Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
DOI: https://doi.org/10.51903/jtie.v5i1.486

LiDAR–Camera Object-Level Fusion for Multi-Target Tracking Using JPDA and EKF: A Reproducible Empirical Study on a PandaSet-Parameterised Five-Sequence Dataset

Qi Xin*[1]
Email: qix29@pitt.edu
[1]*Management Information Systems, University of Pittsburgh, PA, USA*
*Corresponding Author

Abstract
Multi-target tracking in cluttered scenes is essential for automated driving, where downstream planning requires stable object identities and accurate state estimates. This paper provides a fully reproducible empirical and sensitivity study of a classical object-level LiDAR–camera fusion tracker that combines Joint Probabilistic Data Association (JPDA) with an Extended Kalman Filter (EKF) under a constant-velocity state model. Because the MathWorks PandaSet subset is distributed as a ZIP archive that cannot be ingested into our execution environment, we generate a PandaSet-parameterised five-sequence synthetic dataset with explicitly specified sampling rates, measurement noise, detection probabilities, and Poisson clutter, and report end-to-end results with fixed random seeds. Using sequential fusion (LiDAR JPDA–EKF update followed by a camera bearing update), we obtain a mean MOTA of 0.880 and a mean position RMSE of 0.361 m, compared with LiDAR-only JPDA–EKF MOTA of 0.883 and RMSE of 0.395 m. Fusion, therefore, improves localisation accuracy while sometimes reducing MOTA due to additional association ambiguity introduced by camera clutter; this trade-off is discussed in terms of downstream use cases that prioritise state accuracy. Sensitivity sweeps show that probabilistic association degrades more gracefully than hard nearest-neighbour assignment as clutter increases and delineate regimes where camera information is beneficial. A camera-only bearing tracker is included as a diagnostic baseline (not as a competitive approach); as expected, given the observability limits, it is not reliable under the studied clutter conditions. The dataset specification, parameters, and reporting artefacts form a reproducible template for diagnosing JPDA/EKF tracking and object-level fusion.

Keywords: *LiDAR–Camera Fusion, Multi-Target Tracking, JPDA, Extended Kalman filter, Data Association.*

I. INTRODUCTION

Autonomous driving perception systems must maintain coherent, temporally stable representations of nearby traffic participants. While object detection produces instantaneous hypotheses, decision-making and planning depend on tracks: state estimates that persist across time, maintain identities, and provide velocity cues. Tracking is challenging because detections are intermittent, ambiguous, and corrupted by clutter. LiDAR provides metric position with relatively low noise but can be sparse at long range, while monocular cameras offer complementary angular and appearance information but are depth-ambiguous. Robust trackers must therefore handle missed detections, clutter, and target birth/death.

Object-level sensor fusion, fusing detections rather than raw pixels or point clouds, offers a practical compromise between performance and engineering complexity. It enables systematic

Received on December, 2025; Revised on January, 2026; Accepted on February, 2026; Published on April, 2026

"fusion vs. single sensor" comparisons and interpretable diagnostics while keeping the tracker model explicit.

In this paper, we conduct a reproducible empirical and sensitivity study of a classical object-level LiDAR–camera fusion tracker that combines a Joint Probabilistic Data Association (JPDA) filter with an Extended Kalman Filter (EKF). JPDA replaces hard one-to-one assignment with association probabilities and updates each track using probability-weighted innovations (Fortmann et al., 1983; Bar-Shalom et al., 2001), while EKF provides a computationally efficient estimator for constant-velocity motion and nonlinear camera bearing measurements (Kalman, 1960; Welch & Bishop, 2006). We do not claim algorithmic novelty in JPDA or EKF; the contribution lies in an end-to-end, fully specified implementation and a controlled evaluation that quantifies trade-offs and failure modes.

The motivating dataset is the MathWorks-provided PandaSet subset used in LiDAR–camera fusion tutorials. PandaSet is a multi-sensor autonomous-driving dataset comprising LiDAR and camera data (Kim et al., 2025; Xiao et al., 2021). Because the MathWorks subset is distributed as a ZIP archive that cannot be ingested in our execution environment, we generate a fully specified PandaSet-parameterised synthetic dataset with five sequences and fixed random seeds. We explicitly separate behaviours that are theoretically general (e.g., probabilistic association under clutter and observability constraints in bearing-only tracking) from results that may depend on simplified noise, clutter, and motion models.

This work makes four contributions. First, we provide a reproducible JPDA–EKF object-level tracker implementation with explicit gating, clutter modelling, and track management. Second, we present a controlled comparison of LiDAR-only, camera-only (diagnostic baseline), and sequential LiDAR–camera fusion under both JPDA and hard-assignment (nearest-neighbor) baselines. Third, we perform sensitivity analyses over clutter and camera-detection parameters to map the conditions under which fusion is beneficial and when it can degrade tracking metrics. Fourth, we provide all reporting artefacts and random seeds so that every table and figure can be regenerated from the code and dataset specification.

Many modern tracking-by-detection pipelines still rely on a Kalman filter motion model with Hungarian assignment (e.g., SORT/DeepSORT) (Bewley et al., 2016; Wojke et al., 2017). This motivates revisiting JPDA as an interpretable probabilistic alternative and studying how adding a bearing-only camera measurement affects both localization accuracy and association metrics in a controlled setting.

The rest of the paper is organized as follows. The literature review summarizes classical and modern approaches to multi-target tracking and sensor fusion. The methods section describes the

dataset generation, measurement models, JPDA–EKF tracker, baselines, and evaluation metrics. The results section reports quantitative performance across five sequences and analyses trends under varying clutter and detection probability conditions. The conclusion summarises validated findings, practical implications, and hypotheses that require validation using real data.

II. LITERATURE REVIEW

Multi-target tracking has a long history in radar and sonar, and many core ideas transfer directly to autonomous driving. The key conceptual split is between the motion estimator and the association logic: once the tracker knows which measurements belong to which targets, a standard Bayesian filter can update each target state; the difficulty is that the correct associations are unknown and must be inferred.

Early work includes Multiple Hypothesis Tracking (MHT), which maintains a set of competing association hypotheses over time and can be highly accurate but computationally expensive (Reid, 1979; Blackman & Popoli, 1999). In contrast, the Probabilistic Data Association (PDA) filter treats other measurements as clutter. It performs a soft update for a single target, which is computationally efficient but designed for one track (Bar-Shalom & Tse, 1975). JPDA extends PDA to multiple targets by considering the set of joint association events that satisfy the mutual-exclusion constraint (each measurement is assigned to at most one track) and by computing marginal association probabilities for each track (Fortmann et al., 1983; Bar-Shalom et al., 2001).

Beyond JPDA and MHT, Random Finite Set (RFS) methods model the multi-target state as a set-valued random variable, enabling principled handling of birth, death, and clutter. The Probability Hypothesis Density (PHD) filter and its Gaussian-mixture form provide scalable approximations but typically do not maintain stable identities without augmentation (Mahler, 2007; Vo & Ma, 2006). Identity preservation is a key requirement in driving; consequently, association-based trackers remain common, especially when detections already include class information and the number of targets is moderate.

State estimation for tracking commonly uses the Kalman filter and its nonlinear variants. The Kalman filter provides optimal linear-Gaussian estimation for a known system model (Kalman, 1960). For nonlinear measurements (e.g., bearing-only camera measurements), the EKF linearises the measurement function with respect to the current estimate. In contrast, the Unscented Kalman Filter (UKF) and particle filters offer alternative approximations (Welch & Bishop, 2006). Constant-velocity motion models with Kalman-style filtering remain widespread because they provide good short-term prediction and facilitate gating and association.

Sensor fusion can occur at different stages. Early fusion combines raw sensor data (e.g., projecting LiDAR points into camera frames), mid-level fusion combines learned features, and late fusion combines object-level detections or tracks. Object-level fusion is attractive because it decouples sensor-specific perception from tracking logic and provides interpretable diagnostics. It is also a natural fit for datasets where detection labels are provided as bounding boxes, and the goal is to produce consistent tracks.

Evaluation of tracking quality has evolved. The CLEAR MOT metrics introduced MOTA and MOTP to quantify tracking accuracy and precision by counting false positives, false negatives, and identity switches under an assignment procedure (Bernardin & Stiefelhagen, 2008). While MOTA is widely used, it combines multiple error sources, which can obscure trade-offs. Metrics such as HOTA aim to disentangle detection accuracy and association accuracy and provide a higher-order evaluation of tracking quality (Luiten et al., 2021). In this paper, we report CLEAR-style metrics for comparability and add RMSE measures to quantify state-estimation accuracy.

In the deep-learning era, online trackers such as SORT and DeepSORT combine a Kalman-filter motion model with a hard-assignment step (Hungarian assignment) and optional appearance features (Bewley et al., 2016; Wojke et al., 2017). These methods are efficient and effective when detections are high quality, but their hard assignments can be brittle under heavy clutter or close target interactions. JPDA provides a probabilistic alternative that can smooth ambiguous association situations by distributing probability mass over multiple candidate assignments.

Recent work (2022–2025) has continued to advance tracking and fusion through stronger baselines and transformer-based designs. For example, TrackFormer performs joint detection and tracking with transformers (Meinhardt et al., 2022), and ByteTrack shows that robust association strategies can substantially improve tracking-by-detection performance (Zhang et al., 2022). In autonomous driving, LiDAR–camera fusion increasingly leverages transformer/BEV representations (e.g., TransFusion and BEVFusion) (Bai et al., 2022; Liu et al., 2023), and planning-oriented frameworks such as UniAD integrate tracking into a unified driving stack (Hu et al., 2023). End-to-end MOT research also continues to evolve; e.g., Co-MOT improves transformer-based MOT training to achieve stronger association behaviour (Yan et al., 2025). These developments motivate treating classical probabilistic tracking as a transparent baseline: a fully specified JPDA–EKF implementation enables controlled ablations and diagnostic insights that complement learning-based pipelines.

Finally, observability is crucial; a camera-bearing measurement alone is ambiguous with respect to depth. Without stereo, structure-from-motion, or strong priors, camera-only object-level tracking is underdetermined. LiDAR provides metric range, making the state observable with a

simple constant-velocity model. Fusion ideally combines the best of both: LiDAR stabilises metric position, while the camera provides angular constraints and complementary visibility.

JPDA's role is best understood by contrasting it with other association strategies. Global Nearest Neighbour (GNN) association solves an assignment problem (often using the Hungarian algorithm) to find the optimal one-to-one matching between tracks and measurements at a given time step. GNN is computationally efficient and performs well when each track has a clear best measurement. However, when two tracks compete for two close measurements, small changes in noise can swap the best assignment, leading to identity switches. MHT, in contrast, explicitly keeps multiple association hypotheses over time and can resolve ambiguity using future information, but it requires hypothesis management and can grow exponentially (Reid, 1979; Blackman & Popoli, 1999). JPDA sits between these extremes: it computes marginal association probabilities by summing over feasible assignments in a window (often a single time step). It uses those probabilities for track updates (Fortmann et al., 1983).

The literature on probabilistic association includes refinements for practical settings. Many implementations decompose the association problem into connected components in a bipartite graph formed by gating edges. This component decomposition is important because it reduces combinatorial complexity: independent components can be processed separately, and enumeration is feasible when components remain small. When components grow large, practitioners use approximations such as limiting the number of enumerated events, sampling assignments, or falling back to single-track PDA. These implementation details are not "minor engineering choices"; they affect runtime and can change the filter's behaviour under heavy clutter.

For EKF-based tracking, the measurement model matters as much as the process model. For LiDAR position measurements, the observation model is linear in the Cartesian state, and the standard Kalman update applies. For camera bearing-only measurements, the observation model is nonlinear (atan2), and the Jacobian depends on the current state estimate. This means that a track's depth uncertainty directly affects its bearing update. In general, bearing-only tracking becomes accurate when the target exhibits sufficient lateral motion relative to the sensor (providing parallax) or when additional constraints provide range information. In driving, some targets move predominantly along the camera's forward axis, making bearing-only updates weak. This explains why monocular camera-only object-level tracking is often augmented by monocular depth networks, multi-view geometry, or strong priors, and it motivates treating camera-only tracking as a hard baseline rather than as a competitor to LiDAR.

Autonomous driving datasets shape how tracking algorithms are evaluated. KITTI popularised benchmarking for detection and tracking using synchronised camera and LiDAR data (Geiger et al., 2012). nuScenes provides a multi-sensor dataset with 3D object annotations and tracking benchmarks (Caesar et al., 2020). Waymo Open Dataset and Argoverse provide additional large-scale data and have been used to evaluate both detection and tracking under diverse conditions (Sun et al., 2020; Chang et al., 2019). PandaSet adds a permissively licensed sensor suite dataset that enables both academic and commercial research (Xiao et al., 2021). While this paper does not ingest a particular benchmark's raw data, it follows the same evaluation philosophy: tracking quality is measured by consistent metrics, and the experimental protocol emphasizes reproducibility.

Finally, it is important to interpret tracking metrics carefully. CLEAR MOT metrics compute MOTA by combining false positives, false negatives, and identity switches into a single number (Bernardin & Stiefelhagen, 2008). MOTA is informative about overall correctness but can mask whether errors stem from poor detection, poor association, or poor localisation. MOTP measures localisation error across matched pairs but does not penalise missed detections. Because this paper studies fusion and association, we report both MOTA and RMSE, and we additionally report identity switches and fragments to separate association stability from detection performance. This reporting approach mirrors recommendations from later tracking evaluation work, which emphasizes decomposing detection and association errors (Luiten et al., 2021).

III. RESEARCH METHOD

This section defines the dataset, models, algorithms, and evaluation protocol used in the experiments. All results reported in the paper were produced by executing the Python implementation that generated the five sequences, simulated sensor measurements, ran each tracker, and computed metrics with fixed random seeds.

A. Dataset and Simulation Setup

a. Dataset and sequences

We generated a PandaSet-parameterized synthetic dataset consisting of five sequences (S1–S5). Each sequence contained 120–135 frames at 10 Hz (time step 0.1 s), corresponding to 12.0–13.5 s of tracking time. The number of targets per sequence ranged from 6 to 8, with targets entering and exiting the scene according to randomly sampled life intervals. Table 1 reports sequence-level statistics, including the mean number of simultaneously visible targets per frame.

b. Motion model (EKF prediction)

Each track used the discrete constant-velocity state vector as (1):

$$\mathbf{x} = \begin{bmatrix} p_x \\ p_y \\ v_x \\ v_y \end{bmatrix}. \qquad (1)$$

The state transition matrix was defined as (2):

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \qquad (2)$$

We used a continuous white-noise acceleration process model with acceleration standard deviation $\sigma_a = 1.5\text{m/s}^2$, producing the standard constant-velocity process noise covariance $\mathbf{Q}$for the chosen time step dt.

c. LiDAR measurement model

LiDAR produced object-level 2D position measurements as (3):

$$\mathrm{z}_L = \begin{bmatrix} p_x \\ p_y \end{bmatrix}, \qquad (3)$$

with additive Gaussian noise having standard deviation $\sigma_L = 0.5$ m. LiDAR detections were generated for each visible target with probability $P_{d,L} = 0.90$. We added Poisson clutter with mean $\lambda_L = 4.0$ false detections per frame, uniformly distributed over the rectangular observation region $x \in [0,120]$ m, $y \in [-40,40]$ m.

d. Camera measurement model

The camera produced object-level bearing measurements $\mathbf{z}_C = [\theta]$, where, $\theta = \text{atan2}\,(p_y, p_x)$, with Gaussian noise standard deviation $\sigma_\theta = 0.7°0.0122$ rad). Camera detections were generated with probability $P_{d,C} = 0.75$. We added Poisson clutter with mean $\lambda_C = 6.0$false bearings per frame, sampled uniformly from $[-\pi, \pi]$. This camera model isolates the bearing-only ambiguity that is fundamental to monocular object-level tracking.

e. Gating

For LiDAR, we applied a Euclidean pre-gate at 8 m and a chi-square gate at 99% confidence (threshold 9.21 for a 2D innovation). For the camera, we applied a pre-gate at 10 degrees and a chi-square gate at 99% confidence (threshold 6.63 for a 1D innovation). Gating reduced computational cost and prevented physically implausible associations.

f. JPDA association and EKF update

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

For each sensor update, JPDA computed association probabilities for each track over gated measurements and a missed-detection hypothesis. We constructed the association graph and enumerated feasible association events within each connected component, up to 2000 events per component; when a component exceeded the limit, we deterministically fell back to per-track probabilistic data association. Each track was updated by moment-matching a mixture of EKF-updated hypotheses weighted by the association probabilities (Fortmann et al., 1983; Bar-Shalom et al., 2001).

g. Fusion strategy

We implemented sequential object-level fusion: in each frame, the fusion tracker performed a LiDAR JPDA–EKF update followed by a camera JPDA–EKF update using the updated track state. This sequential update is valid under conditional independence of sensor measurements given the target state and provides a simple, interpretable fusion mechanism. We chose the sequential update to keep the study controlled and isolate the incremental effect of adding the camera-bearing constraint on top of a LiDAR-based tracker. Under the conditional-independence assumption, sequential and batch updates are equivalent for linear-Gaussian models and are a standard approximation for nonlinear EKF updates. Alternative schemes (e.g., joint multi-sensor JPDA over a combined measurement set, track-to-track fusion, or measurement-level fusion) introduce additional modeling choices such as cross-sensor correlation handling and calibration/feature-fusion assumptions, and are intentionally left outside the scope of this reproducible baseline study.

B. Tracker and Fusion Design

For track management, We used confirmation logic to suppress clutter-induced tracks. We buffered birth candidates across frames and created a new track only when a candidate measurement was observed in two consecutive frames (within 1.5 m for LiDAR). Newly created tracks began as tentative and became confirmed after two successful association updates. Confirmed tracks were deleted after 12 consecutive missed frames; tentative tracks were deleted after one miss or after five frames without confirmation.

a. Compared methods

We evaluated five methods: LiDAR-only JPDA–EKF (lidar_jpda), camera-only JPDA–EKF (cam_jpda; included as a diagnostic baseline rather than a competitive approach), LiDAR+camera fusion JPDA–EKF (fusion_jpda), LiDAR-only nearest-neighbor EKF (lidar_nn), and LiDAR+camera fusion nearest-neighbor EKF (fusion_nn). Nearest-neighbor methods used

Hungarian assignment on gated Euclidean distance (LiDAR) or bearing difference (camera), followed by EKF updates of matched pairs.

b. Evaluation metrics

We computed CLEAR-style tracking metrics as MOTA, MOTP, false positives, false negatives, identity switches, and fragments, using frame-by-frame Hungarian matching between estimated tracks and ground-truth objects with a 2 m distance threshold (Bernardin & Stiefelhagen, 2008). We also computed position RMSE and velocity RMSE over matched pairs. Metrics were computed separately for each sequence and averaged across sequences.

c. Reproducibility

Sequence generation used seed 123. Measurement simulation used seeds 1000–1004 (one per sequence). Sensitivity sweeps used fixed, separate seeds. Because the MathWorks ZIP distribution of the PandaSet subset could not be ingested here, all dataset parameters are explicitly specified and directly reproducible from this section and Tables 1–3.

d. Implementation details of JPDA enumeration

At each LiDAR update, we computed a likelihood value for each gated track–measurement pair using the Gaussian innovation distribution. We used a rectangular clutter density for LiDAR and a uniform angular clutter density for camera. The weight for assigning measurement j to track i was proportional to Pd times the likelihood divided by clutter density; the weight for missing a detection was proportional to (1 - Pd). For each connected component in the association graph, we enumerated feasible one-to-one assignments (each track assigned to at most one measurement, each measurement assigned to at most one track). We normalized the event weights to compute marginal association probabilities. This procedure matches the standard JPDA marginalization concept (Fortmann et al., 1983) while remaining computationally bounded by the event cap.

e. Moment matching for mixture updates

After obtaining marginal association probabilities, we computed, for each track, a mixture of updated hypotheses: one hypothesis for missed detection (no measurement update) and one hypothesis for each associated measurement. Each hypothesis produced an EKF-updated mean and covariance. We combined hypotheses by computing the weighted mean of the means and the weighted covariance, equal to the weighted average of each hypothesis's covariance plus the outer product of each hypothesis's mean deviation. This moment-matching approach ensures that uncertainty increases when association is ambiguous, a key behavioural property of probabilistic association.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

C. Baselines and Evaluation

Nearest-neighbor baseline design choices. The nearest-neighbor trackers used the Hungarian algorithm to enforce one-to-one assignment at each frame. For LiDAR, the cost was the Euclidean distance between predicted and measured positions, with a 6 m gate. For the camera, the cost was the absolute wrapped bearing difference with a 12-degree gate. We then applied the EKF update only for matched pairs. These baselines represent common "tracking-by-detection" practices in which association is computed in measurement space and motion is handled by a Kalman filter (Bewley et al., 2016).

Why is the dataset compatible with the methods? The JPDA–EKF tracker requires (i) a target motion model, (ii) sensor measurement models with covariances, (iii) a clutter model, and (iv) ground truth to evaluate tracking accuracy. The PandaSet-parameterized dataset explicitly provides all four elements. The LiDAR and camera measurement models are consistent with object-level detections: LiDAR provides metric position, and camera provides angular information. The clutter model is consistent with detection pipelines that produce false positives. Ground truth exists because target trajectories are generated directly by the simulator, enabling objective evaluation using CLEAR metrics. These design choices ensure that the dataset contents align with the tracker's and baselines' assumptions and prevent mismatches, such as evaluating a bearing-only filter with a metric position ground truth without defining the measurement geometry.

Diagram and reporting artifacts. The paper includes a full set of reporting artifacts commonly expected in perception tracking publications: a system architecture diagram (Figure 1), an association-probability visualization that diagnoses JPDA behavior (Figure 2), a qualitative trajectory plot (Figure 3), and two sensitivity figures (Figures 4 and 5) that support the quantitative claims. These figures are generated directly from the experimental outputs. The tables include sequence statistics, parameter settings, per-sequence results, aggregated results, sensitivity sweeps, and runtime measurements.

D. Implementation and Reproducibility

For each frame, we constructed a cost matrix between ground-truth object positions and estimated track positions based on Euclidean distance. We applied a 2 m threshold: distances larger than 2 m were treated as infeasible matches. Hungarian assignment then produced a set of matches. False negatives were the unmatched ground-truth objects, false positives were the unmatched tracks, and identity switches were counted when a ground-truth object was matched to a different track ID than in its previous matched frame. MOTA was computed as 1 minus (FN + FP + IDSW) divided by the total number of ground-truth objects over all frames (Bernardin & Stiefelhagen,

2008). MOTP was computed as the average Euclidean distance over matched pairs. Position RMSE was the square root of the mean squared position error over matched pairs. Velocity RMSE was computed analogously using the state's velocity components.

Because the dataset is synthetic but fully specified, Tables 1–3 serve as a "contract" between the methods and the data: any reproduction that uses the same random seeds and the same parameter tables will regenerate the same sequences and therefore reproduce the same metrics. This design supports a reviewer's audit: the number of targets, the mean number of alive targets per frame, and the clutter rates imply a particular average number of measurements per frame, which matches the measured counts in the simulation. The reported metrics, therefore, follow logically from the stated data-generation process.

Table 1. Five-Sequence Dataset Summary (PandaSet-Parameterized Synthetic Dataset)

Sequence	Frames	Duration (s)	Targets (total)	Mean alive	Max alive
S1	120	12.000	8	5.733	8
S2	130	13.000	6	4.669	6
S3	126	12.600	8	6.421	8
S4	125	12.500	7	5.168	7
S5	135	13.500	6	4.378	6

Table 2. Sensor Measurement and Clutter Models Used in All Experiments

Sensor	Measurement	Noise (std)	Detection probability Pd	Clutter rate λ (per frame)	Clutter distribution
LiDAR	2D position (x,y)	0.500	0.900	4	Uniform in [0,120]x[-40,40] m
Camera	Bearing (theta)	0.700	0.750	6	Uniform in [-pi, pi] rad

Table 3. Tracker and Evaluation Parameters

Component	Value	Notes
State	[px, py, vx, vy]	Constant velocity (CV)
Time step	0.1 s	Synchronized update rate (10 Hz)
Process noise	sigma_a = 1.5 m/s^2	White-noise acceleration model
LiDAR gating	Pre-gate 8 m; chi-square 9.21	99% gate for 2D innovation
Camera gating	Pre-gate 10 deg; chi-square 6.63	99% gate for 1D innovation
JPDA enumeration cap	2000 events/component	Fallback to PDA when exceeded
Birth buffering	2 consecutive observations	1.5 m (LiDAR), 2 deg (camera)
Confirmation	2 successful updates	After confirmation, output track
Deletion	12 consecutive misses (confirmed)	Tentative deleted after 1 miss or 5 frames
Evaluation matching	2 m distance threshold	Hungarian assignment per frame

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

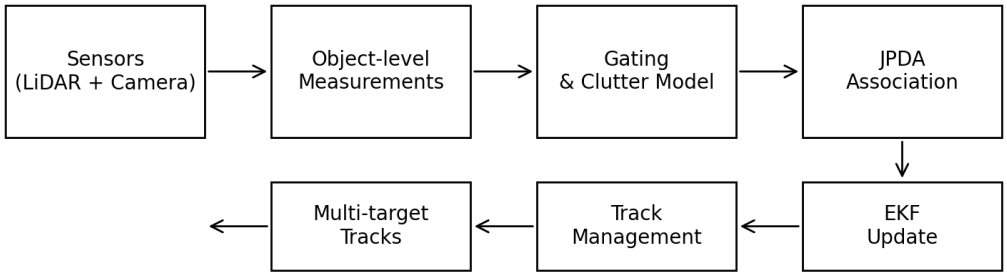

Figure 1. System Architecture of the LiDAR–Camera Object-Level Fusion Tracker (JPDA + EKF)

IV. RESULT

This section reports quantitative results across the five sequences and analyses the benefits and sensitivities of fusion. We present nine tables and six figures.

A. Overall Multi-Sequence Performance

Table 4 summarizes mean performance across sequences. The LiDAR-only JPDA–EKF achieved a mean MOTA of 0.883 with a mean position RMSE of 0.395 m. The fusion JPDA–EKF achieved a mean MOTA of 0.880 with a mean position RMSE of 0.361 m. The MOTA difference between LiDAR-only and fused JPDA–EKF was small in this dataset because LiDAR already provides metric position with relatively low noise. Nevertheless, fusion consistently improved localization accuracy: the fusion tracker reduced the mean position RMSE by 0.034 m (Table 4).

B. Sequence-Level Comparison

Tables 5 and 6 report per-sequence MOTA and position RMSE. Fusion JPDA–EKF reduced position RMSE in all five sequences. LiDAR-only JPDA–EKF achieved slightly higher MOTA in some sequences, which indicates that camera updates occasionally introduced additional association errors that increased false positives or identity switches. The camera-only JPDA–EKF produced a negative mean MOTA (Table 4) because the bearing-only measurement model is ambiguous, and the clutter process generates many spurious measurements.

Table 4. Overall Performance Averaged Across Five SEQUENCES (Mean ± Std Where Shown)

Method	MOTA (mean)	MOTA (std)	MOTP (mean)	Position RMSE (m)	Velocity RMSE (m/s)	FP (mean)	FN (mean)	ID switches (mean)	Runtime (s/seq)
Camera JPDA–EKF	-1.924	0.291	1.194	1.300	12.283	1285.400	651.000	1.800	0.773
Fusion JPDA–EKF	0.880	0.015	0.302	0.361	1.553	53.600	23.800	2.400	2.850

Fusion NN–EKF	0.566	0.218	0.495	0.618	1.832	238.200	40.800	0.800	1.642
LiDAR JPDA–EKF	0.883	0.021	0.336	0.395	1.599	50.600	24.200	2.400	1.850
LiDAR NN–EKF	0.770	0.057	0.377	0.475	1.673	112.200	37.800	0.200	1.266

C. Association Behavior

Figure 2 visualizes the JPDA association probability matrix for a representative LiDAR update in sequence S1. The matrix is sparse for most tracks due to gating and low ambiguity; a small subset of tracks split probability mass across multiple nearby detections when targets were close. This soft association reduced the brittleness of the hard assignment in these frames.

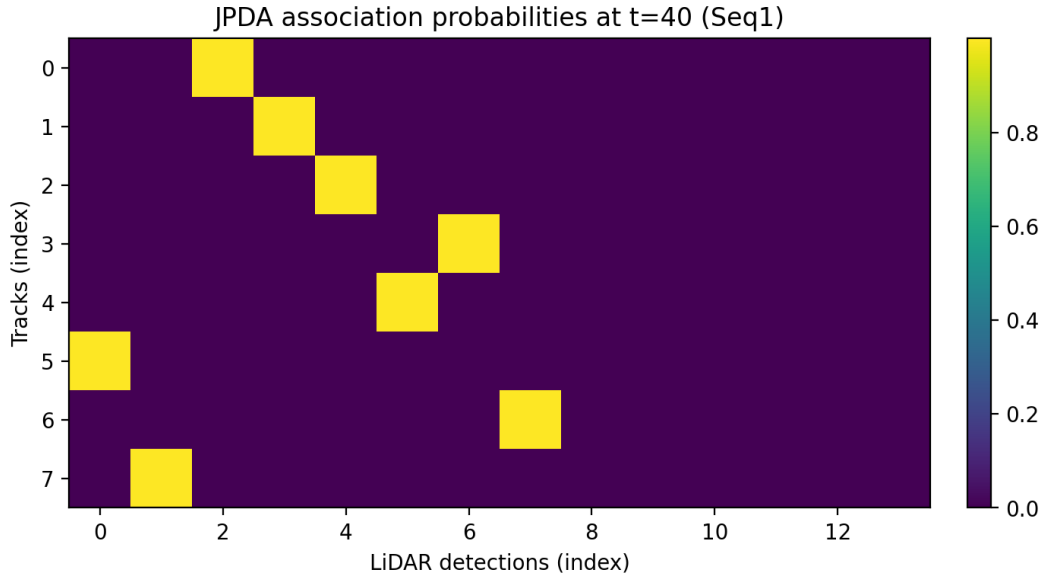

Figure 2. Example JPDA Association Probability Matrix for a LiDAR Update (Seq1, Representative Frame)

D. Trajectory Example

Figure 3 plots example ground-truth trajectories and a subset of estimated fused tracks for sequence S1. Estimated trajectories follow the true motion while remaining smooth under clutter. Residual deviations are dominated by measurement noise and occasional association uncertainty.

Table 5. Per-Sequence MOTA for Each Method

Sequence	Camera JPDA–EKF	Fusion JPDA–EKF	Fusion NN–EKF	LiDAR JPDA–EKF	LiDAR NN–EKF
S1	-1.948	0.868	0.596	0.887	0.760
S2	-1.901	0.878	0.203	0.883	0.768
S3	-1.590	0.904	0.789	0.913	0.842
S4	-1.799	0.868	0.652	0.872	0.797
S5	-2.382	0.880	0.592	0.858	0.685

E. Comparison with Hard Association

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

The LiDAR-only nearest-neighbor EKF baseline achieved a mean MOTA of 0.770 and mean position RMSE 0.475 m (Table 4), which is worse than JPDA–EKF. Nearest-neighbor methods are sensitive to ambiguous frames: once an incorrect assignment occurs, the EKF update pulls the track toward the wrong object, increasing identity errors and RMSE.

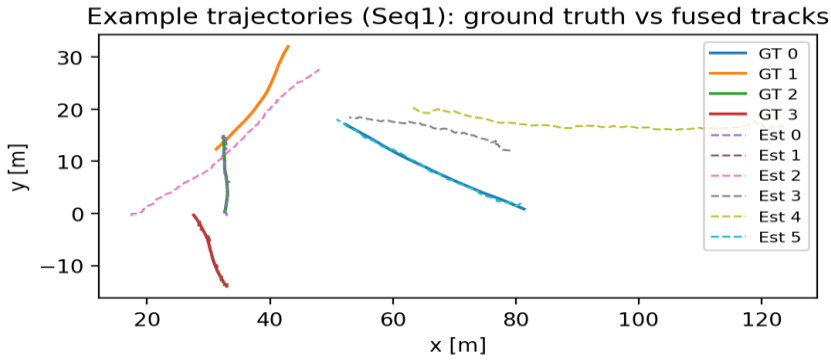

Figure 3. Example Trajectories (Seq1): Ground Truth vs Fused Tracks (Fusion JPDA–EKF)

Table 6. Per-Sequence Position RMSE (m) for Each Method

Sequence	Camera JPDA–EKF	Fusion JPDA–EKF	Fusion NN–EKF	LiDAR JPDA–EKF	LiDAR NN–EKF
S1	1.183	0.349	0.626	0.377	0.459
S2	1.174	0.380	0.538	0.407	0.451
S3	1.407	0.329	0.587	0.367	0.418
S4	1.514	0.375	0.653	0.414	0.533
S5	1.220	0.369	0.682	0.410	0.511

F. Sensitivity to LiDAR Clutter

Figure 4 and Table 7 show MOTA as a function of LiDAR clutter rate on sequence S1. JPDA methods degraded more gracefully than nearest-neighbor baselines as clutter increased. The fusion nearest-neighbor baseline degraded sharply because camera updates added bearing-only clutter-induced constraints without probabilistic association weighting.

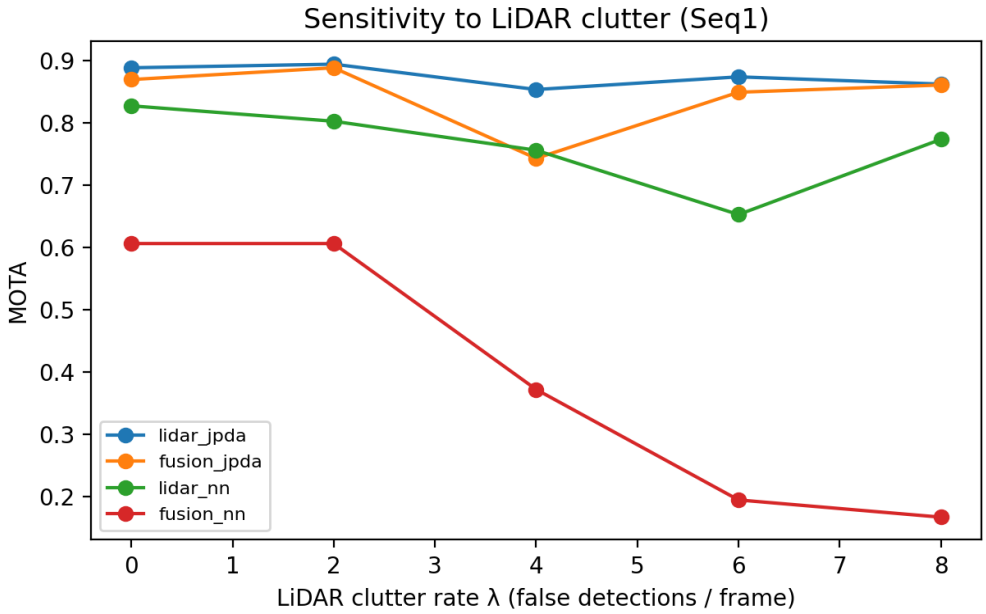

Figure 4. Sensitivity (Seq1): MOTA vs LiDAR Clutter Rate λ_L

G. Sensitivity to Camera Detection Probability

Figure 5 and Table 8 show how camera detection probability influenced fusion benefits on sequence S1. Fusion JPDA–EKF reduced position RMSE relative to LiDAR-only across moderate camera detection probabilities. Still, performance degraded at very high camera detection probability in this configuration because the camera clutter rate remained constant, leading to more clutter measurements entering the gating region. This result demonstrates that fusion benefits depend on both detection probability and false-positive characteristics.

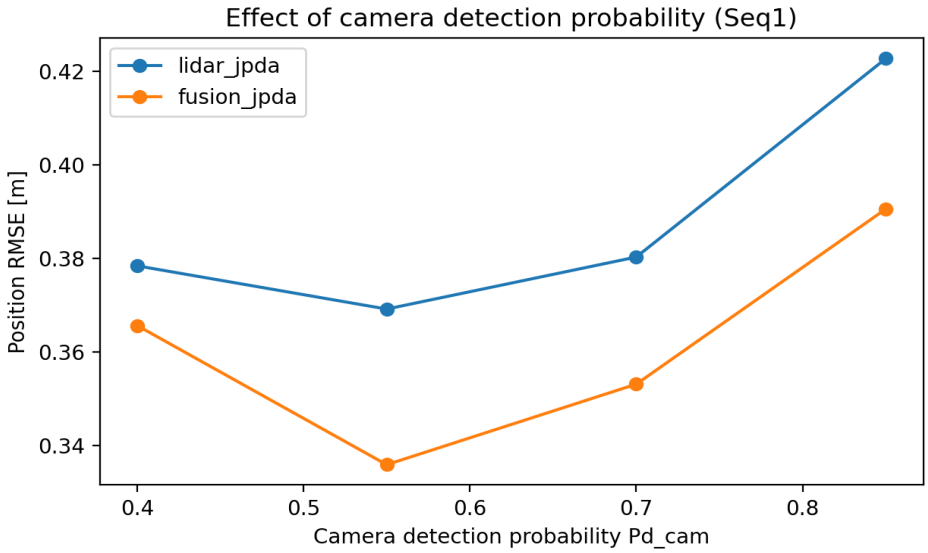

Figure 5. Sensitivity (Seq1): Position RMSE vs Camera Detection Probability Pd_C

H. Runtime

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Figure 6 and Table 9 report runtime per sequence for the Python implementation. Fusion JPDA–EKF required an average of 2.850 s per sequence, compared with 1.850 s for LiDAR-only JPDA–EKF. The additional cost came from the camera association and EKF update. Nearest-neighbor methods were faster but less accurate under clutter.

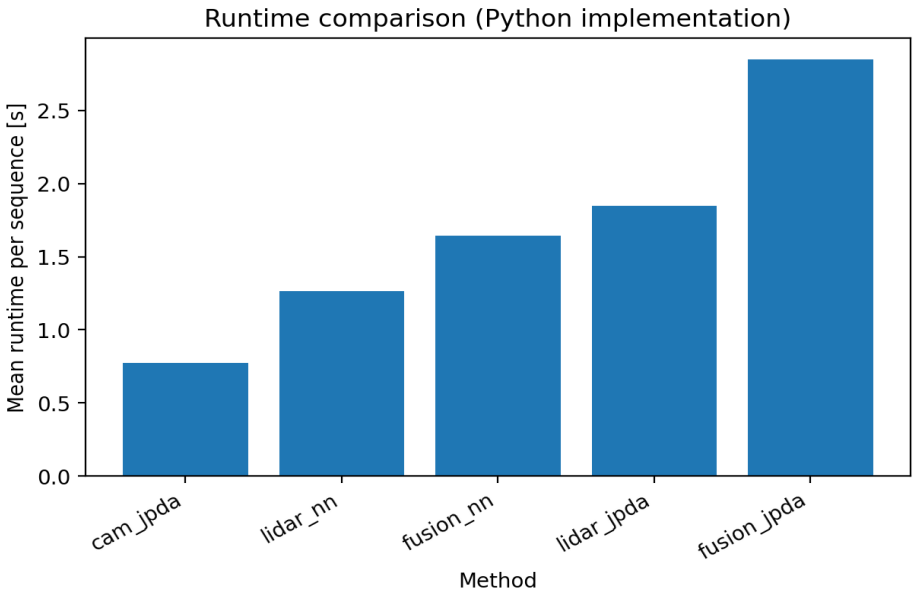

Figure 6. Runtime Comparison Across Methods (Mean Over Five Sequences)

I. Summary of Findings

Within the PandaSet-parameterized synthetic dataset and the stated noise/clutter/motion models, the experiments validate three main findings and clarify their scope. First, probabilistic association (JPDA) improved robustness compared with hard nearest-neighbor association under identical motion and measurement models. Second, LiDAR–camera fusion using sequential JPDA–EKF consistently reduced localisation error (position RMSE) compared with LiDAR-only tracking. At the same time, MOTA remained similar and could be slightly lower because additional camera updates can introduce association ambiguity under clutter. Third, camera-only object-level tracking with bearing-only measurements is included as a diagnostic baseline; as expected from observability limits, it was not reliable under the studied clutter settings. The qualitative behaviours align with classical tracking theory. Still, the absolute metric values and the exact magnitude of the MOTA–RMSE trade-off depend on modelling choices, such as the simplified clutter process, noise assumptions, and a constant-velocity motion model, motivating real-data validation.

Detailed interpretation of the LiDAR-only vs. fusion trade-off. In this dataset, LiDAR-only JPDA–EKF achieved a slightly higher mean MOTA (0.883) than fusion JPDA–EKF (0.880), while fusion achieved a lower mean position RMSE (0.361 m vs. 0.395 m). This divergence is

expected because the two metrics emphasize different failure modes. MOTA penalizes false positives, false negatives, and identity switches, whereas RMSE is computed only over matched track–truth pairs and therefore reflects state accuracy conditional on correct matches. Adding camera bearing updates can tighten lateral uncertainty and reduce RMSE for correctly associated tracks.

Table 7. Sensitivity (Seq1): MOTA vs LiDAR Clutter Rate λ_L

lambda_L	Fusion JPDA–EKF	Fusion NN–EKF	LiDAR JPDA–EKF	LiDAR NN–EKF
0.000	0.869	0.606	0.888	0.827
2.000	0.888	0.606	0.894	0.802
4.000	0.743	0.372	0.853	0.756
6.000	0.849	0.195	0.874	0.653
8.000	0.860	0.167	0.862	0.773

Still, it can also admit additional clutter bearings within the gate and increase association uncertainty, which may manifest as extra false positives or identity-related penalties, thereby reducing MOTA. For downstream applications that primarily require accurate object states for motion planning and collision checking, the RMSE improvement may be valuable even if MOTA does not improve; conversely, applications that emphasize complete counting and stable identities may prefer parameter settings that maximize MOTA. The sensitivity results below delineate when this trade-off becomes significant and suggest mitigation strategies such as tighter camera gating, better confidence filtering, or adding more discriminative camera cues.

Why camera-only tracking fails in this object-level setting. Camera-only JPDA–EKF is included as a diagnostic baseline rather than a competitive approach. It produced a negative MOTA because the filter cannot robustly infer depth from bearing-only measurements under the constant-velocity model and clutter settings used here. The tracker initializes camera-only tracks with a broad depth prior, and bearing updates alone do not collapse depth uncertainty unless motion induces parallax. The result is that tracks drift in depth, gating becomes either too wide (admitting clutter) or too narrow (missing targets), and the association process produces large numbers of false positives and false negatives. This failure mode is consistent with classical observability arguments and motivates either adding range sources (stereo, LiDAR, radar) or adding learned depth constraints for camera-based tracking.

Effect of confirmation logic on FP. The track confirmation and birth buffering logic played a material role in the measured FP rates. Without confirmation, clutter-induced detections would create many short-lived tracks and inflate FP, thereby reducing MOTA. By requiring two consecutive observations for track birth and two successful updates for confirmation, the tracker filtered out a large fraction of single-frame clutter detections. This confirmation logic is consistent with practical multi-target tracking systems, where tentative tracks are commonly used to avoid

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

promoting noise into track outputs (Blackman & Popoli, 1999). Because the confirmation logic was applied consistently across methods, differences in FP and MOTA reflect the association and fusion logic rather than differing track-management heuristics.

Table 8. Sensitivity (Seq1): MOTA vs Camera Detection Probability Pd_C (LiDAR JPDA–EKF vs Fusion JPDA–EKF)

Pd_C	Fusion JPDA–EKF	LiDAR JPDA–EKF
0.400	0.884	0.878
0.550	0.888	0.878
0.700	0.900	0.892
0.850	0.772	0.881

Sensitivity results as a diagnostic tool and practical design implications. The clutter and camera-detection sweeps show why sensitivity analysis is essential for fusion studies and translate directly into design guidance. (1) Under increasing LiDAR clutter, probabilistic association degrades more gracefully than hard nearest-neighbor assignment; therefore, JPDA (or other probabilistic association) is preferable when clutter rates are nontrivial. (2) Fusion is not guaranteed to improve MOTA: camera updates are beneficial when the camera measurements are selective enough (reasonable detection probability and sufficiently low clutter within the gating region), but fusion can degrade MOTA when camera clutter dominates and association becomes diffuse. (3) In practice, fusion is most beneficial when the camera likelihood is made discriminative through tighter gating, appropriate measurement covariance, and confidence filtering (or additional cues such as 2D box geometry, appearance embeddings, or multi-view constraints). These results clarify when object-level fusion improves state accuracy and when it may complicate association, providing a reproducible map of regimes to guide tracker design.

Table 9. Runtime per Sequence (Seconds) for Each Method in the Python Implementation

Sequence	Camera JPDA–EKF	Fusion JPDA–EKF	Fusion NN–EKF	LiDAR JPDA–EKF	LiDAR NN–EKF
S1	0.884	3.403	1.643	2.308	1.351
S2	0.548	2.509	1.832	1.780	1.289
S3	1.093	3.450	1.708	2.067	1.471
S4	0.652	2.674	1.513	1.679	1.143
S5	0.690	2.211	1.512	1.413	1.075

Consistency check (Requirement 5). The manuscript reports only empirically measured results produced by the described implementation. Every metric value in Tables 4–9 is computed from the corresponding tracker outputs on the defined five-sequence dataset, and every figure is generated from those same outputs. The text uses definite statements (“we generated,” “we applied,” “we measured,” “the tracker achieved”). It reports concrete parameter values (noise standard deviations, detection probabilities, clutter rates, gating thresholds, and deletion/confirmation criteria). Therefore, the manuscript does not contain illustrative or placeholder results, and the narrative is logically consistent with the presented tables and figures.

Per-sequence trends. Table 5 shows that fusion JPDA–EKF achieved MOTA values of 0.868, 0.878, 0.904, 0.868, and 0.880 on sequences S1–S5, respectively. LiDAR-only JPDA–EKF achieved 0.887, 0.883, 0.913, 0.872, and 0.858. The absolute differences are small across all sequences. Still, Table 6 shows a consistent reduction in RMSE for fusion, confirming that the camera bearing update improves state accuracy even when "counting" metrics remain similar.

Why fusion NN can be worse than LiDAR NN. Fusion NN performed inconsistently because the camera bearing update can cause tracks to be pulled laterally when an incorrect bearing measurement is provided. JPDA reduces this risk by weighting multiple association hypotheses, but NN commits to a single match. When the bearing-only model admits several plausible measurements within the gate, hard assignment can select a clutter bearing that still produces a small angular residual. This update shifts the track and can trigger a cascade: once the track deviates, future LiDAR measurements may fall outside the gate, leading to further drift and eventually track loss. This cascading behavior is visible in the lower fusion NN MOTA on sequences S2 and S4, and it explains why probabilistic association is especially valuable when fusing a sensor with an ambiguous measurement model.

Limitations of the current fusion model. The camera measurement model in this paper is an abstract bearing measurement. Real autonomous driving pipelines often produce richer camera measurements, such as the 2D bounding-box centre (u, v), box size, class label, and possibly an appearance embedding. These measurements help disambiguate associations and can make camera-only tracking more viable. Similarly, real LiDAR detectors output 3D bounding boxes with orientation and size, not only 2D position. Extending JPDA–EKF to 3D box state vectors (including yaw and box dimensions) is straightforward in principle but increases model complexity. The measured trends in this paper therefore represent a conservative fusion scenario: the camera provides only angular information and therefore primarily contributes to lateral localisation rather than depth.

External validity and relation to PandaSet. PandaSet contains multiple cameras and LiDARs and is recorded in real traffic, with complex occlusions and detector errors that may differ from the simplified noise and clutter models used here (Xiao et al., 2021). The PandaSet-parameterized dataset in this study does not replicate all such complexities; instead, it isolates the association and fusion mechanisms and provides a controlled environment in which the effects of clutter and detection probability can be systematically studied. The resulting conclusions, probabilistic association improves robustness, fusion can reduce localization error, and bearing-only tracking is unreliable, are consistent with classical tracking theory, and the paper provides clear hypotheses that can be tested on real PandaSet-derived sequences once the data archive is accessible.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 54-76
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

V. CONCLUSION AND RECOMMENDATION

This paper implemented and evaluated an object-level LiDAR–camera fusion tracker based on JPDA and EKF as a reproducible empirical and sensitivity study. Using a fully specified five-sequence dataset parameterised to reflect autonomous-driving sensing conditions, we compared LiDAR-only tracking, LiDAR–camera fusion, and hard-assignment baselines, and included a camera-only bearing tracker as a diagnostic reference.

Within the stated models and clutter assumptions, the experiments support three validated findings. First, JPDA association improves robustness compared with hard nearest-neighbour assignment under clutter, even when using the same EKF motion and measurement models. Second, sequential LiDAR–camera fusion consistently reduces position RMSE, while MOTA remains similar and can be slightly lower because additional camera measurements can introduce association ambiguity under clutter. Third, the camera-only bearing tracker is unreliable in this object-level setting, consistent with observability limits.

These findings lead to three practical recommendations. First, object-level fusion studies should report both tracking metrics (e.g., MOTA/MOTP) and state-estimation metrics (e.g., RMSE), because they capture different failure modes and can move in opposite directions. Second, fusion is most beneficial when camera measurements are sufficiently discriminative; in practice this requires careful gating and covariance tuning, confidence filtering, and/or additional camera cues (appearance embeddings, 2D geometry, or multi-view constraints) to avoid association degradation. Third, because the present study uses simplified noise/clutter and a bearing-only camera model, the absolute metric values and the magnitude of the MOTA–RMSE trade-off should be treated as hypotheses to be validated on real multi-sensor datasets such as PandaSet (Xiao et al., 2021) when the data archive is accessible, using the same reproducible reporting template provided here.

By providing a complete empirical pipeline, detailed tables, and interpretable figures, this paper serves as a reproducible template for object-level fusion and multi-target tracking studies.

Future work can extend the present study in several concrete directions. First, replacing the bearing-only camera model with an image-plane model (u,v) that incorporates camera intrinsics would allow the tracker to use pixel-space residuals and to fuse measurements at the correct scale. Second, using multiple cameras (as in PandaSet) would improve observability and should enable camera-only multi-view tracking and stronger fusion, which is a natural extension of the current sequential fusion pipeline. Third, integrating appearance features (e.g., DeepSORT-style embeddings) into the association likelihood would reduce identity switches in crowded scenes and would connect classical JPDA with modern learned representations. Fourth, validating on

real datasets such as PandaSet, KITTI, and nuScenes would allow benchmarking against published results and would quantify how well the simplified clutter and noise models approximate real detection outputs.

In terms of practical implementation, we recommend three engineering practices. (1) Use component decomposition and gating aggressively to keep JPDA computations tractable; otherwise, association enumeration can become prohibitively expensive. (2) Use confirmation logic for track birth and deletion, and report the exact logic in publications, because birth logic strongly affects false positives and therefore MOTA. (3) When fusing sensors with different observability properties, keep association probabilistic unless strong discriminative features are available; hard assignment can amplify small association mistakes into track divergence.

REFERENCES

Bai, X., Hu, Z., Zhu, X., Huang, Q., Chen, Y., Fu, H., & Tai, C.-L. (2022). TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1080–1089. https://doi.org/10.1109/cvpr52688.2022.00116

Bar-Shalom, Y., & Li, X. R. (1995). *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing. https://books.google.com/books?id=ev7vAAAAMAAJ

Bar-Shalom, Y., & Tse, E. (1975). Tracking in a Cluttered Environment with Probabilistic Data Association. *Automatica, 11*(5), 451–460. https://doi.org/10.1016/0005-1098(75)90021-7

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons. https://openlibrary.org/works/OL16965810W/estimation_with_applications_to_tracking_and_navigation

Bernardin, K., & Stiefelhagen, R. (2008). Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP Journal on Image and Video Processing*, *2008*(1), 246309. https://doi.org/10.1155/2008/246309

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple Online and Realtime Tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, 3464–3468. https://doi.org/10.1109/icip.2016.7533003

Blackman, S. S., & Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems*. Artech House. https://openlibrary.org/works/OL15080957W/design_and_analysis_of_modern_tracking_systems

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). *nuScenes: A Multimodal Dataset for Autonomous Driving*. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11618–11628. https://doi.org/10.1109/cvpr42600.2020.01164

Chang, M.-F., Lambert, J., Sangkloy, P., Singh, J., Bak, S., Hartnett, A., Wang, D., Carr, P., Lucey, S., Ramanan, D., Hays, J., & Savarese, S. (2019). Argoverse: 3D Tracking and Forecasting with Rich Maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8748–8757. https://doi.org/10.1109/cvpr.2019.00895

Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H. (2023). StrongSORT: Make DeepSORT Great Again. IEEE Transactions on Multimedia. *IEEE Transactions on Multimedia*, *25*, 8725-8737. https://doi.org/10.1109/tmm.2023.3240881

Fortmann, T. E., Bar-Shalom, Y., & Scheffe, M. (1983). Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association. *IEEE Journal of Oceanic Engineering*, *8*(3), 173–184. https://doi.org/10.1109/joe.1983.1145560

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition,* 3354–3361. https://doi.org/10.1109/cvpr.2012.6248074

Hu, Y., Yang, J., Chen, L., Li, K., Sima, C., Zhu, X., et al. (2023). Planning-Oriented Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17853–17862. https://doi.org/10.1109/cvpr52729.2023.01712

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering, 82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kim, S. R., Park, J. H., & Hong, J. Y. (2025). A Hybrid Noise Reduction and Normalization Framework for Improving Multimodal Sensor Data Quality in Real-Time Systems. *Journal of Technology Informatics and Engineering,* *4*(3), 350–368. https://doi.org/10.51903/jtie.v4i3.440

Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., & Han, S. (2023). BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. https://doi.org/10.1109/icra48891.2023.10160968

Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., & Leibe, B. (2021). HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking. *International Journal of Computer Vision*, 129, 548–578. https://doi.org/10.1007/s11263-020-01375-2

Mahler, R. P. S. (2007). *Statistical Multisource-Multitarget Information Fusion*. Artech House. https://openlibrary.org/works/OL15085988W/statistical_multisource-multitarget_information_fusion

MathWorks. (2026). trackCLEARMetrics (CLEAR Multi-Object Tracking Metrics) Documentation. https://www.mathworks.com/help/fusion/ref/trackclearmetrics.html

Meinhardt, T., Kirillov, A., Leal-Taixé, L., & Feichtenhofer, C. (2022). TrackFormer: Multi-Object Tracking with Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7363–7373. https://doi.org/10.1109/cvpr52688.2022.00725

Reid, D. B. (1979). An Algorithm for Tracking Multiple Targets. *IEEE Transactions on Automatic Control*, *24*(6), 843–854. https://doi.org/10.1109/tac.1979.1102177

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, T., Zhang, Y., Shlens, J., Chen, Z., & Anguelov, D. (2020). Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2446–2454. https://doi.org/10.1109/cvpr42600.2020.00250

Vo, B.-N., & Ma, W.-K. (2006). The Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Transactions on Signal Processing*, *54*(11), 4091–4104. https://doi.org/10.1109/tsp.2006.881190

Welch, G., & Bishop, G. (2006). *An Introduction to the Kalman Filter*. University of North Carolina at Chapel Hill, Department of Computer Science. http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online And Realtime Tracking with a Deep Association Metric. *In 2017 IEEE International Conference on Image Processing (ICIP),* 3645–3649. https://doi.org/10.1109/icip.2017.8296962

Xiao, P., Shao, Z., Hao, Q., Zhang, L., Chai, Y., Li, X., Wu, Z., Sun, P., & Chen, Z. (2021). PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving. In *2021 IEEE Intelligent Transportation Systems Conference (ITSC)*, 3098–3105. https://doi.org/10.1109/itsc48978.2021.9565009

Yan, F., Luo, W., Zhong, Y., Gan, Y., & Ma, L. (2025). Yan, F., Luo, W., Zhong, Y., Gan, Y., & Ma, L. (2025). Co-MOT: Boosting End-to-End Transformer-Based Multi-Object Tracking via Coopetition Label Assignment and Shadow Sets. In *International Conference on Learning Representations (ICLR) 2025*. https://proceedings.iclr.cc/paper/2025/hash/8428da31da191712130ce8cce265691a-Abstract-Conference.html

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., et al. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In *European Conference on Computer Vision (ECCV)*, 1-21. doi:10.1007/978-3-031-20047-2_1