A Touchless Vehicle Control System Using Deep Learning-Based Hand Gesture Recognition: System Design and Real-Time Implementation

Dr Neethu P S[1,2], Dr Manju Bargavi S[3]

[1]Postdoc Researcher, Lincoln University College, Malaysia

[2]Christ University Bangalore

[3]Jain Deemed to be University, Bangalore

ps.neethu@gmail.com

Abstract: The vehicle control system allows drivers to control essential vehicle features such as turning signals, windshield wipers, headlights, etc., by means of pre-defined hand gestures no physical contact is required with any vehicle controls. A Google MediaPipe-based hand area extractor was used to determine hand areas from video images, and a custom-trained YOLOv11 object detection model was used to classify detected hand areas into one of several hand gestures. In addition to the MediaPipe-based hand area extractor and the YOLOv11 object detection model, two additional components were included in order to provide reliability to both the gesture recognition and mapping to vehicle action. These are: a) Temporal Buffering Mechanism (TBM), which provides increased predictive capability for gesture recognition to minimize false positives due to gesture noise or mis-recognized gestures caused by variability in human motion b) Adaptive Confidence Filtering Strategy (ACFS), which filters out weakly recognized gestures to ensure reliable mapping to vehicle action. Results of real time testing demonstrated that the proposed system effectively identifies and executes gesture commands mapped to vehicle action with acceptable latency. The proposed architecture addresses several significant limitations associated with: a) Driver Distraction b) Accessibility c) Smart Mobility and offers a scalable and practical approach to future generation intelligent vehicle control interfaces.

Keywords- Touchless Vehicle Control, Gesture-Based HMI, Intelligent Transportation, Driver Assistance, Human–Vehicle Interaction, Real-Time Control Systems.

I. INTRODUCTION

Modern cars include many additional electronic systems such as infotainment systems, and driver assistance technology which continue to need user interaction. Therefore traditional input mechanisms (physical button inputs, touchscreen inputs, rotary knob inputs) require both the users visual attention and their manual input to provide a means to interact with these systems therefore increase the likelihood of user distraction while operating the vehicle [1] [3].

User distraction while operating a vehicle is one of the most common causes of road accidents worldwide, therefore developing safer, more intuitive HMI interfaces for vehicle operation is an important area of research and development in Intelligent Transportation Systems.

Touch-less gesture based Human Vehicle Interaction (HVI), provides a new way for a user to provide input to a vehicle using natural hand movements while remaining outside of direct physical contact with any part of the vehicle. This type of user interface will allow for the reduction of cognitive loading on the user, maintain the user's visual focus on the road-way, and also enable users with reduced or no mobility to operate a vehicle [4][7].

Recent advancements in Computer Vision and Embedded Processing have made it possible to develop Real-Time Gesture Recognition systems which can be deployed into vehicles as a safe and viable option towards achieving safer and more accessible vehicle operation.

The objective of this paper is to document the Design and Implementation of a Complete System Design for Real-Time Touch-less Vehicle Control. A Six Hand Gestures are mapped to six Essential Vehicle

Functions. The system utilizes a robust pipeline that involves Media Pipe for Hand Segmentation, YOLO V11 for Gesture Classification, Temporal Buffering for Gesture Timing and Class-Specific Confidence Thresholding. The main Contributions of this Work are:

(1) A Complete End-To-End Gesture To Vehicle Function Pipeline Suitable for Real World Deployment,

(2) A Temporal Buffering Mechanism along with Majority Voting to Ensure Stable Command Execution,

(3) Reliable Experimental Validation of Touch-Less Operation Of Vehicle Functions In Real Time

II. RELATED WORK

A. Gesture-Based Vehicle Interaction

The amount of research into gesture-based driving is quite large due to its potential to enhance safety and reduce distractions from manual input. Systems that use computer vision (CNN) and are based on vision have shown good accuracy in real-time classification of the movement of drivers hands[11][20]. Adding multimodal sensor technologies such as depth cameras and thermal imaging cameras with infrared technology, has increased the robustness of these types of systems in various lighting conditions of the interior cabin environment. There have been several studies showing that gesture interfaces can be used effectively for controlling infotainment systems, adjusting lights, and for driver monitoring applications.[7][12]

B. Deep Learning Models for In-Vehicle Gesture Recognition

YOLO-based architectures are especially well-suited for real-time in-vehicle hand-gesture detection because of how fast they can run and do a complete object detection (i.e., one-stage) [5] [6]. Studies that utilized YOLOv4 [1], YOLOv5 [2][3] and YOLOv8 [17] showed acceptable performance for both driver-assistance systems as well as for autonomous vehicles interacting with people [3]. YOLOv11 represents an advancement in all three of these areas by including new attention mechanisms and better feature-extraction capabilities, thus achieving high accuracy at real time frame-rates [16], [18].

C. System Reliability and Safety Considerations

A major issue when implementing a system using hand gestures (to help a driver with controlling their car) is preventing incorrect activation of safety-related systems, so they are able to function properly. Several techniques were found to be useful ways to reduce the possibility of false positives these include temporal smoothing [13] confidence thresholding [14] or use of a majority vote on a buffer to determine the most likely gesture [14]. Additional layers of reliability can also be built into adaptive threshold-based confidence assessment systems; this would allow for different confidence levels to be set for each type of gesture class, which may require higher thresholds for more dangerous actions, i.e., changing the headlights.

III. SYSTEM ARCHITECTURE

A. System Overview

The overall system has been configured to have four operational modules in a modular sequential pipeline format: (1) media pipe-based video capture and extraction of hand area, (2) identification of hand-gesture through the use of yolo-v-11 for training, (3) Buffering of temporal data and validation of the stability of the identified hand-gesture, and (4) Execution of vehicle-control commands with simultaneous monitoring of real-time system-status.

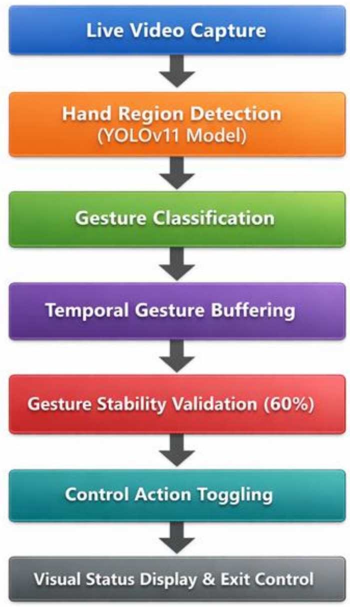

Fig. 1. Block diagram of the proposed real-time hand gesture-based vehicle control system.

B. Gesture-to-Function Mapping

Six hand motions were given unique definitions based on how many fingers are raised (from zero to five). A finger count is a straightforward way to encode the six motions so they can be remembered and easily distinguished from one another. The finger-counting method is summarized in table 1.

Fingers Extended	Gesture Class	Vehicle Function
0 (fist)	Engine Toggle (0)	Engine ON/OFF
1	Indicator Right (1)	Right Turn Indicator
2	Indicator Left (2)	Left Turn Indicator
3	Wiper Control (3)	Wiper Activation
4	Low Beam (4)	Low Beam Headlights
5 (open hand)	High Beam (5)	High Beam Headlights

Table 1. Gesture-to-Vehicle Function Mapping Used in the Proposed System.

C. Hand Region Extraction

The system utilizes a webcam to capture real-time video frames. Each captured frame is then processed utilizing Google MediaPipe's Hand Landmark Detection Module. MediaPipe locates the hand within the video frame and delivers 21 three dimensional hand landmarks for each located hand. These landmarks are used to generate an area of interest (a bounding rectangle) around the hand. Then all non-hand pixels in the video frame are set to black; effectively removing background noise prior to classifying. Preprocessing has proven to be instrumental in increasing the dependability of subsequent gesture classifications. By masking all non-hand pixels to black prior to classification, this ensures that the

YOLOv11 object detection model will process only the hand in question thus decreasing false positives resulting from complex backgrounds.

D. Gesture Classification Pipeline

The Masked Hand Region (MHR) that was detected in the previous stage is then processed using the Gesture Classification module where it is sent to a pre-trained version of the YOLOv11 algorithm for gesture identification. Each MHR will generate a prediction for which of the six possible classes or actions it is most likely representing. Alongside this prediction for which action has been identified, the YOLOv11 algorithm also generates a probability score related to how confident the model is about its prediction. A minimum confidence score for each individual class (i.e., "Engine Toggle" = Class 0, "Right Indicator" = Class 1, "Left Indicator" = Class 2, "Wipe Windshield" = Class 3, "Low Beam Headlights" = Class 4, "High Beam Headlights" = Class 5) is set based on the average level of confidence generated when predicting these actions. This allows non-relevant detections from background noise to be eliminated prior to sending them to the Buffering Module. For example, a confidence score of at least .50 (.5) is required for detection of "Engine Toggle," "Right Indicator," and "Left Indicator." However, because of the similar appearance of the Wiper Gesture and other Gestures in the environment; the confidence threshold for detecting "Wipe Windshield" is set at a lower value of .20 (.2). Finally, a much higher confidence score of .80 (.8) is required for detection of both "Low Beam Headlights" and "High Beam Headlights" to ensure that false activation does not occur when operating safety critical functions such as lighting controls.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Real-Time Implementation Results

The entire system is fully deployed and has been tested in an indoor lab setting that simulates vehicle operation under actual time constraints. The system was developed using the Python programming language as well as Visual Studio Code and tested to various predetermined hand gestures based on session-to-session variations of lighting and background. Figure 2 displays some real-time examples of the systems performance during evaluation by showing how it detects and executes control commands for each of several types of gestures.

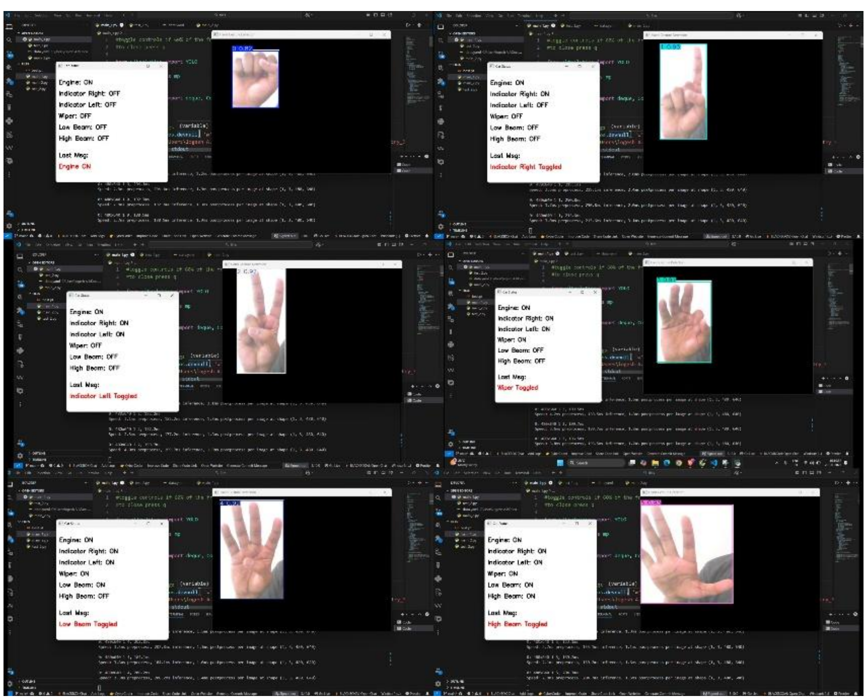

Fig. 2. Real-time implementation results of the proposed hand gesture-based vehicle control system showing gesture detection, bounding box overlay, and vehicle status display.

The system (as depicted in FIGURE 2) has been able to recognize and classify all gestures as expected, using an accurate bounding box and also providing high confidence scores for these classifications. The Vehicle Status Display will update automatically based on the current state of each controlled function. These results confirm that the system will respond appropriately and consistently to the input from a user's gestures with virtually no delay or latency between presenting a gesture and updating the control state. The inclusion of the MediaPipe Preprocessing Stage was found to greatly reduce false detection, resulting from background noise, when utilizing Full-Frame Inference directly.

B. Performance Evaluation

The average precision (mAP@0.5) of the YOLOv11-trained gesture recognition system was about 0.91 in the test phase. Therefore, the reliable classification of gesture control systems has been demonstrated at a system level over all six gesture-vehicle function assignments. All gesture-to-vehicle function assignments have acceptable precision and recall values to be useful in a practical application. The precision value (90%) and the F1 score (85%) were found to be the greatest among all gesture control applications. The lowest precision and F1 scores occurred from Wiper Control, primarily because of ambiguities in visual detection near adjacent gestures. Due to the strict confidence threshold, there is virtually no chance of false activation when using Low Beam or High Beam. This method maintains acceptable recall values and therefore demonstrates a total precision of 86% and F1 score of 78%, which validates the potential use of this system as a means of real time vehicle interaction.

C. System Reliability and Robustness

The 60 percent majority vote mechanism and the Temporal Buffer Mechanism reduced false control commands (i.e. false positive) caused by brief misclassification events (e.g. as a result of movement while repositioning). In the absence of the buffer, short term misclassification would have resulted in transient activation of the control. The System demonstrated consistent and correct execution of the input commands for an extended period of time when using both mechanisms under adverse test conditions; i.e., partial obstruction of the hand by clothing or objects within the cabin, rapid transitions between gestures, and low light levels.

V. CONCLUSION

This research presents a full design and the live test of a touchless vehicle control system based on hand gesture recognition using a deep learning model. In this study we integrated MediaPipe hand segmentation and YOLOv11 gesture classification into the same pipeline as well as adding a temporal buffer, and an adaptive filter for determining the level of confidence in detected gestures. The overall pipeline provides a reliable method for mapping six predefined gestures to fundamental vehicle control inputs. Experimental results demonstrated that the system successfully detects gesture input commands and executes the corresponding vehicle control action(s), while minimizing response time under various light levels and background conditions.

Overall, the entire system achieved a mean Average Precision at 50% recall = 0.91 and a total F1-score = 78%. This demonstrates the ability to identify a wide variety of hand gestures effectively enough to deploy in practice. Additionally, the proposed touchless vehicle control method can enhance driver safety by eliminating the need for manual interaction with vehicle controls; especially, for individuals with mobility issues. The proposed system provides a flexible, modular, and affordable architecture for the integration of future generations of intelligent vehicle systems.

In terms of future research directions, we plan to investigate how to integrate the current prototype with appropriate embedded automotive hardware (i.e., Raspberry Pi, NVIDIA Jetson Nano) to enable deployment within an automobile. We also intend to expand our set of recognized gesture commands and evaluate the feasibility of implementing multimodal voice/gesture control. Finally, we will explore ways to improve system performance over a greater number of diverse real-world driving environments.

REFERENCES

[1] Vijayalakshmi et al., "A Novel Method for Gesture Recognition in Autonomous Driving Using Pose Estimation," 2023.

[2] L. Hao et al., "A lightweight network for driver gesture recognition," Computers & Electrical Engineering, 2025.

[3] Ravali et al., "Hand Gesture Recognition System Using Transfer Learning," IEEE, 2023.

[4] Suresha et al., "Driver Hand Gesture Detection and Recognition in Road Scenarios," 2022.

[5] A. Rangesh et al., "Hidden Hands: Tracking Hands with an Occlusion Aware Tracker," IEEE, 2016.

[6] Y. Zhuang et al., "Real-Time Gesture Recognition Based on YOLOv4," Applied Sciences, 2021.

[7] S. Chua et al., "Hand gesture control using deep learning," 2022.

[8] A. D'Eusanio et al., "Multimodal Hand Gesture Classification for Human–Car Interaction," 2020.

[9] A. Rangkuti et al., "Optimizing Hand Gesture Recognition Using CNN Model," 2023.

[10] Siddharth et al., "Driver hand localization and grasp analysis," 2018.

[11] Pansare et al., "Computer vision techniques for real-time hand gesture detection," 2021.

[12] O. Köpüklü et al., "DriverMHG: Driver Micro Hand Gesture Dataset," IEEE, 2020.

[13] Furquan et al., "Deep Learning for Recognizing Gesture States," 2023.

[14] A. Acevedo-Bringas et al., "YOLOv5 for Efficient Hand Gesture Recognition," 2023.

[15] K. Yuen et al., "ConvNet approach for hand detection in vehicles," IEEE, 2020.

[16] J. Schulte et al., "Gesture Recognition for Autonomous Vehicle Control," 2022.

[17] Setyaji et al., "YOLOv8 Multi-Detection for Vehicle Interaction," IEEE, 2024.

[18] Y. Wang et al., "Online Gesture Recognition for Smart Driving Systems," 2019.

[19] P. Molchanov et al., "Multi-sensor driver hand gesture recognition," IEEE, 2015.

[20] V. John et al., "Touchless automotive gesture user interface using deep learning," IEEE, 2017.