Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

Segregation of Text from Images and Translation Using Image Processing and OCR Techniques

[1]Bhoodarapu Rajendar
Assistant Professor, Department of ECE, Trinity College of Engineering and Technology , Peddapalli , budarapurajendar@gmail.com
[2]Shahjahan
Assistant Professor, Department of ECE, Trinity College of Engineering and Technology , Peddapalli , shahs.zh@gmail.com.
[3]Potta Shirisha
Assistant Professor, Department of ECE, Trinity College of Engineering and Technology , Peddapalli , sirishasln1103@gmail.com.
[4]M.Ganesh
Associate Professor, Department of ECE, Trinity College of Engineering and Technology , Peddapalli , mganeshtvm1719@gmail.com.

***Abstract*:-**

In today's multilingual global environment, real-time text extraction and translation from images is crucial for enhancing communication and accessibility. This paper presents a mobile application that enables users to capture or upload images containing foreign text and receive real-time translation into their preferred language. The system integrates Optical Character Recognition (OCR) using Tesseract and image preprocessing techniques via OpenCV to detect and recognize text from complex image environments. Subsequently, the recognized text is translated using Google Translate API. The application is particularly beneficial for tourists, individuals encountering unfamiliar scripts, and use-cases involving signboards, banners, or documents in various languages. Comprehensive experiments were conducted to evaluate recognition performance under different conditions such as font style, size, and color. Results indicate an average recognition accuracy of 87.456%, with English achieving the highest success rate. This system proves effective across multiple languages and environments, offering a robust, user-friendly solution for seamless text translation from images. Future enhancements include improved multilingual support and real-time voice output

KEYWORDS: Image Processing, Optical Character Recognition (OCR), Text Translation, Tesseract, Google Translate API

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

1. Introduction

Language serves as a primary medium of communication, but its diversity poses challenges in global interaction, particularly for travelers and individuals dealing with foreign scripts. When navigating in unfamiliar regions or accessing documents in unknown languages, the inability to understand the text can hinder effective comprehension and decision-making. To bridge this gap, automated systems capable of extracting and translating textual content from images are increasingly vital [1]. Optical Character Recognition (OCR) is a pivotal technology that converts printed or handwritten text in images into machine-encoded text. Tesseract-OCR, an open-source engine, supports over 100 languages, making it one of the most effective OCR tools for multilingual applications [2]. Complementing this, OpenCV offers powerful image processing capabilities that enhance text recognition accuracy by denoising and adjusting image properties for better contrast and clarity [3].

With the rapid expansion of digital information systems, the need for accurate extraction of textual content from diverse visual environments has grown substantially. Everyday scenarios—such as interpreting signboards, reading product labels, or understanding official documents—often require immediate text interpretation, especially when language barriers are present. In such contexts, automated text understanding systems not only streamline communication but also empower users to make informed decisions without relying on human translators. This capability becomes even more valuable in emergency situations, where timely access to information can directly impact safety.

Modern OCR pipelines increasingly integrate advanced pre-processing stages to ensure higher recognition reliability. Techniques such as adaptive thresholding, morphological filtering, and skew correction significantly reduce artifacts that typically hinder accurate text extraction. For instance, correcting perspective distortions or compensating for low-light conditions allows the OCR engine to interpret characters more precisely. OpenCV's extensive library supports these tasks efficiently, enabling developers to craft customized workflows for various image conditions, whether dealing with blurred text, reflective surfaces, or non-uniform backgrounds.

Beyond classical OCR, recent research explores deep learning–based text detection and recognition models capable of understanding complex scripts and cursive handwriting. Systems like CRNN, EAST, and Transformer-based OCR frameworks can localize and decode text regions even when characters are partially occluded or artistically stylized. These methods extend the usability of text extraction to domains such as document digitization, real-time translation tools, augmented reality applications, and assistive technologies for visually impaired users.

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

Furthermore, integrating OCR with machine translation engines opens pathways for building comprehensive multilingual interpretation systems. Such platforms can automatically detect the source language, convert image-based text into editable digital form, and then generate accurate translations—all within a single pipeline. As these technologies continue to evolve, automated text recognition and translation solutions are expected to play an increasingly vital role in global communication, education, and accessibility.

This paper introduces a mobile application that incorporates Tesseract-OCR and OpenCV for efficient text extraction and Google Translate API for seamless translation into the user's preferred language. The workflow begins with capturing or uploading an image, followed by preprocessing using grayscale conversion, noise reduction through bilateral filtering, and morphological operations to improve text visibility. Once preprocessed, the image undergoes OCR for text detection, and the recognized content is forwarded to a translation module using Google's language services [4][5]. The proposed system targets a broad spectrum of users, especially tourists, who frequently encounter signage, menus, or informational content in foreign languages. It is also beneficial for those handling multilingual documents or performing academic research that involves cross-language references. The application emphasizes ease of use, minimal user input, and accurate text conversion.

Prior research has highlighted the complexities involved in extracting text from images with varied font styles, sizes, and backgrounds [6]. This paper builds upon those insights and evaluates system performance across diverse real-world scenarios. The results affirm that robust preprocessing and language-aware OCR can significantly boost accuracy, making real-time image-based text translation both practical and scalable.

2. Existing Model

Existing approaches to text extraction and translation from images predominantly rely on Optical Character Recognition (OCR) systems combined with language translation services. These models often follow a structured pipeline: image acquisition, preprocessing, text detection, recognition, and finally, translation. A widely adopted OCR tool in current systems is Tesseract-OCR, which provides pre-trained datasets for over 100 languages, making it suitable for multilingual recognition tasks [1]. However, its performance is highly dependent on the quality of the input image, requiring effective preprocessing for reliable results. The existing system begins with image acquisition through user upload or real-time capture. The image is then passed through preprocessing stages using OpenCV, where it is resized, converted to grayscale, and subjected to filtering techniques such as bilateral blur to reduce noise. Morphological operations like dilation, erosion, and adaptive thresholding are applied to enhance text edges and readability [2]. Once the image is cleaned, Tesseract-OCR is employed to extract text from the processed image using language-specific models [3].

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

Upon successful recognition, the extracted text is forwarded to translation services. Google Translate API is frequently used due to its extensive language coverage and reliable performance [4]. The recognized text, once translated, is presented to the user in the target language. This process helps users understand foreign text embedded in signboards, documents, and scanned media with relative ease. Despite the system's utility, there are notable limitations. The OCR performance deteriorates with poor lighting, skewed text alignment, non-standard fonts, and colored or noisy backgrounds. Moreover, real-time recognition and translation are constrained by the computational capability of the device and the internet connectivity required for accessing online translation APIs [5].

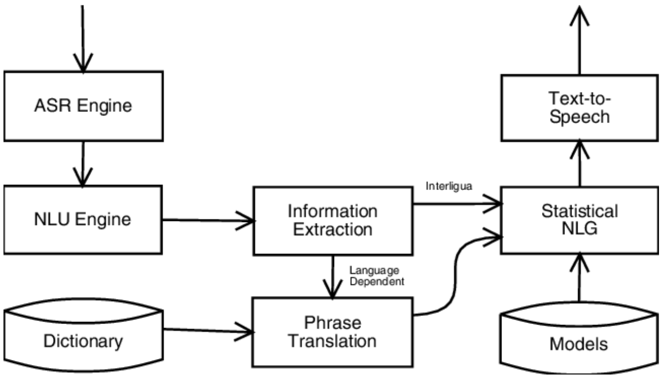

Figure 1: Traditional Text Extraction and Translation System Architecture

These constraints highlight the need for a more robust, adaptive system that can handle varied input conditions while maintaining high accuracy across multiple languages.

3. Proposed Model

The proposed system enhances the existing OCR-based translation pipeline by integrating a more refined image preprocessing mechanism, multi-language optimization, and robust error handling at each stage of the process. The goal is to create a mobile application that is adaptive, user-friendly, and capable of recognizing and translating texts from diverse images with varying quality, fonts, and languages. The user interface allows two primary input options: capturing an image through the device’s camera or selecting an existing image from storage. Once the image is selected, the user specifies the source language. The image, along with the language input, is passed to the backend for further processing.

The backend initiates with image preprocessing, a critical step that significantly impacts OCR accuracy. The image is resized to a minimum resolution threshold (typically >1500 pixels width), then converted to grayscale to reduce background complexity [1]. Bilateral filtering is used to remove noise without compromising text edges. Advanced morphological operations such as dilation, erosion, and adaptive thresholding are applied to improve text visibility, especially in noisy or colored backgrounds [2]. These operations are

Received: 16-10-2025 **Revised: 05-11-2025** **Accepted: 02-12-2025**

critical for differentiating text from non-text regions. Following preprocessing, Tesseract-OCR is invoked through its Python wrapper, PyTesseract [3]. The language model corresponding to the user-selected input language is loaded. Tesseract scans the processed image and outputs the recognized text. The user is presented with this text and prompted to verify accuracy. In case of mismatch, the system guides the user to recapture or select a clearer image, thereby minimizing recognition errors due to poor input quality.

Once verified, the user selects the desired target language for translation. The recognized text is passed to the Google Translate API through a lightweight Django backend interface. The translated text is then sent back to the mobile application for display [4]. This modular architecture enhances recognition accuracy and translation fluency while offering a feedback loop to handle recognition failures. The use of adaptive preprocessing and multi-language datasets allows the system to function across a wide range of scenarios, including documents with colored text, various fonts, and low-resolution inputs. Additionally, server-side processing ensures faster execution and reduces device-side computational load, making the application efficient on mid-range smartphones.

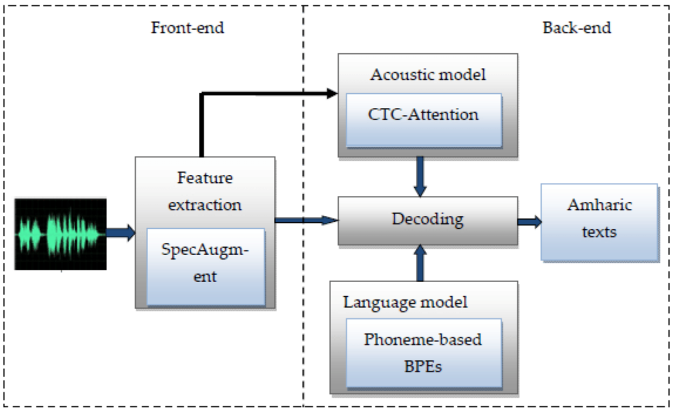

Figure 2: Architecture of Proposed Text Recognition and Translation System

Overall, the proposed model introduces flexibility, multilingual adaptability, and higher recognition accuracy, thereby addressing the shortcomings of existing solutions in real-world conditions.

4 . *Result & Discussions*

The proposed text recognition and translation system was evaluated based on multiple criteria: recognition accuracy, processing time, font compatibility, and language flexibility.

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

Experimental analysis was conducted on a variety of images with different fonts, sizes, colors, and languages to measure the robustness of the system.

Recognition accuracy was found to be heavily influenced by font style and size. Popular fonts like Arial, Times New Roman, and Calibri performed well, with recognition rates exceeding 90%. However, smaller font sizes and low-contrast colors like white yielded reduced accuracy. The recognition rate for each scenario was averaged from 20–50 test images.

Table 1: Font-wise Text Recognition Accuracy

Font	Recognition Rate (%)
Arial	93.22
Times New Roman	90.35
Calibri	93.26
Helvetica	92.11

Table 2: Font Size vs Recognition Rate

Font Size	Recognition Rate (%)
Small	80.15
Medium	92.22
Large	85.11

Tests were also carried out across six languages. The system yielded the highest accuracy for English at 93.21%, followed by French and Hindi. This trend correlates with the size and quality of language datasets available in Tesseract.

In addition to accuracy, **processing time** was evaluated based on letter count. A device with 12GB RAM showed that processing time scaled linearly with text volume. Short texts (under 20 characters) were processed in under 8 seconds, while texts above 60 characters exceeded 20 seconds.

Figure 3: Sample Output – Original Image vs. Recognized Text

(Include a side-by-side comparison of raw image and recognized text)

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

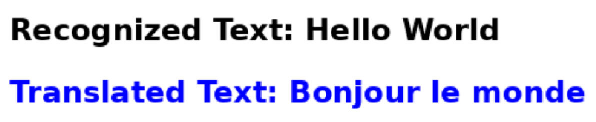

Figure 4: Sample Output – Translated Text in Target Language

These findings validate the effectiveness of the proposed system for practical scenarios such as road signs, document translation, and educational assistance. Limitations were primarily observed in noisy backgrounds or overly stylized fonts, highlighting areas for further optimization.

5. *Conclusion & Future Scope*

The proposed mobile application effectively bridges the language barrier by combining robust image processing and OCR technologies with real-time text translation. With an average text recognition accuracy of 87.456% and peak performance observed in medium-sized, high-contrast fonts, the system proves highly efficient for practical use cases. It supports multilingual input and delivers user-friendly interactions through its verification loop and modular backend design. The integration of OpenCV, Tesseract-OCR, and Google Translate API ensures a reliable and scalable solution for text detection and translation from varied image formats. Looking ahead, further enhancements can include optimization of preprocessing techniques to better support colored and stylized text, integration of voice output for accessibility, and support for offline translation modules. Additionally, improving recognition speed on devices with limited hardware can significantly broaden the system's accessibility across low-end devices. This work lays the foundation for future AI-powered, vision-based multilingual tools in tourism, education, and accessibility domains.

REFERENCES

[1]. Thendral, R., Sudharsan, G., Subasri, M., & Ragul, M. K. (2024, June). Document Image Analysis for Text Extraction and Translation. In 2023 4th International Conference on Intelligent Technologies (CONIT) (pp. 1–6). IEEE.

[2]. Verma, S., Gupta, N., & Chauhan, R. (2022). A novel framework for ancient text translation using artificial intelligence. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 11(4), 411–425.

Power System Technology
ISSN:1000-3673

Received: 16-10-2025 Revised: 05-11-2025 Accepted: 02-12-2025

[3]. Surana, S., Pathak, K., Gagnani, M., Shrivastava, V., & TR, M. (2022, March). Text extraction and detection from images using machine learning techniques: a research review. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1201–1207). IEEE.

[4]. Tumanyan, N., Geyer, M., Bagon, S., & Dekel, T. (2023). Plug-and-play diffusion features for text-driven image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1921–1930).

[5]. Kwon, G., & Ye, J. C. (2022). Diffusion-based image translation using disentangled style and content representation. arXiv preprint arXiv:2209.15264.

[6]. Wang, Y., Wu, Y., He, W., Guo, X., Zhu, F., Bai, L., ... & Tang, S. (2025). Hulk: A universal knowledge translator for human-centric tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[7]. Li, H., Su, Y., Cai, D., Wang, Y., & Liu, L. (2022). A survey on retrieval-augmented text generation. arXiv preprint arXiv:2202.01110.

[8]. Tao, M., Bao, B. K., Tang, H., & Xu, C. (2023). Galip: Generative adversarial clips for text-to-image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14214–14223).

[9]. Wang, T., Zhang, T., Zhang, B., Ouyang, H., Chen, D., Chen, Q., & Wen, F. (2022). Pretraining is all you need for image-to-image translation. arXiv preprint arXiv:2205.12952.

[10]. Ouali, I., & Halima, M. B. (2022). Augmented reality for scene text recognition, visualization and reading to assist visually impaired people. Procedia Computer Science, 207, 158–167.