MULTIMODAL AI FOR AUTONOMOUS CARS

***Project Reference No.:** 48S_BE_5896*

College : B.M.S. College Of Engineering, Bengaluru
Branch : Department Of Artificial Intelligence And Machine Learning
Guide(S) : Dr. Sandeep Varma N
Student(S): Ms. Ilaa S Chenjeri
Ms. Anushka Rajgiri
Ms. Dyuthi B V

Keywords:

Multimodal AI, Autonomous vehicle, Sensor Fusion, RGB-D Input, Object/Obstacle Detection, Path Planning, LIDAR, Depth Sensing, Early Fusion, End-to-End Learning, Explainable AI (XAI), Multisensor Integration, AI Model Evaluation, Navigation in Dynamic Environments, Neural Network-based Driving, Low-latency Systems, Multimodal Input Processing.

Introduction:

The rise of Artificial Intelligence (AI) has significantly impacted various sectors, including transportation. Autonomous vehicles are at the forefront of this revolution, aiming to ensure safer, smarter, and more efficient mobility. The primary challenge in autonomous driving lies in perceiving the environment accurately and making real-time decisions that align with human expectations.

This project introduces a novel approach to autonomous driving by implementing Multimodal Artificial Intelligence—a system that fuses information from multiple sensors such as RGB cameras and LiDAR to develop a deeper contextual understanding of the vehicle's surroundings. The use of multimodal data supports better object detection, trajectory planning, and situational awareness.

A unique aspect of the project is a live demonstration bot that mirrors real-life scenarios such as self-start, obstacle recognition, traffic rule adherence, and destination halting. The bot represents the physical manifestation of the system's capabilities and highlights its real-world potential. This combination of simulation and physical modeling bridges the gap between abstract AI algorithms and tangible engineering applications.

Objectives:

- To design and implement a Multimodal AI system using deep learning, sensor fusion, and computer vision.
- To develop a fully autonomous bot that navigates, detects obstacles, avoids collisions, and responds to environmental changes.
- To improve real-time decision-making and reduce accident likelihood using multimodal sensor data.
- To simulate and test real-world driving conditions in a controlled environment to evaluate system adaptability.
- To foster advancements in intelligent transportation systems by enhancing user trust, safety, and usability.

Methodology:

Phase 1: Dataset Selection and Hardware Setup

Identifying suitable datasets for training multimodal AI models, focusing on RGB and depth data. Selecting the Jetson Nano Developer Kit as the core hardware platform for its GPU capabilities and compatibility with edge AI tasks. Choosing sensors including an RGB camera and a Depth sensor, along with motors, power supply, and basic electronics for the bot. Setting up the hardware infrastructure for both data collection and live deployment.

Phase 2: Early Fusion Strategy Implementation

Attempting early fusion of RGB and Depth data at the input stage. Combining the two modalities into a single input tensor to allow the neural network to learn shared features from both sources. This fusion is enhancing the model’s ability to perceive depth and color context simultaneously, improving environmental awareness.

Phase 3: Simulation on CARLA

Before physical testing, the AI models are being evaluated in the CARLA simulator, which provides high-fidelity autonomous driving scenarios. Testing for various situations like obstacle presence, traffic signal handling, and sharp turns. Simulation results are helping refine the model parameters and improve robustness. YOLOv3 is being used for object detection due to its speed and accuracy. It is processing RGB and Depth inputs to detect obstacles, traffic signs, and other elements in real time. The model is first being tested in the CARLA simulator to ensure reliability before being integrated into the bot for real-world navigation and decision-making.

Phase 4: Bot Design and Building

A physical autonomous bot is being built using the Jetson Nano as the main processing unit. The bot's movement is being controlled using a motor driver connected to DC motors, while an ESP32 microcontroller is handling communication and sensor inputs. Basic components like a chassis, wheels, power supply, and wiring are being assembled to create a stable platform. The bot is being designed to run the trained AI model for real-time navigation, obstacle avoidance, and path following.

Result and Conclusion:

The multimodal approach using RGB-D input yielded significantly better performance in object recognition and decision accuracy compared to single-modal methods. Early fusion enabled the neural network to learn richer features, leading to more accurate path planning and obstacle prediction. The bot successfully demonstrated tasks like autonomous navigation, real-time traffic response, and accurate stopping, validating the project's hypotheses.

In conclusion, Multimodal AI significantly enhances the environmental understanding of autonomous systems. Fusing depth with RGB input increases contextual awareness, resulting in better decision-making and safety. The live bot demonstration proved the feasibility of integrating such systems into small-scale vehicles, indicating promising future applications in full-scale autonomous driving systems.

Future Scope:

- **Real-World Integration:** Expand the model to operate on actual vehicles with real-world road conditions and traffic scenarios.
- **Scalability:** Implement the system in urban autonomous vehicles and adapt to dynamic environments using 5G-enabled communication.

- **Advanced Sensor Fusion:** Integrate GPS, IMU, and radar data to improve redundancy and robustness.
- **Sustainability:** Optimize energy consumption using AI-based route optimization and minimal hardware load.
- **Extended Testing:** Run the bot in diverse terrains and climates to test generalizability.
- **Explainable AI Integration:** Add modules for visualizing AI decisions in real-time to build user trust and transparency.