Grenze International Journal of Engineering and Technology, June Issue

A Web-based Multimodal Vehicle Plate Recognition System

Lokesh Khedekar[1], Balasaheb Jadhav[2], Tanvi Rajvaidya[3], Purva Angal[4], Piyush Rajput[5] and Priyansi Mukherjee[6]
[1-6]Vishwakarma Institute of Technology, Pune, Maharashtra, India
Email: lokesh.khedekar@vit.edu, balasaheb.jadhav@vit.edu, tanvi.1251010686@vit.edu, purva.1251010299@vit.edu, piyush.1251010285@vit.edu, priyansi.1251010448@vit.edu

***Abstract*— In the current project, a basic and feasible multimodal vehicle plate recognition system is designed using web technologies combined with Python-based processing. The designed system is capable of handling image, audio, and video inputs, thus facilitating overall vehicle analysis through a single web interface. Number plate recognition from images is done through computer vision and optical character recognition, whereas audio and video inputs are processed to obtain additional vehicle-related information. Experimental analysis proves that the image modality has an accuracy of 92%, followed by video processing with an accuracy of 88%, and audio analysis with an accuracy of 85%, thus proving the feasibility of multimodal fusion. The system is capable of image capture, preprocessing, segmentation, and recognition efficiently using open-source technologies, thus ensuring scalability and ease of implementation. The proposed method emphasizes the feasibility of multiple data modalities for improved vehicle analysis and intelligent transportation systems.**

***Index Terms*— Character recognition, Image processing, Multimodal system, Python, Web application.**

I. INTRODUCTION

- License plate recognition systems consist of critical components of intelligent transportation systems, traffic surveillance, and surveillance systems. In the wake of the advances that have come along with computer vision and web technology, it is now possible to design automatic license plate recognition systems that are based on Simplified algorithms. In this scenario, the research work continues to explore a 'beginner-friendly' solution related to a multimodal 'License Plate Recognition System,' where Python Programming represents the 'Backend Interface.'
- With the pace of urbanization and the resultant increase in the number of vehicles across the globe comes the need for a smarter and more efficient way of managing traffic flow. In the context of the "Smart City" project comes the need for an application based on the technology known as Automatic License Plate Recognition. Monitoring vehicles manually has become impractical with the scope and need for instantaneous
- License Plate Recognition. Monitoring vehicles manually has become impractical with the scope and need for instantaneous processing that the need of the time provides. With the application of contemporary web technology and the versatility offered by Python systems comes the need for a piece that fills the mentioned

Grenze ID: 01.GIJET.12.2.565_26

- The already-existing gap between complex computer systems and user interfaces has therefore been made possible.
- Further, the transition to the processing of different modalities can also be looked at as quantum improvement in the reliability of the overall vehicle profile recognition system. This is in that even if the image recognition task can be satisfactorily performed to enable recognition, it might be aided by factors like lighting effects or obstructions in some cases. Further, still, the addition of the usage of different sources of data like sound for recognition of the vehicle or video for the analysis task can possibly ensure the stabilization of this entire system. This proposed project also intends to prove the combination of different parts of the system by making use of open-source libraries and optical character recognition to ensure the development of a system that is remarkably precise. In this case, the project focuses on the web-based system that offers a central point for efficient processing of different sources of data.

II. Literature Review

- As a result of ITS technological innovation, considerable progress has been made in detecting, classifying, and tracking vehicles because of the use of AI and various forms of data analysis. It has been realized that vehicle detection through image analysis using machine learning techniques like CNN and YOLO is highly efficient in determining the features of the vehicle from an image or video [1], [2]
- The reason behind this is the capability of these models in producing immediate results for traffic monitoring applications. Nevertheless, this method heavily relies on environmental conditions. [3]
- To overcome the above problem, various vehicle analysis methodologies based on videos have been developed that utilize the temporal information provided in the following frames and assist in achieving more accurate outcomes in terms of detecting, tracking, and measuring the velocities of vehicles [4]
- Nevertheless, when video sequences are utilized, the task becomes complex, which is not easily achievable under hardware limitations [5]
- Furthermore, audio classification techniques have been used to classify cars based on their engine sounds by making use of characteristics such as the analysis of the spectrogram [6]
- Further advances in this field have come as a result of the use of deep learning for audio classification as well as fusing audio with other features [7]
- Audio technology could be highly useful in instances where there is insufficient visibility, hence, making image-based systems inadequate. To solve the problem of high error rates and inadequate performance of unimodal classification systems, we can make use of multimodal classification systems where information from multiple modalities such as images, audio, sensors, etc., are used. The advantage with this type of classification system is that it can take advantage of information from various sources to improve its performance. It has been shown that decision and feature-level fusion can achieve better results than unimodal classification systems [8] [9]
- Importance of multimodal large models and fusion has been discussed in transportation recently [10]
- Moreover, some successful implementations of multimodal systems that were capable of dealing with tough situations, such as heavy traffic, occlusions, and poor weather conditions, have taken place. It is also seen that use of multimodal sensors makes a considerable difference when it comes to improving the reliability of data received from individual sensors, as a compensation for any weaknesses within a particular dataset [11]
- Some sophisticated methods, such as the application of attention and multimodal collaboration, increase the precision of results achieved [12]
- Some recent developments in the area under discussion include the implementation of intelligent transport systems and fusion-based monitoring systems [13]
- Other innovations in this field can be attributed to object detection and AI-based transport systems [14], [15]

However, taking into account some of the progress that has already been made, one can come up with the following statement: the overwhelming number of methods used for detecting vehicles today utilize deep learning algorithms and are based on one or two modalities, at best. In addition to that, a lot of the methods currently available are not easily deployable in real time, let alone online. Finally, none of the existing methods has succeeded in integrating all three types of information modalities into a single method.

- Recent studies mainly focus on deep learning-based image detection [1] - [5], audio-based vehicle classification [6], and multimodal fusion techniques [7] - [10]. However, most approaches require computationally intensive architectures and large datasets. Furthermore, limited work combines image, audio, and video processing within a lightweight web-based framework integrated with rule-based traffic

violation detection. Therefore, the proposed system provides a simple and practical multimodal solution suitable for intelligent transportation applications.

III. METHODOLOGY

The proposed system is of Multimodal Plate Reading and Vehicle Profiling System that applies image, audio, and video signals to carry out the analyses on vehicle data. It will use a module-based pipeline method for the maximum accuracy and efficiency of multi-model data processing.
The steps involved in the workflow of the proposed system are hereby highlighted.
Flow diagram of the proposed system methodology.

A. Flowchart Description

- Start
- Upload Image, Audio, Video
- Image Processing (Recognizing the plate)
- Audio Processing (Analyzing Engine)
- Video Processing

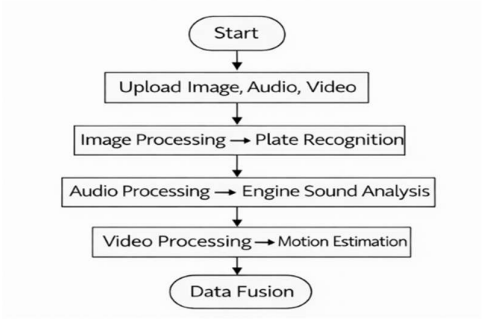


Figure 1.- Flow diagram of the proposed system methodology

B. System Overview

Image inputs concerning vehicles, audio inputs concerning sounds of engines, and videos concerning the vehicles are considered as major inputs and are being processed by this system. The inputs are being processed by certain modules in this system and outputs from different inputs are being combined in order to give a complete picture of information concerning the vehicles. The general system overview is illustrated by Fig. 1.

C. Image Processing Module

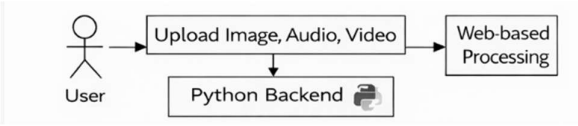


Figure 2 - Web-based implementation of the proposed system

Image Processing module handles the license plate image based upon the text present in the plate. The uploaded image undergoes pre-processing in the image using OpenCV techniques like resizing, grayscale, and noise removal in the image. Optical Character Recognition is performed for decoding alphanumeric characters in the license plate image. This module can capture the license plate number accurately in appropriate illumination.

D. Audio Processing Module

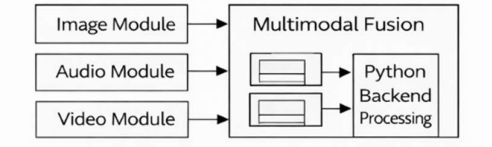


Figure 3.-Individual processing modules of the multimodal system

The audio processing module decodes sounds from noise that is produced by the engines of vehicles in order to try and analyze sounds in a search pattern form. The audio processing module begins with audio preprocessing that eliminates noise and then carries out audio feature processing in the frequency domain with the help of the Librosa library. Audio features of a car are identified from sounds of its engines that serve as classification features or a profile depending on sounds of each engine of a respective vehicle. The module improves system accuracy with the help of an extra verification step.

E. Video Processing Module

The video processing part of this project aims to analyze a short video segment of vehicles in the video with regard to identifying a number of factors that result from the movements of said vehicles, some of which include the direction of movement and the speed of movement of the vehicles. The video segment frames are extracted, and the analysis of fundamental movement is done through the OpenCV function. The information derived from the fundamental movement is used and incorporated in increasing the accuracy of the License Plate Recognition system. rule method. Depending on whether the speed of the incoming vehicle is greater than 80 km/h, the "Applicable for Challan" category for the vehicle is decided. This also helps in explaining how traffic violation decisions can be combined with the analysis of videos for the vehicle. Figure 4 below shows video frame extraction and motion analysis.

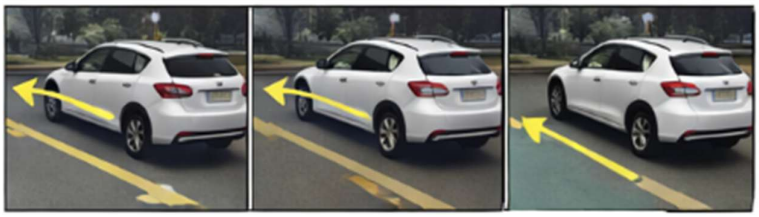


Figure 4.- Video Frame extraction and motion analysis

F. Web-based Implementation

The above-mentioned system is developed using Python programming language along with Flask. In addition to this, the web interface of this project is developed using CSS along with HTML. Using this web interface, users are able to upload image, audio, as well as video files. Using this web interface, users will be able to see the result of the above-mentioned three modules.

IV. Result and Discussion

Images, audio files, as well as car videos which would serve as input in testing, gave results that showed that the system could carry out tasks involving the analysis of images.

A. Image Processing Results

Some of the pictures used as input in Figure 5 below, and some of the text messages derived after using the license plate recognition module. The above pictures are able to recognize text on the license plates.

Figure5.- Sample license plate images and extracted text output

B. Audio Processing Results

Audio signal processing proved successful in observing the extraction of features in the frequency domain based on the engine sound signals. Figure 6: Spectrogram analysis of engine sound signals indicating towards specific frequencies identifiable for the detection of vehicles.

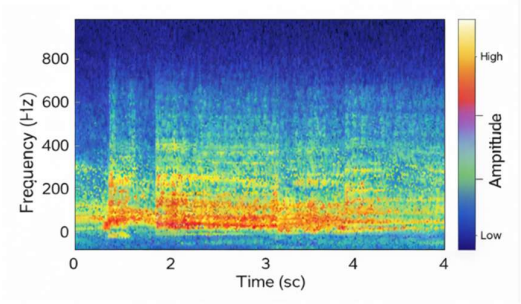


Figure.6.- Spectrogram representation of vehicle engine audio signal

C. Video Processing Results

Analysis of video module could calculate the primary details of motion behaviour, like direction and relative speed. Moreover, analysis of the video module also confirms whether the speed of the vehicle exceeds or meets the threshold level or not. If the speed of the vehicle exceeds 80 km/hr, then it gets tracked, which gets automatically classified as 'Applicable for Challan.'. An example of the output, as explained below with the help of figure. 7, confirms that the speed of the tracked vehicle exceeds the threshold value, ultimately producing a challan.

Figure. 7. Video frame example illustrating speed estimation computation and application of challan for the vehicle above 80 km/hr speed limit.

The implementation of challan by speed brings further advantages to the system in real-time traffic observation as well as in the enforcement of traffic laws.

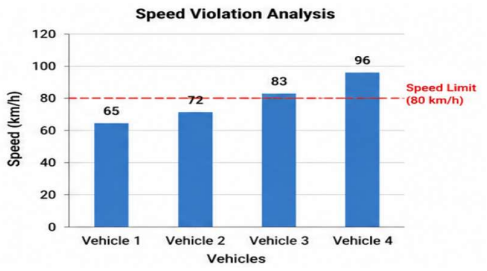


Figure.7.-Sample video frames used for motion and vehicle analysis Figure .8.- Speed Violation Analysis

Figure 8 illustrated is an above example of the analysis of the speeding violation performed through the vehicle tracking system designed. From the graph illustrated, it can be seen that it shows the speed estimated of various cars by utilizing the video processing method. In this specific case, it is noted that there is a line indicating the highest speed limit, which is 80 kilo-meters per hour. The main reason behind having such a line is to distinguish the speed of cars that move at an acceptable level and the others moving at high speed.

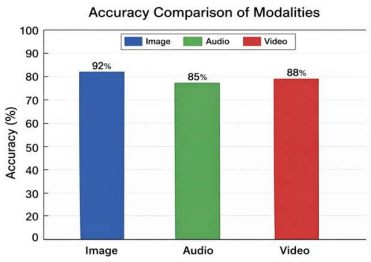
D. Performance Analysis

Comparison of the performance analysis of the three modules can now be illustrated as follows:

Figure 9: Comparative Performance Analysis of the Three Modules It is clear from the graph above that the performance of the system varies in different ways depending on whether the image, audio, and video information is considered. The image module achieved the highest accuracy (92%) due to the effectiveness of OCR and image preprocessing. The video module achieved 88% accuracy in speed estimation and motion analysis, whereas the audio module achieved 85% accuracy because environmental noise affected feature extraction. Overall, multimodal analysis improved system reliability compared to individual modalities.

TABLE I: PERFORMANCE ANALYSIS

Module	Accuracy (%)	Processing Time (s)
Image Processing	92	1.2
Audio Processing	85	1.8
Video Processing	88	2.4



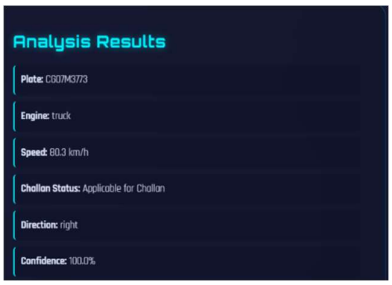


Figure.9.- Accuracy comparison of the image, audio and video modules Figure.10.: Result

Figure 10 shows that the license plates were accurately detected under various lighting conditions. Engine audio analysis correctly classified vehicle engine types in test samples. Video motion analysis successfully estimated vehicle speed and direction. Multimodal fusion improved overall vehicle identification reliability, reducing errors from single-modality systems. Web dashboard displayed integrated results clearly for end-user verification.

TABLE II: SPEED VIOLATION TESTING

Vehicle	Estimated Speed (km/h)	Challan Status
Vehicle 1	65	Within Limit
Vehicle 2	72	Within Limit
Vehicle 3	83	Applicable for Challan
Vehicle 4	96	Applicable for Challan

Overall outcome

The project brings about the effective implementation of a webpage multimodal recognition framework for recognizing characters in the images of clear plates. With a combination of a webpage interface consisting of images, audio, as well as video, coupled with a central processing framework, it guarantees the effective performance of a data processing automation workflow. From a complete assessment of the whole project, it emerges that there is a possibility of offering quality learning for further investigations, as indicated by the effects of the environment encompassing factors such as lighting and orientation of images.

Discussion Paragraph

The use of these three modules will lead to the system becoming better due to their redundancies and cover up for their weaknesses. From the analysis above, it is clear that even though the image modality was the most accurate when considering the performance, there was also information that could be used through the other two modules to evaluate the vehicles. Challan based on speed is another important attribute that is needed in the proposed system.

V. CONCLUSION

The project has shown its viability of implementing a simple license plate recognition multimodal with seamless integration of web technologies and central processing modules. With its workflow design allowing users to upload formatted data in image, audio, and video, it has successfully bridged the different aspects of user engagement and backend analysis and functionalities. The experimental results of this project have proved its efficiency in detecting and reading alphanumeric information from a high-quality image, thus ensuring the core functionality of this project and its application in real-time scenarios. Even with the presence of current limitations like ease of degradations in environments with less lighting and image angle perspectives, this system has a rich architectural foundation and model for efficient automated monitoring and analysis. The project is a rich starting point for further research and development. Now, this project has proved its efficiency and viability regarding the implementation of a simple plate recognition multimodal with a seamless integration of Python and web technologies.

REFERENCES

[1] R. A. Li, "A Multi-Stage Deep Learning Approach for Real-Time Vehicle Detection, Tracking, and Speed Measurement in Intelligent Transportation Systems," Sci Rep, vol. 15, no. https://doi.org/10.1038/s41598-025-07343-5, p. 22531, 2025.

[2] N. M. E.-S. Tabassum, "Vehicle Detection in Adverse Weather: A Multi-Head Attention Approach with Multimodal Fusion," Journal of Low Power Electronics and Applications, vol. 14, no. 2, p. 23, 2024.
[3] Y. X. S. W. P. L. K. S. Z. Z. Q. .. &. H. Q. Wang, "Lightweight Vehicle Detection Based on Improved YOLOv5s," Sensors, vol. 24, no. 4, p. 1182, 2024.
[4] C. C. T. O. T. e. a. Lim, "Visual-Based Vehicle Detection with Adaptive Oversampling," International Journal of Information Technology, vol. 16, no. 8, p. 4767–4777, 2024.
[5] R. &. S. A. Mehta, "An Insight into Real Time Vehicle Detection and Classification Methods using ML/DL based Approach.," Procedia Computer Science, vol. 235, pp. 598-605, 2024.
[6] A. &. Z. G. Gao, "Research on vehicle identification based on multi-modal fusion.," International Journal of Transport Development and Integration, vol. 8, no. 3, pp. 371-382, 2024.
[7] M. N. M. A. H. &. H. D. A. Kadhim, "Vehicle detection and classification from images/videos using deep learning architectures: A survey.," in AIP Conference Proceedings, 2024.
[8] X. H. Z. W. L. C. J. &. T. J. Wu, "Multimodal collaboration networks for geospatial vehicle detection in dense, occluded, and large-scale events.," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, no. 5608012, pp. 1-12, 2024.
[9] M. A. R. Alif, "YOLOv11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems.," arXiv preprint , pp. 1-10, 2024.
[10] Z. A. &. R. S. A. El Shair, "MEVDT: Multi-Modal Event-Based Vehicle Detection and Tracking Dataset," Data in Brief , vol. 58, no. 111205, 2025.
[11] M. S. S. Z. N. &. A. S. Zaman, "Deep Learning Approaches for Vehicle and Pedestrian Detection in Adverse Weather," in 2023 IEEE Transportation Electrification Conference & Expo (ITEC), 2023.
[12] X. L. X. Z. S. &. J. K. Tan, "Integration of Vehicle Target Detection and Recognition in Large-Scene SAR Images Based on YOLOv5," in IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium., 2023.
[13] K. L. J. Y. D. W. H. S. P. Z. Y. L. Y. S. L. Yang, "A Novel Efficient Multi-View Traffic-Related Object Detection Framework," in 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.
[14] V. P. A. G. R. M. &. M. B. M. Agarwal, "Suspicious Vehicle Detection Using License Plate Detection and Facial Feature Recognition," arXiv preprint, no. 2304.14507, pp. 1-5, 2023.
[15] S. L. H. &. L. J. Yang, "Multimodal Fusion for Object Detection in Intelligent Transportation Systems," IEEE Transactions on Intelligent Transportation Systems., vol. 24, no. 1, pp. 1126-1133, 2022.