Urban Mobility Perception System

Detection, Tracking, and Distance Estimation using Camera and LiDAR

Submitted by: Vishwas Tokas
Phone no: 9577744000

Abstract

This project describes the realization of an efficient urban mobility perception system for detection, tracking, and distance estimation of road users based on images and LiDAR data. The system integrates deep learning for object detection, multi-object tracking, and LiDAR distance estimation, and also takes care of real-world data limitations. The system has been tested on the KITTI Tracking benchmark, which has generated an annotated video as an output in almost real-time. The project has focused on robustness, realism, and practicability rather than accuracy.

---

1. Introduction

The traffic scenario in the urban setting consists of various dynamic components. These include automobiles in the city as well as pedestrians crossing the road. For designing intelligent transportation systems as well as self-driving technology, it is necessary to track these components as well as calculate the distance of these components from the self-driving car.
The purpose of this project is to provide a comprehensive perception system that is capable of implementing these fundamental skills while dealing with real-world issues such as missing sensor data. This is done with a focus of developing a system that can work as intended while being easy to understand.

---

2. System Overview

The perception pipeline is composed of four main stages:

1. **Object Detection** – identifies traffic participants in each camera frame
2. **Multi-Object Tracking** – assigns consistent identities across frames
3. **Distance Estimation** – estimates object distance using LiDAR data
4. **Visualization and Output** – displays results and saves annotated video

Each component is implemented in a modular manner, making the system easier to debug, evaluate, and extend.

---

3. Object Detection

Object detection is accomplished with the help of the YOLOv8n pretrained model and the COCO dataset. YOLOv8 models were selected based on the optimal balance between accuracy and efficiency, making them suitable for real-time tasks such as object detection.
Others
The detector deals with generic urban object detection such as cars, trucks, buses, motorcycles, and pedestrians. This ensures the detection of relevant objects in the image, thus yielding a noise-free final output.

---

4. Tracker Selection

Choosing a good track algorithm is very important for keeping track of object identity across images.
SORT was not selected because it is based solely on information about motion and tends to have a problem with a crowded environment or during occlusion, leading to a high chance of ID switches. DeepSORT is an advancement of SORT because it utilizes features that are based on both motion and appearance. However, it is more complex.

ByteTrack has been chosen as an alternative tracker that balances well. It also deals well with short trackers by utilizing high and low-confident detection results without the use of explicit appearance embeddings. ByteTrack can, therefore,be considered suitable for real-time urban mobility applications and for implementation in the project.

---

5. Sensor Fusion Logic

The system performs **late fusion** between camera-based perception and LiDAR-based distance estimation.

In an ideal setup, LiDAR points are projected into the image plane using camera intrinsics and extrinsics. However, the KITTI Tracking dataset does not provide LiDAR–camera extrinsic calibration on a per-sequence basis. As a result, precise 3D-to-2D projection is not feasible.

To address this limitation, an approximate yet stable approach is adopted:

- The forward axis of the LiDAR sensor is used as a proxy for object depth
- Median distance is computed to reduce the impact of noise
- Distance information is associated with objects using persistent tracking IDs

This approach prioritizes robustness and transparency, allowing the system to function reliably despite incomplete calibration.

---

6. Handling Missing LiDAR Data

The data from the LiDAR scanner is intermittent, so not every image scan is accompanied by a scan from the LiDAR scanner. To stabilize its operation, a distance cache is maintained for each object.
In case LiDAR data is not available, then_ranges are set based on the known distance of each targeted object. The distance values are then smoothened using an exponential moving average function so that there are no abrupt variations.

---

7. Results and Observations

The system designed works successfully to identify and track multiple users of the road efficiently by keeping a stable tracking ID. The estimated distances work continuously even without LiDAR data. The system runs near real-time processing on the CPU hardware and provides a clear understanding of the visualization.

---

8. Failure Cases

Several failure cases were observed during experimentation:

- Missing LiDAR frames, addressed through distance caching
- Occasional ID switches during long occlusions
- Approximate distance estimation due to missing calibration
- Reduced detection stability for objects at long range

Despite these limitations, the system remains stable and performs reliably under realistic conditions.

---

9. Conclusion

The proposed project undertakes a comprehensive demonstration of a functional urban mobility perception. The project area of interest has been restricted to the provision of a realistic engineering solution. The perception model has taken into consideration the limitations of the dataset. Thus, the project serves as a good foundation or starting point for further research in the area of autonomous vehicles.
