IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMAGE TEXT TO SPEECH CONVERSION IN DESIRED LANGUAGE

Dr. Prashantha H S[1],A.Akash[2],B.Jaidev[3],Girish.G[4],Jonna Dileep[5]

[1]Professor ,Department of Computer Science and Engineering, KS Institute of Technology Karnataka, India
[2345]Under-Graduate Student, Department of Computer Science and Engineering, KS Institute of Technology Karnataka, India

Abstract: The goal of this proposed work is to create an Android-based image text-to-speech (ITTS) application that enables users to translate text in photographs into spoken in formation in the language of their choice. The ability for users to customize the language in which the synthesized voice is produced is one of the application's standout features. Because of its user-friendly interface, a wide audience can access the Android application. Performance of an Android application is evaluated by precision, reactivity, and ability to customize language. This proposed workcan serve a variety of user demands, such as language learners, visually impaired people, and people looking for portable, effective tools for information consumption.

Keywords*:* Image, Text, Speech, Conversion ,Extraction, Image Processing

I. INTRODUCTION

The convergence of natural language processing and computer vision has produced novel technologies in recent years that have wide-ranging us es in assistive technology, accessibility, and education. In order to develop an image text-to-speech conversion system that accurately extracts text from photos and lets users choose the language they want for the synthesized speech output, the proposed research focuses on improving and expanding current approaches. Numerous mobile applications have been created to aid in reading or supporting individuals with visual impairments. If you've attempted to converse with someone who speaks a different language, you understand the significant challenge it can pose, even with the assistance of state-of-the-art technology. Translation sites where we need to pay lot sums of money to fulfill our task. The creation of image text-to-speech (ITTS) conversion systems, which make it possible to convert text found in images into spoken content, is one suchfield of study. This technical development has enormous potential to meet the demands of various user groups, such as language learners, those with visual impairments, and people interacting with textual content. In this Android-based image text-to-speech conversion proposed work is driven by the imperative to enhance accessibility and cater to diverse user needs. By seamlessly integrating computer vision and natural language processing on a widely-used mobile platform, the proposed work aims to provide a practical and customizable solution for individuals with visual impairments, language learners, and anyone seeking efficient ways to consume textual information in their preferred language.

Metrics such as accuracy in text extraction, responsiveness of the speech synthesis, and the effectiveness of language customization are systematically assessed. Additionally, user feedback and usability testing

contribute to refining and optimizing the application for real-world scenarios. Mobile Phones have become a main source of communication for this digitalized society. We have the capability to easily place calls and send text messages from one location to another. Verbal communication is recognized as the most effective means of delivering and understanding accurate information. In order to assist individuals more efficiently, text-to-speech (TTS) services were initially created to aid the visually impaired by providing a synthesized spoken voice to "read" text to the user. This proposed work will focus on text-to-speech conversion by utilizing Optical Character Recognition.

II.LITERATURESURVEY

Muhammad Ajmal; Farooq Ahmad; Martinez-Enriquez A.M.; Mudasser Naseer; Aslam Muhammad; Mohsin Ashraf ;Image to Multilingual Text Conversion for Literacy Education; 17-20 December 2018.Atthemoment,languageandvisualsworktogethertosupportliteracyinstruction,buttheyarealso vital to the texts we read. An application to translate text coupled with visuals for visual literacy is developed in this research project. Additionally, a thorough review of various methods for multilingual image-to-text translation is conducted. An improved methodology is proposed by filling in the gaps found by thorough examination of the literature. Consequently, there are four main stages involved in the construction of an application: capture, extraction, recognition and translation. Additionally, the Optical Character Recognition method is specifically utilized for high-accuracy character extraction and recognition in variety ofenvironmental settings. Simply taking an image with the user's smart phone's camera allows it to translate text, and the user can choose which language the translation shows in real time on their mobile device. The suggested method would be especially useful for teaching literacy, learning foreign languages, and possibly even serving as visitor’s aid.

H.Waruna H.Premachandra Information Communication Technology Center, Wayamba University of SriLanka, Makandura, SriLanka; Anuradha Jayakody; Hiroharu Kawanaka; Converting high resolution multi-lingual printed document images into editable text using image processing and artificial intelligence;12-13 March 2022. Information, mostly hand written or printed text on paper materials, is converted into a n editable electronic version via the optical character recognition process. The literature claims that few OCR systems are capable of accurately identifying multilingual characters, such as characters that combine English and Sinhala. The primary issue for this study is the absence of suitable technology to identify multilingual text, which is still a challenge that the scientific community as a whole needs to address. The major objective of this project is to create a bilingual character recognition system that can recognize printed S in hala and English scripts simultaneously using artificial neural networks and character image geometry properties. The plan is to enhance the solution to support the three most widely spoken languages in Sri Lanka, with Tamil being added as a later update. Artificial Neural Networks and character geometry features were the main technologies used in this investigation. With a database of over 800 images, separated into 46characters(20 Sinhala and 26English), and each character represented by20 different character images, about 85% of the success rate has been attained thus far. By removing individual character data from printed bilingual documents and sending it into the algorithm, researchers are experimenting with text recognition from printed documents.

Nikolaos Bourbakis; Image understanding for converting images into natural language text sentences; 21-23August2010.Only a summary form is provided. Knowledge discovery, document interpretation, human-computer interaction, and other fields of study greatly benefit from the effective processing, association, and comprehension of multimedia-based events or multi-modal information. The creation of a common platform for integrating many modalities (text, graphics, etc.) into one medium and linking them for effective processing and comprehension is a smart strategy for handling this crucial issue. Thus, this session describes the creation of a system that uses image processing-analysis techniques and graphs with attributes for object detection and picture understanding to automatically convert photos into natural language (NL) text sentences. Itthen transforms NL text sentences from graph representations. Additionally, it offers a process for converting Natural Language(NL) sentences into Graph representations, which are subsequently converted into descriptions using Stochastic Petri-nets (SPN). This provides a shared model for representing multimodal data and also allows for the association of "activities or changes "in image frames for the representation and interpretation of events. The reason the SPN graph model was chosen above other models is that it can effectively express structural and functional information in situations when other models cannot. Simple examples are given to demonstrate the idea that is being discussed here.

CongMa; Yaping Zhang; MeiTu; XuHan; Linghui Wu;YangZhao;YuZhou; Improving End-to-End Text Image Translation From the Auxiliary Text Translation Task; 21-25 August2022 Recent research has focused a great deal of emphasis on end-to-end text image translation (TIT), which attempts to translate the source language encoded in images to the target language. However, the performance of end-to-end text picture translation is limited by data sparsity. An on trivial solution to this issue is multi-task learning, which involves examining knowledge from related activities that are complimentary to one another. In this research, we offer a unique text translation augmented text picture translation method that uses text translation as an auxiliary job to train the end-to-end model. Through multi-task training and sharing of model parameters, our approach fully utilizes the readily accessible large-scale text parallel corpus. Our suggested approach surpasses current end-to-end methods, according to extensive experimental results, and joint multi-task learning with both text translation and recognition tasks produces better out-comes, demonstrating the complementarily of translation and recognition auxiliary tasks**.**

FaiWong; SamChao; WaiKitChan; YiPingLi; Recognition of Chinese character in snap shot translation system; 23-25November 2010 . We introduce Cyclops ,a mobile-based snap shot translation system, in this work. The technology converts an image containing Chinese text into Portuguese, English, or both languages based on the textual content of the image. The underlying principle of the design is to give users a thorough user interface for language translation to ols so they can understand the meaning of non-native content. The system was created using a variety of technologies, such as machine translation, optical character recognition in Chinese, and image processing. In this paper, we mainly describe the character recognition module that represents Chinese character attributes using Peripheral Direction Contributively (PDC).Mostnotably, it has been designed to function on popular mobile devices with storage and memory constraints.

SagarPatil; MayuriPhonde;SiddharthPrajapati;SarangaRane;AnitaLahane; Multilingual Speech and Text Recognition and Translation using Image; 04, April-2016 The aforementioned document outlines the efforts undertaken to identify the text within an image, which is either stored in the system or captured using a camera. This text is then translated into the required language and the translation result is displayed on the system's screen. This model uses the Tesseract OCR engine for extracting the text from the images. Further it splits the text into words and then it is search in the dictionary for translating the text from English to other languages. Finally a speech synthesizer is used for converting the above text to a speech format The VB.net Speech Software development kit is employed to compile the desired program or code module.. Therefore this application was built to automatically reduce the user task for understanding the languages for communication.

Karen Simonyan ; Andrea Vedaldi ; Andrew Zisserman; Max Jaderberg; Reading Text in the Wild with Convolutional Neural Networks; 4 Dec 2014In this study, we introduce a comprehensive system for text spotting, which involves localizing and recognizing text in natural scene images, as well as text-based image retrieval. The system relies on a region proposal mechanism for detection and deep convolutional neural networks for recognition. The automatic detection and recognition of text in natural images, known as text spotting, represents a significant challenge for visual comprehension. The use of region proposals circumvents the computational complexity associated with evaluating an expensive classifier using exhaustive multi-scale, multi-aspect-ratio sliding window searches.We use a combination of Edge Box proposals and a trained aggregate channel features detector to generate candidate word bounding boxes.

NileshJondhale ; Dr. Sudha Gupta; Reading text extracted from an image using OCR and android Text to Speech Volume 03 - Issue 04 || April 2018 || PP. 64-67Extensive research has been conducted in the field of Pattern Recognition, which falls within the domains of Machine Learning and Artificial Intelligence. OCR well known as Optical Character Recognition is one of the leading branch of the Pattern Recognition. Now-a-days Machine learning has become one of the peak of technology. Previously it was not possible to compute data at higher or faster rate, with the help of leading technology it is now possible to process data at higher rate to get optimized hence better result. Pattern recognition, a branch in machine learning is/can be helpful in many different ways. OCR technology is utilized for the high-accuracy recognition of characters. It involves using the camera of a handheld mobile device to capture an image of a printed or handwritten document in order to

extract the text from it. It's worth noting that there are billions of Android devices in operation on a global scale With the help of android device and android text to speech we can convert text into an effective & accurate speech optimally. Keywords: Android, Machine Learning, OCR, Text-to-Speech

SaiHarshith Thanneru1 ; Kajal Kumari1 ; Naresh Kunta1 ; Pavan Kumar Manchalla ;Image to audio, text to audio, text to speech, video to text conversion using,NLP techniquesOften, language bias between communicators can create communication problems. This article discusses a prototype that addresses this issue by enabling users to hear the content of text images. This process entails extracting the text from an image and converting it into speech in the user's chosen language. Moreover, the device can be utilized by individuals with visual impairments. Overall, this device helps users to listen to the content of images being presented. The suggested system allows the user to take a picture, which is then scanned and analysed by the application to read the English text. The acquired information is subsequently transformed into speech, allowing visually impaired individuals to comprehend the text's content. The output is presented in speech format to grant access to the information contained within the document. Natural Language Processing techniques are employed by the system to enhance accuracy and performance.

K. LAKSHMI ; Mr. T. CHANDRA SEKHAR RAO; Design And Implementation Of Text To Speech Conversion Using Raspberry PIThe most fundamental and commonly employed method is Braille. In addition to Braille, other technologies such as Talking Computer Terminals, Computer Driven Braille Printers, Paperless Braille Machines, and Optacon are also utilized in this context. These technologies use different techniques and methods allowing the person to read or convert document to Braille. This passage outlines the advancements in technology for facilitating interactions between computers and individuals with visual impairments. It describes the use of synthesized voice to read content, devices that scan and provide access to documents through tactile interfaces such as Braille or vibrating pegs, and the development of phone applications to aid the visually impaired. Additionally, it introduces a system utilizing Optical Character Recognition (OCR) and Text-to-Speech Synthesizer (TTS) in Raspberry Pi, enabling effective vocal interaction with computers. The system's purpose is to extract text from color images and convert it to voice using OCR technology. It further discusses the device's design, implementation, and experimental results, featuring two key modules: image processing and voice processing, all built on a Raspberry Pi v2 platform with a 900 MHz processor.

M Vaishnavi ; HR DhanushDatta ; VarshaVemuri ; L JahnaviLanguage ; Translator Application ;July2022The development of an android language converter app aims to address the longstanding challenge of language barriers hindering effective information communication. This app seeks to provide an efficient solution for language translation, improving learning processes, and enabling stress-free communication. Additionally, the system is designed to assess language translations to ensure their suitability for everyday conversation, offering the potential to enhance communication across language differences. To develop an android application for language translation that facilitates the user to understand unknown languages.

Sharvari S ; Usha A ; Karthik P ; Mohan Babu C ; Text to Speech Conversion using Optical Character Recognition ;Volume: 07 Issue: 07 | July 2020The increasing digitization of the world has led to the prevalence of phone calls, emails, and text messages as primary modes of communication. To enable effective and efficient message conveyance, various applications have emerged to act as mediators, facilitating the transmission of text to speech signals across vast networks. This project focuses on addressing the challenges faced by individuals with visual impairments and illiteracy. The proposed device aims to convert hard copies of text into speech, providing a solution to these hurdles. Many of these applications utilize functions such as articulators, text-to-speech signal conversion, and language translation. The project will employ different techniques and algorithms to realize the concept of Text to Speech (TST).

ShrutiMankar ; Nikita Khairnar ; MrunaliPandav ; Hitesh Kotecha ; Text-To-Speech Systems Adaptive technologies such as text-to-speech (TTS) have been developed to assist individuals with reading difficulties, illiteracy, and visual impairments. TTS, also known as "read-aloud" technology, converts digital text into audio, making it particularly beneficial for those facing reading challenges. Extensive research has been and continues to be conducted on text-to-speech technology, leading to the proposal and implementation of various

approaches and solutions. This research includes a systematic review of methods employed by active researchers in the field, encompassing technologies, methodologies, and algorithms such as machine learning, neural networks, and optical character recognition.

Augmentative Communication Support For The Vocally Impaired Using Nepali Text-To-Speech TribhuvanUniversiy Institute Of Engineering Pulchowk Campus Department Of Electronics And Computer EngineeringThe year 2016 saw over 147,000 individuals in Nepal facing speech or hearing impairments, highlighting the pressing need for effective communication solutions. Furthermore, a notable shortage of dependable Text-to-Speech (TTS) engines specific to the Nepali language has been observed. In response to these challenges, the Aawaj mobile application has been developed with a specific focus on providing augmentative communication support for the vocally impaired population in Nepal, featuring a dedicated Nepali TTS engine. This initiative aims to significantly enhance communication accessibility and inclusivity for individuals with speech or hearing impairments within the Nepali community. BIt utilizes vocal features such as timbre, prosody, rhythm, etc., to create a natural-sounding TTS engine, based on the open-source Tacotron2 TTS architecture published by Google. Rare conditions such as cerebral palsy, spinal cord injury, muscular dystrophy, and amyotrophic lateral sclerosis (ALS) have also led to a physical impediment in speech generation for a large population. This report further proposes an Augmentative and Alternative Communication (AAC) platform using accessibility features such as text prompt generation that provides accessibility to the intended populace of this mobile application.

CHANDRAKANT PatkarBharatiVidyapeeth's College of Engineering Lavale Pune SuhasPatilBharatiVidyapeeth Deemed University Prasad PeddiJagdishprasadJhabarmalTibrewala University ;Translation of English to AhiraniLanguageThe process initiates with the conversion of the image to grayscale, catering to the requirements of numerous OpenCV functions. Subsequently, noise reduction is accomplished via a bilateral filter. Canny edge detection is then employed on the grayscale image, enhancing contour detection. Warp and cropping operations are executed based on the identified contours, facilitating the extraction of the text-containing region and the elimination of irrelevant background elements. Finally, thresholding is applied to produce an image resembling a scanned document. This is done to allow the OCR to efficiently convert the image to text.

JayasakthiVelmurugan;M ; A. Dorairangaswamy ; Tamil Character Recognition Using Android Mobile Phone; 3 Febraury 2018This project provides an accurate and robust method for detecting Tamil texts in natural scene pictures. In this project a fast and effective pruning algorithm is designed to extract Maximally Stable Extreme Regions (MSERs) as character candidates using the strategy of minimizing regularized variations. Character candidates are merged into text candidates by the single link clustering algorithm, where distance weights and clustering threshold are learned automatically by a novel self-training distance metric learning algorithm. The above project has a precision of 95.32 and could be effectively used for the conversion of Tamil text into English Text message.

Mr. SumitChafale; Ms. PriyankaDighore ;Ms.DipikaPanditpawar ; Mr. KhushalBhagawatkar ;Mr. ShrikantSakhare ; Text to Voice Conversion for Visually Impaired Person by using Camera; June 2020This survey paper consists the work done for text to voice conversion for visually impaired person by using camera.In this project initially the captured colour image is converted into a grey scale image using Open CV functions. Tesseract OCR is used on the pre-processed image to convert it from .png form to a .txt file. Finally the Microsoft's speech synthesizer is used for the conversion of the text to an audio output.For the project to be portable for assisting the visually impaired, the entire application is based on MATLAB.

OlumideOlayinkaObe ; Akinwonmi A. E ; Smart Application For The Visually- Impaired; March-April 2021In this work, an application that would allow recognizing objects from images recorded by the camera of a mobile device is developed.This project uses android as the operating system and eclipse as the integrated development environment (IDE). The application development process incorporated the utilization of the Scale-Invariant Feature Transform (SIFT) to enhance its functionality. To further optimize performance, the Features from Accelerated Segment Test (FAST) algorithm, known for its high speed in corner detection, was integrated. Given that the algorithm was implemented on a smartphone, the OpenCV for Android SDK was

leveraged to facilitate this integration. The cascaded filters approach was used by SIFT to detect scale-invariant characteristic points, where the difference of Gaussians (DoG) was calculated on rescaled images progressively. A blob detector based on the Hessian matrix to find points of interest was used by SURF. To assess local variations around specific points, the application made use of the determinant of the Hessian matrix and selected points based on where this determinant was maximized. Additionally, the determinant of the Hessian was employed by SURF to determine scale. The application facilitated the auditory presentation of object recognition results to blind users by delivering pre-recorded messages. 97% accuracy was recorded in the performance of the system.

ReetaBandhu; Nikhil Kumar Singh; BetawarShashank Sanjay; Offline speech recognition on android device based on supervised learning**;**The Offline Android Smartphone Assistant functions as a virtual personal assistant designed to execute fundamental smartphone tasks using speech commands, even in offline mode. Its capabilities encompass opening apps, toggling Wi-Fi and Bluetooth, making calls, sending messages, adjusting brightness, and activating the flashlight. The application employs Natural Language Processing to interpret voice commands and carry out the specified tasks. Within the Android Studio environment, the android.speech.tts library is utilized for converting text to speech. This library incorporates the TextToSpeech class, enabling the synthesis of speech from text for immediate playback. Notably, the TextToSpeech class features a speak() method for converting text into spoken language. Furthermore, the application offers screen overlay functionality, enhancing its practicality and user experience..

Kuldip K. Paliwal ;Recognition of noisy speech using Dynamic spectral sub band centroids (2004) IEEE Volume11, No. 2.A procedure was proposed to construct the dynamic centroid feature vector that essentially embodies the transitional spectral information.It was demonstrated that in clean speech condition SSCs can produce performance comparable to that of MFCCs. Experiments were performed to compare SSCs with MFCCs for noisy speech recognition. The results showed that the centroids and the new dynamics SSC coefficients are more resilient to noise than the MFCC features.

OkpalaIzunn ;Text-to-Speech Synthesis (TTS) (2014) IJRIT, Volume 2, Issue 5**.** Text-to-Speech (TTS) synthesis is a technology designed to convert written text into spoken speech, offering an accessible means of conveying information for individuals with visual impairments or other reading challenges. The models run on JAVA platform and methodology used were object-oriented analysis and development methodology.With Text-to-Speech synthesis, one can medicate on the capabilities of same as like the handicapped individuals. Actually, in these models it's never been that easy to use Text-to-Speech synthesis at just one click and computer will speak text aloud in a clear and natural soothing voice.

Iain R. Murray ; John L. Arnott ; Norman ALM ; Alan F. Newell ;A Communication system for the disabled with emotional synthetic speech produced by rules. (1991) ICA Volume 1 A system for producing synthesis speech while incorporates vocal emotion effects has been developed. A range of common emotions can be simulated by the TTS system.The system which was made runs on a standard laptop PC, and was enable non vocal persons to express a range of emotions via a high-quality speech synthesizer. And also, conversational speech acts and speaking them with appropriate vocal emotion were developed.

DheeneshPubadi ; AyushBasandri ; Ahmed Mashat ; Ishan Gandhi ;A focus on codemixing and codeswitching in Tamil speech -to-text. (2020). IEEE 2020 8th International conference in Software Engineering Research and Innovation. May 20, 2020. The project aimed to develop an application that converts spoken Tamil language into text, serving to promote the usage and preservation of this classical language. Notably, the application was designed to convert spoken Tamil to text without utilizing autocorrection, thereby striving to accurately represent the spoken language. The research maintains that it is very much important to maintain the utilization of Tamil language via technology to help in preservation of one of the oldest surviving languages in the world from ancient times.The system is extendable to any other of the languages just by changing the language rules, intonations and the database. This research work also emphasized on the indigenous design considerations for such applications.

AyushiTrivedi ; Navya Pant ; Pinal Shah ; SupriyaAgrawal;Speech to text and text to speech recognition

systems (2018) IOSR, Volume 20, Issue 2.Most of the application find the use of function such as articulatory and acoustics based speech recognition, conversion from speech signals to text signals and from text to synthetic speech signals, language translation amongst various other. In this paper different techniques and algorithms were applied to achieve the mentioned function abilities.Hybrid machine translation is widely used due to its inoculation of advantages of both rule-based as well as statistical machine It makes sure that there is a creation of syntactically connected and grammatically correct text while also taking care of smoothness in a text, fast learning ability, data acquisitions which are a parts of SMT.

SunandaMendiratta ; Neelam Turk ; DipaliBansal ; A Robust Isolated Automatic Speech Recognition Systems by using Machine Learning. (August 2019) IJITEE ISSN: 2278-3075, Volume-8 Issue-10. The paper covered architecture of ASR that helps in getting ideas about basic stages of speech recognition system. Also, the techniques of machine learning are used in the model. And artificial neutral networks are also covered. The work is done by using the support of vector machines and artificial networks is also covered. The translation of spoken words into respective written scripts is done by speech recognition and language of speech is identified using Automated Speed Recognition (ASR) system. The work shows that traditional classifier results can be further improved by doing hybridization of it with other optimization algorithms.

III.RESEARCHGAP

The research gap in our proposed work could focus on improving existing image text-to-speech (ITTS) conversion systems, especially in the context of supporting multiple languages. Consider investigating:
Multilingual Support: Evaluate the current systems' effectiveness in handling diverse languages and explore ways to enhance accuracy and fluency across a broader linguistic spectrum. Low-Resource Languages: Investigate methods to extend image TTS capabilities to low-resource languages, addressing the challenges associated with limited linguistic data availability for certain languages. Adaptation to Image Complexity: Explore how well existing systems cope with varying levels of image complexity and investigate methods to improve performance on complex visual content.

IV. PROBLEM STATEMENT

The challenge addressed by this application is the hindrance posed by language barriers and limited accessibility for visually impaired individuals. These barriers impede effective cross-lingual communication and understanding, necessitating an innovative solution that provides seamless multilingual translations and audio support, thereby enhancing inclusivity and inter-language interactions.

V.OBJECTIVES

In order to improve speech recognition and enhance text-to-speech conversion, several measures were implemented. These included the creation of a binary image for image recognition through advanced image processing techniques. Additionally, efforts were made to strengthen the audio output, aiming to optimize sound quality. Furthermore, a concerted focus was placed on establishing a seamless connection between speech and text recognition, ensuring a cohesive and accurate conversion process.
Multilingual Support**:** Enable text-to-speech conversion for images with text in multiple languages. Support for various languages allows users to comprehend content in their preferred language.
Language Selection: Allow users to choose their desired language for text-to-speech conversion. Providing a range of language options enhances accessibility and user customization.
Speed and Efficiency: Optimize the conversion process to be fast and efficient, ensuring quick turnaround times for users. This is particularly important for real-time applications or scenarios where prompt conversion is required.
Integration with Accessibility Tools: Enable integration with accessibility tools and services, ensuring that the converted speech output is accessible to individuals with visual impairments.

VI.METHODOLOGY

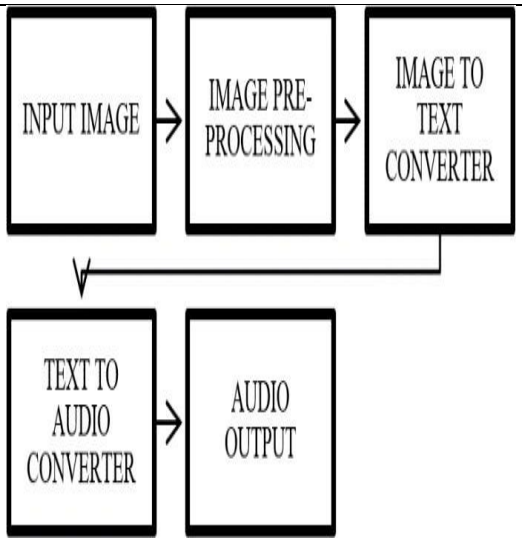

Figure 1

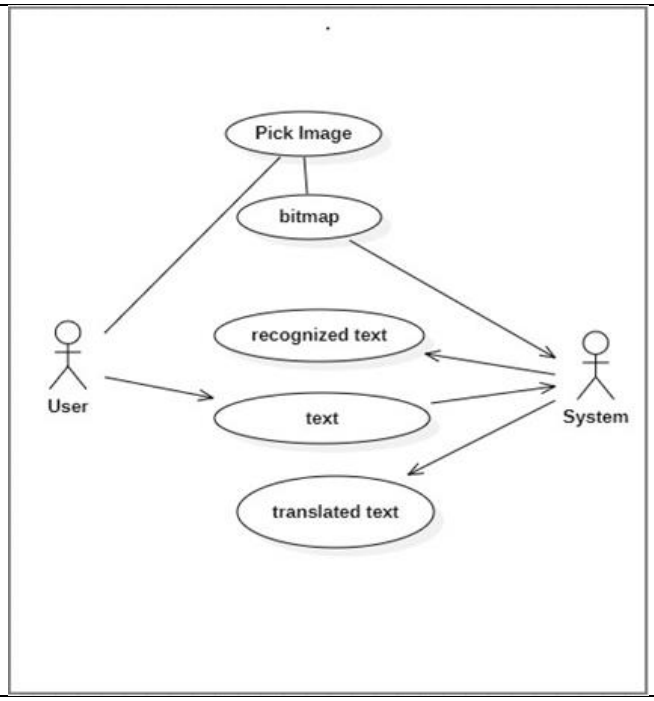

Figure 2

Optical character recognition (OCR) for text extraction from images and text-to-speech(ITTS) synthesis for turning the extracted text into spoken words are two crucial steps inthe methodology for image text-to-speech conversion in desired languages. This is a general process implementation methodology:
Select an API or OCR library: Choose an OCR library or API that can successfully extract text from images and supports a number of languages. OCR tools include Tesseract OCR, Google Cloud Vision API, and Microsoft Azure Computer Vision API.

Image Preprocessing: To improve the quality of text extraction, preprocess the input images. Techniques like resizing, noise reduction, and contrast adjustment might be used for this.

Text Extraction: To extract text from the preprocessed images, use the chosen OCR tool. Consider language support to guarantee precise identification.

Language Configuration: Set the TTS system to pronounce words correctly by using the language that has been detected or specified.

Image processing**:** Books and papers have letters The objective is to extract letters from an image and convert them into a digital format for subsequent recitation. Image processing techniques are utilized to achieve this, involving a series of functions applied to an image format to derive specific information from it. Initially, the image is loaded and converted into a grayscale format, representing the image as pixels within a specific range. This range is then used to discern the individual letters. In grayscale, the image predominantly consists of either white or black content, with white typically denoting spacing between words or blank areas.

Generate Speech: Give the TTS system the extracted text to produce the appropriate speech output. Make sure the TTS system you choose can produce natural- sounding speech and supports the language you want.
User Interface and Interaction: Design User Interface: Provide a user interface where users can choose which languages to use, upload images, and start the text-to-speech conversion process.

User Language Preferences: Give users the option to select the language they want to use for speech synthesis and text extraction.

Integration with Platforms: Include the text-to-speech feature for images in the User Preferences and Customization sections.
User Preferences: To improve the user experience, let users adjust speech preferences like voice pitch, speed, and volume.

Personalization: Create user profiles to store other customization options, such as language preferences, for a more individualized experience.

VII. CONCLUSION

In summary, this creative application combines easy translations, audio support, and image-based text extraction to overcome language barriers. It facilitates inclusive communication by providing a transformative answer for a range of linguistic requirements. As a language bridge, the application facilitates effective cross-lingual communication, which is a significant advancement in removing language barriers.

VIII. REFERENCS

[1]. Muhammad Ajmal; Farooq Ahmad; Martinez-Enriquez A.M.; MudasserNaseer; Aslam Muhammad; Mohsin Ashraf ;Image to Multilingual Text Conversion for Literacy Education ; 17-20 December 2018.

[2]. H.Waruna H. Premachandra Information Communication Technology Center, Wayamba University of Sri Lanka, Makandura, Sri Lanka ; AnuradhaJayakody; HiroharuKawanaka; Converting high resolution multi-lingual printed document images in to editable text using image processing and artificial intelligence; 12-13 March 2022.

[3]. NikolaosBourbakis ;Image understanding for converting images into natural language text sentences; 21-23 August 2010.

[4]. Cong Ma; Yaping Zhang; Mei Tu; Xu Han; Linghui Wu; Yang Zhao; Yu Zhou ;Improving End-to-End Text Image Translation From the Auxiliary Text Translation Task; 21-25 August 2022

[5]. Fai Wong; Sam Chao; Wai Kit Chan; Yi Ping Li; Recognition of Chinese character in snapshot translation system; 23-25 November 2010.

[6] Victor Fragoso, Steffen Gauglitz, Shane Zamora, Jim Kleban, Matthew Turk “TranslatAR: A Mobile Augmented Reality Translator”2010 IEEE.

[7] Ariffin Abdul Muthalib1, Anas Abdelsatar1, Mohammad Salameh1, Juhriyansyah Dalle2 “Making Learning Ubiquitous With Mobile Translator Using Optical Character Recognition (OCR)” 2011 ICACSIS.

[8] Shalin A. Chopra1, Amit A. Ghadge2, Onkar A. Padwal3, Karan S. Punjabi4, Prof. Gandhali S. Gurjar5 “ Optical Character Recognition” International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014.

[9] Hideharu Nakajima, Yoshihiro Matsuo, Masaaki Nagata, Kuniko Saito “Portable Translator Capable of Recognizing Characters onSignboard and Menu Captured by Built-in Camera” 2005 Association for Computational Linguistics/Proceedings of the ACL Interactive Poster and Demonstration Sessions, pages 61–64, Ann Arbor, June 2005..

[10] Nag, S., Ganguly, P. K., Roy, S., Jha, S., Bose, K., Jha, A., &Dasgupta, K. (2018). Offline Extraction of Indic Regional Language from Natural Scene Image using Text Segmentation and Deep Convolutional Sequence. arXiv preprint arXiv:1806.06208.

[11] Yang, C. S., & Yang, Y. H. (2017). Improved local binary pattern for real scene optical character

recognition. Pattern Recognition Letters, 100, 14-21.

[12] Phangtriastu, M. R., Harefa, J., &Tanoto, D. F. (2017). Comparison between neural network and support vector machine in optical character recognition.Procedia Computer Science, 116, 351-357.

[13] Naz S, Hayat K, Razzak MI, Anwar MW, Madani SA, Khan SU. The optical character recognition of Urdu-like cursive scripts.Pattern Recognition. 2014 Mar 1;47(3):1229-48.

[14]https://ieeexplore.ieee.org/document/7919526

[15]https://electronicsworkshops.com/2020/06/24/image-text-to-speech-conversion-using-optical-character-recognition-technique-in-raspberry-pi/