Fusion of Information from 2D LiDAR and RGB Camera

Anoop Kumar Singh
Mechanical Engineering
Indian Institute of Technology Jodhpur
Jodhpur, India
m24mea001@iitj.ac.in

Love Kumar Saini
Mechanical Engineering
Indian Institute of Technology Jodhpur
Jodhpur, India
m24mea006@iitj.ac.in

***Abstract*—The automotive industry is currently undergoing a significant transformation with the rapid integration of autonomous systems. At the heart of safe navigation in these systems lies object detection, where Convolutional Neural Networks (CNNs) have become a fundamental technology. Our project focuses on the fusion of 2D LiDAR and RGB image data specifically for object detection in autonomous vehicles. We have implemented the YOLOv5 algorithm on the KITTI dataset, which provides 2D LiDAR point clouds and RGB images of various road scenes. This model is engineered to deliver high accuracy in detecting and localizing a wide range of objects in real-time. Its detection capabilities extend beyond vehicles to include pedestrians, cyclists, obstacles, and other critical elements essential for safe autonomous navigation. Leveraging the strengths of deep learning, our work aims to advance the state-of-the-art in object detection for self-driving applications.**

***Index Terms*—LiDAR, RGB Camera, Robotics**

I. Introduction

Integrating 2D LiDAR and RGB camera data has emerged as a powerful approach to enhance environmental perception in mobile robotics. While 2D LiDARs have traditionally offered reliable depth measurements, their output often results in sparse spatial maps. Conversely, RGB images provide rich visual context but lack direct depth information. By leveraging the precise depth data from LiDAR and the detailed contextual information from RGB images, this study aims to bridge the gap between the two modalities, ultimately enhancing the density and completeness of LiDAR-based maps.

This fusion technique is geared toward improving the accuracy of depth prediction—an essential requirement for achieving autonomy in both robotics and vehicular systems. A notable advancement in this domain is the introduction of BEVFusion, as proposed in recent research [1]. BEVFusion is a novel LiDAR-camera fusion framework tailored for 3D object detection in autonomous driving applications. Notably, it maintains functionality even in the absence of LiDAR input, offering superior performance and robustness compared to existing methods.

The architecture comprises two parallel streams that independently encode camera and LiDAR data into Bird's Eye View (BEV) features. These features are then fused to enhance perception accuracy. BEVFusion demonstrates strong potential for real-world deployment, with its ability to maintain high performance under typical conditions, robustly handle various LiDAR failures, and generalize across multiple model architectures without requiring additional post-processing.

In another paper, it proposes a novel technique for 3D object detection and localization using only a single image [2]. Unlike existing methods focusing solely on predicting object orientation, this approach combines deep CNN predictions with geometric constraints from 2D bounding boxes to generate precise 3D bounding boxes.

II. Theory

A. LiDAR Sensor

LiDAR sensors work by emitting laser pulses and measuring the time it takes for the pulses to return after hitting objects in the environment. By analyzing the timing and direction of these pulses, LiDAR sensors can generate detailed 3D point clouds, where each point represents a reflection from a surface in the sensor's field of view.

These point clouds provide rich spatial information about the surrounding environment, including the positions and shapes of objects. In the data collection for the KITTI dataset, Velodyne LiDAR sensors are used [3]. Velodyne LiDAR sensors are commonly used in various applications such as autonomous vehicles, robotics, and 3D mapping.

B. YOLO Algorithm

In our project, we have tried using the YOLO algorithm for object detection. When it comes to object detection, we don't only have to identify the object, but we need to locate the object as well. We need to figure out the object's size too; that's why we have b_x, b_y that

describes the centroid of the bounding box and b_h, b_w describe the height and width of the bounding box.

Suppose we want to check whether the given image has a pedestrian, motorcycle, or car. First, we check if any of these objects are present, if so, $p = 1$, otherwise $p = 0$. Then we check which of the classes (c_1, c_2, c_3) is detected. After identification, the label turns out to be:

$$Y = [p_c, b_x, b_y, b_h, b_w, c]$$

To evaluate the object localization, we use the Intersection over Union (IoU) method, where we take the ratio of the intersection of the predicted and ground truth bounding boxes to their union. If IoU is greater than 0.5, the prediction is considered good.

Also, the algorithm may detect the same object multiple times. Non-Max Suppression (NMS) is used to suppress all bounding boxes with lower IoUs than the highest.

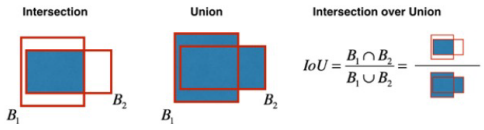

Fig. 1: Intersection over Union

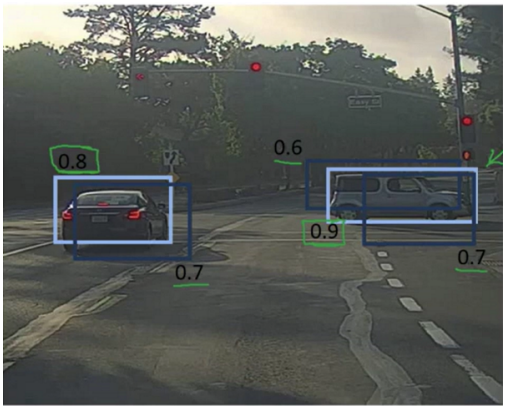

Fig. 2: Non-max suppression

III. Method

A. Data Preprocessing and LiDAR Integration

Our proposed method demonstrates a comprehensive pipeline for processing LiDAR point cloud data and integrating it with object detections from images to draw 3D bounding boxes around detected objects. The process begins with downloading and preparing the KITTI dataset, focusing on a specific dataset containing synchronized raw city data.

Key components include retrieving object detections from images, obtaining LiDAR point clouds, and removing the ground plane via RANSAC. Camera calibration data is used to project LiDAR points onto image space.

B. Object Detection using YOLOv5 and 2D LiDAR

Object detection is performed using YOLOv5, with confidence and IoU thresholds set for optimal detection. LiDAR points are clustered using DBSCAN and transformed into image space. Clusters are refined, 3D bounding boxes drawn, and associated with detected objects. The pipeline is tested on KITTI dataset samples.

IV. Results

Firstly, we loaded the RGB image data and LiDAR sensor data into a shared frame to enable parallel processing and visualization. The resulting fusion of RGB and LiDAR data is shown in **Fig. 3**.

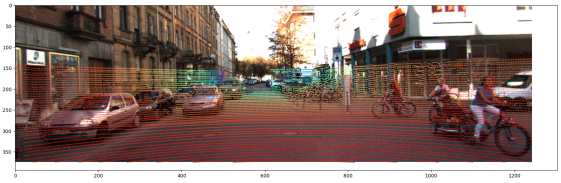

Fig. 3: RGB + LiDAR information obtained from sensors in the same frame.

To reduce noise and eliminate irrelevant data, we applied the **RANSAC** algorithm, which filters out LiDAR points corresponding to the ground plane—areas where no object detection is required. As shown in **Fig. 4**, the ground points have been successfully removed, resulting in reduced computational complexity and improved focus on potential object regions.

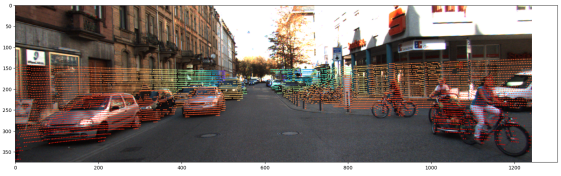

Fig. 4: Output after applying RANSAC on LiDAR data.

Next, the RGB image is passed through a pre-trained **YOLOv5** model. The model performs object detection and generates bounding boxes around identified objects, as shown in **Fig. 5**.

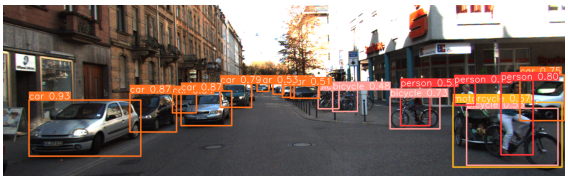

Fig. 5: Object detection using YOLOv5.

To estimate the distance of detected objects from the sensor, we locate the closest LiDAR point near the center of each bounding box and assign its depth value to the corresponding object. This process provides spatial context to the detection results and is visualized in **Fig. 6**.

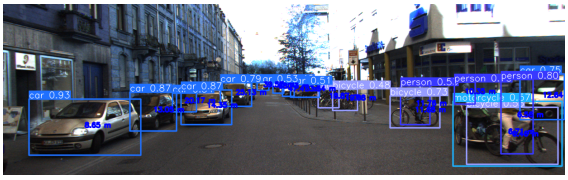

Fig. 6: Distance estimation of detected objects using LiDAR data.

To enhance detection accuracy, we fused the RGB-based detections with LiDAR data using a clustering approach. The **DBSCAN** algorithm was employed to identify meaningful clusters within the LiDAR point cloud. The results of the clustering are shown in **Fig. 7**.

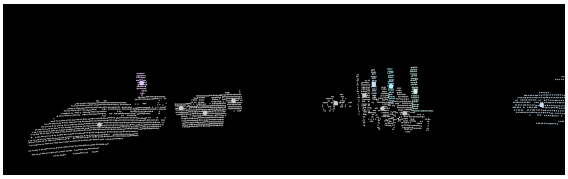

Fig. 7: DBSCAN clustering on LiDAR data.

Subsequently, we evaluated each LiDAR cluster for potential object presence by comparing its centroid with the centroids of YOLO-detected bounding boxes. If the distance between the centroids is below a defined threshold, the cluster is considered associated with the detected object. Additional filtering is applied based on cluster size and shape to discard unlikely candidates. The refined clustering result is presented in **Fig. 8**.

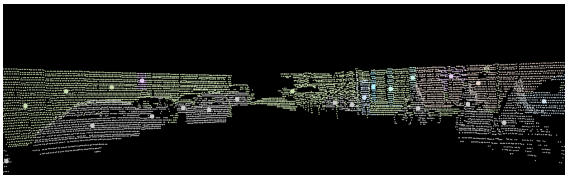

Fig. 8: Updated clustering of LiDAR points in correspondence with image detections.

Finally, using the updated cluster-object associations, we generated **3D bounding boxes** on the image to represent object dimensions and locations more accurately. These enhanced detections are illustrated in **Fig. 9**.

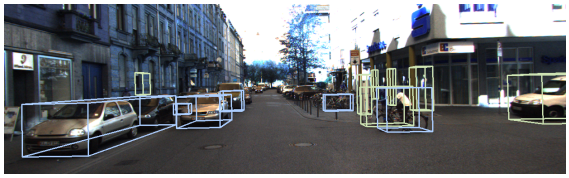

Fig. 9: Updated 3D bounding boxes after sensor fusion.

By removing small, large, and distant clusters, the function reduces the amount of data that needs to be processed in downstream tasks such as object detection or tracking. This can lead to efficiency improvements in algorithm runtime and memory usage. Additionally, by focusing on likely clusters that are close to detected objects, the function improves the accuracy of subsequent processing steps. It reduces the likelihood of false positives and ensures that attention is directed towards clusters that are more likely to represent actual objects in the scene.

- **Video detecting objects using LiDAR and RGB data** – Click here
- **Link to code** – Click here

IV. Conclusion

Integrating LiDAR data with RGB camera data for object detection offers several advantages that can significantly enhance detection performance. LiDAR provides accurate and reliable depth information, enabling precise localization of objects in three-dimensional space, while RGB images contribute rich visual and contextual details in the two-dimensional domain. The fusion of these complementary data sources enhances the system's perception capabilities, improves object classification accuracy, and helps reduce false positives.

Moreover, this multi-modal approach strengthens the robustness of detection algorithms, especially under challenging environmental conditions such as low light, occlusion, or cluttered scenes. The integration also supports better generalization across diverse driving scenarios, which is essential for real-world autonomous driving applications. By leveraging deep learning techniques such as CNNs and advanced fusion frameworks like BEVFusion, the detection system becomes more resilient to sensor failures and variations in input data.

Overall, LiDAR-RGB fusion plays a pivotal role in advancing the reliability, safety, and efficiency of autonomous navigation systems. Future developments may further optimize this fusion through real-time processing, lightweight architectures, and adaptive learning methods

to better handle dynamic and complex road environments.

VI. CONTRIBUTIONS

Anoop Kumar Singh: Handled data preprocessing, LiDAR integration, camera-LiDAR calibration, and project documentation.

Love Kumar Saini: Worked on object detection using YOLOv5, LiDAR-RGB data fusion, clustering, and 3D bounding box generation.

REFERENCES

[1] R. Zhang, Y. Guo, Y. Long, Y. Zhou, and C. Jiang, "Vehicle motion state prediction method integrating point cloud time series multiview features and multitarget interactive information," textitJournal of Advanced Transportation, 2022.

[2] A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, "3D Bounding Box Estimation Using Deep Learning and Geometry," textitArXiv, 2016.

[3] KITTI Vision Benchmark Suite. [Online]. Available: https://www.cvlibs.net/datasets/kitti/

[4] R. Girshick, "Fast R-CNN," textitArXiv, 2015.

[5] J. Redmon, S. Divvala, R. Girshick, "You Only Look Once (YOLO)," textitArXiv.