International Journal of Science and Engineering Applications
Volume 15-Issue 07, 17 – 19, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1507.1004

# Research on Multi-Object Detection Algorithm Based on Decision-Level Fusion of LiDAR and Camera

Peng Tang
Zibo Polytechnic University
Zibo, Shandong, China

**Abstract**: Accurate multi-object perception system is the core component of autonomous driving technology. Aiming at the problem that single sensor is susceptible to environmental interference in complex traffic scenarios, resulting in missed detection and false detection, this paper proposes a multi-object detection algorithm based on decision-level fusion of LiDAR and camera. Firstly, temporal-spatial alignment is performed for LiDAR and camera to ensure the consistency of data in time and space domains. Secondly, PointPillars and YOLOv5 algorithms are adopted to detect objects on preprocessed point cloud data and image data respectively. Finally, Intersection over Union (IoU) matching, D-S evidence theory and weighted box fusion are utilized to realize decision-level fusion of detection results from two sensors. Experimental results demonstrate that the proposed fusion method achieves better detection accuracy than single sensors on both KITTI and nuScenes autonomous driving datasets, which effectively improves the accuracy and robustness of multi-object detection.

**Keywords**: autonomous driving; multi-object detection; multi-sensor fusion; LiDAR; camera; decision-level fusion

## 1. INTRODUCTION

Environment perception is one of the key technologies of autonomous driving systems, and multi-object detection, as the core task of environment perception, directly affects the decision-making and planning capability of autonomous vehicles. In practical application of vehicle-mounted sensors, single sensor has inherent performance limitations: camera can acquire rich color and texture information, yet it cannot obtain accurate depth and distance information of targets; LiDAR can capture precise 3D spatial information to determine target distance and position, but it struggles to identify target categories accurately. Therefore, fusing camera and LiDAR to complement respective advantages has become a critical approach to boost multi-object detection performance.

The rapid development of deep learning has brought new breakthroughs to multi-object detection. In vision-based object detection, YOLO series algorithms are widely adopted due to outstanding real-time performance and detection precision. Sang et al. proposed an improved YOLOv2 vehicle detection framework, which optimizes the length and width of clustered bounding boxes via normalization and applies multi-layer feature fusion to enhance the network's feature extraction capability. In LiDAR-based object detection, algorithms including PointPillars and VoxelNet provide effective solutions for feature extraction and object detection of point cloud data. Ye et al. constructed a hybrid voxel network to fuse multi-scale voxel features at point level, solving the contradiction of unbalanced voxel size.

In terms of multi-sensor fusion, fusion strategies are categorized into data-level fusion, feature-level fusion and decision-level fusion according to fusion stages. Data-level fusion directly merges raw sensor data with high computation cost and strict alignment requirements; feature-level fusion realizes favorable information complementarity at feature extraction stage; decision-level fusion integrates detection outputs, possessing high reliability and real-time performance. This paper adopts decision-level fusion strategy, which retains independent detection advantages of each sensor and further promotes detection performance via fusion algorithm.

This paper proposes a multi-object detection fusion algorithm based on LiDAR and camera decision-level fusion. Temporal-spatial alignment is implemented for two sensors firstly, then PointPillars and YOLOv5 are applied for independent object detection, and finally IoU matching, D-S evidence theory and weighted box fusion are combined to complete decision-level fusion of detection results. Experiments verify the effectiveness of the proposed fusion algorithm in improving detection accuracy and robustness.

## 2. FUSION ALGORITHM DESIGN

The proposed multi-object detection fusion algorithm consists of three core modules: sensor temporal-spatial alignment, single-sensor object detection and decision-level fusion.

### 2.1 Temporal-Spatial Alignment

In actual operation of autonomous vehicles, different sensors hold distinct sampling frequencies and coordinate systems. Direct fusion of unaligned sensor data will severely degrade object detection accuracy. Hence, dual alignment of time and space is required for LiDAR and camera.

#### *2.1.1 Time Alignment*

Time alignment includes hardware time synchronization and software time synchronization. Hardware time alignment provides unified reference clock for all sensors, and each sensor calibrates its own clock based on the benchmark to realize hardware-level synchronization. Software time alignment matches data via timestamps of different sensors, taking the sensor with lower sampling frequency as benchmark and searching data frames with closest timestamps. In practical LiDAR-camera fusion systems, LiDAR timestamps are usually taken as benchmark; camera frames are cached, and the image frame with the most similar timestamp is matched from cache after receiving LiDAR point clouds to complete time alignment.

#### *2.1.2 Spatial Alignment*

Spatial alignment converts coordinate systems of different sensors into a unified reference frame. LiDAR-camera spatial alignment involves coordinate transformation among LiDAR coordinate system, camera coordinate system, image coordinate system and pixel coordinate system. With camera

intrinsic matrix (focal length, principal point coordinates, etc.) and extrinsic matrix (rotation matrix and translation vector) between LiDAR and camera, 3D point cloud in LiDAR coordinate system can be projected onto pixel coordinate system to achieve spatial consistency of two types of sensor data.

## 2.2 Single-Sensor Object Detection

After finishing temporal-spatial alignment, independent object detection is conducted on data from two sensors respectively.

### *2.2.1 Camera-Based Object Detection*

YOLOv5 algorithm is adopted for image object detection in this paper. YOLOv5 contains four components: input terminal, backbone network, neck network and prediction terminal. The input terminal implements Mosaic data augmentation, adaptive anchor box calculation and adaptive image scaling; the backbone network composed of convolution, CSP and SPP structures extracts image target features; the neck network adopts PANet for multi-scale feature fusion, combining low-level location information and high-level semantic information simultaneously; the prediction terminal adopts CIoU_Loss as loss function and retains optimal prediction boxes through Non-Maximum Suppression (NMS). YOLOv5 outputs 2D bounding boxes and category confidence of all targets in images.

### *2.2.2 LiDAR-Based Object Detection*

PointPillars algorithm is utilized for point cloud object detection. PointPillars first divides point clouds into regular pillars, encodes point features within each pillar to generate pseudo-image feature maps, then extracts features and detects objects via 2D convolutional neural networks. In point cloud preprocessing, ground point segmentation is carried out to filter ground points and reduce computation volume; Random Sample Consensus algorithm is adopted for point cloud clustering; principal component analysis is applied to fit 3D bounding boxes of targets. PointPillars outputs 3D bounding boxes and spatial position information of detected objects.

## 2.3 Decision-Level Fusion

After obtaining independent detection results of two sensors, decision-level fusion is adopted to merge bounding boxes. The core idea of decision-level fusion is integrating information from different sensors at detection output stage to leverage respective strengths and compensate defects of single sensor. The decision-level fusion workflow contains three steps as follows:

### *2.3.1 IoU Matching*

Calculate the Intersection over Union (IoU) between camera 2D bounding boxes and projected LiDAR 3D bounding boxes on image plane. Detection box pairs with IoU exceeding preset threshold are judged as matching the same target and enter subsequent fusion process; unmatched boxes are retained or discarded according to confidence scores.

### *2.3.2 D-S Evidence Theory*

For matched detection results, D-S evidence theory is applied to fuse target category decisions. D-S evidence theory can effectively handle uncertain information, generating more reliable category judgment by synthesizing category evidence from different sensors. Specifically, detection confidence of camera and LiDAR for each target category is regarded as basic probability assignment, Dempster combination rule is used for fusion, and the final target category is determined according to fused confidence.

### *2.3.3 Weighted Box Fusion*

After target category confirmation, weighted box fusion is used to integrate position parameters of bounding boxes. Different weights are assigned based on estimation precision of two sensors: LiDAR gains higher position weight for superior spatial measurement accuracy, while camera obtains higher category weight for outstanding classification capability.

The final fused multi-object detection outputs are generated after the above decision-level fusion process, containing both target category information and precise spatial position information.

# 3. EXPERIMENTS AND RESULT ANALYSIS

## 3.1 Experimental Environment and Datasets

To verify the effectiveness of the proposed LiDAR-camera decision-level fusion multi-object detection algorithm, experiments are carried out on two public autonomous driving datasets: KITTI and nuScenes. All targets in datasets are classified into three categories for evaluation: Car, Pedestrian and Cyclist.

KITTI dataset contains PNG-format camera images with resolution 1242×375 and binary bin LiDAR files; point clouds are arranged in N×4 structure (N = number of points, 4 = x, y, z spatial coordinates and intensity). nuScenes dataset adopts JPG-format 1600×900 camera images and pcd LiDAR files; point cloud attributes include x/y/z coordinates, reflection intensity and vertical scan index of radar. Both datasets provide sensor intrinsic and extrinsic matrix files with pre-completed temporal calibration and spatial calibration.

Hardware environment: Intel Core i5-9300H 2.4GHz CPU, NVIDIA GeForce GTX1650 GPU (4GB VRAM), Ubuntu 18.04 operating system; programming language C++ with OpenCV dependency library.

## 3.2 Evaluation Metrics

Three evaluation metrics are selected to assess algorithm performance:

(1) Average Precision (AP) and mean Average Precision (mAP): AP equals the area under Precision-Recall curve for each category; mAP is the average AP value of all target categories, measuring detection precision of each algorithm.

(2) Frames Per Second (FPS): Reflects algorithm running efficiency and real-time performance.

(3) Intersection over Union (IoU) and mean IoU (MIoU): Evaluates overlap degree between predicted bounding boxes and ground truth boxes.

## 3.3 Performance Comparison

### *3.3.1 Single-Sensor Detection*

Comparative experiments of YOLOv5 and PointPillars are conducted on KITTI and nuScenes datasets to test single-sensor detection performance. YOLOv5 achieves competitive detection performance on KITTI dataset. However, detection accuracy declines significantly on nuScenes dataset due to more complex and diverse traffic scenes, which proves the performance limitation of single visual sensor under complicated driving environments.

#### *3.3.2 Fusion Algorithm*

Comparative experiments between the proposed fusion algorithm and single-sensor algorithms are implemented to validate the superiority of decision-level fusion. The IoU values of all three target categories of fusion algorithm outperform single-sensor algorithms, which verifies that decision-level fusion can effectively combine respective advantages of camera and LiDAR to improve detection accuracy and robustness.

### 3.4 Visualization Result Analysis

A complex crossroad traffic scene is selected for visualization comparison to intuitively demonstrate detection performance of the fusion algorithm. YOLOv5 and the proposed fusion algorithm are separately applied for multi-object detection.

Experimental results show that single-camera YOLOv5 can detect most targets, yet it suffers defects in distant objects and small targets, with predicted boxes failing to fit real target contours precisely. In contrast, the proposed fusion algorithm detects more complete targets with bounding boxes fitting real object outlines better, acquiring more accurate spatial position information. This sufficiently illustrates the effectiveness of introducing LiDAR data to promote target localization precision.

## 4. CONCLUSION

This paper proposes a multi-object detection algorithm based on decision-level fusion of LiDAR and camera. Hardware-software combined strategy realizes time alignment of two sensors, and coordinate transformation completes spatial alignment. PointPillars and YOLOv5 are used for independent detection on point cloud and image data respectively. Finally, IoU matching, D-S evidence theory and weighted box fusion are integrated to realize decision-level fusion of detection outputs.

Future research directions are summarized as follows: first, explore Transformer-based multi-modal feature fusion to further boost detection precision of distant and tiny targets; second, optimize real-time performance of fusion algorithm to meet low-latency strict requirements of autonomous driving systems; third, deploy and validate the algorithm on real vehicle platforms to test robustness in actual road scenarios.

## 5. REFERENCES

[1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego: IEEE, 2005: 886-893.

[2] SANG J, WU Z, GUO P, et al. An Improved YOLOv2 for Vehicle Detection[J]. Sensors, 2018, 18(12): 4272.

[3] YE M, XU S, CAO T. Hvnet: Hybrid Voxel Network for LiDAR based 3D Object Detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020.

[4] ZHOU Y, TUZEL O. Voxelnet: End-to-End Learning for Point Cloud Based 3D Object Detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.

[5] SHI S, GUO C, JIANG L, et al. Pv-rcnn: Point-voxel Feature Set Abstraction for 3D Object Detection[C]// 2020 IEEE/CVF Conference on Computer Vision and CHEN Xiaofeng, LI Yufeng, WANG Chuansong, et al. Research on target detection system for autonomous vehicles based on multi-sensor fusion[J]. Laser Journal, 2025, 46(5): 94.

[6] CHEN Xiaofeng, LI Yufeng, WANG Chuansong, et al. Research on target detection system for autonomous vehicles based on multi-sensor fusion[J]. Laser Journal, 2025, 46(5): 94.