Multimodal Sensor Fusion for Real-Time Object Detection

Nethra K[1*], *Kavya* Nayak[1], *Sinchana*[1], *Nithish* N[1], and *Raghavendra* M Shet[1]

[1]Department of Electronics and Communication Engineering, Mangalore Institute of Technology and Engineering, Moodabidri, India

Abstract. This paper describes a multimodal sensor fusion developed on the Raspberry Pi platform for real-time object detection and distance estimation. This architecture has a camera and a 24 GHz mmWave radar sensor for achieving the vision and a range sensing. A pretrained YOLO model is used for identifying classes such as persons from live video frames to carry out real-time object detection. The radar provides the distance measurement for which a 1-D linear Kalman filter is applied to get a smooth and accurate estimate, and then fused with the camera data. The result of this experiment showed that the fused system offers significantly higher stability in object detection and distance estimation as compared to single sensor readings. A final configuration where radar measurement is activated only when the detected object is at the centre of the frame, which achieved near-accurate results with less noise. The proposed system is lightweight and also cost-effective for real-time perception for low-cost embedded applications in autonomous vehicles and intelligent surveillance.

1 Introduction

Clear perception and comprehension of the environment are essential for safe and effective machine fundamental interaction in the rapidly expanding field of intelligent systems. For these sensory systems, object detection is essential. Object detection results in the real-time identification, tracking, and location of objects by machines. But, when object detection is dependent on a single sensor, its accuracy and dependability will decrease. High-resolution visual data from the camera makes it possible to identify the shapes, colours, and diversity of objects. But, unfavourable environmental factors like sunlight, glare, rain, or fog make them function poorly. Even in severe weather and low light, radar performance is effective. Radar offers accurate measures of distance and speed, but it lacks the visual information required to differentiate between various types of objects. Sensor fusion is selected as a practical approach to address these problems. The system offers more comprehensive and lucid view of the surroundings by combining data from several sensors. The detection steadiness, flexibility, and decision-making ability will be enhanced by combining the visual information from the camera with the radar's distance precision. The goal of this project is to develop a real-time object detection system that integrates radar and camera data. The

* Corresponding author: nethrak05052004@gmail.com

Kalman Filter algorithm is One-Dimensional (1D) linear. This recursive estimate approach computes object position and movement over time with accuracy by fusing noisy sensor signals. For embedded applications, the lightweight 1D linear Kalman filter variant works well. The platform used by the system is a Raspberry Pi. It makes use of a millimetre Wave (mmWave) radar sensor for distance sensing and a Pi camera to record visual data. To guarantee precision and reliable real-time object identification, data from both sensors will be merged using Kalman filtering before being displayed on a desktop or Personal Computer (PC). This study demonstrates how reliability of the perception can be improved through the merging of camera and radar sensor data. Future applications in robots, autonomous cars, military, and intelligent monitoring systems depends heavily on reliable perception.

2 Literature review

Nabati et al. proposed a middle-fusion radar-camera system that combines radar point clouds and RGB images for detection and distance estimation in autonomous vehicles. They introduced of Radar Object Proposal (ROP) and Radar Proposal Refinement (RPR) networks and tested on nuScenes (a large-scale autonomous driving benchmark dataset) [1]. Shi et al. examined tracking and detection by radar-camera fusion. This study reviewed fusion stages and datasets for object detection. Also, showcased future work that can be done on the projects in deep learning and adaptive fusion for real applications [2]. Khodarahmi et al. reviewed variants of the Kalman filter for nonlinear, uncertain systems, such as Extended Kalman Filter (EKF), Unscented Kalman filter (UKF), Cubature Kalman Filter (CKF), Interacting Multiple Model (IMM), and Multiple Model Adaptive Estimation (MMAE). These advanced models improve the estimation, but they increase the computational complexity [3].

Ogunride et al. developed Camera-Radar YOLO network (CR-YOLOnet) to enhance object detection in foggy conditions, which involves a radar and camera fusion model based on YOLOv5. Radar data mapped to image space with an attention mechanism, which improved the detection of small, distant objects with an accuracy of 0.849 at 69 frames per second. This paper mentions that fusion is very important in complex situations [4]. Lei et al. did an experiment on 3D object detection and velocity prediction on nuScenes and achieved 67.4 NDS using Hybrid Vision Detection Fusion (HVDetFusion), which combines radar and camera. Radar priors help to improve the quality of data features, which improves accuracy. Flexible input is provided by the modular framework [5].

Barbosa et al. carried out a survey on perceptions of the cameras and radar for Autonomous Vehicles and Advanced Driver Assistance System (ADAS), and observed that cameras are sensitive to weather conditions, whereas radar is good, but has limited shape information. They identified issues with annotation and usage after reviewing datasets, deep learning-based detection, and metrics [6]. A review on millimetre Wave (MMW) radar and camera fusion performed by Yong Zhou et al. They classified different approaches for better detection in all weather conditions [7].

Sensor and sensor-fusion technologies in autonomous vehicles were examined by Yeong et al. They underlined the interaction of vision, LiDAR, and radar. Fusion strategies, necessary calibration, and efficient real-time perception were also discussed by them [8]. A review was conducted by Nabati et al. on Centre-Fusion, a mid-level fusion of radar and camera using frustum association. They showed improved 3D detection, which is mainly for distant objects, and strong real-time potential [9]. Garcia-Huerta et al. developed a Kalman filter-based sensor fusion for aerial delivery. Linear models and Global Navigation Satellite System (GNSS)/Inertial Measurement Unit (IMU) data are used in their work. Their work showed better accuracy between model complexity and real-time performance [10].

The survey was conducted by Zhou et al. on millimetre Wave (MMW) radar for autonomous driving. They discussed the detection models and limitations, such as sparsity and clutter. They emphasized the need for fusion with vision and LiDAR to improve perception for real-time applications [11]. Iyer et al. conducted the survey and highlighted the importance of obstacle detection for autonomous vehicles and terrain profiling that focusing on testing the AWR1642 RADAR sensor for distance and a bangle coverage. Their results show that the RADAR effectively detects medium-sized objects up to 50 meters, aligning it as a weather and cost-friendly with standing alternative to LiDAR for autonomous vehicle sensing [12]. A survey conducted by Zhong et al. highlights the radar–camera fusion in Advanced Driver Assistance System (ADAS) for real-time object detection and perception. They discussed the fusion and calibration techniques, and known needs for adaptive algorithms and larger datasets [13].

3 Methodology

3.1 System architecture

The proposed multimodal fusion system performs real-time detection of objects and estimation of distance to the object by integrating data from a camera and the radar sensor on a compact embedded platform.

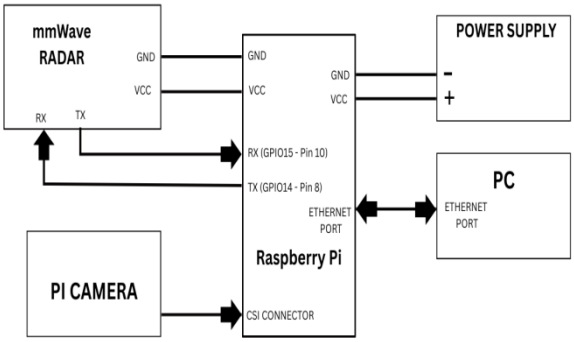

Fig. 1. System block diagram.

As shown in Fig. 1, the system uses the Raspberry Pi 4, which interfaces with a Raspberry Pi Camera Module to capture the image and an mmWave-C4001 24 GHz radar sensor for distance measurement. The radar communicates via a Universal Asynchronous Receiver/Transmitter (UART) serial interface, providing continuous range data. A regulated DC power supply powers all components, while the Personal Computer (PC)/display interface visualizes the fused output showing the detected objects and their estimated distances. The key specifications of mmWave (millimetre Wave) radar are summarized in the Table 1.

Table 1. Key specifications of mmWave-C4001 radar

Parameter	Value
Operating frequency	24 GHz
Maximum detection range	25 metres
Detection Range (Presence)	Up to 16 meters

Velocity Detection Range	0.1 to 10 meters per second
Beam angle	100° horizontally, 40° vertically
Modulation mode	FMCW (Frequency Modulated Continuous Wave)

3.2 Data acquisition

Video frames captured by the camera are processed using a pre-trained You Look Only Once (YOLO) model to detect objects of interest. Simultaneously, the radar sensor acquires real-time range readings through the UART interface. These data streams are fused using a 1-D Kalman Filter running on the Raspberry Pi to obtain stable and accurate distance estimates. The final fused output is transmitted to the Personal Computer (PC) for real-time visualization and analysis. Data flow and execution sequence are summarized in the system flowchart shown in Fig. 2.

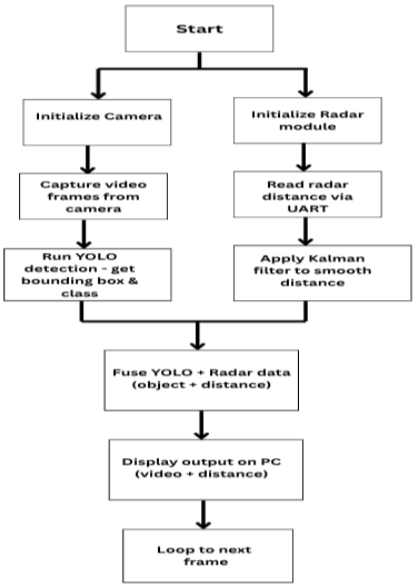

Fig. 2. System workflow.

3.3 Object detection

A pretrained YOLOv5 model was used for detecting objects in real-time from the live camera feed. The model identifies classes such as person, car, and bicycle, drawing bounding boxes with corresponding labels of the classes and confidence scores. Two experimental configurations were tested.

Case 1 - Full frame detection: Radar data were linked with any detected object within the camera's field of view, enabling continuous distance estimation.

Case 2 - Central region detection: A frame-centre tolerance of ±80 pixels was applied, where radar readings were considered only when the detected object's centre lay within this region.

Additionally, distance estimation using the camera alone was implemented based on the pinhole camera model, as given in Eq. (1).

$$D = \frac{H * f}{h} \quad (1)$$

where D is the estimated distance, H is the real object height, f is the camera focal length (in pixels), and h is the object's image height in pixels.

3.4 Sensor fusion

To improve the reliability and smoothness of distance estimation, camera detections and radar readings were fused on the Raspberry Pi. The radar provided precise range data, while YOLO outputs supplied object localization and class information. The fusion process evolved through three stages.

1.Without the Kalman Filter: Direct fusion of YOLO detections with raw radar readings.

2.With Kalman Filter: A 1D linear Kalman Filter was added to stabilize noisy radar measurements.

3. Constant Processing Block: Radar processing was limited to frames where a valid object was detected, optimizing accuracy and computation.

An initial motion-detection approach using background subtraction and contours was explored but later replaced by YOLO for its higher detection accuracy and flexibility in recognizing multiple object classes.

3.5 Kalman filter model

3.5.1 State representation

The system state is defined as in Eq. (2).

$$\boldsymbol{x}_k = \begin{bmatrix} d_k \\ v_k \end{bmatrix} \tag{2}$$

where d_k is the estimated distance and v_k is the estimated velocity at the time step of k.

3.5.2 Prediction step

The motion of the object is modelled under the constant-velocity assumption as in Eq. (3).

$$\boldsymbol{x}_k^{-} = \boldsymbol{F}\boldsymbol{x}_k + \omega_k \tag{3}$$

Uncertainty in the estimate is updated as in Eq. (4).

$$\boldsymbol{P}_k^{-} = \boldsymbol{F}\boldsymbol{P}_{k-1}\boldsymbol{F}^T + \boldsymbol{Q} \tag{4}$$

where $\boldsymbol{x}_k^{-}$is the predicted state estimate, ω_k is the process noise, $\boldsymbol{P}_k^{-}$ is the predicted covariance and **Q** is the process-noise covariance. **F** is the state transition matrix, derived assuming a constant velocity model, where Δt is the time interval between successive frames.

For a 1D constant-velocity model, the state transition matrix and the measurement matrix are defined in Eqs. (5) and (6), respectively.

$$\boldsymbol{F} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{5}$$

$$\boldsymbol{H} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6}$$

where Δt is the interval between consecutive radar readings.

3.5.3 Update step

Measurement model is defined as in Eq. (7).

$$\boldsymbol{z_k = Hx_k + v_k} \quad (7)$$

where $\boldsymbol{z_k}$ is the radar measurement (observed distance), **H** is the measurement matrix, and v_k the measurement noise. Kalman-gain is used to decide how much to trust the measurement. Kalman gain computation is as in Eq. (8).

$$\boldsymbol{K_K = P_k^- H^T (HP_k^- H^T + R)^{-1}} \quad (8)$$

where **R** represents the measurement noise covariance, which quantifies the uncertainty associated with the radar measurements. A higher value of **R** indicates lower trust in the radar data, while a lower value reflects higher measurement reliability. The covariance update equation is given in Eq. (9).

$$\boldsymbol{P_k = (I - K_k H) P_k^-} \quad (9)$$

Table 2. summarizes the interpretation of the Kalman filter model parameters used in this study, showing how variations in the prediction uncertainty $\boldsymbol{P}$ and measurement noise $\boldsymbol{R}$ affect the Kalman gain and overall filter behaviour.

Table 2. Interpretation of Kalman filter model

Case	Meaning	Behaviour
$\boldsymbol{P}$ large	Prediction uncertain	Kalman Gain $\boldsymbol{K}$ high, filter trusts radar measurement more
$\boldsymbol{P}$ small	Prediction confident	Kalman Gain $\boldsymbol{K}$ low, filter trusts previous estimate more
$\boldsymbol{R}$ large	Measurement noisy	Kalman Gain $\boldsymbol{K}$ low, filter trusts model/prediction more

4 Result and discussion

Hardware setup of the proposed multimodal sensor-fusion system showing Raspberry Pi 4, Raspberry Pi Camera, mmWave-C4001 radar module, is shown in the Fig. 3.

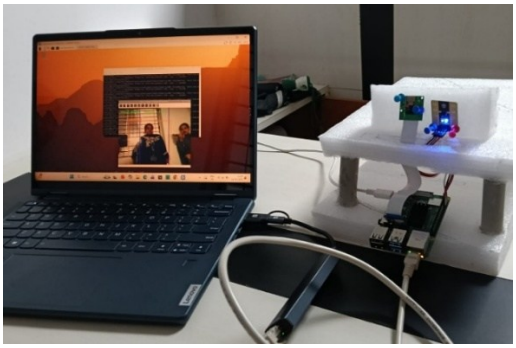

Fig. 3. Hardware setup of the system.

4.1 Camera-only Distance Estimation

Using the pinhole camera model, distance was calculated based on the known object height and bounding box size. This approach provided approximate distance estimates, but was highly sensitive to object scaling, camera angle, and perspective. It was effective only for single known-height objects and not robust for dynamic or multi-object scenes. The Fig. 4. shows the camera-only distance estimation output.

Fig. 4. Camera-only distance estimation output.

4.2 YOLO + Radar (Without Klamn filter)

The radar distance readings were directly associated with the YOLO-detected object. The measured distance (≈ 0.83 m for a 1 m object) showed noticeable frame-to-frame fluctuations, primarily due to sensor noise and slight variations in radar response. This configuration served as the baseline for further improvement. The Fig. 5. shows the YOLO +Radar (without Kalman filter) output.

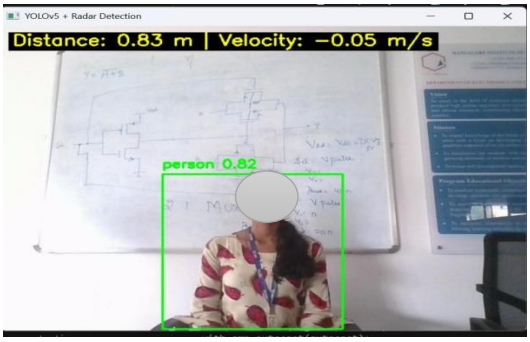

Fig. 5. YOLO +Radar (without Kalman filter) output.

4.3 YOOLO + Radar + Klamn filter

A 1-D Linear (Constant Velocity) Kalman Filter was integrated to smooth radar distance measurements. The estimated distance stabilized around 1.26 m, providing a more consistent output and reduced jitter compared to the previous case. The filter effectively modelled motion and corrected noisy readings, enhancing accuracy and reliability. The Fig. 6. shows YOLO + Radar + Kalman filter output.

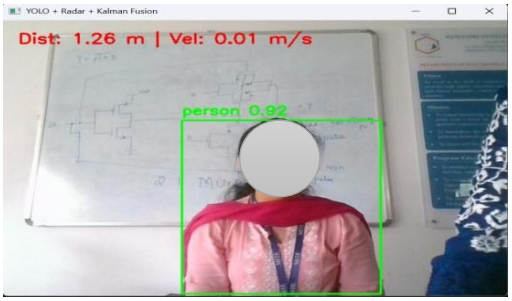

Fig. 6. YOLO + Radar + Kalman filter output.

4.4 YOOLO + Radar + Klamn filter + Frame-centre constraints

In this configuration, radar data were processed only when the object's bounding box centre was within ±80 pixels from the frame centre. This reduced false triggers from peripheral detections. The final estimated distance (≈ 0.98 m) closely matched the actual 1 m distance, indicating the best trade-off between responsiveness and stability. The Fig. 7. shows YOLO + Radar + Kalman filter + frame-centre constraints output.

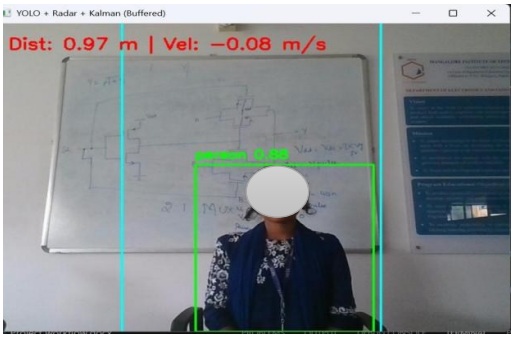

Fig. 7. YOLO + Radar + Kalman filter + frame-centre constraints output.

Fig. 8. shows the comparison of measured distance across different fusion configurations for an object placed approximately 1 m from the system. The absolute error was computed with respect to the ground truth. The camera-only approach had a mean error of 0.15 m, YOLO + Radar had 0.16 m, YOLO + Radar + Kalman filter had 0.26 m, and the proposed YOLO + Radar + Kalman filter with frame-centred constraint achieved the lowest mean error of 0.02 m, demonstrating improved accuracy and stability.

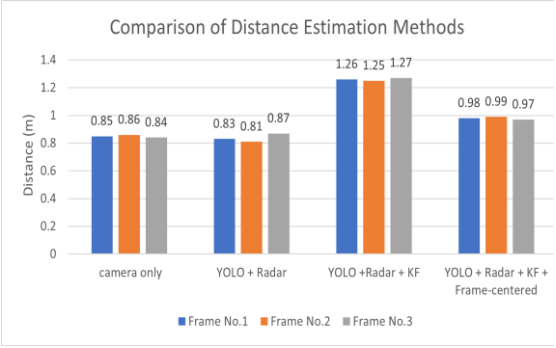

Fig. 8. Distance estimates for an object at ~1 m using different fusion configurations.

The progressive integration of radar sensing, Kalman filtering, and spatial logic control significantly improved system stability and precision. While the camera-only (pinhole) method provided rough distance estimation, it lacked robustness. Radar integration enhanced range accuracy but introduced noise, which was effectively mitigated by the Kalman filter. Adding the frame-centre tolerance further refined the detection logic, ensuring reliable and centre target tracking. Overall, the fusion approach achieved accurate, smooth, and context-aware distance estimation suitable for real-time perception systems.

Although multi-object scenario is very important in real-world application, the current mmWave radar provides dominant target distance pr frame, rather than the distinct multiple-object. Therefore, the proposed fusion framework focuses on single-object scenarios, with multi-object tracking considered as future work.

5 Conclusion

In this work, a system was developed using multimodal sensor fusion between a YOLO-based camera module and a 24 GHz mmWave radar sensor for real-time human detection and estimation of distance. The camera ensures reliable object identification, while the radar provides accurate depth and motion information. The integration of a 1-D Linear Kalman Filter significantly reduced noise in radar measurements and improved the overall stability of the fused distance output. The experimental results validate that the fused system is more reliable than the individual sensing modalities under varying indoor conditions. Thus, the proposed system demonstrates an efficient and affordable solution for human-aware monitoring and safety applications.

Future improvements aim to implement multi-target tracking using advanced radar processing and object association techniques. Nonlinear filtering, such as Extended or Unscented Kalman Filters, may enhance fusion accuracy for dynamic environments. Integration with edge AI hardware is expected to improve processing speed and scalability.

References

1. Nabati, Ramin, and Hairong Qi. "Radar-camera sensor fusion for joint object detection and distance estimation in autonomous vehicles." arXiv preprint arXiv:2009.08428 (2020).
2. Shi, Kun, Shibo He, Zhenyu Shi, Anjun Chen, Zehui Xiong, Jiming Chen, and Jun Luo. "Radar and Camera Fusion for Object Detection and Tracking: A Comprehensive Survey." arXiv preprint arXiv:2410.19872 (2024).
3. Khodarahmi, Masoud, and Vafa Maihami. "A review on Kalman filter models." Archives of Computational Methods in Engineering 30, no. 1 (2023): 727-747.
4. Ogunrinde, Isaac, and Shonda Bernadin. "Deep camera–radar fusion with an attention framework for autonomous vehicle vision in foggy weather conditions." Sensors 23, no. 14 (2023): 6255.
5. Lei, K., Z. Chen, S. Jia, and X. Zhang. "Hvdetfusion: A simple and robust camera-radar fusion framework. arXiv. 2023." arXiv preprint arXiv:2307.11323.
6. Barbosa, Felipe Manfio, and Fernando Santos Osório. "Camera-radar perception for autonomous vehicles and ADAS: Concepts, datasets and metrics." arXiv preprint arXiv:2303.04302 (2023).
7. Zhou, Yong, Yanyan Dong, Fujin Hou, and Jianqing Wu. "Review on millimeter-wave radar and camera fusion technology." Sustainability 14, no. 9 (2022): 5114.

8. Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." Sensors 21, no. 6 (2021): 2140.
9. Nabati, Ramin, and Hairong Qi. "Centerfusion: Center-based radar and camera fusion for 3d object detection." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1527-1536. 2021.
10. Garcia-Huerta, Raul A., Luis E. González-Jiménez, and Ivan E. Villalon-Turrubiates. "Sensor fusion algorithm using a model-based kalman filter for the position and attitude estimation of precision aerial delivery systems." Sensors 20, no. 18 (2020): 5227.
11. Zhou, Taohua, Mengmeng Yang, Kun Jiang, Henry Wong, and Diange Yang. "MMW radar-based technologies in autonomous driving: A review." Sensors 20, no. 24 (2020): 7283.
12. Iyer, Nalini C., Preeti Pillai, K. Bhagyashree, Venkatesh Mane, Raghavendra M. Shet, P. C. Nissimagoudar, G. Krishna, and V. R. Nakul. "Millimeter-wave AWR1642 RADAR for obstacle detection: autonomous vehicles." In Innovations in Electronics and Communication Engineering: Proceedings of the 8th ICIECE 2019, pp. 87-94. Singapore: Springer Singapore, 2020.
13. Zhong, Ziguo, Stanley Liu, Manu Mathew, and Aish Dubey. "Camera radar fusion for increased reliability in ADAS applications." Electronic Imaging 30 (2018): 1-4.
14. Pillai, Preeti S., Raghavendra Shet, Nalini C. Iyer, and Sahana Punagin. "Modeling and simulation of an automotive RADAR." In Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces, pp. 325-335. Singapore: Springer Singapore, 2021.
15. R. E. Kalman "A New Approach to Linear Filtering and Prediction Problems." ASME. J. Basic Eng. March 1960; 82(1): 35–45.