

Explanation of Gesture Recognition and Vehicle Control Mapping

The provided system implements real-time gesture recognition using a webcam and maps the detected gestures to vehicle control commands in the CARLA simulator. Gesture detection is powered by **MediaPipe Face Mesh**, which tracks 468 facial landmarks. Key facial features such as the **mouth**, **eyebrows**, **eyes**, **cheeks**, and **nose tip** are used to infer specific gestures. **For example, mouth openness is measured by the distance between the upper and lower lips, frowning is detected based on the distance between the eyebrows and the eyes, and head tilt is calculated using the vertical position difference between the cheeks. Nodding gestures (up or down) are inferred by monitoring the vertical change in the position of the nose tip over time.**

Each gesture is then mapped to a specific vehicle control action in CARLA. **Mouth open** triggers acceleration (throttle), **frowning** activates the brake, and **head tilt** to the left or right is translated into left or right steering, with two sensitivity levels depending on the degree of tilt. Nodding **up** increases the gear, and **nodding down** decreases it, enabling forward and reverse driving modes. These commands are sent to the vehicle in real-time using CARLA's Python API, and a **VehicleControl** object is used to update the vehicle's state.

Additionally, the system logs all detected gestures and control states (throttle, brake, steer, gear) to a CSV file, providing traceability and aiding in debugging or analysis. This tight loop of perception and actuation creates a responsive and interactive driving experience controlled entirely through facial expressions and head movements.