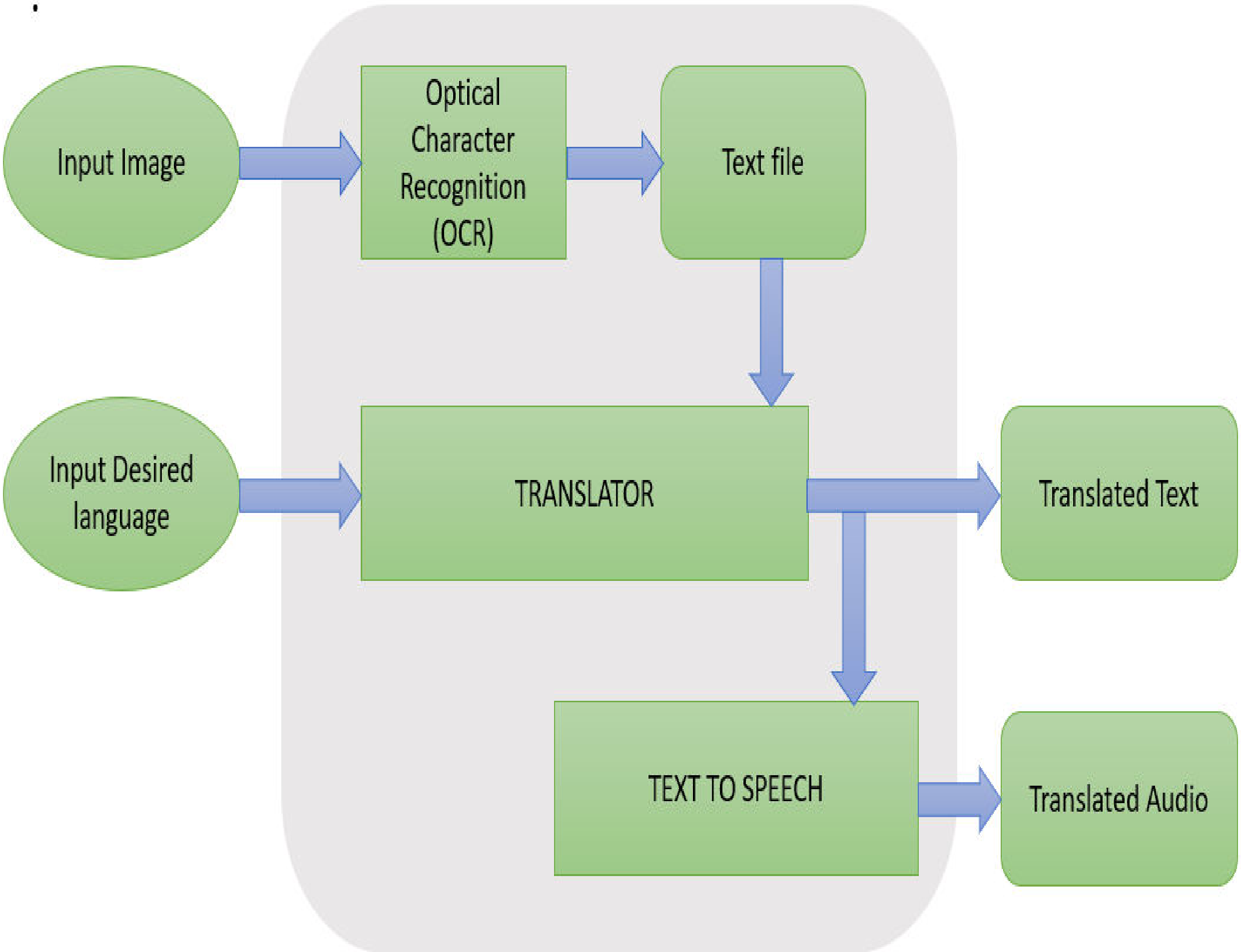


Fig : Block daigram of proposed model

In the proposed system we have used Tesseract OCR, it is a type of OCR engine with matrix matching. The selection of Tesseract engine is because of its flexibility and extensibility of machines and the fact that many communities are active researchers to develop this OCR engine and also because Tesseract OCR can support 149 languages. Before feeding the image to the OCR, it is converted to a binary image to increase the recognition accuracy. Image binary conversion is done by using Image magic software, which is another opensource tool for image manipulation. The output of OCR is the text, which is stored in a file (speech.txt). Machines still have defects such as distortion at the edges and dim light effect, so it is still difficult for most OCR engines to get high accuracy text. It needs some supporting and condition in order to get the minimal defect. Pre-processing is the first step in the processing of scanned image. The scanned image is checked for noise, skew, slant etc. There are possibilities of image getting skewed with either left or right orientation or with noise such as Gaussian. Here the image is first converted into grayscale and then into binary. After pre-processing, the noise free image is passed to the segmentation phase, where the image is decomposed into individual characters. The binarized image is checked for inter line spaces. If inter line spaces are detected then the image is segmented into sets of paragraphs across the interline gap. The lines in the paragraphs are scanned for horizontal space intersection with respect to the background. Histogram of the image is used to detect the width of the horizontal lines. Then the lines are scanned vertically for vertical space intersection. Here histograms are used to detect the width of the words. Then the words are decomposed into characters using character width computation. Feature extraction follows the segmentation phase of OCR where the individual image glyph is considered and extracted for features. First a character glyph is defined by the following attributes like height of the character, width of the character. Classification is done using the features extracted in the previous step, which corresponds to each character glyph. These features are analyzed using the set of rules and labeled as belonging to different classes. This classification is generalized such that it works for single font type. The height of the character and the width of the character, various distance metrics are chosen as the candidate for classification when conflict occurs. Similarly, the classification rules are written for other characters. This method is a

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |
|| Volume 9, Issue 7, July 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0907057 |

genericonesinceitextractstheshapeofthecharactersandneednotbetrained.Whenanewglyph is given to this classifier block it extracts the features and compares the features as perthe rules andthen recognizes the character and labels it. Finally in the proposed system after applying the preprocessing to the images and extracting text we give the text to a translation and text to speech engines. Theses engines produce translated text and the text to speech engine reads the text aloud so that user can even listen and understand without the need to read the text.

IV. RESULTS

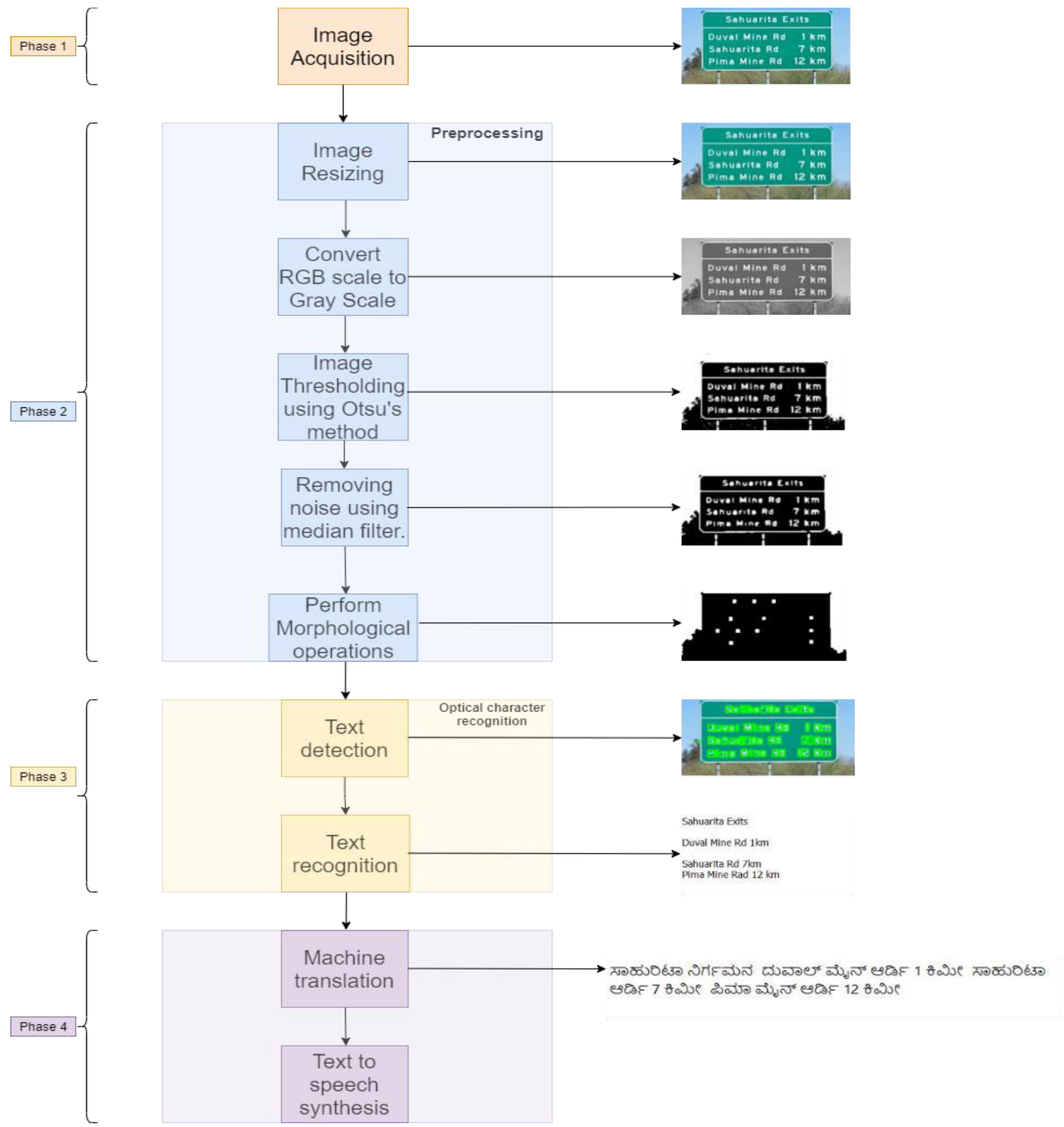


Fig:Flowchart of implementation

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

The proposed project has been executed in 4 phases:

Phase 1: Image acquisition, Phase 2: Image pre-processing, Phase 3: Text detection and text extraction, Phase 4: Machine translation and text to speech analysis.

Phase 1: Image acquisition: Digital imaging or digital image acquisition is the creation of a representation of the visual characteristics of an object, such as a physical scene or the interior structure of an object. The term is often assumed toimply or include the processing, compression, storage, printing, and display of such images.

Phase 2: Image pre-processing: Image resizing: Image resizing is necessary when you need to increase or decrease the total number of pixels, whereas remapping can occur when you are correcting for lens distortion or rotating an image. Interpolation works by using known data to estimate values at unknown points. Image interpolation works in two directions, and tries to achieve a best approximation of a pixel's intensity based on the values at surrounding pixels. Common interpolation algorithms can be grouped into two categories: adaptive and non-adaptive. Adaptive methods change depending on what they are interpolating, whereas non-adaptive methods treat all pixels equally. Non-adaptive algorithms include: nearest neighbour, bilinear, bicubic, spline, sinc, lanczos and others. Adaptive algorithms include many proprietary algorithms in licensed software such as: Qimage, Photo Zoom Pro and Genuine Fractals**. RGB to Grayscale:** A pixel colour in an image is a combination of three colours Red, Green, and Blue (RGB). The RGB colour values are represented in three dimensions XYZ, illustrated by the attributes of lightness, Chroma, and hue. Quality of a colour image depends on the colour represented by the number of bits the digital device could support. The basic colour image represented by 8 bits, the high colour image represented using 16 bits, the true colour image represented by 24 bits, and the deep colour image is represented by 32 bit.**Image thresholding using Otsu's method:** Otsu's Thresholding concept: Automatic global thresholding algorithms usually have following steps: Process the input image,obtain image histogram (distribution of pixels), Compute the threshold value,replace image pixels into white in those regions, where saturation is greater than and into the black in the opposite cases. **Remove Noise by using median filter:** Median filter is one of the well-known order-statistic filters due to its good performance for some specific noise types such as "Gaussian," "random," and "salt and pepper" noises. The median filtering process is accomplished by sliding a window over the image. The filtered image is obtained by placing the median of the values in the input window, at the location of the centre of that window, at the output image. Medianfilters are useful in reducing random noise, especially when the noise amplitude probability density has large tails, and periodic patterns.

Phase 3: Text detection and text extraction; Boundingboxes: Bounding boxes are one of the most popular—and recognized tools when it comes to image processing for image and video annotation projects. A bounding box is an imaginary rectangle that serves as a point of reference for object detection and creates a collision box for that object. Data annotators draw these rectangles over images, outlining the object of interest within each image by defining its X and Y coordinates. This makes it easier for machine learning algorithms to find what they're looking for, determine collision paths, and conserves valuable computing resources. Bounding boxes are one of the most popular image annotation techniques in deep learning. Compared to other image processing methods, this method can increase annotation efficiency. **Text extraction:** In this phase of our tool, we use an OCR tool to extract the text from the pre-processed images and use it for further translation and for speech synthesis.

Phase 4: Machine translation and text to speech synthesis: Machine translation (MT) is the task to translate a text from a source language to its counterpart in a target language. The neural approach uses neural networks to achieve machine translation. Graphics processing unit (GPU) is used for a faster artificial neural network. It is used to implement the matrix multiplication of a neural network to enhance the time performance of a text detection system. In this phase we finally use the extracted text and give it to the translator API which converts the given text language to some other desired language. A Text-to-speech synthesizer is an application that converts text into spoken word, by analysing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3 file. The development of a text to speech synthesizer will be of great help to people with visual impairment and make making through large volume of text easier. The mp3 file gets stored in the project workspace and then it is finally played using the play sound library in python.The proposed system ensures the following:

1. It takes an image as input, pre-process it and extracts text from it.
2. The extracted text is translated to desired language.
3. The translated text is converted to audio and read aloud.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

A Graphical User Interface was built for a user to interact with the model.

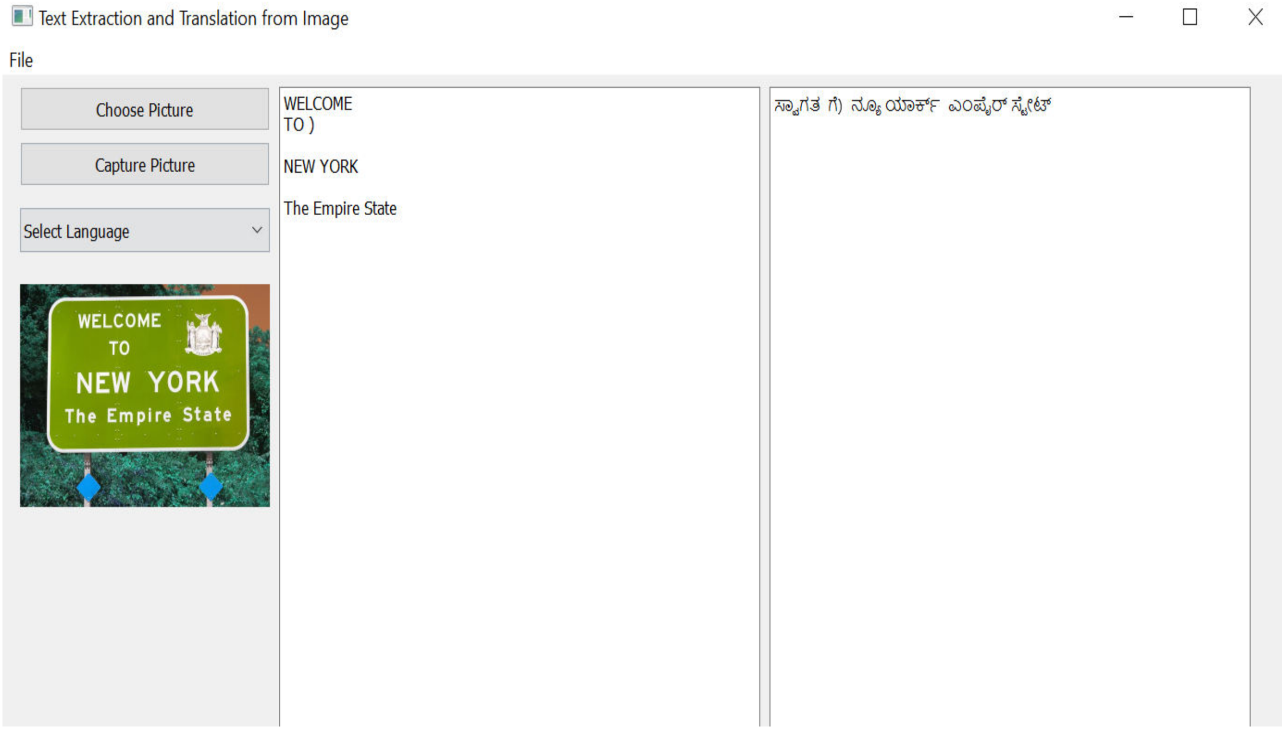


Fig: GUI with outputs

V. Evaluation of the Proposed Model

To evaluate the application, 25 images from the database were selected and provided as input to the software. It has been ensured that the collected images fed to the software are unique. Best efforts have been made to ensure unbiased evaluation since this evaluation will prove to be an important feedback on the project.To evaluate the project, Accuracy is considered as the main parameter and is calculated in 2 stages. Calculating accuracy in 2 Stages is a good way to get detailed information of the evaluation.

In stage 1 the accuracy of the extracted text is calculated.

$$\text{Formula used: Extraction\_Accuracy} = \frac{\textit{number of words extracted correctly}}{\textit{total number of words}} * 100$$

In stage 2 the accuracy of the translated text is calculated.

$$\text{Formula used: Translation\_Accuracy} = \frac{\textit{number of words translated correctly}}{\textit{total number of words}} * 100$$

After evaluating the application using 25 images the average accuracy was found to be 90.1 % for stage 1 i.e., Extraction accuracy and the average accuracy for stage 2 was found to be 81.8 % i.e., Translation accuracy.

VI. Conclusion

Finally describing the conclusion of our project, it mainly intents to describe the approach of recognizing and extracting text in an image, then translating it to other language. The output of the project is presented by a Graphical User Interface (GUI) where the user can easily interact. The main aim of our project is to solve the problem of language and help the user to read the text in his known language, the User Interface also reads the translated text, hence even an uneducated/illiterate user can also know about the text in the image. Currently the proposed project is capable of translating text from English to Kannada, but for future enhancement of our project various other languages can be added, the user interface has an option of addition of languages in it. Further the project can be converted to mobile application or an App where it can be launched on app stores and can serve a broader range of users.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907057 |

REFERENCES

1. Karishma Tyagi, Vedant Rastogi Department of Computer Science & Engineering, IET Alwar, Rajasthan, India "**Survey on Character Recognition using OCR Techniques**" International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue2, February-2014.
2. Othman Saleh Mahdy Mohammed, Ph.D. Scholar Shaikh Suhel Samad, Assistant Professor Hassan Saleh Mahdi, Associate Professor, University of Bisha, Saudi Arabia "**A Review of Literature of Computer-Assisted Translation**" India's Higher Education Authority UGC Approved List of Journals Serial Number49042.
3. G.Vamvakas, B.Gatos, N.Stamatopoulos, and S.J.Perantonis Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", "**A Complete Optical Character Recognition Methodology for Historical Documents**".
4. Ankush Garg, Mayank Agarwal Department of Computer Science University of Massachusetts Amherst, "**Machine Translation: A Literature Review**" rXiv:1901.01122v1 [cs.CL] 28 Dec2018.
5. Li Peng Foreign Language School East China Jiaotong University, Nanchang330013, China "**A Survey of Machine Translation Methods**" TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7125~7130 e-ISSN:2087-278X.
6. Karmel, A & Sharma, Anushka & pandya, Muktak & Garg, Diksha. (2019). IoT based Assistive Device for Deaf, Dumb and Blind People. Procedia Computer Science. 165. 259-269. 10.1016/j.procs.2020.01.080.
7. Nwakanma, Ifeanyi & Oluigbo, Ikenna & Izunna, Okpala. (2014). Text – To – Speech Synthesis (TTS). 2. 154-163.
8. Ta Yang Goh, Shafriza Nisha Basah, HanizaYazid, Muhammad Juhairi Aziz Safar, Fathinul Syahir Ahmad Saad School of Mechatronic Engineering, University Malaysia Perlis, 02600 Pauh Putra, Perlis, Malaysia "Performance analysis of image thresholding: Otsu technique".
9. Digital image processing, 3rd edition by Rafael C Gonzalez, University of Tennessee and Richard E Woods, Pearson International Edition prepared by Pearson Education.
10. Google cloud for understanding text to speech translation.
11. Fundamentals of Digital Image Processing by Roger L Easton, Jr. 22 November 2010.
12. I. Marosi, "Industrial OCR approaches: architecture, algorithms and adaptation techniques", Document Recognition and Retrieval XIV, SPIE Jan 2007, 6500-01.
13. A Alcina, S Victoria, J Granell Translation Technology Skills Acquisition Perspectives: Studies in Translatology, volume 15, issue 4, p. 230 – 244.
14. L Bowker Computer-aided translation: Translator training The Routledge Encyclopaedia of Translation Technology, p. 88 – 104.