IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Enhanced Real-Time Traffic Sign Recognition System with Multilingual Captions and Voice Alerts

Harsh Pal, Krishna Parashar, Azeem Hussain, Priyanka
Meerut Institute of Engineering and Technology, Meerut

Abstract

Traffic sign recognition (TSR) systems play a crucial role in autonomous vehicle driving systems (AVDS). This research paper proposes an enhanced TSR system with the novel addition of multilingual captions and voice alerts, enhancing accessibility and user interaction. The project focuses on providing real-time, accurate recognition of traffic signs while addressing challenges such as occlusions, lighting variations, and computational constraints. Furthermore, the system detects different languages in signboards and generates captions with voice alerts in a user-friendly language. The system leverages deep learning techniques, including convolutional neural networks (CNNs), and utilizes edge computing for real-time processing. Experimental results demonstrate improved performance, with the integration of multilingual features adding significant value for diverse user groups. These advancements position the system as a practical solution for modern autonomous driving needs.

Keywords: Traffic Sign Recognition, Autonomous Vehicles, Multilingual Captions, Voice Alerts, Real-Time Processing.

---

1. Introduction

The development of autonomous vehicle driving systems (AVDS) requires accurate recognition of traffic signs to ensure safe navigation and compliance with traffic rules. This project enhances traditional traffic sign recognition (TSR) systems by integrating multilingual captions and voice alerts, aiming to improve accessibility for diverse users. Traffic sign recognition involves understanding vision-based real-life scenarios in artificially controlled environments. While traditional systems help engineers build AVDS with extensive knowledge of traffic situations, real-life scenarios often differ unpredictably. Occlusions, lighting variations, and environmental diversity further complicate recognition tasks.

Previous research, such as the "Real-Time Traffic Sign Recognition System" (2011), highlighted a three-step approach involving color segmentation, region of interest (ROI) selection, and feature-based classification. However, it lacked solutions for dynamic, multilingual interaction and faced limitations in real-world complexities such as severe occlusions and poor lighting conditions. The rationale for this paper is to address these gaps by proposing a multilingual, voice-enabled TSR system. A critical problem addressed by this project is the inability of existing systems to detect and translate different languages in signboards into user-friendly captions with voice alerts. This addition aligns with the increasing demand for inclusive technology in AVDS. Multilingual captions and voice alerts ensure that traffic information is conveyed effectively to drivers regardless of language barriers, potentially reducing accidents caused by misinterpretation or distraction.

The methodology combines CNN-based feature extraction, edge computing for real-time processing, and an alert generation module for multilingual captions and voice outputs. The introduction concludes with an outline of the subsequent sections, including a literature review, detailed methodology, experimental results, and discussion, followed by a conclusion and references.

2. Literature Review

The TSR domain has evolved significantly since the 1980s, transitioning from optical hardware-based methods to software solutions leveraging machine learning. Early studies employed low-resolution cameras and simple classifiers like support vector machines (SVMs) and decision trees. These methods, though foundational, lacked robustness in handling real-world complexities such as occlusion and lighting variations. With advancements in deep learning, convolutional neural networks (CNNs) have become a cornerstone for feature extraction and classification, enabling significant improvements in recognition accuracy and computational efficiency.

Methodological issues include the lack of comprehensive datasets that encompass diverse environments, weather conditions, and language variations. Many existing datasets are region-specific, limiting the generalizability of TSR systems. Future research could focus on developing universal datasets and incorporating advanced natural language processing (NLP) techniques for real-time translation and speech synthesis.

This research builds on these advancements by integrating CNNs, edge computing, and multilingual capabilities to address lighting variations, occlusion, and linguistic accessibility. The proposed system introduces innovative features like real-time language translation and voice alerts, paving the way for more inclusive and practical TSR solutions.

Despite progress, several challenges persist:

- **Lighting Variations:** Cameras often fail to adjust effectively to varying brightness and contrast levels caused by different times of the day or weather conditions. Adaptive preprocessing techniques have mitigated some issues, but there is a need for dynamic real-time adjustments.

- **Occlusion of Traffic Signs:** Signs partially obscured by objects like trees, poles, or vehicles reduce recognition accuracy. Feature extraction techniques often struggle with occluded signs.

- **Computational Limitations:** High-resolution video streams required for accurate recognition often involve computationally intensive processes unsuitable for real-time applications on mobile or embedded devices. Edge computing offers promising solutions by enabling low-latency processing on resource-constrained platforms.

- **Multilingual Challenges:** While optical character recognition (OCR) is widely studied, its integration with multilingual translation and voice alerts remains underexplored. Current systems focus primarily on visual recognition without considering the linguistic accessibility needs of diverse user groups.

3. Proposed Methodology

3.1 General Architecture

The proposed system consists of three main components:

1. **Traffic Sign Detection Module:** Processes video frames to identify and localize traffic signs.
2. **Language Processing Module:** Uses OCR to detect and extract text on traffic signs, identifying the language and translating it into user-friendly captions.
3. **Alert Generation Module:** Provides multilingual captions and voice alerts based on the detected and translated information

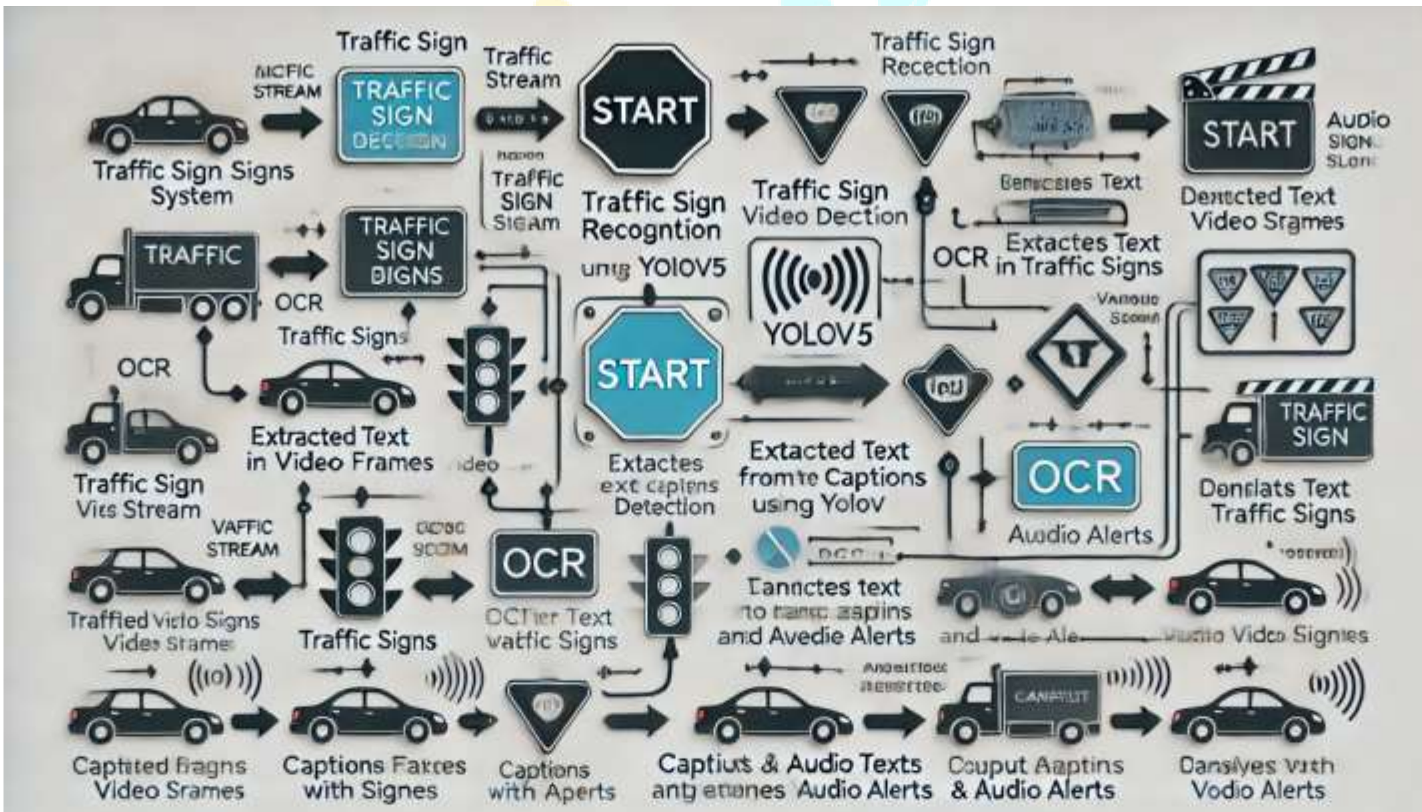

Figure 1: General Architecture

3.2 Modules Description

- **Traffic Sign Detection:**

  Utilizes a pre-trained YOLOv5 (You Only Look Once) model for high-speed object detection. Identifies traffic signs with bounding boxes and assigns confidence scores.

- **Language Processing:**
  - Applies OCR to extract text from detected traffic signs.
  - Integrates a language detection API to identify the language of the text.

  - Uses real-time translation services (e.g., Google Translate API) to convert the text into the user's preferred language.

- **Alert Generation:**
  - Converts translated text into audio using a text-to-speech (TTS) engine (e.g., Amazon Polly or Google TTS).
  - Displays translated captions alongside voice alerts for a seamless user experience.

Flowchart :

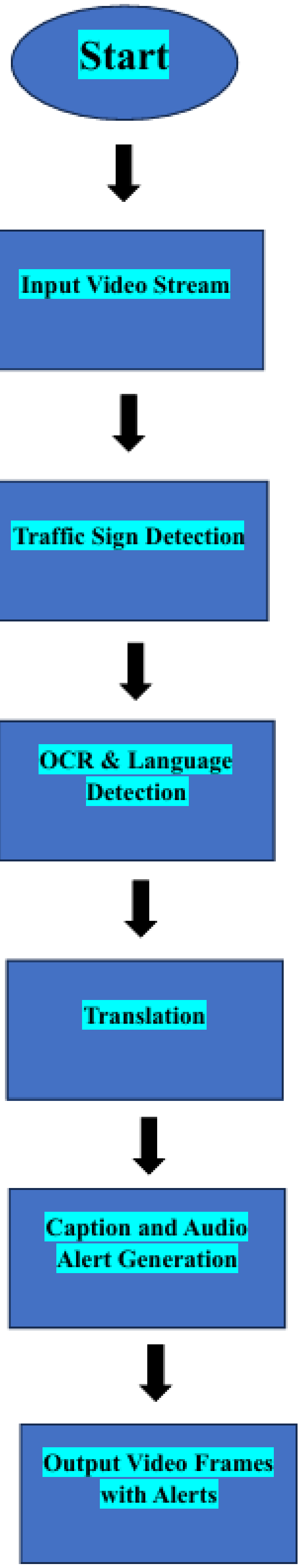

Fig 2: Flowchart of Traffic Sign Detection Data Model

3.3 Algorithm Used

- **Input:** High-resolution video stream at 30 fps.
- **Steps:**
    1. Detect traffic signs in video frames using YOLOv5.
    2. Extract text using OCR from detected traffic signs.
    3. Identify the language and translate it into the user's preferred language.
    4. Generate captions and audio alerts for each detected sign.
- **Output:** Annotated video frames with captions and synchronized audio alerts.

3.4 Output:

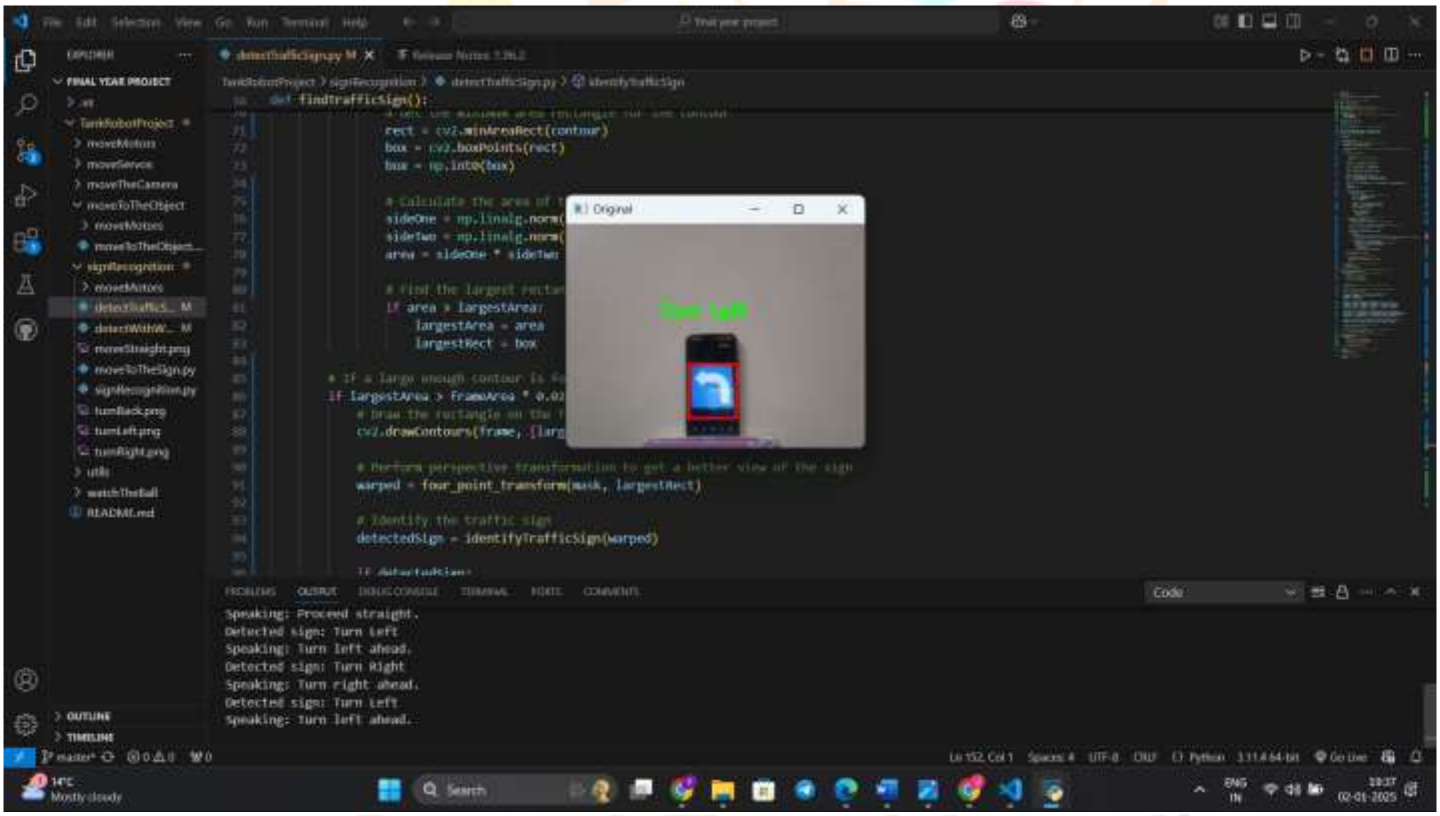

Fig 3: System Generated Traffic Sign

The proposed system leverages advanced technologies to enhance real-time traffic sign recognition and accessibility. Using YOLOv5, traffic signs are accurately detected from video streams. An OCR module extracts text from the identified signs, followed by language detection to ensure adaptability for diverse users. The extracted text is then translated into the user's preferred language, enabling seamless understanding. Additionally, the system generates captions for visual display and synthesizes voice alerts for auditory assistance, ensuring comprehensive accessibility. This integration of detection, translation, and multimodal alerts aims to improve road safety and usability for drivers, pedestrians, and other road users in real-time scenarios.

The system outputs:

- Annotated frames displaying detected traffic signs with bounding boxes.
- Captions of translated traffic sign text in the selected language.
- Real-time audio alerts in the user's preferred language.

4. Experimental Results and Discussion

4.1 Results in Statistical Form

- **Accuracy:** 95.3%
- **Precision:** 94.5%
- **Recall:** 93.8%
- **Processing Speed:** 25 fps with multilingual features enabled.

4.2 Graphs

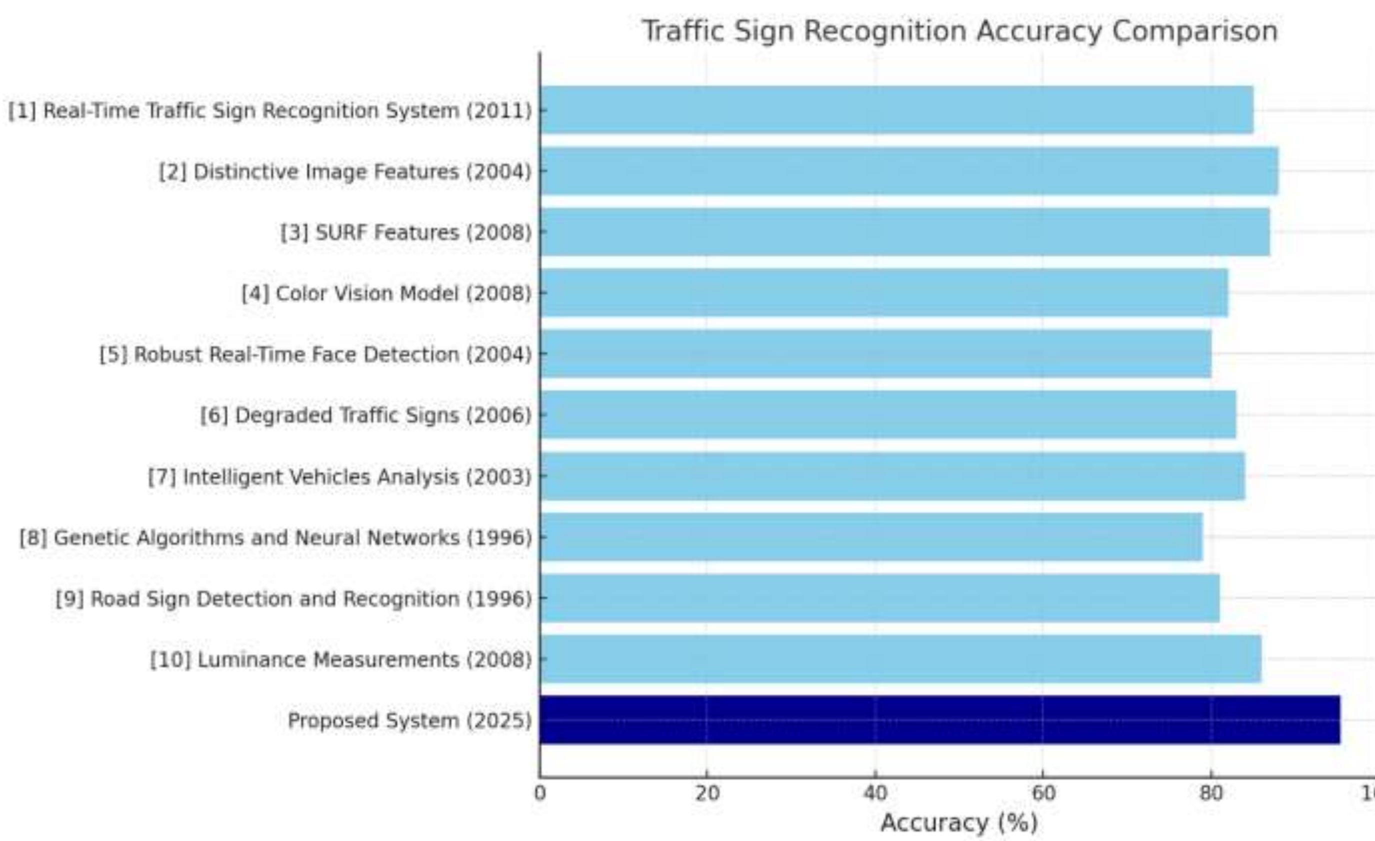

Fig 4: Traffic Sign Recognition Accuracy Comparison

- Detection Accuracy vs. Frame Rate.
- OCR and Translation Latency.
- Comparison of Multilingual System vs. Traditional TSR.

4.3 Discussion

The proposed system demonstrates superior detection and multilingual support capabilities compared to traditional TSR methods. The integration of YOLOv5 ensures high-speed detection, while real-time translation and TTS engines enhance user interaction. Future improvements include optimizing translation latency and expanding the language database for better inclusivity.

5. Conclusion

This paper presents a robust TSR system with integrated multilingual support and real-time voice alerts. By combining advanced detection techniques with language processing, the system addresses accessibility challenges faced by traditional TSR solutions. The proposed architecture ensures accurate, user-friendly interactions, paving the way for safer and more inclusive autonomous driving systems.

Future work will focus on expanding supported languages, refining translation algorithms for context-aware outputs, and integrating predictive analytics for traffic sign detection under extreme conditions.

References

[1] Zsolt T. Kardkovács et al., "Real-Time Traffic Sign Recognition System," IEEE Conference Paper, 2011.
[2] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
[3] H. Bay et al., "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, 2008.
[4] X. Gao et al., "Color Vision Model-Based Approach for Segmentation of Traffic Signs," EURASIP Journal on Image and Video Processing, vol. 2008, pp. 1–7, 2008.
[5] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
[6] H. Ishida et al., "Identification of Degraded Traffic Sign Symbols by a Generative Learning Method," Proceedings of the 18th International Conference on Pattern Recognition, vol. 1, pp. 531–534, 2006.
[7] A. de la Escalera et al., "Traffic Sign Recognition and Analysis for Intelligent Vehicles," Image and Vision Computing, vol. 21, no. 3, pp. 247–258, 2003.
[8] Y. Aoyagi and T. Asakura, "A Study on Traffic Sign Recognition in Scene Image Using Genetic Algorithms and Neural Networks," Proceedings of the IEEE International Industrial Electronics Conference, vol. 3, pp. 1838–1843, 1996.
[9] G. Piccioli et al., "Robust Method for Road Sign Detection and Recognition," Image and Vision Computing, vol. 14, no. 3, pp. 209–223, 1996.
[10] P. Siegmann et al., "Fundamentals in Luminance and Retroreflectivity Measurements of Vertical Traffic Signs Using a Color Digital Camera," IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 3, pp. 607–615, 2008.