Multi-Object Tracking and LiDAR–Camera Fusion on KITTI Dataset

1. Goal of the Project

The goal of this project was to build a simple system that:

- finds cars and people in camera images
- gives each object an ID so we can follow it across frames
- uses LiDAR data to measure how far each object is
- shows everything together in one output video

This type of system is used in self-driving cars.

2. Dataset and Camera Setup

I used the **KITTI Tracking Dataset**.
Each frame has:

- a camera image
- a LiDAR scan
- calibration data (to link LiDAR and camera)

The camera is mounted on a moving car.
The LiDAR gives 3D points of the world.
The calibration information helps match LiDAR points to pixels in the image.

Only LiDAR points that are in front of the car and inside the image were used.

3. How the System Works

The system has four parts.

3.1 Detecting Objects

I used YOLOv8 (a pre-trained object detection model) to detect objects in each image.

I only kept:

- **Cars**
- **Persons / Pedestrians**

Other classes (like bicycle, truck, dog, etc.) were removed.

Very small or low-confidence detections were also removed to reduce noise.

3.2 Tracking Objects

To keep the same ID across frames, I used a simple tracking method.

It compares bounding boxes between frames.
If a box in the new frame matches a box in the last frame, it keeps the same ID.
If not, a new ID is created.

This keeps track of moving objects over time.

3.3 Estimating Distance Using LiDAR

LiDAR points are projected into the camera view.

For each object box:

- I take the LiDAR points inside that box
- I estimate the distance using those points

This gives a real-world distance in meters.

Outliers and impossible values are ignored.

3.4 Final Output Video

In the final video, each object shows:

- a bounding box
- an ID
- a label (Car or Pedestrian)
- distance in meters

For example:

ID 3 | Car | 14.2 m

ID 1 | Pedestrian | 7.8 m

4. Results

What Worked Well

- Cars and pedestrians were detected correctly in most frames
- Objects usually kept the same ID for several frames
- Distance values looked realistic
- The final video clearly shows detection + tracking + distance

Performance

The system runs in real time or near real time, depending on hardware.

5. Limitations

This is a simple baseline system.
Some problems still happen:

- Sometimes IDs change when objects overlap
- There are still some false detections
- Distance is less accurate far away
- The tracker is simple and does not use motion history
- The model is not trained on KITTI, so accuracy is not perfect

These issues are normal for this type of project.

6. Possible Future Improvements

If I had more time, I would:

- use a stronger tracker like ByteTrack or DeepSORT
- smooth distance estimates over time
- fine-tune the detector on KITTI
- improve handling of occlusion
- use 3D bounding boxes

7. Conclusion

This project shows a full working pipeline for:

- detecting cars and pedestrians
- tracking them with IDs
- using LiDAR to measure distance
- showing the results in a video

Even though it is not perfect, it clearly demonstrates how camera and LiDAR can work together for self-driving perception.

The main learning goals of the assignment were achieved.

8. References

- KITTI Dataset
- Ultralytics YOLOv8
- Public tracking research papers