

**RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES-SRIKAKULAM**

**MINI PROJECT**

**Multilingual OCR and Speech Synthesis Platform**

**Under the Guidance of:**

Mrs.Seetha Durga mam

**Project Team:**

- 1.Appana Srinivas - S191041
- 2.Byreddy Sireesha - S190649
- 3.Basireddy Sreehitha - S190605

# ABSTRACT

Multilingual OCR and Speech Synthesis Platform is a task where it takes picture as input and extract text and translate into another language. Using advanced technology like Optical Character Recognition (OCR) and Natural Language Processing (NLP), our system accurately extracts and translates text, making it easier to understand. What sets our platform apart is that it provides both written and spoken translations. This makes it simple for people to access and understand text in many languages, whether by reading or listening. Our system represents a big step forward in making communication inclusive and accessible to all.



# INTRODUCTION

Multilingual OCR and Speech Synthesis Platform project uses OCR and NLP to turn text from images into written and spoken words. For students, it helps by translating and reading out information from books, notices, and other materials, making learning easier and more accessible in different languages.

For travelers, this system translates and vocalizes text from signs, menus, and other sources, helping them understand and navigate new places. It breaks language barriers, providing essential information in an easy-to-understand spoken form, making travel more convenient and enjoyable.



# GOAL OF THE PROJECT

Despite technological advancements, accessing text-based information remains challenging for travelers and students. Traditional OCR systems can recognize text but often can't convert it into both written and spoken forms, limiting their usefulness. This gap restricts students' language learning and makes it difficult for travelers to understand important texts like signs and menus.

To solve these issues, we need an advanced system that combines OCR and Natural Language Processing (NLP). This system would recognize text in images and convert it into written and spoken formats. For students, this means easier access to educational materials and better support for learning new languages. For travelers, it means seamless translation and vocalization of text, helping them navigate and understand information in foreign languages more easily, enhancing their travel experience.

# LITERATURE REVIEW

Old methods for text extraction:

## 1.GOCR:

**Overview:** Early open-source OCR software.

**Weaknesses:** Lower accuracy, fewer languages, outdated algorithms.

**References:** "GOOCR - A Free Optical Character Recognition Program." - GOCR Official Website.

## 2.OCRopus:

**Overview:** Open-source, modular OCR system with machine learning focus.

**Weaknesses:** Complex setup, less stable.

**References:** Breuel, T. M. "The OCROpus Open Source OCR System." (Document Recognition and Retrieval XV Conference, 2008). - ResearchGate website

# Old methods for language translation:

## 1. SYSTRAN:

SYSTRAN is a pioneering machine translation system developed in the 1960s, utilizing rule-based and statistical methods.

**Limitations:** Less accuracy, Fewer language Support

**Reference:** Hutchins, W. J. "Machine Translation: A Concise History." Oxford University Press, 2003.

## 2. IBM ViaVoice:

IBM ViaVoice was a popular text-to-speech (TTS) system developed in the late 1990s and early 2000s. It converted text into synthetic speech.

**Limitations:** Lack of Naturalness, Limited availability of languages and voices

**Reference:** P. M. Bagshaw, "The Development of IBM ViaVoice: A Brief Overview," IBM Technical Report, 1999

# **New method for text extraction:**

Tesseract is a powerful tool for extracting text from images, leveraging advanced techniques like text detection, localization, segmentation, and binarization. Its high accuracy and open-source nature make it a valuable resource for various applications requiring text extraction from digital images.

## **Advantages of Tesseract:**

- High Accuracy
- It supports Over 100 languages.

**References:** Smith, R. "An Overview of the Tesseract OCR Engine." (ICDAR 2007).

# New method for language translation:

## Advantages of Google Translate:

### For Text-to-Text Translation:

- Google Translate uses state-of-the-art neural networks for accurate and contextually appropriate translations.
- Extensive Language Support
- User-Friendly Interface

### For Text-to-Speech Translation:

- Natural Sound
- Multilingual Support
- Seamless integration with devices and platforms, making it convenient for on-the-go translations.
- Real-Time Speech Translation

# WORK FLOW

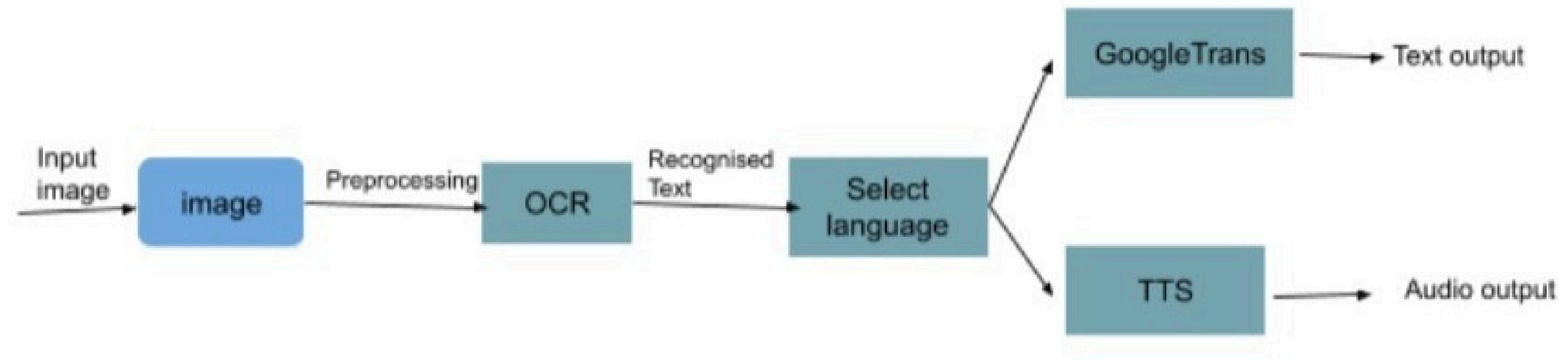


Fig. Image to text and speech conversion flow

## **Image Input:**

At this stage, the process begins with the input of an image containing text. This image could be sourced from various places such as images, photographs, or screenshots.

## **Optical Character Recognition (OCR):**

The image containing text is passed through an Optical Character Recognition (OCR) system. OCR software processes the image to recognize and extract the text embedded within it. This step is crucial for converting the visual information of the text into a machine-readable format.

## **Language Selection:**

Now select your desired option by clicking the provided options.

## **Googletrans:**

One of the best Python packages to help you with this task is Googletrans. The Google Translate API allows developers to easily access translations, allowing them to translate text from one language to another. The googletrans package provides a simple API for translating text from one language to another.

## **TTS:**

TTS, an acronym for Text-to-Speech, is speech synthesis technology that converts written text to spoken words. Note that it synthesizes words rather than playing back pre-recorded messages A

## **Language Translation:**

The extracted text, which is typically in one language, is then translated into multiple languages. Language translation services or APIs are used for this purpose. The translated text allows for broader accessibility and understanding across different linguistic backgrounds

## **Translated text:**

After translation, the text is available in multiple languages. Each translated version represents the original text in a different language. This step enables the content to be comprehensible to speakers of various languages, thus enhancing its reach and impact.

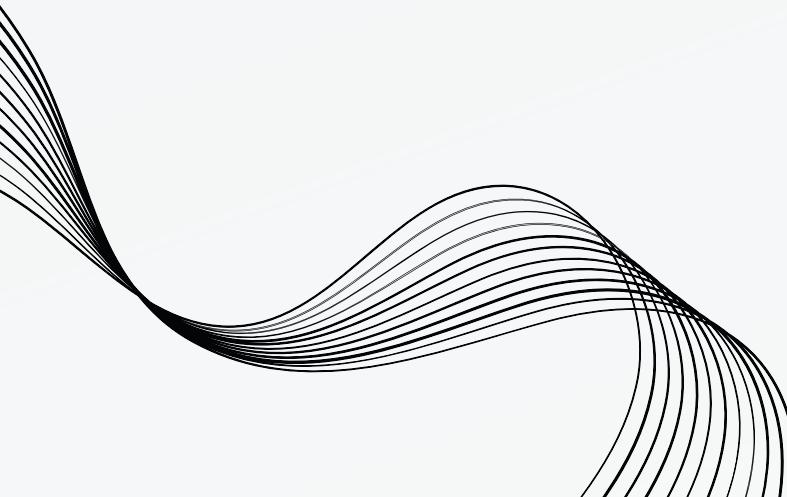
## **Audio Output:**

Finally, the audio output containing the synthesized speech is generated. This audio output represents the translated text in spoken form. It can be seamlessly integrated into audio playback systems or incorporated into applications, offering enhanced listening convenience and accessibility.

Overall, this block diagram outlines the process of extracting text from an image, translating it into multiple languages, and converting it into audio. Each step contributes to making the content more accessible and understandable to a diverse audience with different language preferences and accessibility needs.

# **ALGORITHMS USED:**

1. Optical Character Recognition (OCR)
2. Text to Speech Algorithms



# FUTURE WORK

## 1. Enhanced Handwriting Recognition:

Improving the ability of the system to accurately understand and convert handwritten text by using advanced neural networks.

## 2. Expanded Offline Capabilities:

Allowing the system to perform language translation even without an internet connection, enhancing accessibility in diverse environments.

# **CONCLUSION:**

The project aimed to develop a comprehensive system for recognizing text from images, translating it into multiple languages, and converting it to speech. This system serves as a valuable tool for both students and individuals, offering enhanced accessibility and usability in accessing and comprehending textual content.

The project utilized advanced technologies such as Optical Character Recognition (OCR), language translation APIs, and text-to-speech synthesis to achieve its objectives. By leveraging OCR algorithms, text could be accurately extracted from images, enabling users to access information from various sources such as circulars, notices, and books. The integration of language translation APIs facilitated the seamless translation of text into multiple languages, catering to diverse linguistic needs and preferences..

**THANK YOU!**