

# AI-Based Traffic Police Gesture Recognition System

Mr. Shreyan Jain<sup>1</sup>, Gurukiran C S<sup>2</sup>, Darshan A J<sup>3</sup>, Lakshmish K S<sup>4</sup>, Gowtham D K<sup>5</sup>

<sup>1</sup>*Assistant Professor, Artificial Intelligence & Machine Learning, Bahubali College of Engineering*

<sup>2,3,4,5</sup>*Computer Science & Engineering, Bahubali College of Engineering, Shravanabelagola.*

**Abstract**— Manual traffic control using hand gestures is still widely practiced in urban areas, especially during peak hours, emergencies, or signal malfunctions. Interpreting these gestures automatically is essential for intelligent transportation systems and autonomous vehicles. This project presents a real-time traffic police hand gesture recognition system developed using deep learning techniques. MediaPipe is employed to extract human pose and skeletal keypoints from live video streams, while a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units analyzes temporal motion patterns for accurate gesture classification. The proposed system demonstrates reliable real-time performance with low latency and high recognition accuracy under varying environmental conditions. The results indicate that the system can effectively support intelligent traffic management and autonomous driving applications.

**Index Terms**—Traffic Police Gesture Recognition, Computer Vision, MediaPipe, RNN, LSTM, Deep Learning, Intelligent Transportation Systems, Real-Time Processing.

## I. INTRODUCTION

Gesture recognition is a key domain within computer vision that focuses on enabling machines to interpret human movements and actions. It has gained significant importance due to its applications in areas such as human-computer interaction, surveillance systems, healthcare monitoring, and autonomous driving. In real-world traffic environments, traffic police officers frequently use hand gestures to regulate vehicle flow when automated signals are unavailable or insufficient.

Despite advancements in vision-based systems, accurately recognizing traffic police gestures remains challenging due to differences in gesture execution, occlusion, lighting variations, and complex

backgrounds. Traditional image-based methods often fail to capture the temporal nature of gestures. To address these challenges, deep learning-based sequence models are increasingly being adopted. This project proposes an AI-based traffic police hand gesture recognition system that operates in real time. By combining MediaPipe-based pose estimation with an RNN-LSTM architecture, the system captures both spatial and temporal characteristics of gestures. The objective is to provide a reliable solution that can assist intelligent traffic control systems and autonomous vehicles, thereby improving road safety and traffic efficiency.

## II. LITERATURE SURVEY

Recent studies in gesture recognition have demonstrated the effectiveness of deep learning models for analyzing human motion patterns. Pose estimation frameworks such as MediaPipe and OpenPose have been widely used to extract skeletal keypoints, enabling robust gesture representation independent of background noise.

Several researchers have shown that Recurrent Neural Networks, particularly LSTM networks, are well-suited for sequential data analysis due to their ability to retain temporal information. These models have been successfully applied to human activity recognition and hand gesture classification tasks. Additionally, attention mechanisms have been introduced in some works to enhance model focus on critical joints and motion segments, resulting in improved recognition accuracy. In the context of intelligent transportation systems, gesture recognition has been explored as a complementary control mechanism for autonomous vehicles and smart traffic monitoring. Existing research highlights the need for real-time, accurate, and computationally efficient

models capable of handling diverse real-world conditions. These findings motivate the adoption of MediaPipe and LSTM-based architectures in the proposed system.

### III. METHODOLOGY

The proposed system follows a structured pipeline consisting of data acquisition, keypoint extraction, gesture classification, and result visualization. The overall workflow is designed to ensure real-time performance and robustness

#### A. System Architecture

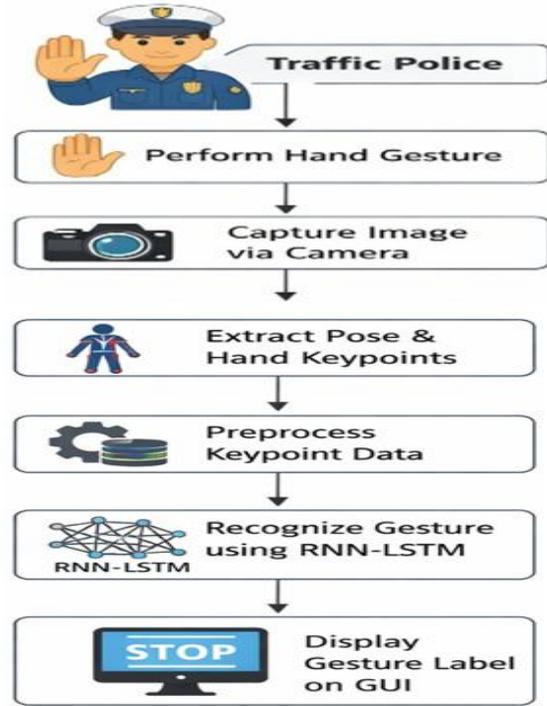
A camera captures live video input containing traffic police gestures. Each frame is processed using MediaPipe to detect human pose landmarks and skeletal keypoints. The extracted keypoints are normalized and arranged into sequential data representing motion over time. An RNN with LSTM layers processes the sequential keypoint data to learn temporal dependencies associated with different gestures. A multi-head attention mechanism is incorporated to emphasize important joints and motion patterns. The trained model outputs the predicted gesture class, which is forwarded to the user interface for display.

#### B. Voice Announcement & GUI

The graphical user interface (GUI) provides real-time visualization of recognized gestures. After gesture classification; the predicted label is sent to the frontend through a backend server. The GUI displays the gesture name clearly for user understanding. The interface updates dynamically for each recognition cycle. User interaction is simplified through start and reset controls. This visual feedback enables easy monitoring and verification of gesture recognition results.

### IV. UNITS

The methodology follows a systematic approach to integrate real-time detection with



### VI. TESTING

Test Case No	Input	Expected Output	Actual Output	Remarks
1	Hand gesture shown to camera	Gesture detected	Gesture detected	Pass
2	Stop gesture performed	“Stop” label displayed	“Stop” displayed	Pass
3	Invalid gesture	No classification	No classification	Pass
4	Multiple frames input	Stable prediction	Stable prediction	Pass
5	Continuous operation	No system crash	No system crash	Pass

## VII. FUTURE ENHANCEMENT

Future work may involve extending the system to recognize a broader range of traffic police gestures. Integration with smart traffic infrastructure and autonomous vehicle control systems can further enhance real-world applicability. Additionally, incorporating multimodal inputs such as depth sensors may improve accuracy under complex scenarios.

## VIII. CONCLUSION

This project presents an AI-based traffic police hand gesture recognition system using MediaPipe and RNN-LSTM networks. By effectively modeling temporal gesture patterns, the proposed approach achieves accurate and real-time recognition. The system demonstrates strong potential for deployment in intelligent traffic management systems and autonomous driving environments, contributing to improved road safety and traffic efficiency.

## REFERENCES:

- [1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE TPAMI*, 2019.
- [2] WF. Zhang et al., "Human gesture recognition using deep learning," *Pattern Recognition Letters*, 2020.
- [3] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
- [4] Xu, L., Liu, J., Zhao, H., Zheng, T., Jiang, T., & Liu, L. 2024. "Autonomous Navigation of Unmanned Vehicle Through Deep Reinforcement Learning" *arXiv:2407.18962*.
- [5] Bathla, Gourav, Kishor Bhadane, Rahul Kumar Singh, Rajneesh Kumar, Rajanikanth Aluvalu, Rajalakshmi Krishnamurthi, Adarsh Kumar, R. N. Thakur, and Shakila Basheer. "Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities." *Mobile Information Systems* 2022, no. 1 (2022):7632892.
- [6] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2021). OpenPose: Realtime MultiPerson 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172-186.
- [7] de Miguel, M. Á., García, F., & Armingol, J. M. 2020. "Improved LiDAR Probabilistic Localization for Autonomous Vehicles Using GNSS." *Sensors* 20(11): 3145.
- [8] Ibrahim, H. A., Azar, A. T., Ibrahim, Z. F., & Ammar, H. H. 2020. "A Hybrid Deep Learning-Based Autonomous Vehicle Navigation and Obstacles Avoidance." In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, 296-307. Springer International Publishing.
- [9] T. Kim, Y. Lee, J. Park and H. -T. Choi, "A study on the effectiveness of MOOS-IvP for Unmanned Surface Vehicle," 2017 IEEE.
- [10] Jasim, A. M., & Mohd, B. D. (2019). A Survey of Hand Gesture Recognition Using Machine Learning Techniques. *International Journal of Advanced Research in Computer Science*, 10(2), 24-34.
- [11] A. Anis, "Development of an ultrasonic cane as a navigation aid for the blind people," in *Proc. Int. Conf. Control, Instrum., Commun. Comput. Technol. (ICCICCT)*, Jul. 2014, pp. 475-479.
- [12] C. T. Patel, V. J. Mistry, L. S. Desai, and Y. K. Meghrajani, "Multisensor- based object detection in indoor environment for visually impaired people," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2018, pp. 1-4.
- [13] Z. Bauer, A. Dominguez, E. Cruz, F. Gomez-Donoso, S. Orts- Escolano, and M. Cazorla, "Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors," *Pattern Recognit. Lett.*, vol. 137, pp. 27-36, Sep 2020.