IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

A Web Application to Automate Multiple Task like Text Recognition from Speech, Speech Recognition, Machine Translation

Ms.Disha.D.Naik[1], Ms.Gauri.V.Mochemadkar[2], Ms.Triveni.D.Parulelar[3], Mr.Mhir.A.Sawant[4]

Prof.Kate.P.D[5]

Department of Computer Engineering, Yashwantrao Bhonsale Polytechnic, Sawantwadi, India

ABSTRACT: The aim of our project to automate the application to overcome from the language barrier among countries and also states within the country, the above-mentioned application will perform the various features in the application. Language technology has become pervasive in everyday life. Neural networks area key component in this technology thanks to their ability to model large amounts of data. Text recognition in images is a research area which attempts to develop a computer system with the ability to automatically read the text from images. Speech is the first important primary need, and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. Machine translation is a domain of computational linguistics, which explores the use of software to translate text from one language to another. Machine translation simply performs substitution of words in one language for words in other, but that may not assure good translation. In the Project we combine all three domains in one system. Here we integrate Text extract from image, Speech recognition and Text translation from one language to another in one web application.

KEYWORDS: Text Extraction, SpeechRecognition, Machine Translation

I. INTRODUCTION

We introduced this web application for language barriers in communication nowadays create problems for successful communication. A web application is a client-server program. It means that it has a client-side and a server-side. The term "client" here refers to the program the individual uses to run the application. It is part of the client-server environment, where many computers share information. Text recognition is a wide area of research in the field of image processing and pattern recognition. With the growing computational power character recognition methodologies have been improved and increasing its demand in various applications. Speech recognition are mainly used for converting the speech to text for understanding the language which are spoken by user during communication, because of this person can recognize the speech are spoken by another person. Machine translation (MT) is a domain of computational linguistics, which explores the use of software to translate text from one language to another. Machine translation simply performs substitution of words in one language for words in other.

Problem Statement

Mobile device usage is on the rise. Many machine learning applications like multilingual translation, text extraction, Speech Recognition, etc. are developed for mobile users. But these applications are separate for each region. We are going to build a one single application which includes Text extraction from images, Speech recognition and

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

Translation of text from one language to another

Objective

Our main objective is to combine all different tasks such as extract text from image, speech recognition or Speech to text conversion, text translation from one language to another all embedded in one so that we get a user-friendlyweb application.

Scope

We develop a user-friendly web application. Here we are integrating the text extract from image, Speech recognition or speech to text andtext translation from one language to another inone system so user doesn't have to download for the different application and doesn't refer for different web applications.

II. LITERATURE REVIEW

In this paper, The aim of the project is to automate the application to overcome from the language barrier among countries and also states within the country, the above mentioned application will perform the various features inthe application. The application recognizes speech (human matter) in one language to another user defined language to communicate in expressive manner. It includes 4 modules voice recognition, translation and speech synthesis and image translation and gives audioof the translated language. Also the applicationaccepts text written and converts it into the language needed. Application is able to recognize the text present in the image which stored in system or captured using camera and translate the text into the language needed and display the translation result back on to the screen of system. [1] In this paper, Text present in image provides important information for automatic annotation, indexing and retrieval. Therefore, its extraction is a well-knownresearch area in computer vision. However, variations of text due to differences in orientation, alignment, font, size, low imagecontrast and complex background make the problem of text extraction extremely challenging. In this paper, we propose a texture- based text extraction method using DWT with K-means clustering. First, the edges are detected from image by using DWT. Then, a small size overlapped sliding window is used toscan high frequency component sub-bands from which texture features of text and non-textregions are extracted. Based on these features, K-means clustering is employed to classify theimage into text, simple background and complex background clusters. Finally, voting decision process and area based filtering are used to locate text regions exactly. Experimentation is carried out using public dataset ICDAR 2013 and our own dataset for English, Hindi and Punjabi text images for different number of clusters. The results show that the proposed method gives promising results with different languages in terms of detection rate (DR), precision rate (PR) and recall rate (RR). [2] In this review paper, In present industry, communication is the key element to progress. Passing on information, tothe right person, and in the right manner is veryimportant, not just on a corporate level, but alsoon a personal level. The world is moving towards digitization, so are the means of communication. Phone calls, emails, text messages etc. have become an integral part of message conveyance in this tech-savvy world. In order to serve the purpose of effective communication between two parties withouthindrances, many applications have come to picture, which acts as a mediator and help in effectively carrying messages in form of text, orspeech signals over miles of networks. Most ofthese applications find the use of functions suchas articulatory and acoustic-based speech recognition, conversion from speech signals to text, and from text to synthetic speech signals, language translation amongst various others. Inthis review paper, we'll be observing different techniques and algorithms that are applied to achieve the mentioned functionalities. [3] This paper discusses, Machine Translation is the translation of text or speech by a computer with no human involvement. It is a popular topic in research with different methods being created, like rule-based, statistical and example Machine Translation using Deep Learning

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

(Neural Machine Translation) is a newlyproposed approach to machine translation. Theterm Machine Translation is used in the sense of translation of one language to another, with no human improvement. It can also be referredto as automated translation. Unlike the traditional sta, the neural machine translation aims at building a single neural network that can be jointly tuned to maximize the translation performance. The models proposed recently for neural machine translation often belong to a family of encoder–decoders and encode a source sentence into a fixed-length vector fromwhich a decoder generates a translation. In thisapproach Neural Machine Translation technique will be used to translate Japanese language into English language. Tanaka corpuswill be used in this approach. Japanese will be translated into English using improved Recurrent Neural Network (RNN).

III. PROPOSED METHOD

In the proposed system, we will develop a web application to automate multiple tasks. A web application is a computer program that utilizes web browsers and web technology to perform tasks over the Internet. Web applications use a combination of server-side scripts (PHP and ASP) to handle the storage and retrieval of the information, and client-side scripts (JavaScriptand HTML) to present information to users. In this web application we combine Text extraction from images, Speech recognition andText translation from one language to another in one system. So, user don't have to use different web application. For our system, it has3 modules that is text extraction from image, Speech recognition or speech to text conversionand Text translation from one language to another.

Text Recognition from image

In this proposed system, we are taking the Textextract from images module. This module usesTesseract, OCR and OpenCV to extract text from images. Tesseract is an open-source text recognition Engine. Tesseract can be used directly via command line, or (for programmers) by using an API to extract printed text from images. It supports a wide variety of languages. Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposedon an image. OpenCV is the huge open-source library for the computer vision, machine learning, and image processing and now it playsa major role in real-time operation which is veryimportant in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human.When it integrated with various libraries, such as NumPy, python is capable of processing theOpenCV array structure for analysis. To Identify image pattern and its various features we use vector space and perform mathematicaloperations on these features

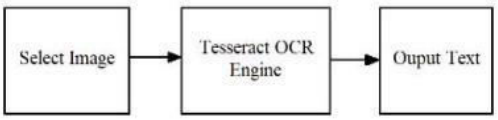

Figure 1: Text extraction from image

Speech Recognition

Our next module is Speech Recognition orSpeech to text translation. Speech recognition, or speech-to-text, is the ability for a machine orprogram to identify words spoken aloud and convert them into readable text. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. More sophisticated software has the ability to accept natural speech, different accents and languages. Speech Recognition Systems can be classified on basis of the following parameters:

- **Speaker:** All speakers have a different kind of voice. The models hence are either designedfor a specific speaker or an independent speaker.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

- **Vocal Sound:** The way the speaker speaks also plays a role in speech recognition. Some models can recognize either single utterances orseparate utterance with a pause in between.

- **Vocabulary**: The size of the vocabulary plays an important role in determining the complexity, performance, and precision of the system. In this proposed system we use web speech api to convert for human voice input to text form. The Web Speech API has two parts: SpeechSynthesis (Text-to-Speech), and SpeechRecognition (Asynchronous SpeechRecognition)

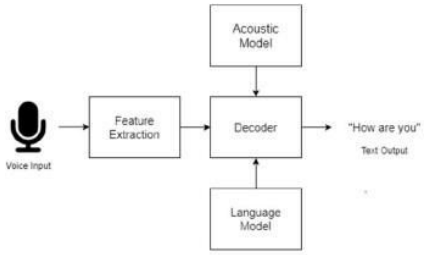

Figure 2: Speech Recognition System

Image To Pdf

Our third module is Image To Pdf convertor from multiple image to pdf. Image Convertor, sometimes referred to by the abbreviation MT (not to be confused with computer-aided convertor, machine-aided human translation or interactive conversion), is a sub-field of computational linguistics that investigates the use of software to convert image are convert into pdf format .Image Translator performs substitution of words in one Image to Pdf in another, but that alone rarely produces a good translation because recognition of whole phrases and their closest counterparts in the target is needed. For Image convertor model use to save time .

Figure 3: Text Translation

IV. FUTURE WORK

Here we implement a web application system,so in future we can implement this system for mobile phone and desktops separately. So, users can more efficiently use this system by one click of mobile or desktop for conversion of languages without using any browser. In future developer can add some modules like text extraction from scanned pdf, textsummarization, etc.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

V. RESULT AND DISCUSSION

Figure 4: Text Recognition from Image

Figure 5: Speech Recognition System

IV.CONCLUSION

In our proposed system, we have implemented a system for the user who is language problems. So that the user interface is also user friendly sothat user can easily interact with this web application. This system eliminates the need for the user to use separate web applications to extract text from image, speech recognition and text translation. Therefore, the user's burden ofunderstanding the language of communication will be reduced automatically.

V. ACKNOWLEDGMENT

As the outset we offer our sincere thanks to our honorable guide Mrs Surekha Khot for her guidance and also encouraging us with her knowledge and experience for the development process of the project. We also value her eagerness and enthusiasm in encouraging us todevelop our technical and creative ideas, whichultimately led to success of our project. Our special thanks to faculty members of Information Technology Engineering Department for their great support and kind co-operation to provide us with whatever we require for our project.

REFERENCES

1. S. Patil1, M. Phonde, S. Prajapati, S. Rane, and A. Lahane, “Multilingual speech and text recognition and translation using image”, InternationalJournal of Engineering Research, vol. 5, no. 04, 2016.
2. Vemula, Vamshidhar Reddy. (2022). Integrating Zero Trust Architecture in DevOps Pipeline: Enhancing Security in Continuous Delivery Environments.
3. D. Ghai, D. Gera, and N. Jain, “A new approach to extract text from images based on dwt and k-means clustering, “International Journal of Computational Intelligence Systems,vol. 9, no. 5, pp. 900–916, 2016.
4. Trivedi, N. Pant, P. Shah, S. Sonik, and S. Agrawal, “Speech to text and text to speech recognition systems-areview”, IOSR Journal of ComputerEngineering, vol. 20, no. 2, p. 39, 2018.
5. PR Vaka, et al., "Anthem Health Insurance Breach or Ransomware Attacks," International Scientific

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204024 |

Journal of Contemporary Research in Engineering Science and Management, 2(1), pp. 41-49, 2017.

6. M. V. S. Rishita, M. A. Raju, and T. Harris, "Machine translation usingnatural language processing," in MATEC Web of Conferences, vol. 277, p. 02004, EDP Sciences, 2019.
7. J. R. Chaudhary and A. C. Patel, "Bilingual machine translation usingRNN based deep learning", Int J Sci Res Sci Eng Technol, vol. 4, no. 4, pp. 1480–1484, 2018.
8. Kartheek, Pamarthi (2022). Applications of Big Data Analytics for Large-Scale Wireless Networks. Journal of Artificial Intelligence, Machine Learning and Data Science 1 (1):920-926.[7]
9. A. B. Nassif, I. Shahin, I. Attili, M.Azzeh, and K. Shaalan, "Speech recognition using deep neural

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details