Stefano Ruggiero – Project 2

Sensor Fusion and Multi-Object Tracking with Kalman Filter and Parameter Optimization on NuScenes

Objective

The goal of this project is to design, implement, and evaluate a multi-object tracking system that fuses LiDAR and camera data using a Kalman filter. The proposed pipeline processes data from the NuScenes dataset to estimate and track the positions of multiple objects in dynamic urban environments. The project also includes the optimization of parameters to maximize tracking accuracy.

Pipeline Overview

1. Data Loading and Calibration

The system operates on the NuScenes mini dataset, which provides synchronized LiDAR point clouds and camera images for each frame. A calibration procedure transforms the LiDAR point cloud into the camera coordinate system using extrinsic and intrinsic parameters.

2. Bounding Box Extraction and Point Clustering

For each frame, 3D bounding boxes for all annotated objects are projected onto the camera image. LiDAR points are then projected into the same image space, and for each bounding box, a cluster is formed by selecting points falling within its polygonal area. This allows associating LiDAR measurements to camera-detected objects.

3. Measurement Fusion

For each object, a fused measurement is computed by taking a weighted average of the camera bounding box center and the centroid of the corresponding LiDAR cluster. The weight is determined by the density of LiDAR points within the box: when many points are present, the estimate relies more on LiDAR, otherwise it defaults toward the camera. This adaptive weighting makes the fusion robust to sensor sparsity.

4. Multi-Object Tracking

Each detected object is assigned a tracker based on a constant velocity Kalman filter with a 2D state (position and velocity). At each time step, the filter predicts the next state for all existing trackers. Measurements are associated to trackers using the Hungarian algorithm. Trackers are updated with matched measurements, new trackers are created for unmatched measurements, and trackers are deleted if they are not updated for several consecutive frames.

5. Parameter Optimization

The most relevant hyperparameters of the Kalman filter (process noise, measurement noise, initial covariance), as well as thresholds for measurement association and fusion weighting, are optimized using Optuna, a Bayesian optimization framework. The objective function minimizes the average RMSE (Root Mean Squared Error) between predicted tracker positions and reference object positions across all scenes.

6. Evaluation

Tracking performance is assessed for each scene by computing RMSE between the positions estimated by the trackers and the ground-truth bounding box centers at each frame (both per-frame and overall per-scene RMSE are reported to quantify quality of predictions and robustness). The implementation includes visualization tools to display LiDAR points, projected bounding boxes, and tracker predictions on camera images. These tools have been essential for qualitative assessment.

Results

The following table summarizes the overall RMSE for each scene:

Scene	RMSE
0	7.04
1	7.74
2	10.05
3	7.28

Scene	RMSE
4	32.27
5	2.34
6	14.37
7	2.65

Scene	RMSE
8	10.62
9	11.20

The system achieves good tracking in the majority of scenes, with several sequences (e.g., scenes 5 and 7) showing low RMSE and stable behavior. Some scenes (e.g., scene 4 and scene 6) show higher RMSE due to increased object density, occlusions, or challenging association scenarios. For each scene, the per-frame RMSE is also computed to highlight tracking consistency and the presence of outliers.

Analysis

- The fusion approach dynamically adjusts the confidence placed in each sensor, improving overall reliability in diverse conditions.
- Automatic parameter optimization via Optuna significantly improves performance, demonstrating that careful tuning is crucial for real-world, multi-sensor tracking.
- Performance degrades in scenes with many overlapping objects, sparse LiDAR data, or severe occlusions, primarily due to data association challenges or lack of strong measurements for the filter.
- The approach is modular and could be extended with more advanced measurment-level fusion strategies (different type of sum between LiDAR and Camera for Measurement Fusion) to further improve robustness in complex scenarios, or using an Extend Kalman Filter (for non linearity of the system).

Conclusion

The developed pipeline fuses LiDAR and camera data for multi-object tracking using a Kalman filter. By leveraging calibration, adaptive measurement fusion, and parameter optimization, the system achieves good performance in the majority of the urban driving scenarios. The results highlight both the strengths and current limitations of this approach.