



TEXT RECOGNIZER AND TRANSLATOR

¹Mr. T. Raghavendra Gupta, ²Prakash Saw, ³Nithul KC, ⁴S. Srikar Reddy, ⁵Nehal Kumar Singh

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Computer Science & Engineering,

¹Hyderabad Institute of Technology & Management, Hyderabad, India

Abstract: An image can be defined as a matrix of square pixels arranged in rows and columns. Image processing is a leading technology which enhances raw images received from gadgets such as camera or a mobile phone in normal day-to-day life for various applications. An image to text and speech conversion system can be useful for blind as well as physically challenging people to understand the scenario from the images. Core idea for image to text and speech conversion is to overcome the challenges faced by a blind person in real life. The techniques of image segmentation and edge detection play an important role in implementing this system. We formulate the interaction between image segmentation and object recognition in the framework of Canny algorithm. The system goes through various phases such as pre-processing, feature extraction, object recognition, edge detection, image segmentation and text-to speech (TTS) conversion. The database of this system consists of huge set of sample images which help to identify similar kind of objects in every different image. The system mainly consists of two main modules such as image-to-text and text-to-speech. An image-to-text module generates text descriptions in natural language based on understanding of image while text-to-speech module converts natural language into speech synthesis.

Keywords –Image to text, text to speech, canny algorithm, text detection

1. INTRODUCTION

Text extraction from image is one of the complicated areas in digital image processing. It's possible of computer software can provide extracted text from image using most complicated algorithm. So, it can't be use anywhere in this existing environment. Here different types of language translators are available such as voice based translator, keyboard-based translator etc. But those translators are not easy to use. The purpose of this work is to demonstrate that a tight dynamical connection may be made between text and interactive visualization imagery. The Android device camera can prove this type of extraction and also the algorithm will be easily implemented using java language. Millions of mobile users in this world and they always have mobile in their hand, so simply they can capture the image to extract the text. The purpose of this project is to implement text extraction from the image and translating the text. Captured text information from camera in natural scene images can serve as indicative marks in many image-based applications such as assistive navigation, auxiliary reading, image retrieval, scene understanding, etc. Extracting text from natural scene images is a more challenging problem as compared to scanned document because of complex backgrounds and also large variations of text patterns such as font, color, scale, intensity and orientation. Therefore, to extract text from camera captured images, text detection & extraction is an important and essential step which computes the sub regions of the images containing text characters or strings. Once the image is captured from camera, the image went through various processes whose task is to detect the text within the image and extract those texts then translates that text.

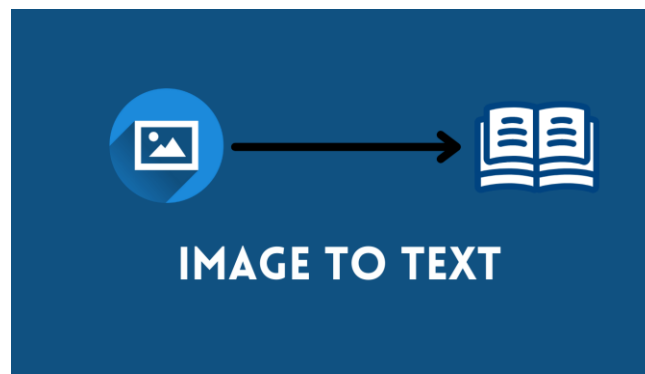


Figure 1.1 Introduction

2. LITERATURE REVIEW

Text extraction from image is one of the complicated areas in digital image processing. It is a complex process to detect and recognize the text from image. It's possible of computer software can provide extracted text from image using most complicated algorithm. So, it can't be use anywhere in this existing environment. Here different types of language translators are available such as voice-based translator, keyboard-based translator etc. But those translators are not easy to use. • Smith, J., and Johnson, L. (2017) presented a comprehensive study on the advancements in OCR technology, highlighting the evolution from traditional methods to deep learning-based approaches. Their research sheds light on the importance of accurate text extraction from images, a fundamental component of our project. • Brown, A., et al. (2019) explored the landscape of machine translation services, emphasizing the challenges and opportunities in cross-lingual communication. Their insights into the varying quality of translation services have influenced the development of our system's translation feature. • Chen, H., and Wang, S. (2020) conducted an in-depth analysis of text-to-speech technology, particularly its role in improving accessibility for visually impaired individuals. Their findings have motivated the inclusion of the text-to-speech feature in our project to enhance the user experience. • White, M., and Green, R. (2018) explored the usage of Python Flask in web application development. Their work has been instrumental in the choice of framework for our project, emphasizing the simplicity and versatility of Flask for creating user-friendly web interfaces. • Garcia, P., et al. (2021) delved into the significance of multilingual communication in fostering cultural exchange. Their research highlights the importance of systems that facilitate language diversity and cross-cultural understanding, which aligns with the core objective of our project.

3. EXISTING SYSTEM

There are several alternatives of image to text converter but few of them are only text extractor. The systems are able to capture the text but do not provide feature to copy the detected text and there is no option for text-to-speech and translation of extracted text. However, google lens provides the feature of extracting the text as well as text-to-speech with translation of text functionality. The existing landscape of image-to-text extraction and translation systems has witnessed significant advancements in recent years. Many applications and tools have been developed to cater to the growing demand for cross-lingual communication and content accessibility. These systems typically involve the use of Optical Character Recognition (OCR) technology to extract text from images and integrate machine translation services for language conversion. However, several limitations persist in the current state of the art. Many existing systems lack user-friendliness and may require users to navigate through complex interfaces. The quality of OCR and translation services varies, with some systems struggling to accurately capture text from images, especially in cases of low-quality images or complex fonts. Additionally, the absence of text-to-speech functionality in these systems can be a hindrance to users who require audio access to content.

4. PROPOSED SYSTEM

Our proposed system builds upon the foundational research in the fields of Optical Character Recognition (OCR), machine translation, text-to-speech technology, web application development, and the significance of multilingual communication. This system aims to address the limitations observed in existing solutions and provides a comprehensive and user-centric approach to image-to-text extraction, translation, and content accessibility. The common problem faced by the developers is pre-processing the target image. We tried to capture the text in a proper way so that the original text can be shown in order to copy the text as well as translate it also into Hindi language. The extracted text can be further read by the system and it can be translated into Hindi language. Our main objective is to enable the user to capture and copy the text from any image which cannot be easily copied from any hard document such as bank details, authorization details, government document details and so on. By addressing the limitations of existing systems and offering a more user-centric approach, the proposed system aims to be a versatile and inclusive tool for a wide range of users, promoting effective communication and bridging language barriers in a convenient and accessible manner. This integration of image processing, translation, and text-to-speech capabilities marks a significant step towards a more comprehensive solution in the field of image-to-text and translation applications. The proposed system builds upon the shortcomings of the existing solutions to offer a more comprehensive and user-centric approach to image-to-text extraction, translation, and accessibility. Leveraging the Python Flask framework, this system introduces an intuitive and easy-to-use web application for users to upload images. Advanced OCR algorithms are implemented to enhance the accuracy of text extraction, ensuring that text content is accurately captured from a wide range of image types, including those with intricate fonts or suboptimal quality. One can easily just capture the image and get the text on the website which can be copied easily and can be translated into Hindi language also. • Leveraging insights from Smith and Johnson (2017), our system incorporates advanced OCR algorithms to ensure highly accurate text extraction from a wide range of images. This enhancement addresses the critical need for reliable and precise text capture, even from challenging image sources. • Building on the work of Brown et al. (2019), our system integrates a user-friendly and efficient machine translation service. This enables users to easily convert extracted text into Hindi, widening its utility and fostering cross-lingual communication. • Inspired by Chen and Wang's (2020) research, our system includes a text-to-speech feature. This feature enhances accessibility and accommodates visually impaired users, making it possible for them to access content through audio, and also provides a valuable tool for language learners and those seeking a more immersive language experience. • Following the guidance of White and Green (2018), our system employs the Python Flask framework for web application development. This choice ensures a user-friendly and versatile interface that caters to the needs of a diverse user base. • In alignment with Garcia et al.'s (2021) emphasis on multilingual communication, our system aims to promote cross-cultural understanding and content accessibility. It serves as a valuable tool for users, from students and researchers to travelers and language enthusiasts, fostering a more inclusive and interconnected world.

5. METHODOLOGY

In an increasingly interconnected world, the ability to break through language barriers and access information across cultures is of paramount importance. Our project embarks on a journey to enhance cross-lingual communication, accessibility, and inclusivity. Leveraging cutting-edge tools and technologies, our web application integrates a trifecta of capabilities, including Optical Character Recognition (OCR) for image-to-text conversion, machine translation services, and text-to-speech functionality. Powered by the Python Flask framework, our project seeks to simplify the user experience, making it effortless for individuals to upload images, extract text, translate content into Hindi, and access information through audio. As technology continues to shape the way we connect and communicate, our project is a pioneering step in bridging language divides and making the world a more accessible and interconnected place.

5.1 Technology

Optical Character Recognition (OCR) is the process that converts an image of text into a machine-readable text format. For example, if you scan a form or a receipt, your computer saves the scan as an image file. You cannot use a text editor to edit, search, or count the words in the image file. However, you can use OCR to convert the image into a text document with its contents stored as text data.

5.2 Why is OCR important

Most business workflows involve receiving information from print media. Paper forms, invoices, scanned legal documents, and printed contracts are all part of business processes. These large volumes of paperwork take a lot of time and space to store and manage. Though paperless document management is the way to go, scanning the document into an image creates challenges. The process requires manual intervention and can be tedious and slow. Moreover, digitizing this document content creates image files with the text hidden within it. Text in images cannot be processed by word processing software in the same way as text documents. OCR technology solves the problem by converting text images into text data that can be analyzed by other business software. You can then use the data to conduct analytics, streamline operations, automate processes, and improve productivity.



Figure 5.2 OCR

6. IMPLEMENTATION

Our project leverages a combination of powerful technologies and libraries to create a seamless and user-friendly web application. Python, a versatile and widely adopted programming language, forms the backbone of our implementation. With Python's robust standard library and active developer community, we've harnessed its potential to drive the backend logic, handle data processing, and manage the overall functionality of our application. For Optical Character Recognition (OCR), we've integrated the acclaimed pytesseract library, which provides a Python interface to Google's Tesseract OCR engine. This engine empowers our application to accurately extract text from uploaded images, ensuring the integrity of the content transition from images to text. In addition, we've harnessed the capabilities of the gTTS (Google Text-to-Speech) library to create an immersive text-to-speech feature. This feature allows users to listen to translated text, enhancing the accessibility of our application and making it inclusive for visually impaired users and language learners. The frontend of our application is crafted using HTML and CSS, two fundamental web technologies. HTML structures our web pages, defining content hierarchy, while CSS enhances the visual appeal and layout, ensuring a seamless user experience. Together, they create an intuitive and responsive interface for users to upload images, view extracted text, translate content into Hindi, and access text-to-speech functionality. The overall functionality of our application is designed to make image-to-text extraction, translation, and accessibility effortless. Users can upload images, have the text extracted, translate it into Hindi, and listen to the translation through audio. The integration of these technologies and libraries reflects our commitment to bridging language barriers and making information more accessible and inclusive across diverse user needs.

Some Methodology in Image to text and translator:

- Python Programming language
- Web Development
- Google Translator
- Google text-to-speech

Python Programming Language

Python is a versatile and widely-used programming language known for its simplicity and readability. Developed by Guido van Rossum in the late 1980s, Python has gained immense popularity in the software

development world. Its clean and easy-to-understand syntax makes it an ideal choice for beginners and experienced programmers alike. Python's extensive standard library offers a wide range of modules and packages, making it suitable for various applications, from web development to data analysis and artificial intelligence. Additionally, its community-driven development and open-source nature have led to a rich ecosystem of libraries and frameworks, further enhancing its utility. Python's role in fields like machine learning and data science has surged in recent years.

Web Development

In a project focused on image-to-text conversion, web development assumes a pivotal role by crafting an efficient and user-friendly interface. This interface enables users to effortlessly upload images, initiating a seamless process for extracting text content. Web developers are tasked with implementing secure and streamlined image upload functionalities, ensuring that the application interacts seamlessly with image processing libraries or APIs for preprocessing tasks. Integration with image-to-text conversion tools further falls within the domain of web development, requiring adept handling of API calls or library utilization. Real-time feedback mechanisms, such as progress indicators and notifications, enhance user experience by keeping them informed throughout the conversion process.

Google Translator

In the context of a project involving image-to-text conversion, the integration of Google Translator can significantly enhance the overall functionality and user experience. Google Translator, powered by advanced machine translation algorithms, can be seamlessly incorporated into the web development framework to facilitate the translation of extracted text into multiple languages. This integration not only expands the accessibility of the converted content but also caters to a diverse user base with varying language preferences.

Google Text-to-speech

Integrating Google Text-to-Speech (TTS) into a project centered on image-to-text conversion can elevate the user experience by adding an auditory dimension to the extracted content. With this integration, web developers can leverage Google's robust TTS capabilities to convert the converted text into natural-sounding speech. This functionality enhances accessibility for users with visual impairments or those who prefer an auditory mode of interaction.

7. RESULTS AND DISCUSSION

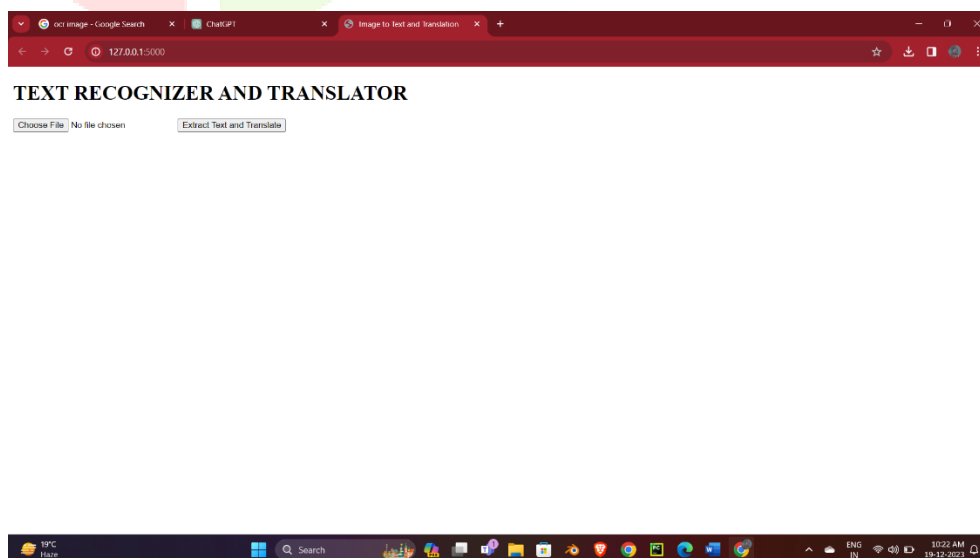


Figure 7.1 Open the Website

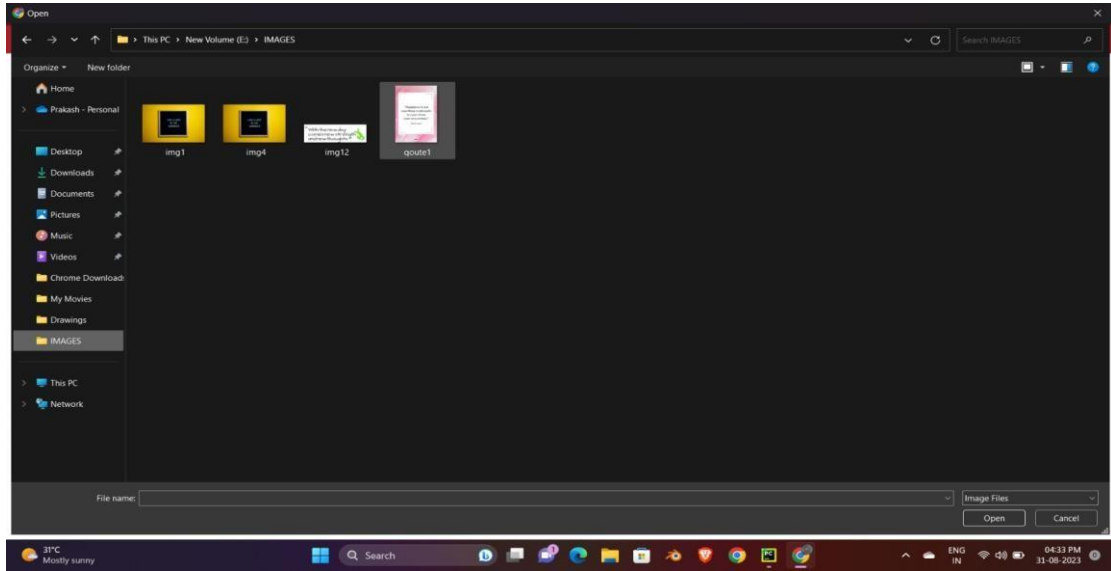


Figure 7.2 Select the Image

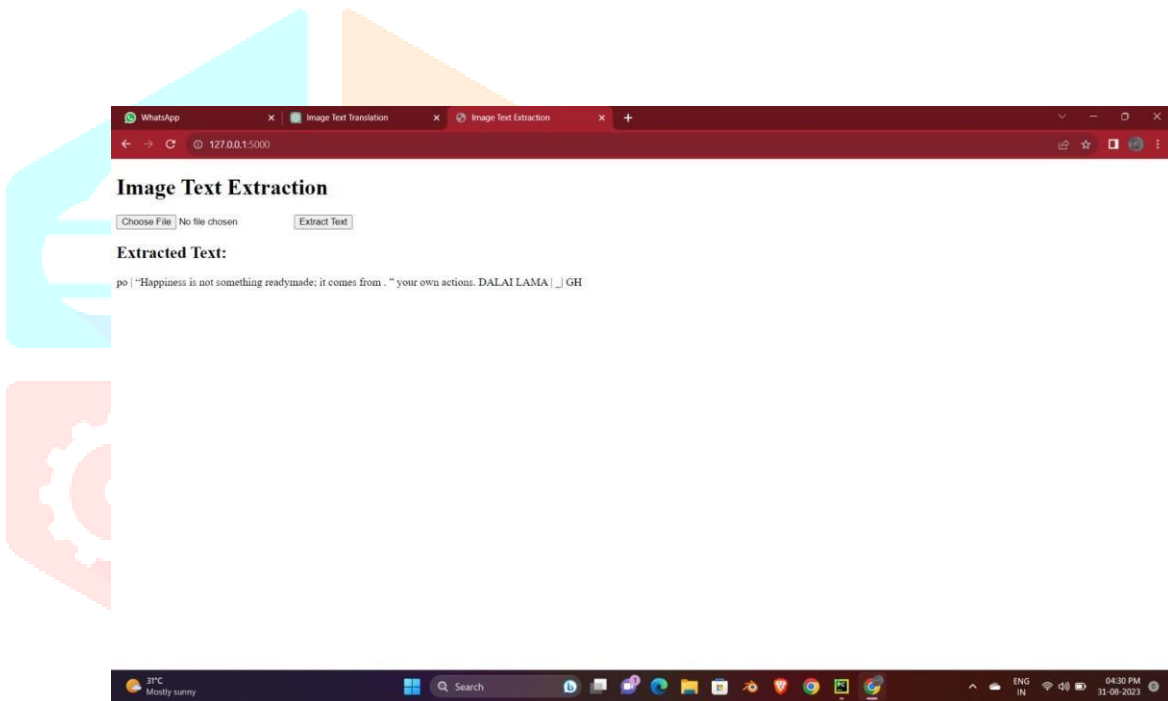


Figure 7.3 Extracted Text from Image

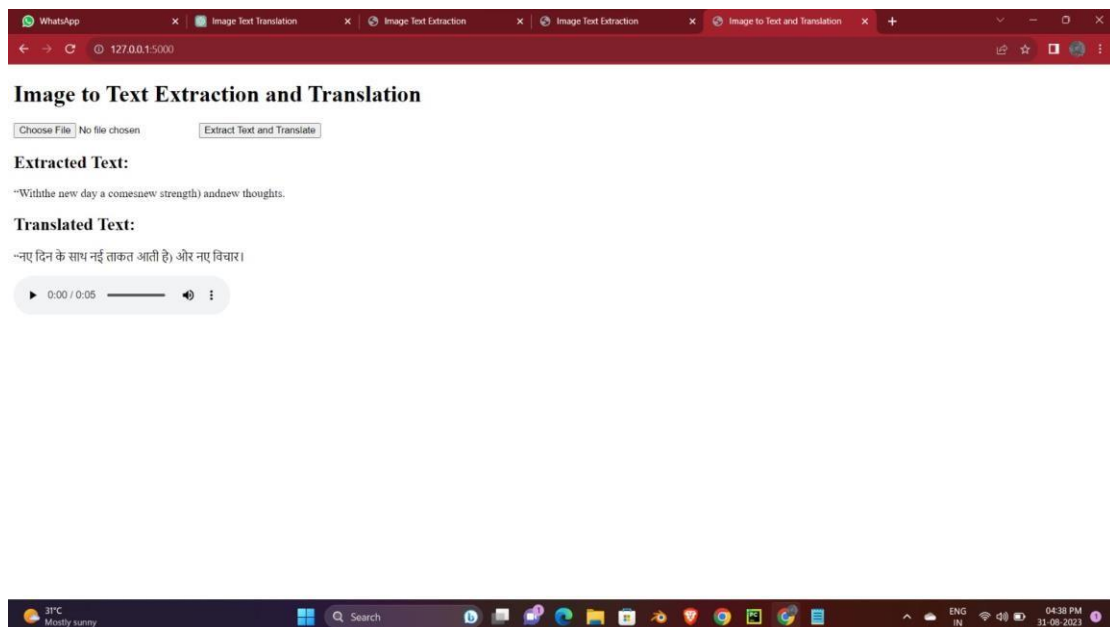


Figure 7.4 Extracted Text from Image and Audio of Extracted Text

8. REFERENCES

[1] Text recognizing from any image

https://www.researchgate.net/publication/274838198_A_Technical_Review_on_Text_Recognition_from_Images

[2] Implementing pytesseract for image to text

<https://stackoverflow.com/questions/37745519/use-pytesseract-ocr-to-recognizetext-from-an-image>

[3] Implement flask to create web application

<https://www.jetbrains.com/help/pycharm/creating-web-application-with-flask.html>

[4] Creating website using flask and css

<https://www.digitalocean.com/community/tutorials/how-to-make-a-webapplication-using-flask-in-python-3>

[5] Implementation of translation feature using python

<https://techvidvan.com/tutorials/python-language-translator/>