IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

OPTICAL CHARACTER RECOGNITION,TRANSLATION AND SPEECH GENERATION

[1]Garikipati.Chinmayeeswari , [2]Monduri.Sion kumari , [3]Gollu.Revathi ,
[4]Gottapu.Harsha priya , [5]Guntuku.Pravalika, [6]K N S Nrusimha Saraswati Lolla

[1]Student,[2]Project Guide,[3]Student,[4]Student,[5]Student,[6]Student,
[1]Department Of Computer Science and Systems Engineering,
[1]Andhra University College Of Engineering For Women, Visakhapatnam, India

Abstract: In today's digital environment, effective interpersonal interaction is vital, facilitated through various channels like text messages, emails, and phone calls. However, illiteracy and language limitations present significant obstacles, hindering smooth knowledge sharing and potentially leading to manipulation. Our approach aims to enhance accessibility, reduce misuse, and overcome linguistic barriers by converting text visuals into speech. By integrating text-to-speech (TTS) systems with optical character recognition (OCR) technology, we seek to create a versatile translation system proficient in efficiently translating text across languages, promoting linguistic inclusivity and intercultural communication.

Index Terms **- Optical Character Recognition, Language Translation, Text-to-Speech, Multifaceted Approach, Cross-Cultural Communication, Accessibility, Assistive Technologies.**

I. INTRODUCTION

Optical character recognition (OCR) automates text extraction from images and documents, streamlining data digitization and analysis without manual entry errors. OCR engines, developed over decades, transcribe typed, handwritten, and signage texts.This project an end-to-end OCR pipeline. It inputs images, extracts text using OCR, translates it into various languages, and converts it to speech. The system unlocks image text, offering multilingual spoken access. Applications range from aiding visually impaired users to automating document translation and audiobook creation. Combining OCR, neural machine translation, and text-to-speech synthesis in a Python application showcases its powerful capabilities.

II. LITERATURE SURVEY

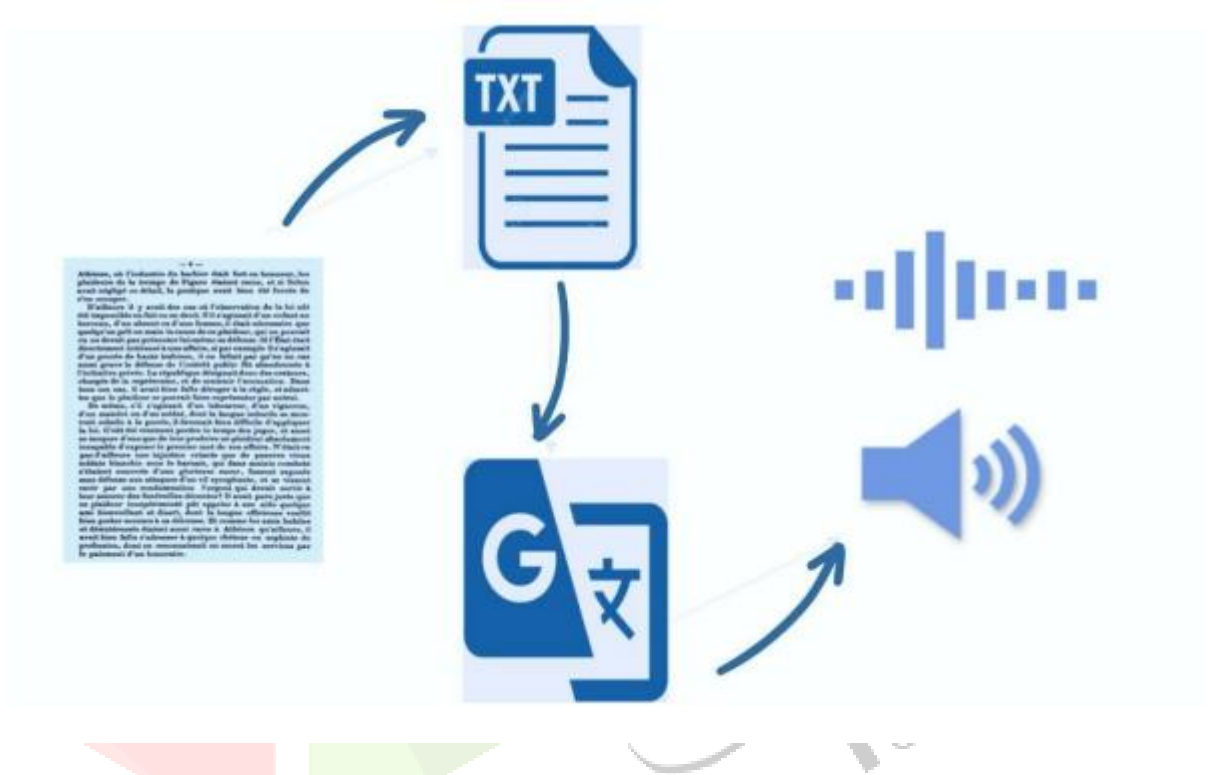
Exploring text image-to-speech conversion involves conducting a comprehensive literature review encompassing optical character recognition (OCR) and text-to-speech (TTS) synthesis, as well as their integration. The significance of OCR and TTS technologies lies in their pivotal roles in enhancing accessibility and facilitating information retrieval, particularly for aiding visually impaired individuals and improving document accessibility. Investigation into OCR techniques, spanning from traditional methods like template matching to contemporary deep learning approaches, reveals a wide array of methodologies for extracting text from images and documents. Analysis of OCR and TTS integration uncovers numerous research projects aimed at achieving seamless text image-to-speech conversion, with a keen emphasis on methodologies, architectures, and performance metrics. Real-world applications across various domains underscore the versatility of text image-to-speech converters, while insights into future directions highlight the ongoing evolution and potential impact of OCR-TTS integration in promoting accessibility and inclusivity

III. Proposed System

3.1 Methodology

The OCR translation and speech generation system functions by taking a text image as input and producing an audio file in the preferred language as output. This process comprises three primary steps: text extraction, translation, and speech generation.

Initially, the system extracts textual content from the input text image through optical character recognition (OCR) technology. Subsequently, the extracted text undergoes translation from its original language to the desired language. Finally, the translated text is synthesized into spoken audio using text-to-speech (TTS) technology, culminating in the creation of an audio file. There are three steps: Text Extraction, Translation and Speech Generation.



Phases of speech generation

3.1.1 Text Generation

The system initiates the process by utilizing the Optical Character Recognition (OCR) component, which relies on the capabilities of the Pytesseract library. This library acts as an interface to the Tesseract OCR engine, well-known for its accuracy and efficiency in text extraction from images. Pytesseract provides access to advanced OCR functionalities, ensuring precise and reliable extraction of textual content from diverse image sources. The Tesseract OCR engine employs sophisticated algorithms to analyze visual patterns within input images, accurately identifying and interpreting characters, words, and sentences. Consequently, text is extracted from the input images while preserving the original layout and formatting as closely as possible. Integrating Pytesseract into the system streamlines the text extraction process, facilitating seamless conversion of image-based text into a machine-readable format. This guarantees that subsequent steps such as translation and speech generation can operate on accurately extracted textual content, establishing a foundation for effective communication and information dissemination.

3.1.2 Translation

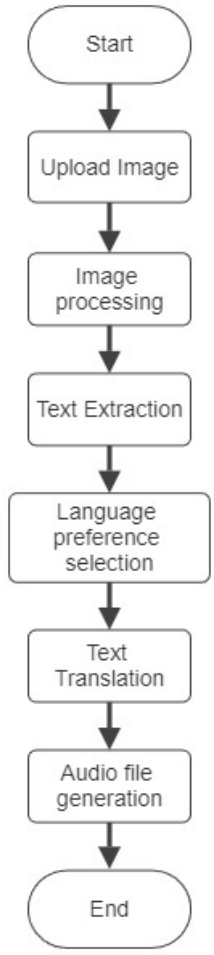
In the translation step, the system determines the user's preferred language from a list of available options. To achieve this, it utilizes the Google Translate API, a robust tool for language translation. Integration with the Google Translate API is facilitated through the googletrans(3.1.0a0) Python library, providing a convenient

interface for accessing the API's functionalities. By leveraging this library, the system seamlessly communicates with the Google Translate API, enabling precise translation of text between languages. The Google Translate API supports a wide range of languages and offers robust translation capabilities, ensuring accuracy and reliability.

3.1.3 Speech Generation

In the third step, the translated text is converted into an audio file, allowing users to hear the text in the translated language directly on the page. This functionality is achieved using the gTTS (Google Text-to-Speech) library, which harnesses Google's speech synthesis capabilities to produce natural-sounding audio output. By employing gTTS, the system seamlessly transforms the translated text into spoken audio, enhancing user experience and accessibility.

Moreover, alongside the integration of gTTS, Flask serves as the backbone for the web application, facilitating user-friendly interaction through an intuitive web interface. Flask enables smooth communication between the front-end interface and the backend functionality, ensuring the seamless integration of the audio generation process into the web application.



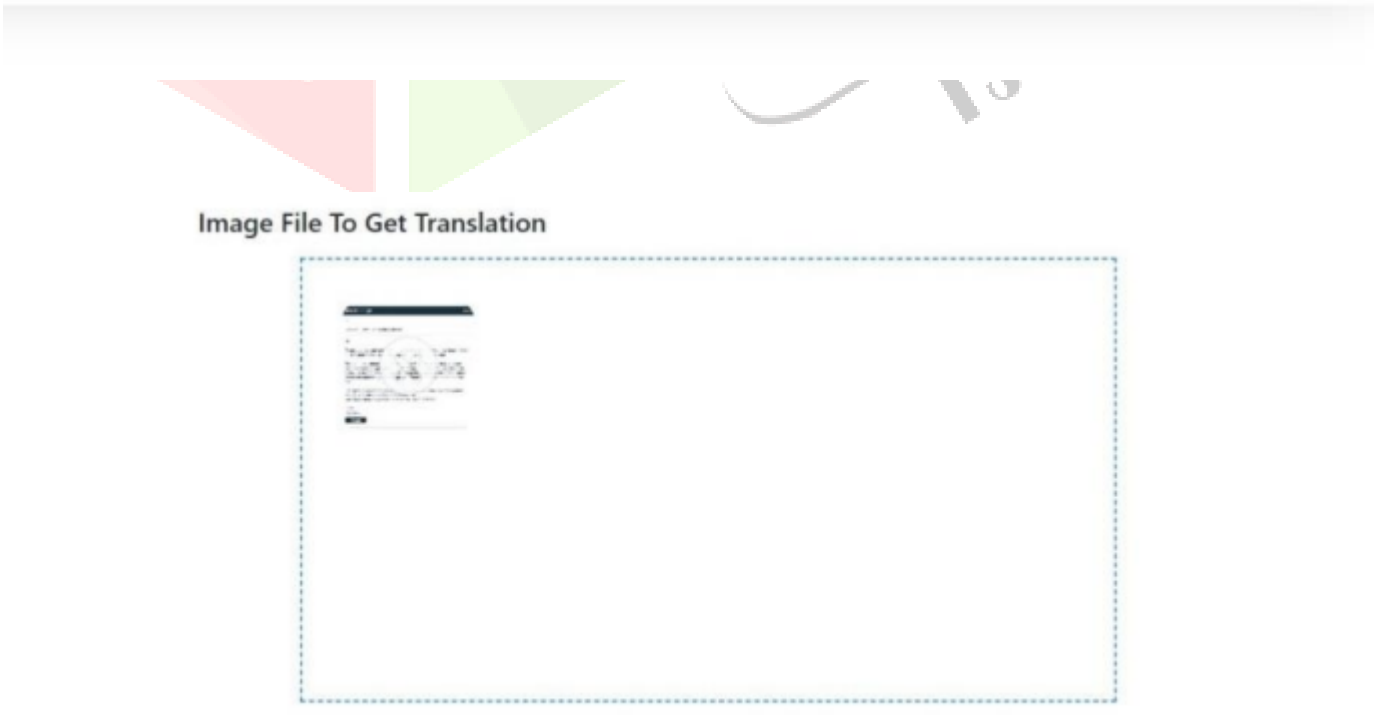
flow chart

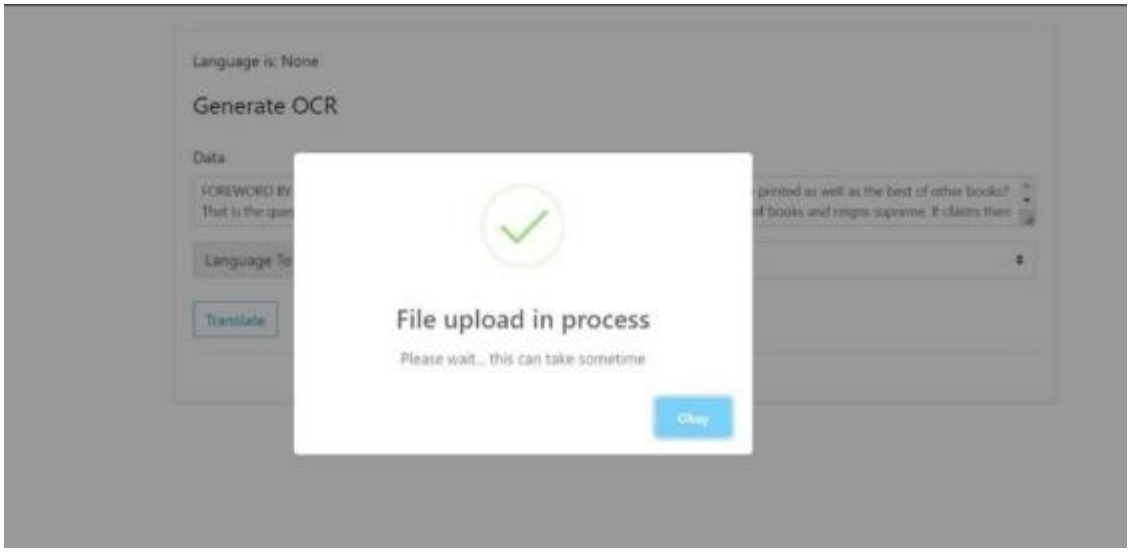
Furthermore, the project incorporates auxiliary libraries such as Pillow for image processing, NumPy for numerical operations, and OpenCV for computer vision tasks. These libraries enhance the system's capabilities in handling image input, preprocessing, and text extraction, contributing to the overall robustness and efficiency of the OCR translation and audio generation system.

Additionally, the deployment environment is meticulously managed within a virtual environment to ensure package compatibility and isolation. By deploying the system within a virtual environment, potential conflicts between dependencies are minimized, guaranteeing smooth execution and reliable performance of the application.

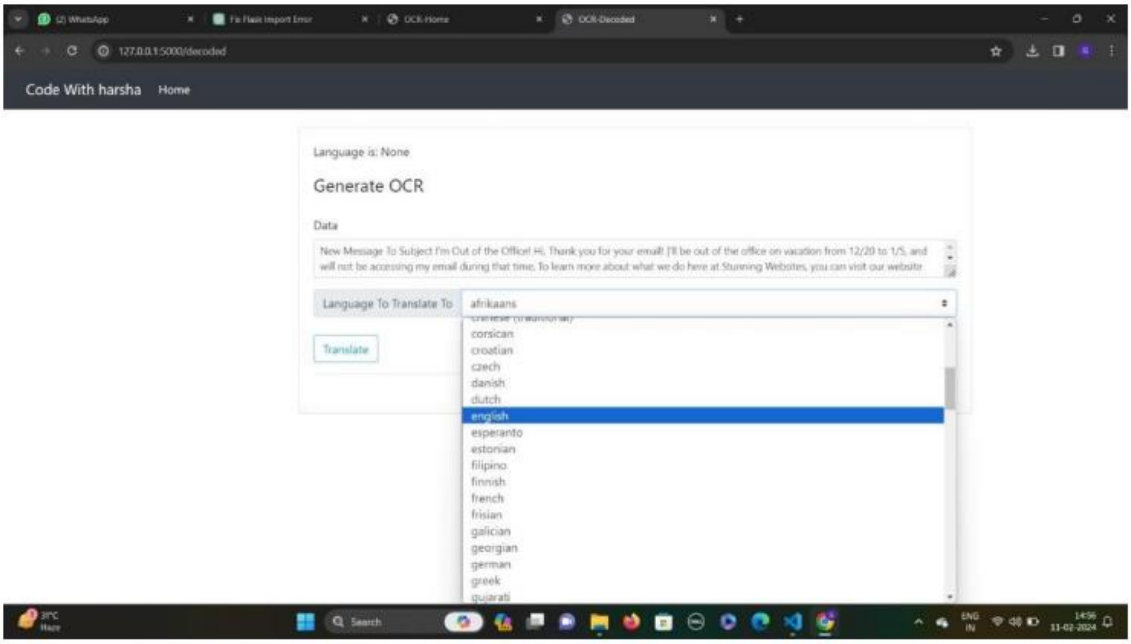
IV. Results and Discussion

4.1 Input




Language is: None
Generate OCR
Data
Translate
File upload in process
Please wait... this can take sometime
Okay

4.2 Output


Code With harsha
Home
Language is: None
Generate OCR
Data
Language To Translate To
afrikaans
Translate
corsican
croatian
czech
danish
dutch
english
esperanto
estonian
filipino
finnish
french
frisian
galician
georgian
german
greek
gujarati

V. Conclusion

This project established an end-to-end pipeline for processing images containing text, leveraging optical character recognition (OCR) to extract text, translating it into various languages, and synthesizing speech for audible output. Python facilitated seamless integration across image processing, OCR, translation, and speech synthesis modules, thanks to its extensive libraries. The project showcased the integration of cutting-edge OCR, NLP, and speech synthesis technologies to automate the extraction and reading of textual information. Applications range from digitizing books to aiding visually impaired users and streamlining document workflows. It provided valuable hands-on experience in Python, OCR, machine translation, and text-to-speech synthesis, crucial for future technologies. However, there's room for improvement such as enhancing OCR models, expanding language support, utilizing cloud APIs for scalability, and developing mobile/web interfaces. These enhancements can make the system more versatile, robust, and applicable in real-world scenarios, laying a strong foundation for applied OCR solutions.

References

[1] Chinmaya misra,P K Swain,Jibendu Kumar Manthri. "Text Extraction and Recognition from Image using Neural Network." International Journal of Computer Applications (0975 – 8887) Volume 40– No.2, February 2012 13

[2] .Ranjan Jana,Amrita Roy Chowdhury and Mazharul Islam. "Optical Character Recognition from Text Image." International Journal of Computer Applications Technology and Research Volume 3– Issue 4, 239 - 243, 2014, ISSN: 2319–8656

[3] V. Wu, R. Manmatha and E. M. Riseman, " Finding Text in Images", In Proc. of Second ACM International Conference on Digital Libraries, Philadelphia, PA, pp. 23-26, 1997.

[4] C. Garcia and X. Apostolidis, "Text Detection and Segmentation in Complex Color Images", In Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP2000), Istanbul, Vol. 4, pp. 2326- 2330, 2000

[5] A Unified Framework for Multi-oriented Text Detection and Recognition‖ IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 11, NOVEMBER 2014 4737.

[6] Teofilo E.,de Campos,&Bodla Rakesh Babu. (2002). Optical Character Recognition (OCR) Technology. In: IIIT Hyderabad Conference**.**

[7] ARCHANA A. SHINDE, D. 2012.Text Pre-processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology, pp. 810-812.

[8] SMITH, R. 2007. An Overview of the Tesseract OCR Engine. In proceedings of Document analysis and Recognition.. ICDAR 2007. IEEE Ninth International Conference.