Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Sensor Fusion Approach for Object Detection and Distance Estimation in Autonomous Vehicle Perception using Camera and Lidar

Gouri Pandey[1*], Sesha Surya Sai Pullapanthul[1†] and Surender Kannaiyan[1†]

[1]*Electronics and communication, VNIT, Ambazari road, Nagpur, 220010, Maharashtra, India.*

**Corresponding author(s). E-mail(s): gouripandey1210@gmail.com; Contributing authors: seshasuryasai.pullapanthula@gmail.com; ksurender@ece.vnit.ac.in;*

ARTICLE INFO

Received: 01 Aug 2024

Revised: 15 Sept 2024

Accepted: 25 Sept 2024

ABSTRACT

The real-time integration of LiDAR and camera sensor data plays a vital role in the field of autonomous driving. This fusion enables accurate depth estimation and object detection at various distances, significantly enhancing a vehicle's perception capabilities. This study proposes an effective method to estimate distances between a self-driving vehicle and surrounding objects such as other vehicles, pedestrians, and traffic signs through LiDAR-camera data fusion. The methodology begins with applying Rigid Body Transformations (rotation and translation) to align the coordinate frames of the LiDAR and camera systems. This is followed by projecting 3D LiDAR points onto the 2D camera image plane using Homogeneous Coordinate Transformation (matrix multiplication). The fused data is then processed with the YOLOv5 deep learning model for object detection. Distance estimation involves associating the nearest bounding box coordinates of detected objects in the camera images with the corresponding LiDAR points. Depth values are extracted from the LiDAR data, and the Euclidean distance from the ego vehicle is computed. This sensor fusion approach is rigorously evaluated using both real-world scenarios and simulated environments. The analysis includes both quantitative and qualitative assessments. The results demonstrate significant improvements in environmental perception, with consistent and reliable depth information that supports safe autonomous navigation. The technique achieved an accuracy of approximately 82%, with a mean absolute error of 4.69 and RMSE of 5.863 in highway road scenarios. These results highlight the robustness and potential of sensor fusion for enhancing perception systems in autonomous vehicles.

Keywords: Deep Learning, Object Detection, Camera and LiDAR Sensor Fusion

Introduction

Autonomous vehicles (AVs) are transforming modern transportation by reducing acci- dent risk and enhancing road safety. They also promise to reduce emissions, improve traffic flow, and drive economic growth [1]. AVs are designed to perceive their surround- ings and navigate with little or no human input. According to Precedence Research, the global AV market reached around 6,500 units in 2019 and is projected to grow at a CAGR of 63.5% from 2020 to 2027 [2]. Object detection and distance estimation are key to AVs' safe navigation in dynamic environments. Recently, Multi-Source and Heterogeneous Information Fusion (MSHIF) has emerged as a powerful strategy for enhancing AV perception. By combining inputs from various sensors, MSHIF addresses limitations of individual sensors and provides a richer, more accurate environmen- tal understanding [3]. Sensors like cameras, LiDAR, radar, sonar, GPS, IMU, and odometers are integral to autonomous driving. Among them, camera-LiDAR fusion has gained attention due to its complementary strengths—cameras provide high-resolution visual data, while LiDAR offers accurate depth information [4]. 3D LiDAR sensors are widely used for their wide field of view, precise depth capabilities, and long-range detection, even at night [5]. However, point cloud sparsity at greater distances can reduce classification accuracy [6]. In contrast, cameras excel at object classification due to their high resolution and

Research Article

the recent progress in deep learning-based image recog- nition [7]. Typical object detection methods involve generating region proposals (e.g., sliding window, edge box, selective search) and using convolutional neural networks (CNNs) for classification [8]. Sensor data fusion is achieved through various techniques, including rule-based methods, probabilistic models like Kalman filters, and Bayesian inference [9]. More recently, machine learning and deep learning fusion techniques have shown promise, especially CNNs and recurrent neural networks (RNNs), which pro- cess multi-modal data for improved perception, localization, and decision-making [10]. Integrating sensor data in autonomous vehicles is challenged by calibration, synchro- nization, and real-time processing. Environmental factors like weather and lighting add complexity, requiring robust fusion algorithms. This thesis aims to enhance percep- tion and reliability through advanced LiDAR-camera fusion, benefiting autonomous driving and related fields. In this study, we introduce a novel approach that combines camera and LiDAR data with the YOLOv5 deep learning model for integrated object detection and distance estimation. This fusion leverages the strengths of both sen- sor cameras for rich visual context and LiDAR for accurate depth information. By processing the combined data through YOLOv5, we achieve reliable object detection and precise distance measurements. Experimental evaluations across various real-time datasets demonstrate that this approach delivers strong accuracy and robust perfor- mance in diverse driving scenarios. The objective is to create a robust perception system that enhances the vehicle's autonomous navigation capabilities in complex environments. The project is entitled as "Sensor Fusion for Object Detection and Dis- tance Estimation in Autonomous Vehicle Perception," which succinctly summarizes its purpose. Camera–LiDAR fusion plays a crucial role in enhancing autonomous vehicle perception, which is highly relevant for Industry and advanced manufacturing envi- ronments. By combining the rich visual details from cameras with the precise depth information from LiDAR, sensor fusion enables accurate object detection and reli- able distance estimation. In industrial settings such as automated warehouses, smart factories, and material transport systems, this capability ensures safer navigation, col- lision avoidance, and efficient task execution. Moreover, robust perception through camera–LiDAR fusion supports the development of intelligent robotic platforms and autonomous guided vehicles (AGVs), which are key components in Industry, ultimately improving productivity, safety, and automation in manufacturing processes.

Related Work

Autonomous vehicles (AVs), or self-driving cars, operate with minimal human input by processing data from multiple sensors for safe navigation [7]. They perform two key perception tasks: environmental perception (using RGB/thermal cameras, LiDAR) and localization (using GNSS, IMU, INS, odometry, and LiDAR) [11]. Key perception modules include object detection, tracking, and SLAM, where cameras detect objects and road signs, and LiDAR provides accurate depth information [12]. Combined, these sensors support mapping and localization by extracting features used in SLAM or matching with HD maps [13].Sensor fusion enhances perception by integrating outputs from various sensors to reduce uncertainty and improve decision-making [14]. Tradi- tional fusion algorithms include statistical, probabilistic, knowledge-based, evidence reasoning, and interval analysis methods [15]. Deep learning has significantly advanced sensor fusion, with models like CNNs, RNNs, DBNs, and AEs applied to perception tasks [16]. Fusion of LiDAR and camera data is widely used to improve detection. For example, Asvadi et al. [17] fused LiDAR-generated depth and reflectance maps with RGB images, processed via YOLO, using decision-level fusion for vehicle detection. Another approach [18] projects LiDAR ROIs onto camera images for enhanced detec- tion. Fusion methods are categorized as early fusion (e.g., projecting 3D LiDAR onto 2D images) and late fusion (e.g., merging 2D and 3D bounding boxes post-detection) [19]. Object detection and distance estimation techniques include camera-based (e.g., Fast R-CNN, YOLO) [20], LiDAR-based (e.g., PointRCNN, VoxelNet) [21], and multi- sensor-based methods. While cameras offer rich visuals, they struggle in poor lighting. LiDAR provides precise 3D data but becomes sparse with distance. Combining sensors enhances results. Object-centric fusion (using Bird's Eye View and front view), Point- Painting [22], and radar fusion (e.g., CenterFusion) help improve robustness, especially in challenging weather conditions. Based on the above research, our focus is to enhance sensor integration for autonomous driving and improve perception and localization for accurate object detection and distance estimation. This approach addresses key chal- lenges

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

such as sensor misalignment and environmental variability, while also enhancing the accuracy of object localization and distance measurements. The combination of deep learning and late fusion thus forms a strong foundation for safer and more efficient autonomous vehicle navigation.

Methodology

1.1 Dataset

The KITTI dataset, collected in Germany's rural and urban areas, contains six hours of traffic data captured at 10–100 Hz using a GPS/IMU system, a 64-beam Velodyne LiDAR, and high-resolution grayscale and color stereo cameras. Data were recorded in daylight and good weather to minimize illumination effects. The dataset provides raw and processed stereo sequences (0.5 MP, PNG), LiDAR point clouds (100k points/frame), GPS/IMU metadata (location, speed, acceleration), and calibration files (camera, LiDAR, IMU) for accurate sensor fusion and depth estimation. It also includes 3D object tracking labels for vehicles, pedestrians, cyclists, and others, along with synchronized timestamps. setup for sensors on car is shown in Figure1.

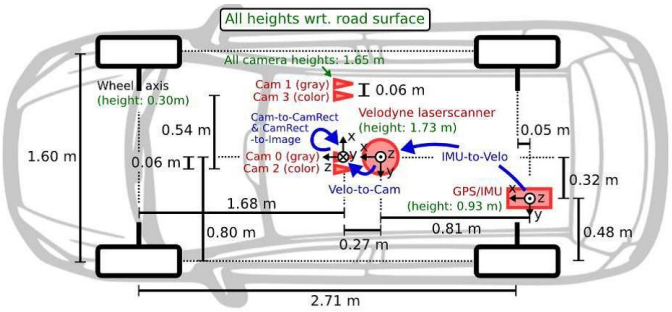


Fig. 1: KITTI dataset data collection platform with multi-sensor setup including stereo cameras, LiDAR, and GPS/IMU [23]

1.2 Yolo v5 Model

YOLOv5 is an efficient object detection model with four variants (small to extra- large) offering a trade-off between speed and accuracy. Its model variants are shoen in Figure 2. It uses a CSP-Darknet53 backbone with SPP and PANet for robust feature extraction, and its head predicts bounding boxes, classes, and scores. Key improve- ments include SiLU activation, BCE and CIoU loss functions, optimized bounding box equations for edge accuracy, and a Focus Layer that reduces FLOPs and memory use. Implemented in PyTorch, YOLOv5 achieves high accuracy and speed, making it well-suited for real-time detection.

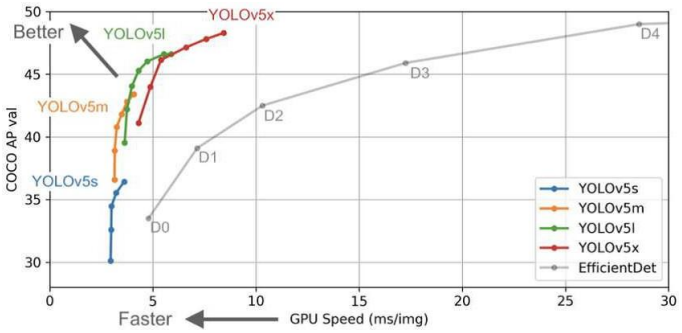


Fig. 2: YOLO V5 Performance Comparison with Other models [18]

1.1 Proposed Method

The proposed method is illustrated in Figure 3. We used the KITTI dataset [23], which provides pre-calibrated LiDAR and camera data along with publicly available calibration files. These files contain the intrinsic and extrinsic parameters required to transform and project LiDAR points onto image planes, ensuring accurate sensor alignment. Intrinsic parameters of the camera are obtained using the checkerboard cal- ibration method, while

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

extrinsic parameters between LiDAR and camera are estimated with a planar 3D marker. Finally, rigid transformations (translation and rotation) are applied to align the coordinate systems using equations (1)–(4).

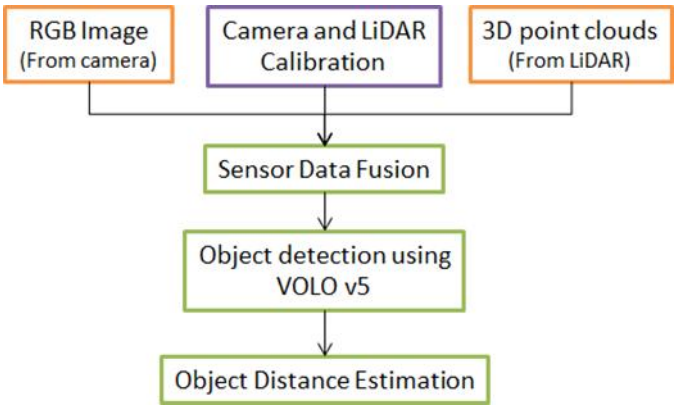


Fig. 3: Process flow of the algorithm of the proposed method

a. Rotation Matrix : A rotation matrix R can rotate a point (x,y,z) in 3D space to a new position (x′,y′,z′). It is a 3x3 matrix . where R is 3 x 3 rotational matrix.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathrm{R}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1)$$

$$\mathrm{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (2)$$

b. Translation Vector : Translation involves moving the coordinate system by a vector t = [tx, ty, tz]T . If (x, y, z) is a point in 3D space, the translated point (x′, y′, z′) is given by:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathrm{R}\begin{bmatrix} x + tx \\ y + ty \\ z + tz \end{bmatrix} \quad (3)$$

The translational vector t is :

$$\mathrm{t} = \begin{bmatrix} tx \\ ty \\ tz \end{bmatrix} \quad (4)$$

c. Homogeneous Transformation: Rotation and translation are often combined into a single 4 × 4 homogeneous transformation matrix T, which includes both components:

$$\mathrm{T} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (5)$$

d. Projection onto Image Plane: The overall transformation matrix from the LiDAR frame to the rectified camera frame projects 3D LiDAR points (X, Y,Z) onto the 2D image plane (u, v) as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K(R * \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T) \quad (6)$$

where K is the 3 × 3 intrinsic camera matrix, R is the 3 × 3 rotation matrix, T is the 3 × 1 translation vector, and (X, Y, Z) are the 3D LiDAR points. The fused data is processed using YOLOv5, which applies confidence and IoU thresholds to detect objects with bounding boxes, class labels, and confidence scores. Each detection is then associated with the nearest LiDAR point cloud, enabling calculation of the Euclidean distance (Equation 7.) between the object's 3D coordinates (x_p, y_p, z_p) and the ego vehicle's reference coordinates (X_e, Y_e, Z_e). This ensures accurate object detection and distance estimation.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$d = \sqrt{(xp - Xe)^2 + (yp - Ye)^2 + (zp - Ze)^2} \quad (7)$$

1.3 Evaluation Metrics

An evaluation matrix is vital for assessing the performance and effectiveness of a model or system. They provide quantitative measures to determine how well the system meets its objectives.

1.3.1 Evaluation Metrics for Object Detection

To assess and compare the predictive capabilities of the different object detection models, it's essential to rely on some standardized quantitative measures. Among the most prevalent evaluation metrics are the Intersection over Union (IoU) and the Average Precision (AP).

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (8)$$

$$precision = \frac{TP}{TP + FP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

Where, TP = True Positive

FP = False Positive

TN = True negative

1.3.2 Evolution matrix for Distance Estimation

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are common metrics for evaluating prediction accuracy. MAE measures the average absolute differ- ence between predicted and actual values, while RMSE squares errors before averaging, giving more weight to large discrepancies and penalizing outliers.

$$MAE = \frac{|(yi - yp)|}{n} \quad (11)$$

$$RMSE = \sqrt{\sum \frac{(yi - yp)^2}{n}} \quad (12)$$

Where, y_i = actual value, $\hat{y}_i$ = predicted value, and n = number of observations.

Results

This section presents visualizations, GPU usage, and model performance. The model utilized only 26% GPU memory, showing efficient processing, and achieved a maximum accuracy of 81.82% with detailed distance error analysis across scenarios. These results confirm its robustness and real-world applicability.

4.2 Visualization Results

The visual results for 3 different visual scenarios are shown in the following Figure 4, which includes results for a) 3D point cloud projected on 2D camera image, b) Yolo Object detection with IoU score, c) Detection and Distance Estimation with the proposed method, and d) Bird's Eye View with surrounding vehicles.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

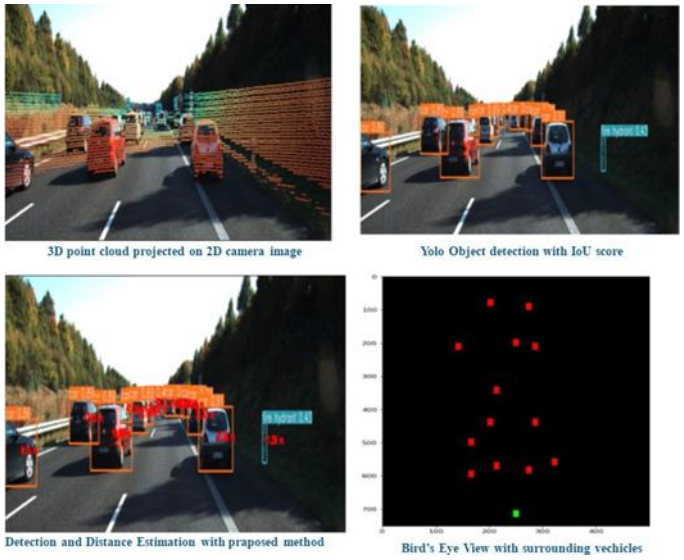


Fig. 4: Results for scenario 1 Highway Road

Figure 5 shows a combined image of LiDAR + camera view together with detection and distance for a road scenario on the highway.

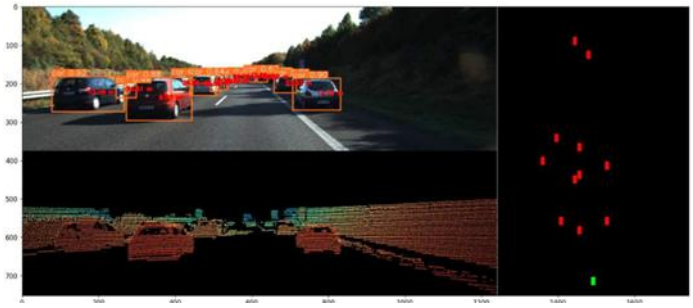


Fig. 5: LiDAR+Camera+BEV together on a single frame for highway road Figures 6 and 7 show a combined image of LiDAR + Camera View together with Detection and Distance for a traffic road scenario and a Residential Road, respectively.

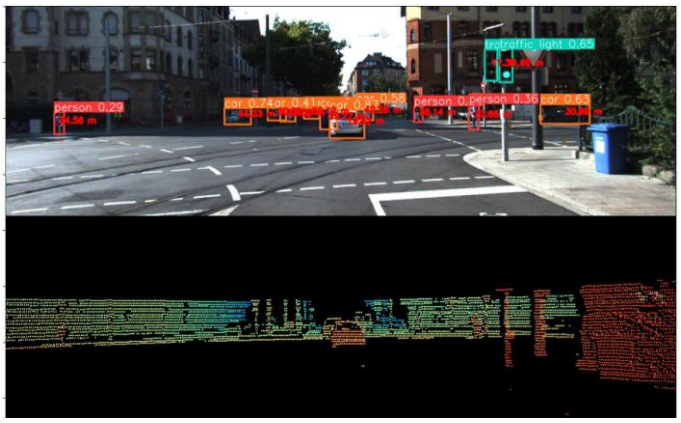


Fig. 6: LiDAR+Camera+BEV together on a single frame for a traffic road scenario

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

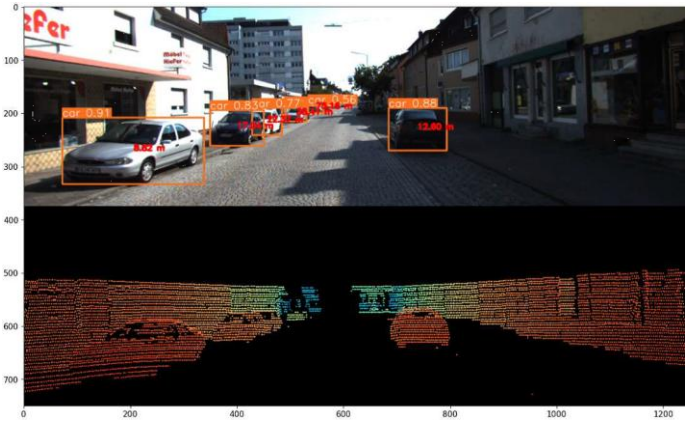


Fig. 7: LiDAR+Camera+BEV together on single frame Residential Road scenario

4.3 Performance Result

The proposed system achieves a frame rate of 10, as illustrated in Figure 8. The results for different scenarios are listed in Table 1.

Table 1: Performance results for different scenarios

Sr. No.	Scenario	MAE (meters)	RMSE (meters)	Model Accuracy (%)
1	Highway Road	0.7852	0.9773	81.82
2	Traffic Signal Road	0.7264	0.8996	77.27
3	Residential Road	0.8434	1.0570	72.73

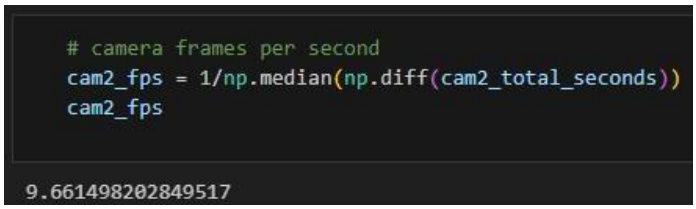
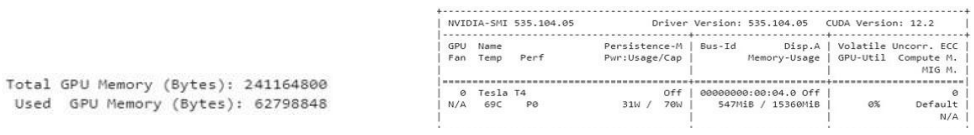


Fig. 8: FPS for Proposed model

We have run our model on both CPU and cloud GPU. CPU is an INTEL I-5 7th Gen Processor. GPU details are in Figure 9. We have observed that the GPU utilization of my model is around 26



(a) GPU score (b) Memory utilization

Fig. 9: GPU score and Memory utilization of system

Conclusion

In this research work, we comprehensively reviewed the deep learning-based sensor fusion approach by integrating the early fusion of LiDAR and Camera information. We used the Coordinates transformation at

the early stage of projecting the 3D point cloud onto 2D image and then transferred the data to the YOLO v5 object detection block and got the bounding box predictions. We have got box prediction along with the LiDAR point cloud information inside. We have x, y, z coordinates of the point cloud. We then used IMU data, which was earlier used during sensor calibration, provided by the DATASET author, to localize the EGO vehicle and measure the distance of the EGO vehicle from the other objects using the nearest point cloud information. We have performed out model in both CPU and Cloud GPU. We have worked on different scenarios like highway roads, residential roads, traffic signals etc. The proposed method achieved an accuracy of 82%, with a mean absolute error of 4.69 and an RMSE of 5.863 on the highway road scenario, demonstrating the effectiveness of sensor fusion in enhancing autonomous vehicle perception.

Future Scope

The field of autonomous driving is growing day by day. There is a wide range of research opportunities. The proposed method can be further improved by working with different large datasets and under different weather conditions. We can develop a vehicle tracking application and a navigation application, and improve the perception.

We can also impart the fusion of RADAR data, thereby improving the object distance estimation.

References

[1] Fayyad, J., Jaradat, M.A., Gruyer, D., Najjaran, H.: Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors 20(15), 4220 (2020) https://doi.org/10.3390/s20154220
[2] Yeong, D.J., Velasco-Hernandez, G., Barry, J., Walsh, J.: Sensor and sensor fusion technology in autonomous vehicles: A review. Sensors 21, 2140 (2021) https://doi.org/10.3390/s21062140
[3] Fung, M.L., Chen, M.Z.Q., Chen, Y.H.: Sensor fusion: A review of methods and applications. In: Proc. 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, pp. 3853–3860 (2017). https://doi.org/10.1109/CCDC.2017.7979175
[4] Kumar, G.A., Lee, J.H., Hwang, J., Park, J., Youn, S.H., Kwon, S.: Lidar and camera fusion approach for object distance estimation in self-driving vehicles.Symmetry 12(2), 324 (2020) https://doi.org/10.3390/sym12020324
[5] Asvadi, A., Garrote, L., Premebida, C., Peixoto, P., Nunes, U.J.: Multimodal vehicle detection: Fusing 3d-lidar and color camera data. Pattern Recognition Letters 115, 20–29 (2018)
[6] Zhang, Y., Wang, J., Wang, X., Dolan, J.M.: Road-segmentation based curb detection method for self-driving via a 3d-lidar sensor. IEEE Transactions on Intelligent Transportation Systems 19(12), 3981–3991 (2018)
[7] Li, K., Wang, X., Xu, Y., Wang, J.: Density enhancement-based long-range pedestrian detection using 3-d range data. IEEE Transactions on Intelligent Transportation Systems 17(5), 1368–1380 (2016)
[8] Dollar, P., Appel, R., Belongie, S., Perona, P.: Fast feature pyramids for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(8), 1532–1545 (2014)
[9] Chen, X., Kundu, K., Zhu, Y., Ma, H., Fidler, S., Urtasun, R.: 3d object proposals using stereo imagery for accurate object class detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(5), 1259–1272 (2018)
[10] Gruyer, D., Crouzil, A., Belaroussi, R.: Vehicle detection and tracking by col- laborative fusion between laser scanner and camera. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan (2013)
[11] Katrakazas, C., Quddus, M., Chen, W.H., Deka, L.: Real-time motion planningmethods for autonomous on-road driving: State-of-the-art and future research directions. Transportation Research Part C: Emerging Technologies 60, 416–442 (2015)
[12] Li, J.: Fusion of lidar 3d points cloud with 2d digital camera image. PhD thesis, Oakland University, Rochester, MI, USA (2015)
[13] Gruyer, D., Belaroussi, R., Revilloud, M.: Accurate lateral positioning from map data and road marking detection. Expert Systems with Applications 43, 1–8 (2016)

Research Article

[14] Gruyer, D., Magnier, V., Hamdi, K., Claussmann, L., Orfila, O., Rakotoni- rainy, A.: Perception, information processing and modeling: Critical stages for autonomous driving applications. Annual Reviews in Control 44, 323–341 (2017)
[15] Castanedo, F.: A review of data fusion techniques. The Scientific World Journal 2013, 1–19 (2013)
[16] Wang, H., Lou, X., Cai, Y., Li, Y., Chen, L.: Real-time vehicle detection algorithm based on vision and lidar point cloud fusion. Journal of Sensors (2019) https: //doi.org/10.1155/2019/8473980
[17] Melotti, G., Asvadi, A., Premebida, C.: Cnn-lidar pedestrian classification: Com- bining range and reflectance data. In: 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Madrid, Spain, pp. 1–6 (2018)
[18] Mastin, A., Kepner, J., Fisher, J.: Automatic registration of lidar and optical images of urban scenes. In: Proceedings of the Computer Vision and Pattern Recognition, Miami, FL, USA, pp. 2639–2646 (2009)
[19] Yin, H., Yang, X., He, C.: Spherical coordinates based methods of ground extraction and objects segmentation using 3-d lidar sensor. IEEE Intelligent Transportation Systems Magazine 8(1), 61–68 (2016)
[20] Magnier, V.: Multi-sensor data fusion for the estimation of the navigable space for the autonomous vehicle. PhD thesis, University Paris Saclay and Renault, Versailles, France (2018)
[21] Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR) (2013)
[22] Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1231–1237 (2013). https://doi.org/10.1109/ICCV.2013.150