JOURNAL OF AEROSPACE INFORMATION SYSTEMS
Vol. 21, No. 4, April 2024

AFJPDA: A Multiclass Multi-Object Tracking with Appearance Feature-Aided Joint Probabilistic Data Association

Sukkeun Kim,* Ivan Petrunin,† and Hyo-Sang Shin‡
Cranfield University, Cranfield, England MK43 0AL, United Kingdom

https://doi.org/10.2514/1.I011301

This study addresses a multiclass multi-object tracking problem in consideration of clutters in the environment. To alleviate issues with clutters, we propose the appearance feature-aided joint probabilistic data association filter. We also implemented simple adaptive gating logic for the computational efficiency and track maintenance logic, which can save the lost track for re-association after occlusion or missed detection. The performance of the proposed algorithm was evaluated against a state-of-the-art multi-object tracking algorithm using both multiclass multi-object simulation and real-world aerial images. The evaluation results indicate significant performance improvement of the proposed method against the benchmark state-of-the-art algorithm, especially in terms of reduction in identity switches and fragmentation.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

I. Introduction

MULTI-TARGET tracking (MTT) is a technique used to track multiple targets using diverse sensors for a range of applications. Many approaches, such as the Kalman filter (KF), joint probabilistic data association (JPDA) filter [1–4], multiple hypothesis tracking (MHT) filter [5], and particle filter (PF) [6], have been utilized for many MTT applications [7,8]. Multi-object tracking (MOT) can be considered as a special case of MTT that is usually using image sensors for pedestrian tracking, and it has been studied actively in recent years [9–12]. Many MOT algorithms showed significant improvements in the aspect of not only its tracking performance, but also algorithm speed [13,14]. Some MOT approaches adopted appearance features (named re-identity (re-ID) feature in some articles) from the objects on the image. Adoption of appearance features such as color intensity or histogram [15–17] or convolutional features [18–20] helps to overcome the drawbacks of image tracker that can happen owing to occlusions or missed detection [21,22]. Besides utilizing appearance features, Cao et al. [23] proposed observation-centric strategies to tackle the problems after occlusion in MOT. These generate virtual trajectories for re-updating and consider the consistency of the observation to enhance performance in nonlinear scenarios. However, the observation-centric approach proposed in OC-SORT presumes that the observation is reliable, and this algorithm continues to rely on a hard association based on the intersection over union (IoU) metric, implying that its performance could diminish in cluttered, multiclass situations. This is another issue to be addressed: multiclass multi-object tracking (MCMOT) and the clutter generated by the sensor, specifically, the detection and classification algorithm in this instance.

Even though the real-life scenario has multiclass objects, most of the existing MOT algorithms are not directly applicable to multiclass scenarios. This is because most of the previously proposed algorithms are focused only on single class (pedestrian) tracking problem, which is known as the MOT challenge [9,10]. In the early stage of MCMOT algorithm development, detectors based on conventional image processing were introduced. Notably, Bose et al. [24] proposed the blob-based detection and target-set tracking algorithm with hard association, which associates one measurement to only one track, capable of handling fragmentation and grouping problems. This algorithm, however, is not suitable for tracking multiclass objects with vastly different sizes due to the inherent limitation of the blob-based detector. Another MCMOT algorithm proposed by Spinello et al. [25] utilized an implicit shape model (ISM) detector and two-step association tracker using the extension of Munkres' method [26] for rectangular assignment. This algorithm made a use of both vision and laser range sensors for the measurement update, and hence the algorithm is not directly comparable with other MCMOT algorithms. In addition, Zhang et al. [27] proposed a multiclass tracking algorithm based on the background subtraction with Gaussian mixture, histogram of oriented gradient features, and KF with hard association. However, this algorithm focused more on detection and classification rather than tracking, and assigning the track identity (ID) is not considered in their algorithm. Most of the recent MCMOT algorithms utilized convolutional neural network (CNN)-based detectors, owing to the robustness of CNN-based detectors in various image conditions and their ability to produce classification results simultaneously. For example, Lee et al. [28] proposed a tracker based on an ensemble of Faster R-CNN and Kanade–Lucas–Tomasi feature tracker, along with a Markov chain Monte Carlo–based Bayesian filter. While this algorithm uses changing point detection based on the observation likelihood to detect ID switches or fragmentation, it is not capable to continuing the track; it can only merge fragmented tracks later. For another example, Jo et al. [29] proposed a tracker based on YOLOv2 and KF with Hungarian algorithm. More recently, Micheal and Vani [20] proposed a tracker based on the tiny-deeply supervised object detector, bidirectional-forward long short-term memory, and Hungarian algorithm for the aerial image application. Regardless of the type of detector, the majority of approaches rely on hard association algorithms, such as the Hungarian algorithm [30]. The issue with this hard association is that the performance could be significantly degraded when there exist clutters in the environment, which is likely prevalent in real operational environments: the environment with clutters could result in a large number of ID switches, fragmentation, or both. A multiclass extended version of the state-of-the-art MOT algorithm, FairMOT [18], also exists [31], but the limitation of this approach is that association is done only within the same class by the Hungarian algorithm. This implies that the ID of the objects can be fragmented or switched if the classification result changes.

To overcome the limitation of hard association in multiclass tracking scenarios with clutter and missed detection/classification, we proposed a MCMOT algorithm, called appearance feature-aided joint probabilistic data association (AFJPDA) filter. The proposed AFJPDA leverages the soft association concepts, which associate multiple candidate measurements with its association probability within a gating area to tracks, into the extended version of FairMOT [31]. We have chosen the extended version of FairMOT as our

Received 22 April 2023; revision received 10 October 2023; accepted for publication 24 October 2023; published online Open Access 29 December 2023. Copyright © 2023 by Sukkeun Kim, Ivan Petrunin, and Hyo-Sang Shin. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. All requests for copying and permission to reprint should be submitted to CCC at www.copyright.com; employ the eISSN 2327-3097 to initiate your request. See also AIAA Rights and Permissions www.aiaa.org/randp.

*Ph.D. Researcher, School of Aerospace, Transport and Manufacturing.
†Reader, School of Aerospace, Transport and Manufacturing.
‡Professor, School of Aerospace, Transport and Manufacturing; h.shin@cranfield.ac.uk. Member AIAA (Corresponding Author).

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

foundational algorithm, given that it is the most up-to-date algorithm that is directly applied to MCMOT. Considering the balance between tracking performance and computational complexity, the JPDA filter is chosen as the soft association filter over the other algorithms. Data association is performed based on the distance between two data points, and the distance utilized in JPDA is typically a probabilistic distance such as Mahalanobis distance. Since the adoption of the appearance feature can improve practical issues of image tracker, which are caused by occlusions or missed detection, this study also incorporates the appearance feature similarity distance into the JPDA filter. By integrating the JPDA filter and appearance feature simultaneously, the proposed AFJPDA algorithm significantly reduces the number of ID switches and fragmentation and improves other performance metrics consequently in cluttered environments. Using multiple measurements for association, however, can increase the computation significantly, so we proposed a simple adaptive gating logic to alter the size of the gating area. Furthermore, we proposed a modified track maintenance logic to re-associate the missed detection when the objects reappear after occlusion. The tracking performance was evaluated against the extended version of FairMOT, using both simulation (Carla [32]) and real-world (VisDrone [33]) images taken from an unmanned aerial vehicle (UAV).

The main contributions of this paper are as follows:

1) We propose a tracking filter that is directly applicable for multiclass multi-object tracking and significantly reduces fragmentation and ID switch by using a JPDA filter and appearance feature.

2) A simple adaptive gating logic to reduce the computational complexity and a modified track maintenance logic for re-association are proposed.

3) The proposed filter with the adaptive gating logic and the track maintenance logic was tested in both simulation and real-world aerial videos.

The rest of this paper is composed as follows: Sec. II recaps the JPDA filter and Sec. III explains the AFJPDA implementation with adaptive gating logic and modified M/N logic. Section IV discusses the simulation results of the MOT algorithms and compares the performance. Finally, in Sec. V we conclude this paper.

II. Joint Probabilistic Data Association Filter

The JPDA filter is one of the most popular data association techniques for MTT based on the KF, which can be divided into two steps: prediction and update steps [34]. The key difference of the JPDA filter as a soft association filter from the widely used hard association such as global nearest neighbor (GNN) [35] is that the JPDA utilizes association probabilities and innovations of all measurements within the gating area to calculate weighted innovation [3,36]. Here, the innovation is the difference between the measurement and the predicted state, which will be defined in Eq. (8), and the weighted innovation is the weighted sum of innovations based on association probability, which will be defined in Eq. (9). Compared to hard association algorithms that may wrongly associate a track with clutter, the JPDA filter offers an advantage in cluttered scenarios by considering multiple measurements within a gating area and jointly associating them based on their association probabilities. Therefore, when the Euclidean distances between one track and multiple measurements are similar, the JPDA filter is more advantageous than hard association approaches. Additionally, in scenarios where classification results are ambiguous, detectors and classifiers can produce clutter. For example, the detection algorithm can detect two objects (a car and a bus) in an image when only one object (a bus) exists in reality, as shown in Fig. 1.

For completeness, let us briefly recap the JPDA filter for the following linear Gaussian Markov system:

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k \tag{1}$$

$$z_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{2}$$

where $\boldsymbol{x}_k$, $\boldsymbol{F}_k$, and $\boldsymbol{w}_k \sim \mathcal{N}(0, \boldsymbol{Q}_k)$ are state vector, transition matrix, and process noise with covariance matrix $\boldsymbol{Q}_k$ at time step k, respectively. Also, z_k, $\boldsymbol{H}_k$, and $\boldsymbol{v}_k \sim \mathcal{N}(0, \boldsymbol{R}_k)$ are measurement vector, measurement matrix, and measurement noise with covariance matrix $\boldsymbol{R}_k$ at time step k, respectively. From the prediction step, the predicted state and covariance of ith track can be derived as follows:

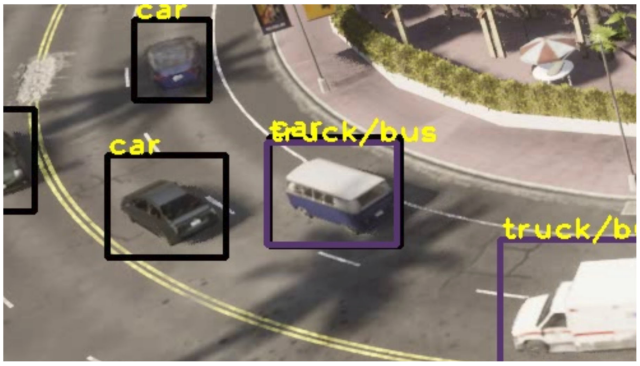


Fig. 1 Clutter example with ambiguous classification result.

$$\hat{\boldsymbol{x}}^i_{k|k-1} = \boldsymbol{F}_k \hat{\boldsymbol{x}}^i_{k-1|k-1} \tag{3}$$

$$\boldsymbol{P}^i_{k|k-1} = \boldsymbol{F}_k \boldsymbol{P}^i_{k-1|k-1} \boldsymbol{F}^T_k + \boldsymbol{Q}_k \tag{4}$$

where $\hat{\boldsymbol{x}}^i_{k|k-1}$ is the predicted state and $\boldsymbol{P}^i_{k|k-1}$ is the predicted covariance of ith track at time step k with measurement up to time step $k-1$. When measurements are received, the predicted state and covariance can be updated as follows:

$$\hat{\boldsymbol{x}}^i_{k|k} = \hat{\boldsymbol{x}}^i_{k|k-1} + \boldsymbol{K}^i_k \boldsymbol{y}^i_k \tag{5}$$

$$\boldsymbol{P}^i_{k|k} = \boldsymbol{P}^i_{k|k-1} - (1 - p^{i0}_k) \boldsymbol{K}^i_k \boldsymbol{H}_k \boldsymbol{P}^i_{k|k-1} + \boldsymbol{P}^i_k \tag{6}$$

where $\hat{\boldsymbol{x}}^i_{k|k}$ and $\boldsymbol{P}^i_{k|k}$ are the updated state and updated covariance. Here, p^{i0}_k is the probability that no measurement is assigned to ith track and $\boldsymbol{P}^i_k$ is the correction term, which can be calculated as follows:

$$\boldsymbol{P}^i_k = \boldsymbol{K}^i_k \left[\sum_{j=1}^{J} p^{ij}_k \boldsymbol{y}^{ij}_k \boldsymbol{y}^{ijT}_k - \boldsymbol{y}^i_k \boldsymbol{y}^{iT}_k \right] \boldsymbol{K}^{iT}_k \tag{7}$$

Here, the association probability of jth measurement corresponding to ith track, p^{ij}_k, will be derived in Eq. (15). Refer to [36,37] for the details. In Eqs. (5–7), the innovation $\boldsymbol{y}^i_k$ and the Kalman gain $\boldsymbol{K}^i_k$ can be derived as follows:

$$\boldsymbol{y}^{ij}_k = z^j_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}^i_{k|k-1} \tag{8}$$

$$\boldsymbol{y}^i_k = \sum_{j=1}^{J} p^{ij}_k \boldsymbol{y}^{ij}_k \tag{9}$$

$$\boldsymbol{S}^i_k = \boldsymbol{H}_k \boldsymbol{P}^i_{k|k-1} \boldsymbol{H}^T_k + \boldsymbol{R}_k \tag{10}$$

$$\boldsymbol{K}^i_k = \boldsymbol{P}^i_{k|k-1} \boldsymbol{H}^T_k (\boldsymbol{S}^i_k)^{-1} \tag{11}$$

where $\boldsymbol{y}^{ij}_k$ is the innovation of ith track corresponding to jth measurement, z^j_k, p^{ij}_k is the association probability of jth measurement corresponding to ith track, and $\boldsymbol{S}^i_k$ is the innovation covariance matrix of ith track. Note that there is no minimum number of measurements for reliable calculation of $\boldsymbol{y}^i_k$ since JPDA will utilize only one innovation if it has been associated with one measurement and will be considered as lost track if no measurement associated to track based on M/N logic, which will be discussed in Sec. III.

The JPDA filter, unlike the classic linear KF, calculates the weighted innovation in Eq. (9) using the innovation of the jth measurement corresponding to the ith track, $\boldsymbol{y}_k^{ij}$, and the corresponding association probability, p_k^{ij}, which serves as a weighting parameter. Joint computation of the association probability for all candidate measurements within a gating area, γ_G, is performed by the JPDA filter, and this gating is based on the Mahalanobis distance of the jth measurement as follows:

$$d_{M,k}^{ij} = \sqrt{(\boldsymbol{y}_k^{ij})^T (\boldsymbol{S}_k^i)^{-1} \boldsymbol{y}_k^{ij}} \tag{12}$$

If the Mahalanobis distance between the jth measurement and the predicted state of the ith track, $d_{M,k}^{ij}$, is less than the gating threshold γ_G, the measurement is considered a potential candidate for association with the track.

With the gated measurements, the association can be divided into three cases to calculate the association probability: The case when jth measurement is assigned to ith track, $g_{ij}P_D$; the case when no measurement is assigned to ith track, $1 - P_D$; and the case when none of the measurements assigned to any tracks (all measurements are false alarm), β. Here, P_D is the probability of detection, β is the probability of false alarm, and g_{ij} is Gaussian likelihood of jth measurement assigned to ith track which can be denoted as follows:

$$g_k^{ij} = \frac{e^{-D_{M,k}^{ij}/2}}{(2\pi)^{\dim/2}\sqrt{|S_k^i|}} \tag{13}$$

where $D_{M,k}^{ij}$ is the squared Mahalanobis distance, $D_{M,k}^{ij} = (d_{M,k}^{ij})^2 = (\boldsymbol{y}_k^{ij})^T (\boldsymbol{S}_k^i)^{-1} \boldsymbol{y}_k^{ij}$, *dim* is dimension of measurement, and $|\boldsymbol{S}_k^i|$ is the determinant of $\boldsymbol{S}_k^i$. From the above equations, the probability of assignment of all objects to all measurements, P_a, can be calculated as follows:

$$P_a = \prod_{j \text{ to } i} g_k^{ij} P_D \prod_{\text{null to } i} (1 - P_D) \prod_{j \text{ to null}} \beta \tag{14}$$

Finally, the association probability of jth measurement corresponding to ith track, p_k^{ij}, can be calculated by summing up all probability, which includes the cases that assigning jth measurement to ith track as follows:

$$p_k^{ij} = \sum_{a=1}^{A} \{P_a | \text{includes cases assigning } j\text{th measurement to } i\text{th track}\} \tag{15}$$

where A is the number of all possible cases of assignment between all tracks and all measurements. Refer to [1,2,36] for further details.

III. Appearance Feature-Aided JPDA Filter

A. Appearance Feature and Fused Distance

A number of studies have proposed using appearance features (also known as re-ID features) to improve the association and tracking performance for the MOT challenge [18,22,31]. These approaches fuse the Mahalanobis distance and cosine distance to represent the physical probabilistic distance and appearance similarity of objects, respectively. The Mahalanobis distance is calculated in the same manner as in Eq. (12), and the cosine distance between two appearance feature vectors $\boldsymbol{f}_k^i$ and $\boldsymbol{f}_k^j$ can be calculated as follows:

$$\text{Cosine similarity} = \frac{\boldsymbol{f}_k^i \cdot \boldsymbol{f}_k^j}{\|\boldsymbol{f}_k^i\| \cdot \|\boldsymbol{f}_k^j\|} \tag{16}$$

$$d_{C,k}^{ij} = 1 - \text{Cosine similarity} \tag{17}$$

The appearance features in Eq. (16) are 128-dimensional vectors that are similar to vectors for classification. Since there are similarities in color or shape of the ground vehicles, designing of feature vector based on those features would be subject to our future study. Nonetheless, as there is a certain tradeoff between performance and computational cost, care should be taken in designing such feature vectors. More detailed information on the appearance feature (or re-ID feature) vector used in Eq. (16) can be found in [18]. We propose a new approach, called appearance feature-aided JPDA (AFJPDA), which uses a fusion of Mahalanobis distance and cosine distance between objects in image coordinates to improve the performance of association after occlusion or missed detection. The fused distance is calculated as a linear combination of the two distances, as defined in previous research [18,22,31] and used for the base algorithm, FairMOT [18], as well:

$$d_{\text{fused},k}^{ij} = \lambda d_{C,k}^{ij} + (1 - \lambda) d_{M,k}^{ij} \tag{18}$$

where λ is the weighting parameter, and we selected $\lambda = 0.6$ empirically. The sensitivity analysis of this weighting parameter λ will be discussed in Sec. IV. The fused distance in Eq. (18) is used not only in gating but also for association probability calculation in Eq. (13), then the new likelihood can be calculated as Eq. (13) but simply substitute the $D_{M,k}^{ij}$ into squared fused distance, $D_{\text{fused},k}^{ij} = (d_{\text{fused},k}^{ij})^2$. In this manner, the AFJPDA filter can utilize the appearance feature of the objects for the association. It is worth mentioning that the calculation of Mahalanobis distance and fused distance in Eq. (18) will not increase the computational complexity compared to the base algorithm since the base algorithm also utilizes this fused distance for its data association with the Hungarian algorithm. Refer to Sec. IV.D and the last column of Tables 3 and 4 for the computation comparison.

B. Adaptive Gating Logic

The computational complexity of the JPDA filter could dramatically increase if too many measurements are considered as potential assignments. To avoid this computational burden, it is important to filter out the measurements not relevant to the object using gating [35]. Choosing a proper gating size that can contain the possibly assigned measurements but considering the computational complexity is the point of the gating. In the case we are focusing on, using a vision sensor from a UAV, applying the same gating area for all objects may degrade the association performance for the objects moving close to the camera. To address this issue, we propose an adaptive gating logic that can change the gating size depending on the size of the bounding box, which can represent the distance from the camera implicitly while considering computational efficiency. The proposed gating size γ_G is obtained as follows:

$$\gamma_G = G_{\text{scr}} * l_{\text{bb}} \tag{19}$$

where G_{scr} is the scaling parameter, which we set to 0.0265 empirically, and l_{bb} is the diagonal length of the bounding box of the detected object. In addition, we set the lower bound of the gating size to $\sqrt{9.488}$ based on the chi-square distribution table for a 0.95 probability of detection with four degrees of freedom to prevent the gating area from being too small. This approach allowed us to achieve computational costs that are reasonably comparable to those of the GNN.

C. Track Maintenance Logic

The track maintenance logic is important for stable tracking in the case where clutter or missed detection exists. For track maintenance, the M/N logic and the score-based maintenance methods are widely utilized for tracking problems. In this study, we applied M/N logic, which is widely appreciated thanks to its simplicity ([2] p. 104) since score-based logic can increase the computation and M/N can perform well in the low-clutter environment (e.g., image sensor). We modified M/N for track maintenance as missed measurements can affect the stability of tracking. This logic confirms a track if measurements are associated M times within N time steps (frames in this case), and terminates the track if there are no associations for M_{term} time steps. Unlike standard M/N logic, the modified version saves lost tracks as tentative tracks if there are no associations for M time

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

steps. Following an occlusion or missed detection event, the AFJPDA filter uses a track maintenance logic that attempts to re-establish the association between lost objects and tentative tracks using newly detected measurements. This re-association process is achieved by considering the appearance feature in addition to the association probability, thereby allowing more stable tracking performance and enhancing the robustness of the tracker with occlusion and missed detections. In this study, we set $M = 2$, $N = 3$, and $M_{\text{term}} = 30$, meaning that the logic confirms a track if measurements are associated two times within three frames, considers it a tentative track if there are no associations for two frames, and terminates the track if there are no associations for an additional 30 frames.

D. Algorithm of the AFJPDA Filter

The proposed AFJPDA filter with the adaptive gating logic and the track maintenance logic can be formulated as Algorithm 1. The flowchart of the entire algorithm, including the detection, classification, and appearance feature extraction parts, is described in Fig. 2 in Sec. IV.

Algorithm 1: AFJPDA filter with adaptive gating logic and modified M/N logic

Input: $\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1}, \boldsymbol{z}_k, \boldsymbol{f}_k, \boldsymbol{F}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{R}, \lambda, G_{scr}, p_D, \beta, N, M, M_{term}$
Output: $\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k}, List_{conf,k}, List_{tent,k}, List_{term,k}$

for i, j in I, J **do**
  $\hat{\boldsymbol{x}}^i_{k|k-1} = \boldsymbol{F}\hat{\boldsymbol{x}}^i_{k-1|k-1}$
  $\boldsymbol{P}^i_{k|k-1} = \boldsymbol{F}\boldsymbol{P}^i_{k-1|k-1}\boldsymbol{F}^T + \boldsymbol{Q}$
  $\boldsymbol{S}^i_k = \boldsymbol{H}\boldsymbol{P}^i_{k|k-1}\boldsymbol{H}^T + \boldsymbol{R}$
  $\boldsymbol{K}^i_k = \boldsymbol{P}^i_{k|k-1}\boldsymbol{H}^T(\boldsymbol{S}^i_k)^{-1}$
  $\boldsymbol{y}^{ij}_k = \boldsymbol{z}^j_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}^i_{k|k-1}$
  $d^{ij}_{M,k} = \sqrt{(\boldsymbol{y}^{ij}_k)^T(\boldsymbol{S}^i_k)^{-1}\boldsymbol{y}^{ij}_k}$
  $d^{ij}_{C,k} = 1 - \frac{\boldsymbol{f}^i_k \cdot \boldsymbol{f}^j_k}{\|\boldsymbol{f}^i_k\| \cdot \|\boldsymbol{f}^j_k\|}$
  $d^{ij}_{fused,k} = \lambda d^{ij}_{C,k} + (1-\lambda)d^{ij}_{M,k}$
  $\gamma_G = G_{scr} * l_{bb}$
  if $\gamma_G \leq \sqrt{9.488}$ **then**
    $\gamma_G \leftarrow \sqrt{9.488}$
  end if
  if $d^{ij}_{fused,k} \leq \gamma_G$ and P_a contains association of j to i **then**
    $p^{ij}_k = \Sigma^A_{a=1} P_a$
  end if
  if $\max(p^{ij}_k) == (1 - p_D)\beta$ **then**
    $Con^i_{tent} \leftarrow Con^i_{tent} + 1$
    $Con^i_{term} \leftarrow Con^i_{term} + 1$
  end if
  $\boldsymbol{y}^i_k = \sum^J_{j=1} p^{ij}_k \boldsymbol{y}^{ij}_k$
  $\hat{\boldsymbol{x}}^i_{k|k} = \hat{\boldsymbol{x}}^i_{k|k-1} + \boldsymbol{K}^i_k\boldsymbol{y}^i_k$
  $\boldsymbol{P}^i_{k|k} = \boldsymbol{P}^i_{k|k-1} - (1 - p^{i0}_k)\boldsymbol{K}^i_k\boldsymbol{H}_k\boldsymbol{P}^i_{k|k-1} + \boldsymbol{P}^i_k$
  if Number of association j to i>M within N frames **then**
    Append i^{th} track to $List_{conf,k}$
    $Con^i_{tent} \leftarrow 0$
    $Con^i_{term} \leftarrow 0$
  end if
  if $Con^i_{tent} \geq M$ **then**
    Append i^{th} track to $List_{tent,k}$
  end if
  if $Con^i_{term} \geq M_{term}$ **then**
    Append i^{th} track to $List_{term,k}$
  end if
  $\boldsymbol{Z}^i_k = \{\boldsymbol{z}^j_k | d^{ij}_{fused,k} < \gamma_G\}$
  if $\boldsymbol{z}^j_k \notin \boldsymbol{Z}^i_k$ **then**
    Initiate $\boldsymbol{z}^j_k$ as a new track
  end if
end for

The AFJPDA filter operates in several steps. First, it predicts the state and covariance of the objects being tracked based on the state and covariance of the previous frame, using a given set of dynamics. Next, the filter calculates the Kalman gain and innovation between the track and the measurements received, and uses this to calculate the Mahalanobis and cosine distances, which are fused to obtain a final distance measure. Third, the filter calculates an adaptive gating area based on the size of the bounding box and uses this to derive the association probability of the jth measurement corresponding to the ith track. The filter then updates the tentative and terminates condition variables for unobserved tracks and associates tracks with measurements. Finally, tracks are confirmed, appended to tentative tracks, terminated, or initiated based on an M/N logic.

IV. Simulation Results and Discussion

A. AFJPDA Filter Implementation

The proposed AFJPDA filter was implemented for MCMOT using the image sensor in both simulation and real-world scenarios, with multiple multiclass objects. The algorithm was based on the framework of a one-shot multiclass multi-object tracker [31], which is an extended version of FairMOT [18]. DLA-34 is used in both algorithms to extract detection and re-ID features from the acquired images, which is built on top of CenterNet [38]. The network of the extended version of FairMOT was modified to allow for classification since the original FairMOT did not consider multiclass tracking. The proposed algorithm takes the center points, size of the bounding box, and appearance feature vectors of the objects from the DLA-34 as input. The AFJPDA filter associates these measurements with tracks for tracking. The flowchart of the proposed algorithm is described in Fig. 2.

The proposed AFJPDA filter estimates the 2D image coordinates of objects by including their center position and size of the bounding box, and their velocity. Therefore, the state of the ith track at time step k in Eq. (3) can be defined as $\boldsymbol{x}^i_k = [x^i_k, v^i_{x,k}, y^i_k, v^i_{y,k}, w^i_k, v^i_{w,k}, h^i_k, v^i_{h,k}]^T$. This state includes the object's center position in the x and y directions, the width and height of the bounding box, as well as their velocity in those directions. Given that the objects in the scenario are moving with nearly constant velocity, a constant-velocity model is assumed for the transition matrix. The transition matrix and process noise covariance matrix in Eqs. (3) and (4) can be written as follows:

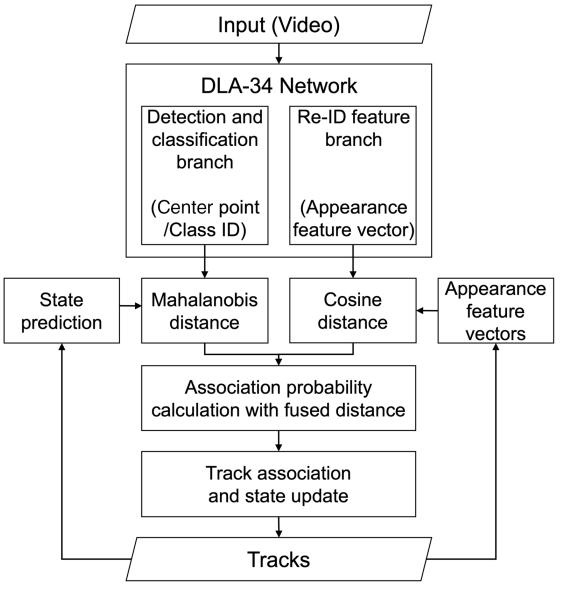


Fig. 2 Flowchart of the proposed algorithm.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

$$\boldsymbol{F} = \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

$$\boldsymbol{Q} = \sigma_p \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\Delta t^3}{2} & \Delta t^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\Delta t^3}{2} & \Delta t^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\Delta t^3}{2} & \Delta t^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\Delta t^3}{2} & \Delta t^2 \end{bmatrix} \tag{21}$$

where Δt is the time step and σ_p is the standard deviation of the process noise.

The AFJPDA filter takes the center, width, and height of the bounding box as a measurement for the update step in the AFJPDA filter, and the jth measurement at time step k in Eq. (8) can be defined as $\boldsymbol{z}_k^j = [x_{k,\text{meas}}^j, y_{k,\text{meas}}^j, w_{k,\text{meas}}^j, h_{k,\text{meas}}^j]^T$. The measurement matrix and measurement noise covariance matrix in Eqs. (8) and (10) can be written as follows:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{R} = \sigma_m \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

where σ_m is the standard deviation of the measurement noise.

B. Performance Evaluation Metrics

The performance of the proposed algorithm is evaluated with the following evaluation metrics and compared with the multiclass extended version of the FairMOT and the JPDA filter. Metrics can be defined as follows [39–41]:

1) IDF1(↑): Identity-F1 score, defined in Eq. (23)
2) MOTA(↑): Multi-object tracking accuracy, defined in Eq. (24)
3) MT(↑): Mostly tracked objects, more than 80% of life span tracked
4) FN(↓): False negative, number of missed detections
5) IDSW(↓): Identity switch, number of ID switches to another ID that is already allocated to a different track
6) Frag(↓): Fragmentation, number of ID changes due to missed detection
7) FPS(↑): Frame per second, number of frames processed in 1 s

$$\text{IDF1} = \frac{2 \cdot \text{IDTP}}{2 \cdot \text{IDTP} + \text{IDFP} + \text{IDFN}} \tag{23}$$

$$\text{MOTA} = 1 - \frac{\sum_t (\text{FN}_t + \text{FP}_t + \text{IDSW}_t)}{\sum_t \text{GT}_t} \tag{24}$$

where IDTP, IDFP, and IDFN are the number of true positive IDs, false-positive IDs, and false-negative IDs, respectively [39,40]; t is the frame of the image sequence; and GT is the number of ground truth objects [40]. In the above metrics, (↑) indicates a higher score showing better performance and (↓) indicates a lower score showing better performance. The metrics IDF1, MOTA, and MT were selected for this study as they are commonly used in evaluating MOT performance. They offer a comprehensive view of the tracker's overall performance, giving readers valuable insights into its efficacy. Furthermore, we also considered FN, IDSW, and fragmentation metrics because they are especially sensitive to variations in performance in the context of MCMOT within cluttered environments. Note that we are focusing on IDF1 more than MOTA since the proposed AFJPDA algorithm uses the same detector with FairMOT, and the MOTA score is highly dependent on detector performance. In other words, MOT algorithms with a good detector can achieve a high MOTA score even if it has a big number of IDSW or fragmentation [42,43].

C. Simulation Scenarios

The simulations were designed to evaluate the MCMOT performance. We assumed a scenario in which two classes of moving ground objects (cars and buses/trucks) were tracked by a hovering UAV with a single image sensor. To generate the synthetic image, we used the Carla simulator [32], where the ground vehicles in the simulator were controlled using autopilot mode. The details of the scenarios are described in Table 1, and sample pictures from the scenarios are shown in Fig. 3.

In addition, AFJPDA is tested in real-world dataset and compared with the extended version of the FairMOT. For the aerial images, some scenarios from VisDrone [33] were tested. The real-world scenarios have more number of objects and clutter in the images. The details of the selected and tested scenarios are described in Table 2, and sample pictures from the scenarios are shown in Fig. 4.

Table 1 Details of simulation scenarios: Carla simulator

Scenario	FPS	Resolution	Length	No. of objects
Carla-1	30 FPS	1920 × 1080 pixels	600 frames (00′20″)	47
Carla-2			1200 frames (00′40″)	29
Carla-3			1800 frames (01′00″)	72
Total			3600 frames (02′00″)	148

Fig. 3 Sample pictures from three scenarios of Carla simulator.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

Table 2 Details of simulation scenarios: VisDrone dataset

Scenario	FPS	Resolution	Length	No. of objects
VisDrone-1	30 FPS	1920 × 1080 pixels	677 frames (00′22″)	100
VisDrone-2			144 frames (00′5″)	53
VisDrone-3			426 frames (00′14″)	87
VisDrone-4			548 frames (00′18″)	22
VisDrone-5			341 frames (00′11″)	118
Total			2137 frames (01′14″)	380

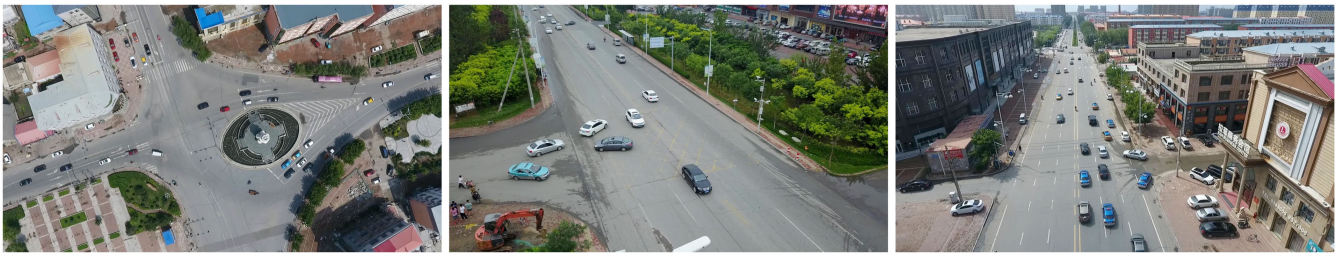


Fig. 4 Sample pictures from three scenarios of VisDrone dataset [33].

D. Results and Discussion

The performance was evaluated using the scenarios in Tables 1 and 2, as well as the metrics described in Sec. IV. The evaluated performance of the multiclass extended version of the FairMOT (referred to as "Base" in the tables), JPDA filter, and AFJPDA filter is presented in Tables 3 and 4. Note that we chose the extended version of the FairMOT algorithm as the baseline because most existing MCMOT algorithms with vision sensors rely on a track-by-detection scheme that employs CNN detectors, e.g., YOLO, and GNN algorithms such as the Hungarian algorithm. The extended

Table 3 Tracking result of the Carla simulation scenarios

Scenario	Algorithm	IDF1 ↑	MOTA ↑	MT ↑	FN ↓	IDSW ↓	Frag ↓	FPS ↑
Carla-1	a) Base	80.62%	80.55%	78.72%	2425	121	172	**18.58**
	b) JPDA	82.63%	**82.13%**	**80.85%**	2110	20	104	**18.58**
	c) AFJPDA	**83.42%**	82.12%	**80.85%**	**2107**	**19**	**100**	17.67
Carla-2	a) Base	**77.63%**	76.01%	86.21%	1068	100	127	20.22
	b) JPDA	72.57%	75.91%	86.21%	950	36	77	**20.23**
	c) AFJPDA	74.34%	**76.52%**	86.21%	**888**	**31**	**74**	19.77
Carla-3	a) Base	79.87%	74.35%	56.94%	6207	153	329	**19.51**
	b) JPDA	80.28%	77.47%	62.50%	5755	69	192	18.93
	c) AFJPDA	**81.05%**	**77.84%**	**66.67%**	**5668**	**50**	**190**	18.60
Total	a) Base	79.68%	76.37%	69.60%	9700	374	628	**19.58**
	b) JPDA	79.56%	78.51%	72.97%	8815	125	373	19.28
	c) AFJPDA	**80.52%**	**78.81%**	**75.00%**	**8663**	**100**	**364**	18.81

The best score is shown as boldfaced value.

Table 4 Tracking result of the VisDrone real-world scenarios

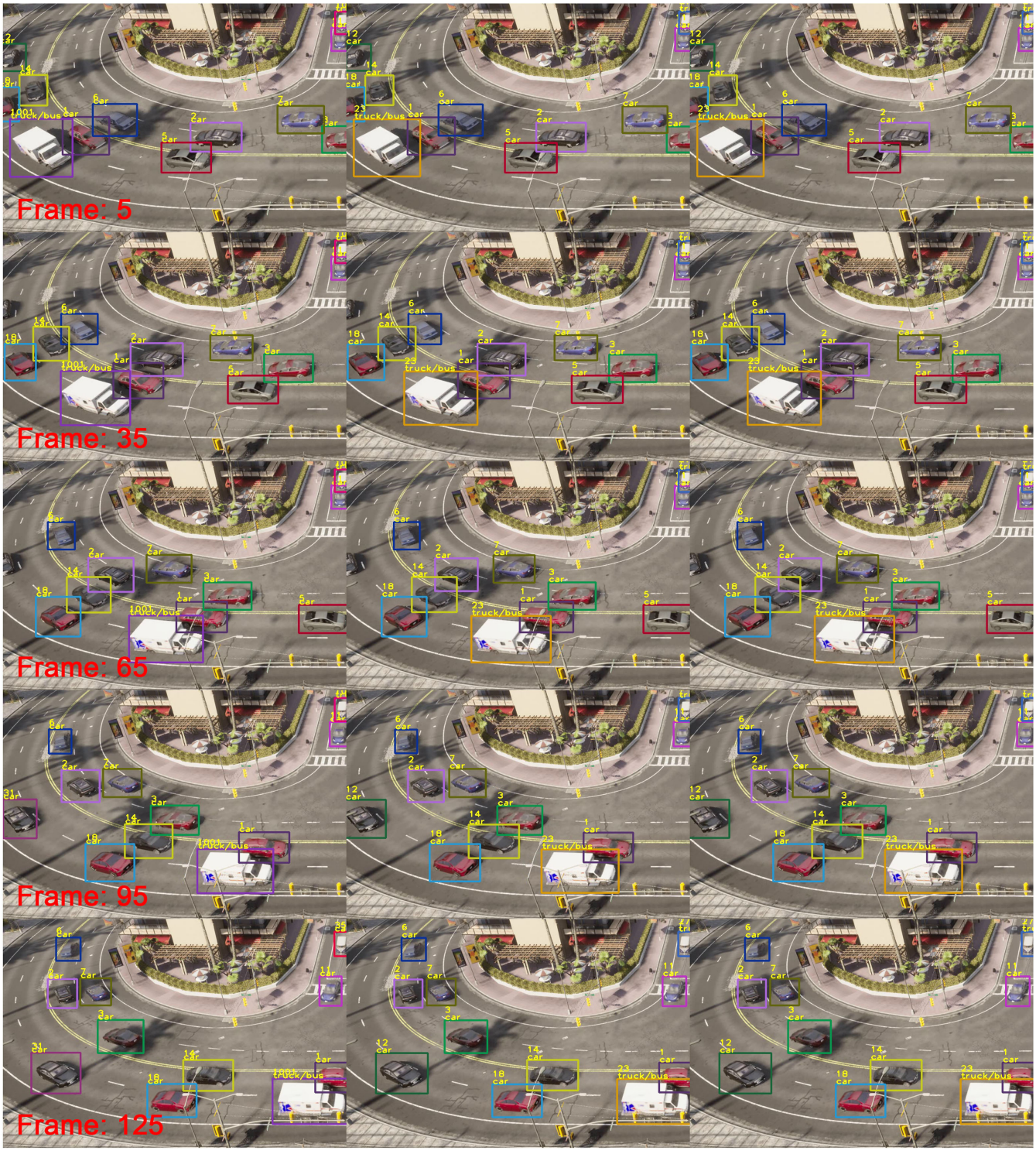
Scenario	Algorithm	IDF1 ↑	MOTA ↑	MT ↑	FN ↓	IDSW ↓	Frag ↓	FPS ↑
VisDrone-1	a) Base	86.16%	85.83%	80.00%	2628	347	529	**16.76**
	b) JPDA	**92.35%**	88.74%	85.00%	2113	30	160	16.00
	c) AFJPDA	92.18%	**89.08%**	**86.00%**	**1994**	**21**	**159**	15.30
VisDrone-2	a) Base	66.03%	**36.41%**	58.49%	1746	22	91	17.12
	b) JPDA	**67.46%**	34.40%	**60.38%**	1563	10	44	**17.25**
	c) AFJPDA	66.80%	34.23%	**60.38%**	**1559**	**6**	**41**	16.37
VisDrone-3	a) Base	80.97%	**67.93%**	77.01%	3508	133	382	**16.39**
	b) JPDA	78.47%	63.68%	67.82%	4032	58	166	11.48
	c) AFJPDA	**81.29%**	66.80%	**79.31%**	**3449**	**25**	**158**	13.43
VisDrone-4	a) Base	62.83%	44.63%	27.27%	2076	12	46	19.70
	b) JPDA	70.07%	53.16%	**31.58%**	1219	**5**	**22**	**21.03**
	c) AFJPDA	**71.74%**	**55.64%**	**31.58%**	**1154**	6	23	19.96
VisDrone-5	a) Base	**80.75%**	**62.56%**	**66.10%**	1887	72	211	**16.57**
	b) JPDA	78.51%	57.62%	62.71%	1846	**30**	**83**	14.93
	c) AFJPDA	78.27%	57.00%	64.41%	**1804**	31	91	14.69
Total	a) Base	82.26%	71.20%	68.95%	11,845	586	1269	**17.35**
	b) JPDA	83.26%	70.99%	67.91%	10,773	133	475	15.64
	c) AFJPDA	**83.98%**	**72.05%**	**71.35%**	**9960**	**89**	**472**	15.78

The best score is shown as boldfaced value.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

version of FairMOT is one of the most state-of-the-art MOT algorithms employing those concepts, and it is the most up-to-date algorithm that is directly applied to MCMOT. Furthermore, as the majority of generic MOT algorithms are not suitable for multiclass MOT, the proposed algorithm is exclusively compared with the multiclass extended version of FairMOT.

The result was evaluated using an identically trained dataset for the same detection and classification algorithm on the PC with Intel i7-10700 CPU and NVIDIA RTX A5000 GPU. This shows that the proposed algorithm, JPDA filter, and AFJPDA filter for multiclass showed better performance than the extended version of FairMOT in most of the metrics, including IDF1, MOTA, and MT and significant improvement in reducing FN, IDSW, and fragmentation. This improvement was able to be achieved because of the characteristic of the JPDA filter that can handle clutters using association probabilities of multiple measurements in the gating area, while the extended version of FairMOT uses only one nearest measurement (hard association). Furthermore, by combining appearance features in the association probability calculation in the JPDA filter, we could achieve further improvement in reducing FN, IDSW, and fragmentation for most of the scenarios. In terms of computational cost, the proposed AFJPDA filter showed average speeds of 18.81 FPS and



a) Extended FairMOT b) JPDA c) AFJPDA

Fig. 5 Tracking results of a) extended FairMOT, b) JPDA filter, and c) AFJPDA filter in the first scenario.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

15.78 FPS, while the base algorithm showed average speeds of 19.85 FPS and 17.35 FPS in the simulation environment and real-world environment, respectively. This means that the proposed AFJPDA filter exhibits comparable computational efficiency with better tracking performance compared with the benchmark algorithm. The main reason behind the slight degradation in computational efficiency in the proposed algorithm compared with the benchmark algorithm is a soft association, i.e., an association of multiple candidates to a track, which results in additional calculations such as calculation of weighted innovation, y_k^i in Eq. (9). Tracking results in the simulation environment are described in Figs. 5–7.

Most objects are well tracked in the sequence of frames as shown in Figs. 5–7. However, a fragmentation example can be observed in a) extended FairMOT (first column). The object with track ID 12 (green box on the left top) was lost from frame 35 to 65 in the extended FairMOT. After the object was detected again in frame 95, the track ID fragmented into 31 in the extended FairMOT, while the JPDA filter and the AFJPDA filter tracked the object with the same track ID. Another case of IDSW and fragmentation is shown in Fig. 6. While the c) AFJPDA filter showed neither fragmentation nor IDSW, the a) extended FairMOT showed both, and the b) JPDA filter showed FN. When a new object enters from the bottom of the image

a) Extended FairMOT b) JPDA c) AFJPDA

Fig. 6 Tracking results of a) extended FairMOT, b) JPDA filter, and c) AFJPDA filter in the second scenario.

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

a) Extended FairMOT b) JPDA c) AFJPDA

Fig. 7 Tracking results of a) extended FairMOT, b) JPDA filter, and c) AFJPDA filter in the third scenario.

in frame 110, it was initially assigned as track ID 18, but this track fragmented into track ID 19 in frame 114 and ID 1012 in frame 118. In frame 126, a new object entered the image, but this object was assigned as track ID 19, which had already been assigned to another track (IDSW). Although the b) JPDA filter did not show any fragmentation or IDSW, it showed FN in frames 110, 122, and 126. An example of IDSW and fragmentation after occlusion is shown in Fig. 7. In this example, objects assigned as track IDs 109 and 1041 in frame 1020 in the a) extended FairMOT fragmented to track IDs 119 and 1046 in frames 1040 and 1100, respectively, after occlusion. Also, the car waiting on the left bottom side, which had been assigned to track ID 83 in frame 1040, fragmented to track ID 1043 in frame 1060. In the case of the b) JPDA filter, the bus with track ID 101 (purple box on the right) fragmented to track ID 105 in frame 1100 after occlusion. However, in the case of the c) AFJPDA filter, no fragmentation occurred for either the waiting car or the bus passing behind the tree. The car with track ID 77 continued to be tracked with track ID 77, as shown in frames 1020 and 1040, and the bus with track ID 81 in frame 1020 was tracked with the same ID in frame 1100 after occlusion by the tree.

The proposed AFJPDA algorithm outperformed the extended version of the FairMOT algorithm, with significant improvements

in reducing FN, IDSW, and fragmentation and slight improvements in IDF1, MOTA, and MT. This supports our main motivation behind the proposition of the AFJPDA filter and confirms its validity. It was achieved by using the soft association in the JPDA filter and by further incorporating appearance features into the JPDA filter. The result showed that the proposed AFJPDA filter with the adaptive gating logic and the track maintenance logic is beneficial in the case of the multiclass multi-object tracking scenario in a cluttered environment and can reduce the FN, IDSW, and fragmentation significantly without much degradation of computation.

E. Sensitivity Analysis

The sensitivity analysis on parameter λ for weighting between the physical distance (Mahalanobis distance) and appearance similarity (cosine distance) in Eq. (18) is discussed here. A total of six λ values are evaluated: $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$, and 0.98 for the Carla-2 scenario. The first case, $\lambda = 0.0$, is the case where the appearance feature is not used in distance calculation, same as the JPDA filter in Table 3, and $\lambda = 0.6$ is the value we used in AFJPDA filter in Table 3. The last case, $\lambda = 0.98$, is the same value with a base algorithm, FairMOT, and multiclass extended version of FairMOT in [18,31]. Note that $\lambda = 1.0$ is excluded for better readability of the graph since it has a significantly huge error compared to other λ cases. The key performance metrics IDF1, MOTA, IDSW, and fragmentation are compared according to different λ values as shown in Fig. 8.

The four key performance metrics, IDF1, MOTA, IDSW, and fragmentation, are shown as blue line with circle marker, yellow line with square marker, orange line with diamond marker, and gray line with triangle marker, respectively. For all four metrics, the algorithms showed the best performance with $\lambda = 0.6$. From $\lambda = 0.2$ to $\lambda = 0.6$ the performance improved as the λ increases, but after $\lambda = 0.6$, the performance degraded as λ decreases. It can be observed that the effect of the weighting parameter λ on MOTA and fragmentation is comparatively less than that on IDF1 and IDSW. This phenomenon is owing to the nature and characteristics of those metrics. MOTA, derived from FN, FP, and IDSW as described in Eq. (24), is predominantly influenced by the detector's performance. Also, the numerical scale of IDSW is smaller than FN or FP in Eq. (24), making its impact on MOTA less pronounced. Consequently, even with improvements in IDSW or minor improvements in FN and FP, the MOTA remains largely unaltered by the choice of weighting parameter. Fragmentation, which arises from missed detections or clutters, is similarly affected primarily by the detector. While its performance can be enhanced by adjusting the filtering approach, without detector improvements, further enhancement in this metrics with the choice of weighting parameter remains restricted as the missed detection is unable to be solved with filtering. On the other hand, metrics like IDSW and IDF1, which emphasize identity assignment consistency and precision over time, are closely tied to the tracker's filtering mechanism and the association's performance. As such, changes in the weighting parameter can notably alter these metrics' outcomes.

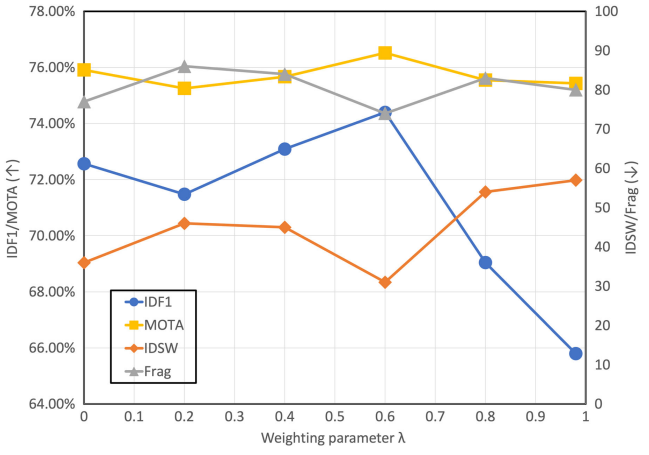


Fig. 8 Sensitivity analysis on weighting parameter λ.

V. Conclusions

In this study, we proposed an appearance feature-aided JPDA (AFJPDA) filter for multiclass multi-object tracking in scenarios with clutter. The AFJPDA filter utilizes the appearance feature from the classification algorithm in the association probability calculation of the JPDA filter. Additionally, we introduced a simple adaptive gating logic and modified track logic to improve the association performance. The proposed algorithm was tested in both simulation and real-world aerial images and demonstrated improved performance compared to the extended version of FairMOT for multiclass tracking. This performance improvement is shown in most metrics, including IDF1, MOTA, and MT and notably significant improvement in reducing FN, ID switch, and fragmentation. This improvement was achieved because of the characteristic of the JPDA filter, which uses multiple candidates for association, unlike hard association, and the augmentation of the appearance feature in the algorithm. In future work, we plan to investigate the potential benefits of incorporating additional techniques into our filter, such as adapting the noise covariance, in order to further improve its performance.

Some of the open issues in multiclass multi-object tracking are still remaining, however, such as ambiguity problems and tracking highly nonlinear objects. Our future work would extend the proposed algorithm to associate with multiple heterogeneous sensors such as radar or light detection and ranging (LiDAR) sensors to show further improvement. In addition to the strength of the vision sensor (appearance feature), depth or dimension information from radar or LiDAR could provide additional information about the object for the detection and distinction of ambiguous objects. This will improve the general performance and overcome the ambiguity issue in MOT. Furthermore, an extension of the proposed work in consideration of targets with highly nonlinear dynamics is subject to future work.

Acknowledgment

This research was supported by the UK Research and Innovation-funded project HADO (project number 10024815).

References

[1] Bar-Shalom, Y., *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, MA, 1990.

[2] Bar-Shalom, Y., and Li, X.-R., *Multitarget-Multisensor Tracking: Principles and Techniques*, Vol. 19, YBs, Storrs, CT, 1995.

[3] Fortmann, T., Bar-Shalom, Y., and Scheffe, M., "Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association," *IEEE Journal of Oceanic Engineering*, Vol. 8, No. 3, 1983, pp. 173–184. https://doi.org/10.1109/JOE.1983.1145560

[4] Rezatofighi, S. H., Milan, A., Zhang, Z., Shi, Q., Dick, A., and Reid, I., "Joint Probabilistic Data Association Revisited," *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE Computer Soc., Washington, D.C., 2015, pp. 3047–3055. https://doi.org/10.1109/iccv.2015.349

[5] Blackman, S., "Multiple Hypothesis Tracking for Multiple Target Tracking," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 19, No. 1, 2004, pp. 5–18. https://doi.org/10.1109/MAES.2004.1263228

[6] Wang, X., Li, T., Sun, S., and Corchado, J. M., "A Survey of Recent Advances in Particle Filters and Remaining Challenges for Multitarget Tracking," *Sensors*, Vol. 17, No. 12, 2017, Paper 2707. https://doi.org/10.3390/s17122707

[7] Smith, D., and Singh, S., "Approaches to Multisensor Data Fusion in Target Tracking: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 12, 2006, pp. 1696–1710. https://doi.org/10.1109/TKDE.2006.183

[8] Vaidehi, V., Chitra, N., Chokkalingam, M., and Krishnan, C., "Neural Network Aided Kalman Filtering for Multitarget Tracking Applications," *Computers & Electrical Engineering*, Vol. 27, No. 2, 2001, pp. 217–228. https://doi.org/10.1016/S0045-7906(00)00013-6

[9] Leal-Taixé, L., Milan, A., Reid, I., Roth, S., and Schindler, K., "MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking," 2015. https://doi.org/10.48550/ARXIV.1504.01942

[10] Milan, A., Leal-Taixe, L., Reid, I., Roth, S., and Schindler, K., "MOT16: A Benchmark for Multi-Object Tracking," 2016. https://doi.org/10.48550/ARXIV.1603.00831

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

[11] Voigtlaender, P., Krause, M., Osep, A., Luiten, J., Sekar, B. B. G., Geiger, A., and Leibe, B., "Mots: Multi-Object Tracking and Segmentation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Inst. of Electrical and Electronics Engineers, New York, 2019, pp. 7942–7951.

[12] Wang, Z., Zheng, L., Liu, Y., Li, Y., and Wang, S., "Towards Real-Time Multi-Object Tracking," *Computer Vision—ECCV 2020*, edited by A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Springer–Verlag, Cham, Switzerland, 2020, pp. 107–122.

[13] Ciaparrone, G., Luque Sánchez, F., Tabik, S., Troiano, L., Tagliaferri, R., and Herrera, F., "Deep Learning in Video Multi-Object Tracking: A Survey," *Neurocomputing*, Vol. 381, March 2020, pp. 61–88, https://www.sciencedirect.com/science/article/pii/S0925231219315966. https://doi.org/10.1016/j.neucom.2019.11.023

[14] Emami, P., Pardalos, P. M., Elefteriadou, L., and Ranka, S., "Machine Learning Methods for Data Association in Multi-Object Tracking," *ACM Computing Surveys (CSUR)*, Vol. 53, No. 4, 2020, pp. 1–34. https://doi.org/10.1145/3394659

[15] Nguyen, H., and Smeulders, A., "Fast Occluded Object Tracking by a Robust Appearance Filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 8, 2004, pp. 1099–1104. https://doi.org/10.1109/TPAMI.2004.45

[16] Wu, S., Tan, Y., Das, S., Broaddus, C., and Chiu, M.-Y., "Multiple-Target Tracking via Kinematics, Shape, and Appearance-Based Data Association," *Signal and Data Processing of Small Targets 2009*, edited by O. E. Drummond, and R. D. Teichgraeber, Vol. 7445, International Soc. for Optics and Photonics, Bellingham, Washington, D.C., 2009. https://doi.org/10.1117/12.829656

[17] Al-Shakarji, N. M., Seetharaman, G., Bunyak, F., and Palaniappan, K., "Robust Multi-Object Tracking with Semantic Color Correlation," *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Inst. of Electrical and Electronics Engineers, New York, 2017, pp. 1–7. https://doi.org/10.1109/AVSS.2017.8078507

[18] Zhang, Y., Wang, C., Wang, X., Zeng, W., and Liu, W., "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking," *International Journal of Computer Vision*, Vol. 129, Nov. 2021, pp. 3069–3087. https://doi.org/10.1007/s11263-021-01513-4

[19] Mahmoudi, N., Ahadi, S. M., and Rahmati, M., "Multi-Target Tracking Using CNN-Based Features: CNNMTT," *Multimedia Tools and Applications*, Vol. 78, No. 6, 2019, pp. 7077–7096. https://doi.org/10.1007/s11042-018-6467-6

[20] Micheal, A. A., and Vani, K., "Deep Learning-Based Multi-Class Multiple Object Tracking in UAV Video," *Journal of the Indian Society of Remote Sensing*, Vol. 50, No. 12, 2022, pp. 2543–2552. https://doi.org/10.1007/s12524-022-01615-7

[21] Bae, S.-H., and Yoon, K.-J., "Robust Online Multi-Object Tracking Based on Tracklet Confidence and Online Discriminative Appearance Learning," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Inst. of Electrical and Electronics Engineers, New York, 2014, pp. 1218–1225. https://doi.org/10.1109/CVPR.2014.159

[22] Wojke, N., Bewley, A., and Paulus, D., "Simple Online and Realtime Tracking with a Deep Association Metric," *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962

[23] Cao, J., Pang, J., Weng, X., Khirodkar, R., and Kitani, K., "Observation-Centric Sort: Rethinking Sort for Robust Multi-Object Tracking," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Inst. of Electrical and Electronics Engineers, New York, 2023, pp. 9686–9696.

[24] Bose, B., Wang, X., and Grimson, E., "Multi-Class Object Tracking Algorithm that Handles Fragmentation and Grouping," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Inst. of Electrical and Electronics Engineers, New York, 2007, pp. 1–8. https://doi.org/10.1109/CVPR.2007.383175

[25] Spinello, L., Triebel, R., and Siegwart, R., "Multiclass Multimodal Detection and Tracking in Urban Environments," *International Journal of Robotics Research*, Vol. 29, No. 12, 2010, pp. 1498–1515. https://doi.org/10.1177/0278364910377533

[26] Bourgeois, F., and Lassalle, J.-C., "An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices," *Communications of the ACM*, Vol. 14, No. 12, 1971, pp. 802–804. https://doi.org/10.1145/362919.362945

[27] Zhang, Q., Sun, Y., Yang, J., and Liu, H., "Real–Time Multi–Class Moving Target Tracking and Recognition," *IET Intelligent Transport Systems*, Vol. 10, No. 5, 2016, pp. 308–317. https://doi.org/10.1049/iet-its.2014.0226

[28] Lee, B., Erdenee, E., Jin, S., Nam, M. Y., Jung, Y. G., and Rhee, P. K., "Multi-Class Multi-Object Tracking Using Changing Point Detection," *Computer Vision—ECCV 2016 Workshops*, edited by G. Hua, and H. Jégou, Springer–Verlag, Cham, Switzerland, 2016, pp. 68–83.

[29] Jo, K., Im, J., Kim, J., and Kim, D.-S., "A Real-Time Multi-Class Multi-Object Tracker Using YOLOv2," *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Inst. of Electrical and Electronics Engineers, New York, 2017, pp. 507–511. https://doi.org/10.1109/ICSIPA.2017.8120665

[30] Kuhn, H. W., "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, Vol. 2, Nos. 1–2, 1955, pp. 83–97. https://doi.org/10.1002/nav.3800020109

[31] "MCMOT," 2020, https://github.com/CaptainEven/MCMOT.

[32] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V., "CARLA: An Open Urban Driving Simulator," *Conference on Robot Learning*, PMLR, Cambridge, MA, 2017. https://doi.org/10.48550/ARXIV.1711.03938

[33] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., and Ling, H., "Detection and Tracking Meet Drones Challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 11, 2021, pp. 7380–7399. https://doi.org/10.1109/TPAMI.2021.3119563

[34] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, No. 1, 1960, pp. 35–45. https://doi.org/10.1115/1.3662552

[35] Blackman, S., and Popoli, R., *Design and Analysis of Modern Tracking Systems (Book)*, Artech House, Norwood, MA, 1999.

[36] Xu, S., Savvaris, A., He, S., Shin, H.-S., and Tsourdos, A., "Real-Time Implementation of YOLO+JPDA for Small Scale UAV Multiple Object Tracking," *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, Inst. of Electrical and Electronics Engineers, New York, 2018, pp. 1336–1341. https://doi.org/10.1109/icuas.2018.8453398

[37] He, S., Shin, H.-S., and Tsourdos, A., "Information-Theoretic Joint Probabilistic Data Association Filter," *IEEE Transactions on Automatic Control*, Vol. 66, No. 3, 2021, pp. 1262–1269. https://doi.org/10.1109/TAC.2020.2989766

[38] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q., "CenterNet: Keypoint Triplets for Object Detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6568–6577. https://doi.org/10.1109/ICCV.2019.00667

[39] Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C., "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking," *European Conference on Computer Vision*, Springer–Verlag, Cham, Switzerland, 2016, pp. 17–35.

[40] Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., and Leal-Taixé, L., "MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking," *International Journal of Computer Vision*, Vol. 129, No. 4, 2021, pp. 845–881. https://doi.org/10.1007/s11263-020-01393-0

[41] Bernardin, K., and Stiefelhagen, R., "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP Journal on Image and Video Processing*, Vol. 2008, No. 1, 2008, p. 1. https://doi.org/10.1155/2008/246309

[42] Karthik, S., Prabhu, A., and Gandhi, V., "Simple Unsupervised Multi-Object Tracking," arXiv, 2020. https://doi.org/10.48550/arxiv.2006.02609

[43] Zhou, X., Koltun, V., and Krähenbühl, P., "Tracking Objects as Points," *European Conference on Computer Vision*, Springer–Verlag, Cham, Switzerland, 2020, pp. 474–490.

M. J. Kochenderfer
Associate Editor

Downloaded by Cranfield University on June 12, 2024 | http://arc.aiaa.org | DOI: 10.2514/1.I011301

https://dspace.org/

AFJPDA: a multiclass multi-object tracking with appearance feature-aided joint probabilistic data association

Kim, Sukkeun; Petrunin, Ivan; Shin, Hyo-Sang

2024-04-01
American Institute of Aeronautics and Astronautics (AIAA)
https://doi.org/10.2514/1.I011301
Kim S, Petrunin I, Shin HS. (2024) AFJPDA: a multiclass multi-object tracking with appearance feature-aided joint probabilistic data association. Journal of Aerospace Information Systems, Volume 21, Issue 4, April 2024, pp. 294-304

Downloaded from a DSpace repository. https://dspace.org/