Article

Long Distance Moving Vehicle Tracking with a Multirotor Based on IMM-Directional Track Association

Seokwon Yeom

School of ICT Eng., Daegu University, Gyeongsan 38453, Korea; yeom@daegu.ac.kr

Abstract: The multirotor has the capability to capture distant objects. Because the computing resources of the multirotor are limited, efficiency is an important factor to consider. In this paper, multiple target tracking with a multirotor at a long distance (~400 m) is addressed; the interacting multiple model (IMM) estimator combined with the directional track-to-track association (abbreviated as track association) is proposed. The previous work of the Kalman estimator with the track association approach is extended to the IMM estimator with the directional track association. The IMM estimator can handle multiple targets with various maneuvers. The track association scheme is modified in consideration of the direction of the target movement. The overall system is composed of moving object detection for measurement generation and multiple target tracking for state estimation. The moving object detection consists of frame-to-frame subtraction of three-color layers and thresholding, morphological operation, and false alarm removing based on the object size and shape properties. The centroid of the detected object is input into the next tracking stage. The track is initialized using the difference between two nearest points measured in consecutive frames. The measurement nearest to the state prediction is used to update the state of the target for measurement-to-track association. The directional track association tests both the hypothesis and the maximum deviation between the displacement and directions of two tracks followed by track selection, fusion, and termination. In the experiment, a multirotor flying at an altitude of 400 m captured 55 moving vehicles around a highway interchange for about 20 s. The tracking performance is evaluated for the IMMs using constant velocity (CV) and constant acceleration (CA) motion models. The IMM-CA with the directional track association scheme outperforms other methods with an average total track life of 91.7% and an average mean track life of 84.2%.

Keywords: drone surveillance; moving object detection; multiple target tracking; directional track association; state estimation

1. Introduction

Small unmanned aerial vehicles (UAVs) or drones are very useful for many applications such as security and surveillance, search and rescue mission, traffic monitoring, and asset and environmental inspection [1,2]. A multirotor drone can hover or fly while capturing video from a distance [3]. This capture is cost effective and does not require highly trained personnel.

Studies on target tracking with a small drone can be categorized into visual, non-visual, or combined methods. Various deep learning-based trackers were compared with a camera motion model [4]. Small objects, large numbers of targets, and camera motion degraded tracking performance even using high-end CPUs [5]. A multi-object tracking and 3D localization scheme was proposed based on the deep learning-based object detection [6]. Usually, the deep learning-based object detector and tracker require heavy computation with massive training data [7,8], thus real-time processing can often be an issue to solve. In [9], the moving objects are detected by background subtraction, and the continuously adaptive mean-shift tracking and the optical flow tracking were applied to the video sequences acquired by the UAV. The particle filtering combined with the mean-shift method

was developed to track a small and fast-moving object [10]. A fast target tracking method was developed based on the SIFT feature with multiple hypothesis tracking [11]. The visual trackers need to transmit high resolution video streams to the ground or impose high computing on the drone. The initial location, target size, and number of targets are often assumed to be known to the visual tracker. Indeed, the visual trackers rarely evaluate the kinematic state of the target.

The kinematic state of the target can be obtained with non-imaging sensing data such as radar signals [12]. In [13], aerial video processing has been combined with the high-precision GPS data from vehicles. Bayesian fusion of vision and radio frequency sensors were studied for ground target tracking [14]. However, in the non-visual approach, the drone should be equipped with high-cost sensors that add more payload to the drone, or infrastructure is required on the ground or in the vehicle.

Multiple targets are tracked by continuously estimating their kinematic state such as position, velocity, and acceleration. The Kalman filter is known to be optimal under independent Gaussian noise assumption in estimating the target state [15]. The inter-acting multiple model (IMM) estimator developed in [16] consists of the several mode-matched Kalman filters to handle different maneuvers of multiple targets. The track initial-ization corresponding to measurement-to-measurement association starts the track [17]. Measurement-to-track association combines measurements with a track to update the state of a target. Track-to-track association (abbreviated as track association) combines tracks to maintain a unique track for a target in a frame [18].

In the paper, an IMM-directional track association scheme is proposed for tracking multiple targets captured by a multirotor. The strategy developed in [19–21] extracts measurements during the object detection, and then establishes the tracks based on those measurements. This method does not directly use the intensity information for target tracking, thus there is no need to store or transmit high-resolution video streams. In [21], the track association based on the Kalam filter was developed; however, the IMM filter can efficiently handle the various maneuvers of a large number of targets. The track association only depends on the statistical distance of the two estimates of different tracks, but the spatial direction of the tracks can be considered to reduce false track association.

In order to detect moving objects, a subtraction is performed between the current frame and the previous frame in all three-color layers, and thresholding follows to generate a binary image. Then, the morphological operation, closing (dilation and erosion) is applied to the binary image. The dilation operation connects the fragmented regions of one object while the erosion operation shrinks the dilated boundary. Finally, false detections are eliminated using the actual object size and two shape properties: squareness and rectangularity [22]. The centroid of the extracted area becomes the measured position for the next tracking stage.

Tracking is performed by the IMM filter to estimate the kinematic states of the targets. Two discrete time kinematic models are used and compared. One is a (nearly) constant velocity (CV) motion model, and the other is a (nearly) constant acceleration (CA) motion model [23]. The CV model is equivalent with the white acceleration model since the accelerations are modeled as white noise. The Wiener-sequence acceleration model is adopted for the CA model assuming that the acceleration increment is a white noise process. The discretization of the continuous model is performed for the discrete model [24], but the Wiener sequence acceleration model is developed in discrete time directly [25]. With high quality measurements, the CA model can produce the accurate estimate of the acceleration [26]. The IMM with a combined CV and CA scheme was contrieved to track a single target [23]. The IMM-CV scheme was applied to track various maneuvering 120 aerial targets for the aerial early warning system [27].

A track is initialized by the two-point differencing initialization following the maxi-mum speed gating. For the measurement-to-measurement and track association tasks, the gating process is first performed based on the chi-square hypothesis testing. Then, the mea-surement (or track) selection and update (fusion) are followed. The nearest neighbor (NN)

measurement–to–track association is the most effective in computing and has been successfully applied to target tracking based on multiple frames captured by a drone [19–22]. The track fusion method in multi-sensor environments was developed assuming a common process noise of the target [18]. In the previous work [21], a practical approach for the track association was proposed; after the hypothesis testing, only the track of the smallest determinant of the covariance matrix is maintained and the other is terminated. The current update of the selected track is replaced by the fusion estimate. In the paper, the directional track association is proposed to be equipped with the directional gating process. The directional gating tests the maximum deviation in the directions of the tracks and the direction of the displacement vector between the tracks. In other words, two tracks can be fused into one track only when the direction of the motion is close to the direction of the displacement.

Several criteria exist for track termination. One is the maximum number of consecutive updates without measurements, and the other is the minimum target speed. The criterion of the minimum target speed is very effective as shown in [22] when high clutter occurs on false targets that are not in motion. Finally, the track is confirmed as a valid track if it continues longer than the minimum track life length. Figure 1 shows a block diagram of detecting and tracking vehicles by a multirotor at a long distance.
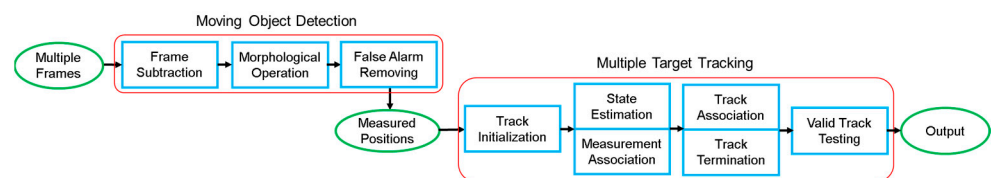

Figure 1. Block diagram of moving object detection and multiple target tracking.

In the experiments, a drone captured a video while flying at a height of 400 m. The drone camera pointed directly downwards. A total of 55 moving vehicles appeared or disappeared for about 20 s. Figure 2 shows sample objects of the video. The object resolution is very low and sometimes they are occluded or move close to each other. The background can include the road lane and mark.

Figure 2. Sample objects in the video captured by a multirotor.

Four methods are compared: IMM-CV with track association and directional track association, and IMM-CA with track association and directional track association. The IMM-CA with the directional track association outperforms other methods with an average total track life (TTL) of 91.7% and an average mean track life (MTL) of 84.2%.

The remainder of the paper is organized as follows: moving object detection is discussed in Section 2. Section 3 demonstrates multiple target tracking. Section 4 presents experimental results, and the conclusion follows in Section 5.

2. Moving Object Detection

The moving object detection consists of frame-to-frame subtraction (abbreviated as frame subtraction) followed by thresholding, morphological operation, and false alarm removing. In the previous works, only gray-scaled images were considered and the threshold for false alarm rejection were heuristically decided. In the paper, all three-color components are used, and the threshold is based on the actual object size and shape properties. First, the frame subtraction and thresholding in each color layer and the logical OR operation generate a binary image as:

$$
B_k(m,n) = \left\{ \begin{array}{l} 1, if \left[\begin{array}{c} \left| I_k^R(m,n) - I_{k-1}^R(m,n) \right| > \theta_T \\ OR \\ \left| I_k^B(m,n) - I_{k-1}^B(m,n) \right| > \theta_T \\ OR \\ \left| I_k^G(m,n) - I_{k-1}^G(m,n) \right| > \theta_T \end{array} \right] \\ 0, otherwise \end{array} \right\}, k = 1, 2, \ldots, N_k, \quad (1)
$$

where $I_k^R(m,n)$, $I_k^G(m,n)$, and $I_k^B(m,n)$ are the red, green, and blue components, respectively, at frame k, N_K is the total number of frames, and θ_T is a threshold value. The morphological operation, closing, is applied to the binary image. Finally, we test the size, squareness, and rectangularity of the base rectangle to eliminate false alarms [22]. A base rectangle is defined as the smallest rectangle that contains objects. The squareness is defined as the ratio between the minor and major axes of the basic rectangle, and the rectangularity is defined as the ratio between the object size and the basic rectangle size. The maximum values for both squareness and rectangularity are equal to one. Although the object resolution is very low as in Figure 2, it will be shown that the object detection based on the frame subtraction can generate measurements for target tracking.

3. Multiple Target Tracking

A block diagram of multiple target-tracking using the new track association scheme is shown in Figure 3. Each step on the block diagram is described in the following subsections.
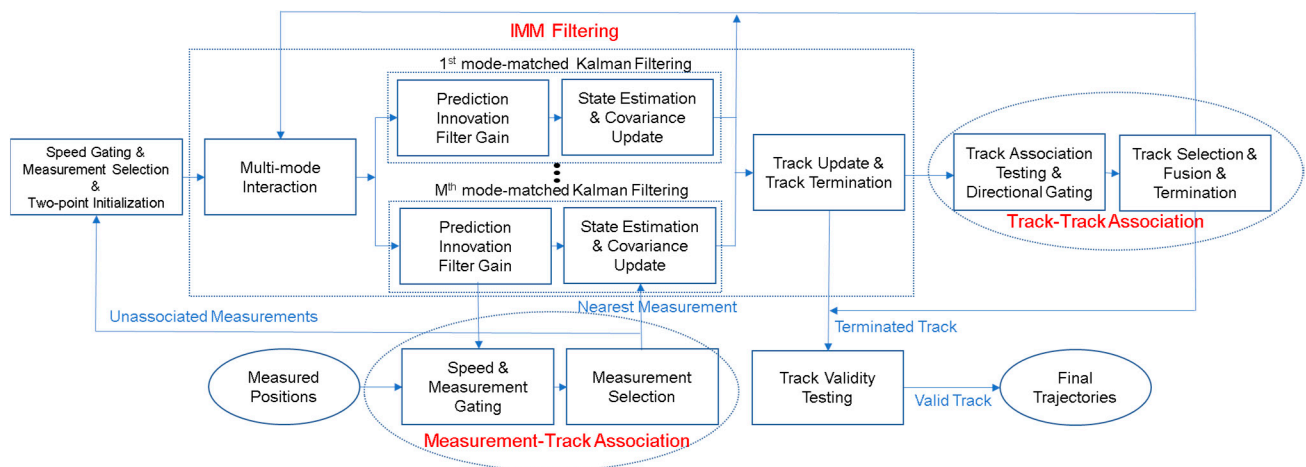

Figure 3. Block diagram of multiple target tracking.

3.1. System Modeling

It is assumed that the kinematic state of a target follows CA motion. The uncertainty of the process noise, which follows the Gaussian distribution, controls the kinematic state of the target. The discrete state equation of the CA model for the IMM estimator is as follows:

$$\boldsymbol{x}^t(k+1) = F(\Delta)\boldsymbol{x}^t(k) + q(\Delta)\boldsymbol{v}_j(k), \; j = 1, \ldots, M, \tag{2}$$

$$F(\Delta) = \begin{bmatrix} 1 & \Delta & \frac{\Delta^2}{2} & 0 & 0 & 0 \\ 0 & 1 & \Delta & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \Delta & \frac{\Delta^2}{2} \\ 0 & 0 & 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \; q(\Delta) = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ \Delta & 0 \\ 1 & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \\ 0 & 1 \end{bmatrix} \tag{3}$$

where $\boldsymbol{x}^t(k) = [x^t(k) \; v^{tx}(k) \; a^{tx}(k) \; y^t(k) \; v^{ty}(k) \; a^{ty}(k)]^T$ is a state vector of target t at frame k, $x^t(k)$ and $y^t(k)$ are positions in the x and y directions, respectively, $v^{tx}(k)$ and $v^{ty}(k)$ are velocities in the x and y directions, respectively, $a^{tx}(k)$ and $a^{ty}(k)$ are accelerations in the x and y directions, respectively, Δ is the sampling time, M is the number of modes, and $\boldsymbol{v}_j(k)$ is a process noise vector, which is Gaussian white noise with the covariance matrix $Q_{vj} = diag\left(\left[\sigma_{jx}^2 \; \sigma_{jy}^2\right]\right)$. The measurement vector for target t consists of the positions in the x and y directions. The measurement equation is as follows:

$$\boldsymbol{z}^t(k) = \begin{bmatrix} z^{tx}(k) \\ z^{ty}(k) \end{bmatrix} = H\boldsymbol{x}^t(k) + \boldsymbol{w}(k), \tag{4}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{5}$$

where $\boldsymbol{w}(k)$ is a measurement noise vector, which is Gaussian white noise with the covariance matrix $R = diag\left(\left[r_x^2 \; r_y^2\right]\right)$. The superscript t in Equations (2) and (4) indicates a target number. The target number may differ from the track number due to missing targets, broken tracks, and redundant tracks. From the next subsection, the superscript t will denote the track number.

3.2. Two Point Differencing Intialization

In the paper, the two-point (current and past) differencing initialization in the CV model is extended to the CA model. The initial state vector and covariance matrix for track t are, respectively, calculated as:

$$\hat{x}_{ini}^t(k-1|k-1) = \begin{bmatrix} z^{tx}(k-1) \\ \frac{z^{tx}(k-1)-z^{tx}(k-2)}{\Delta} \\ 0 \\ z^{ty}(k-1) \\ \frac{z^{ty}(k-1)-z^{ty}(k-2)}{\Delta} \\ 0 \end{bmatrix}, \; P_{ini}^t(k-1|k-1) = \begin{bmatrix} r_x^2 & \frac{r_x^2}{\Delta} & \frac{r_x^2}{\Delta^2} & 0 & 0 & 0 \\ \frac{r_x^2}{\Delta} & \frac{2r_x^2}{\Delta^2} & \frac{3r_x^2}{\Delta^3} & 0 & 0 & 0 \\ \frac{r_x^2}{\Delta^2} & \frac{3r_x^2}{\Delta^3} & \frac{6r_x^2}{\Delta^4} & 0 & 0 & 0 \\ 0 & 0 & 0 & r_y^2 & \frac{r_y^2}{\Delta} & \frac{r_y^2}{\Delta^2} \\ 0 & 0 & 0 & \frac{r_y^2}{\Delta} & \frac{2r_y^2}{\Delta^2} & \frac{3r_y^2}{\Delta^3} \\ 0 & 0 & 0 & \frac{r_y^2}{\Delta^2} & \frac{3r_y^2}{\Delta^3} & \frac{6r_y^2}{\Delta^4} \end{bmatrix}, \; k = 3, \ldots, N_K + 1. \tag{6}$$

The state is confirmed as the initial state of a track with the following maximum speed gating: $\|\boldsymbol{z}^t(k-1) - \boldsymbol{z}^t(k-2)\| \leq \Delta \cdot V_{max}$, where $\|\cdot\|$ denotes vector norm, and V_{max} is the maximum speed of a target. Since the initial acceleration cannot be obtained with two points, it is set to zero in $\hat{x}_{ini}^t(k-1|k-1)$, thus the target is assumed to move at a constant velocity. Other remaining measurements become past measurements and can be associated to measurements in the next frame. In the following frame, they will vanish unless they

are associated for the initialization; the size of the sliding window for the initialization is two, but it is expandable.

3.3. Multi-Mode Interaction

The mode state and covariance at the previous frame $k-1$ are mixed to generate the interacted state and covariance of target t for mode j as:

$$\hat{x}_{0j}^t(k-1|k-1) = \sum_{i=1}^{M} \hat{x}_i^t(k-1|k-1)\mu_{i|j}^t(k-1|k-1), \ t = 1, \ldots, N_T(k), \tag{7}$$

$$
\begin{aligned}
&P_{0j}^t(k-1|k-1) \\
&= \sum_{i=1}^{M} \left\{ \mu_{i|j}^t(k-1|k-1) \left\{ \begin{array}{c} P_i^t(k-1|k-1) + \\ \left[\hat{x}_i^t(k-1|k-1) - \hat{x}_{0j}^t(k-1|k-1)\right]\left[\hat{x}_i^t(k-1|k-1) - \hat{x}_{0j}^t(k-1|k-1)\right]^T \end{array} \right\} \right\},
\end{aligned}
\tag{8}
$$

$$\mu_{i|j}^t(k-1|k-1) = \frac{p_{ij}\mu_i^t(k-1)}{\sum_{i=1}^{M} p_{ij}\mu_i^t(k-1)}, \ i, j = 1, \ldots, M, \tag{9}$$

where $\hat{x}_i^t(k-1|k-1)$ and $P_i^t(k-1|k-1)$ are, respectively, the mode state and covariance of target t for mode i at frame $k-1$, $N_T(k)$ is the number of tracks at frame k, $\mu_i^t(k-1)$ is the mode probability of track t for mode i at frame $k-1$, and p_{ij} is the mode transition probability from mode i to mode j. This interaction part is essential in the IMM estimator, mixing the mode state and covariance to produce the interacted state and covariance. When the track is initialized, the mode state and covariance, $\hat{x}_i^t(k-1|k-1)$ and $P_i^t(k-1|k-1)$, are replaced by the initial state and covariance, $\hat{x}_{ini}^t(k-1|k-1)$ and $P_{ini}^t(k-1|k-1)$, respectively, and $\mu_i^t(k-1)$ is the initial mode probability, $1/M$. In consequence, $\hat{x}_{0j}^t(k-1|k-1)$ and $P_{0j}^t(k-1|k-1)$ are the same with $\hat{x}_{ini}^t(k-1|k-1)$ and $P_{ini}^t(k-1|k-1)$, respectively, when the track is initialized. The number of tracks is $N_T(k)$ because the newly initialized tracks can be added to $N_T(k-1)$.

3.4. Mode Matched Kalman Filtering

The Kalman filter is individually performed for all modes. The state and covariance predictions of track t for mode j at frame k are computed as:

$$\hat{x}_j^t(k|k-1) = F\hat{x}_{0j}^t(k-1|k-1), \tag{10}$$

$$P_j^t(k|k-1) = FP_{0j}^t(k-1|k-1)F^T + Q_j, \tag{11}$$

$$Q_j = q(\Delta)Q_{vj}q(\Delta)^T, \tag{12}$$

where $\hat{x}_j^t(k|k-1)$ and $P_j^t(k|k-1)$, respectively, are the state and the covariance prediction of track t at frame k, and T denotes the matrix transpose. The residual covariance $S_j^t(k)$ and the filter gain $W_j^t(k)$ of target t for mode j are, respectively, obtained as:

$$S_j^t(k) = HP_j^t(k|k-1)H^T + R, \tag{13}$$

$$W_j^t(k) = P_j^t(k|k-1)H^T S_j^t(k)^{-1}. \tag{14}$$

3.5. Measurement-to-Track Association

The measurement-to-track association is the process of assigning the measurements to the established tracks. The NN association rule assigns the $\hat{m}_j^t(k)$-th measurement to track t, which is obtained as:

$$\hat{m}_j^t(k) = argmin_{m=1,\dots,N_M(k)} \|v_{mj}^t(k)^T \left[S_j^t(k)\right]^{-1} v_{mj}^t(k)\|, \tag{15}$$

$$v_{mj}^t(k) = z_m(k) - H\hat{x}_j^t(k|k-1), \tag{16}$$

where $N_M(k)$ is the number of measurements at frame k, and $z_m(k)$ is the m-th measurement vector at frame k. The measurement gating is performed by the chi-square hypothesis testing assuming Gaussian measurement residuals as [18]:

$$v_{\hat{m}_j^t(k)j}^t(k)^T \left[S_j^t(k)\right]^{-1} v_{\hat{m}_j^t(k)j}^t(k) \leq \gamma_f, \tag{17}$$

where γ_f is the gate threshold for the measurement association. The maximum speed gating is also applied to the nearest measurement as:

$$\|z_{\hat{m}_j^t(k)}(k) - H\hat{x}_j^t(k-1|k-1)\| \leq \Delta \cdot S_{max}, \tag{18}$$

where S_{max} is the maximum speed of the target. The measurements that fail to associate with the track go to the initialization step in Section 3.2.

3.6. Mode State Estimate and Covariance Update

The state and covariance of target t for mode j are updated as:

$$\hat{x}_j^t(k|k) = \hat{x}_j^t(k|k-1) + W_j^t(k)v_{\hat{m}_j^t(k)j}^t(k), \tag{19}$$

$$P_j^t(k|k) = P_j^t(k|k-1) - W_j^t(k)S_j^t(k)W_j^t(k)^T. \tag{20}$$

If no measurement is associated, they merely become the predictions of the state and the covariance as:

$$\hat{x}_j^t(k|k) = \hat{x}_j^t(k|k-1), \tag{21}$$

$$P_j^t(k|k) = P_j^t(k|k-1). \tag{22}$$

The mode probability is updated as:

$$\mu_j^t(k) = \frac{\Lambda_j^t(k) \sum_{i=1}^{M} p_{ij}\mu_i^t(k-1)}{\sum_{j=1}^{M} \Lambda_j^t(k) \sum_{i=1}^{M} p_{ij}\mu_i^t(k-1)}, \tag{23}$$

$$\Lambda_j^t(k) = N\left(\mathbf{0}; v_{\hat{m}_j^t(k)j}^t(k), S_j^t(k)\right), \tag{24}$$

where N denotes Gaussian probability density function. If no measurement is associated, the mode probability becomes:

$$\mu_j^t(k) = \sum_{i=1}^{M} p_{ij}\mu_i^t(k-1). \tag{25}$$

Finally, the state vector and covariance matrix are updated as:

$$\hat{x}^t(k|k) = \sum_{j=1}^{M} \hat{x}_j^t(k|k)\mu_j^t(k), \tag{26}$$

$$P^t(k|k) = \sum_{j=1}^{M} \mu_j^t(k) \left\{ P_j^t(k|k) + \left[\hat{x}_j^t(k|k) - \hat{x}^t(k|k) \right] \left[\hat{x}_j^t(k|k) - \hat{x}^t(k|k) \right]^T \right\}. \tag{27}$$

The procedures from Equations (7)–(27) repeat until the track is terminated.

3.7. Directional Track-to-Track Association

If multiple measurements on a single target are continuously detected, multiple tracks can be created for the target. The track association scheme was developed to eliminate redundant tracks when the Kalman filter is applied in [21]. In the paper, the track association method is improved by considering the moving direction of the target in the IMM framework. The multiple tracks on the same target have the error dependencies on each other, thus the following track-association hypothesis testing [18] is preceded as:

$$\begin{aligned} \left[\hat{x}^s(k|k) - \hat{x}^t(k|k) \right]^T \left[T^{st}(k) \right]^{-1} \left[\hat{x}^s(k|k) - \hat{x}^t(k|k) \right]^T \leq \gamma_g, \\ s, t = 1, \ldots, N_T(k), s \neq t \end{aligned} \tag{28}$$

$$T^{st}(k) = P^s(k|k) + P^t(k|k) - P^{st}(k|k) - P^{ts}(k|k) \tag{29}$$

$$P^{st}(k|k) = [I - b^s(k)W^s(k)H] \Big[FP^{st}(k-1|k-1)F^T + Q \Big] [I - b^t(k)W^t(k)H], \tag{30}$$

where $\hat{x}^s(k|k)$ and $\hat{x}^t(k|k)$ are the state vector of track s and t, respectively, at frame k, $P^s(k|k)$ and $P^t(k|k)$ are the covariance matrix of track s and t, respectively, at frame k, γ_g is the gate threshold for track association, $b^s(k)$ and $b^t(k)$ are binary numbers that become one when track s or t is associated with a measurement, otherwise they are zero. It is noted that $T^{st}(k)$ in Equation (29) is meaningless if its determinant is not positive. The fused covariance in Equation (30) is a linear recursion and its initial condition is set at $P^{st}(0|0) = [0]_{6\times 6}$, and $W^t(k|k)$ is obtained as the combined filter gain of track t as:

$$W^t(k|k) = \sum_{j=1}^{M} W_j^t(k|k)\mu_j^t(k). \tag{31}$$

The directional gating process is newly developed for the directional track association. It tests the maximum deviation in the directions of velocity of tracks and the direction of a displacement vector between the positions of two tracks as:

$$max\left(\cos^{-1} \frac{<\hat{d}^{st}(k|k),\ \hat{v}^s(k|k)>}{\|\hat{d}^{st}(k|k)\|\|\hat{v}^s(k|k)\|},\ \cos^{-1} \frac{<\hat{d}^{st}(k|k),\ \hat{v}^t(k|k)>}{\|\hat{d}^{st}(k|k)\|\|\hat{v}^t(k|k)\|} \right) \leq \theta_g, \tag{32}$$

$$\hat{d}^{st}(k|k) = \begin{bmatrix} \hat{x}^s(k) - \hat{x}^t(k) \\ \hat{y}^s(k) - \hat{y}^t(k) \end{bmatrix},\ \hat{v}^s(k|k) = \begin{bmatrix} \hat{v}^{sx}(k) \\ \hat{v}^{sy}(k) \end{bmatrix}, \hat{v}^t(k|k) = \begin{bmatrix} \hat{v}^{tx}(k) \\ \hat{v}^{ty}(k) \end{bmatrix}, \tag{33}$$

where $< \cdot >$ denotes the inner product operation, θ_g is the maximum deviation angle, $\hat{d}^{st}$ is a displacement vector between tracks s and t, and $\hat{v}^s(k|k)$ and $\hat{v}^t(k|k)$ are the velocity vectors of tracks s and t, respectively. The directional gating process is illustrated in Figure 4. Two tracks s and t are set in one target and a track u is set in another target. The angles between $\hat{d}^{st}(k|k)$ and $\hat{v}^s(k|k)$, and $\hat{d}^{st}(k|k)$ and $\hat{v}^t(k|k)$ are almost zero degree, while the angles between $\hat{d}^{tu}(k|k)$ and $\hat{v}^t(k|k)$, and $\hat{d}^{tu}(k|k)$ and $\hat{v}^u(k|k)$ are almost 90 degree. Thus, tracks s and t are associative, but tracks t and u are not associative.
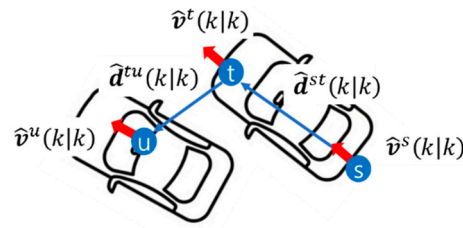
Figure 4. Illustration of directional gating process.

When the track association hypothesis and the directional gating process are satisfied, the most accurate track is selected, then the current state of the selected track is replaced with a fused estimate, and the non-selected track is immediately terminated. The selection process is based on the determinant of the covariance matrix because the more accurate track has less error (covariance). A track is selected and fused as:

$$\hat{c} = \underset{s,t}{argmin}\left[\left|P^s(k|k)\right|, \left|P^t(k|k)\right|\right], \tag{34}$$

$$\hat{x}^{\hat{c}}(k|k) = \hat{x}^s(k|k) + \left[P^s(k|k) - P^{st}(k|k)\right]\left[P^s(k|k) + P^t(k|k) - P^{st}(k|k) - P^{ts}(k|k)\right]^{-1}\left[\hat{x}^t(k|k) - \hat{x}^s(k|k)\right], \tag{35}$$

$$P^{\hat{c}}(k|k) = P^s(k|k) - \left[P^s(k|k) - P^{st}(k|k)\right]\left[P^s(k|k) + P^t(k|k) - P^{st}(k|k) - P^{ts}(k|k)\right]^{-1}\left[P^s(k|k) - P^{ts}(k|k)\right]. \tag{36}$$

3.8. Track Termination and Validity Testing

There are three cases of track termination. One is associated but not selected during the track association. Other cases are when the track is continuously updated without measurements over the maximum searching number, or the track speed is below the minimum target speed. After the track is terminated, its validity is tested with the track life length. The track life length is the number of frames between the initial updated frame and the last frame updated by a measurement. If the track life length is shorter than the minimum track life length, the track is removed as a false track.

The tracking performance is evaluated in terms of the TTL and the MTL [27]. The TTL and MTL are defined, respectively, as:

$$\text{TTL} = \frac{\text{Sum of lengths of tracks which have the same target ID}}{\text{Target life length} - 1}, \tag{37}$$

$$\text{MTL} = \frac{\text{TTL}}{\text{Number of tracks associated in TTL}}. \tag{38}$$

A track's target ID is defined as the target with the most measurements on the track. The denominator of the TTL is slightly modified as the target life length minus one because two frames are required for the initialization. The MTL is equal or less than the TTL due to track breakage or overlap. Figure 5 illustrates three tracks originating from one target. The target *t* is occluded at frame *k*. The target *t*'s life length is 9. The TTL and MTL of target *t* are 0.75 (=6/8) and 0.25 (=0.75/3), respectively. It is noted that if there are multiple tracks on one target in a frame, only one frame is counted for the TTL. Tracks 1–3 in Figure 5 have track life lengths of 3, 5, and 3, respectively. In this illustration, the minimum track life length for the track validity is set to 3 frames.
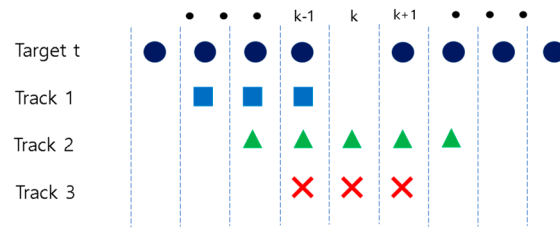
Figure 5. Illustration of three tracks associated with one target.

4. Results

In this section, experimental results will be detailed through video description, parameter setting, and moving vehicle tracking along with the proposed strategy.

4.1. Video Description and Moving Object Detection

The video was captured at a frame rate of 30 fps by a Zenmuse X5 mounted on an Inspire 2. The urban and suburban environment in the video includes a highway interchange, a toll gate, structures, and farms. The drone hovered at a height of 400 m with the camera pointing directly at the ground. There was a total of 55 moving vehicles for approximately 20 s. The frame size is 3840×2160 pixels, and the space to pixel ratio is 6 pixel/m. The actual frame rate is 10 fps as every third frame was processed for efficient detection processing; a total of 202 frames were considered, and N_K is 201 because the detection process starts with two frames ($k = 0$ and $k = 1$). It is noted that the last frame ($k = 201$) corresponds to the 604th frame of the video. Figure 6a shows Targets 1–33 at $k = 1$. As shown in Figure 6b, a total of 46 targets appeared until $k = 101$; targets 1, 6, 20, 32, 33, and 43 either disappeared or were occluded. A total of 53 targets appeared until $k = 151$ as shown in Figure 6c; targets 1, 6, 20, 32, 33, 43, 45, and 46 are missing in Figure 6c. A total of 55 targets appeared until the last frame, $k = 201$ as shown in Figure 6d; targets 1, 3, 6, 20, 26, 31–33, 43, 45, and 52 are not shown in Figure 6d. It can be seen that the bridge above the road in the center and the toll gate at the bottom right occlude the vehicles.
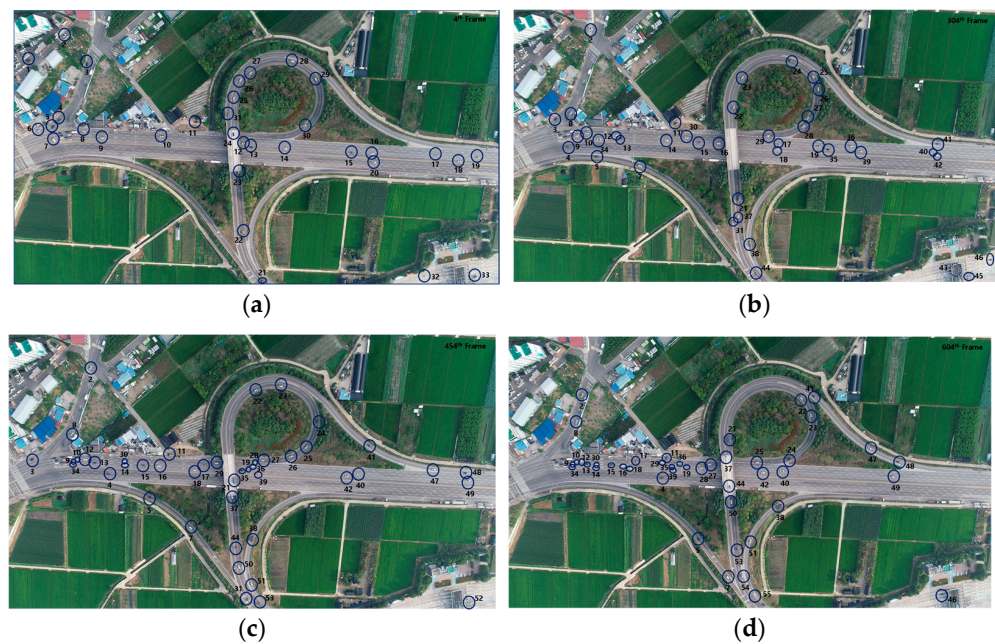

Figure 6. (**a**) 33 targets at $k = 1$; (**b**) 40 targets at $k = 101$; (**c**) 45 targets at $k = 151$; and (**d**) 44 targets at $k = 201$.

For the object detection, θ_T in Equation (1) is set to 30. The structure element for the morphological operation is set at $[1]_{2 \times 2}$. The minimum size of the basic rectangle is set

to 3 m^2, which corresponds to 108 pixels. The minimum squareness and rectangularity are set to 0.2 and 0.3, respectively. Figure 7 shows the intermediate results of the detection process of Figure 6a. Figure 7a is a binary image generated by the frame subtraction and thresholding. Figure 7b shows the object area after applying the morphological operation to Figure 7a. Figure 7c shows the basic rectangles shown as blue contours after removing false alarms in Figure 7b. Figure 7d shows the centroid of the basic rectangle of the detected objects. A total of 120 objects were detected, of which 92 detections were false alarms. Figure 8a–c are the detection results of Figure 6b–d, respectively. The detection counts are 57, 92, and 60 in Figure 8a–c, respectively. Among them, the number of false alarms is 13, 45, and 13, respectively. Figure 9 shows the detections for all frames. Most of the detections were along roads that coincided with the vehicle's trajectory. However, some of them are found in the non-road area such as structures and farms.
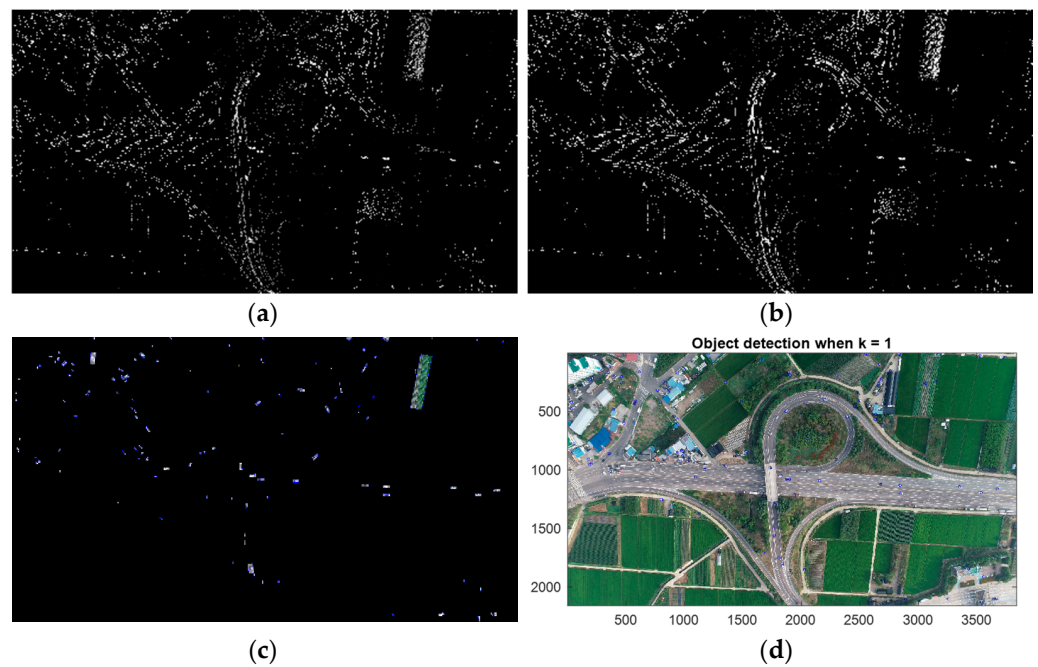

Figure 7. Moving object detection of Figure 6a: (**a**) frame subtraction and thresholding; (**b**) morphological operation; (**c**) basic rectangles after false alarm removing; and (**d**) 120 detections circled in blue, including 97 false alarms.
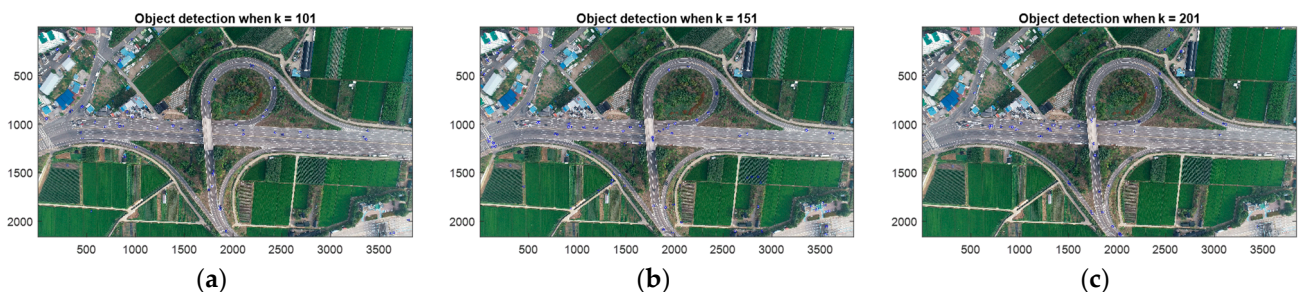

Figure 8. Moving object detection of (**a**) Figure 6b; (**b**) Figure 6c; and (**c**) Figure 6d.
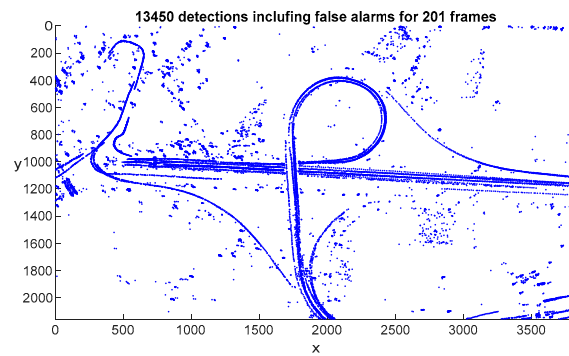
Figure 9. Moving object detection for all frames.

4.2. Multiple Target Tracking

This subsection shows the results of target tracking based on the detections in Figure 9. The sampling time in Equations (2) and (3) is 0.1 s since every third frame is processed. The track association and the directional track association are compared in the case of IMM-CV and IMM-CA. The details of IMM-CV can be found in [22]. The parameters for target tracking are shown in Table 1; θ_g is required only for the directional track association.

Table 1. Parameters for target tracking.

Parameters		IMM-CV	IMM-CA
Sampling time		0.1 s	
Max. target speed for initialization, V_{max}		30 m/s	
Process noise variance	$\sigma_{1x} = \sigma_{1y}$	1 m/s^2	0.01 m/s^2
	$\sigma_{2x} = \sigma_{2y}$	10 m/s^2	0.1 m/s^2
Mode transition probabiltiy p_{ij}		$\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$	
Measurement noise variance, $r_x = r_y$		1.5 m	
Measurent association	Gate threshold, γ_f	8	
	Max. target speed, S_{max}	35 m/s	
Track association	Gate threshold, γ_g	70	
	Angle threhold, θ_g	$\frac{\pi}{9}$ rad	
Track termination	Max. searching number	20 frames (=2 s)	
	Min. target speed	1 m/s	
Min. track life length for track validity		20 frames (=2 s)	

In the case of the IMM-CV, a total of 61 and 65 valid tracks are generated when the track association and the directional track association were used, respectively. They are shown in Figure 10a,b. A total of 84 and 64 valid tracks are generated in the case of IMM-CA as shown in Figure 11a,b, respectively. Four supplementary multimedia files (MP4 format) for Figures 10 and 11 are available online. The first is the IMM-CV with the track association (Supplementary Material Video S1), and the second is the IMM-CV with the directional track association (Supplementary Material Video S2). The third is the IMM-CA with the track association (Supplementary Material Video S3), and the last one is the IMM-CA with the directional track association (Supplementary Material Video S4). The black squares and the numbers in MP4 files are position estimates and track numbers in the order they were initialized. The blue dots are the measured positions including false alarms. Every third frame was processed but the movie was recorded at 30 fps, thus the movies are three times faster than the original video.
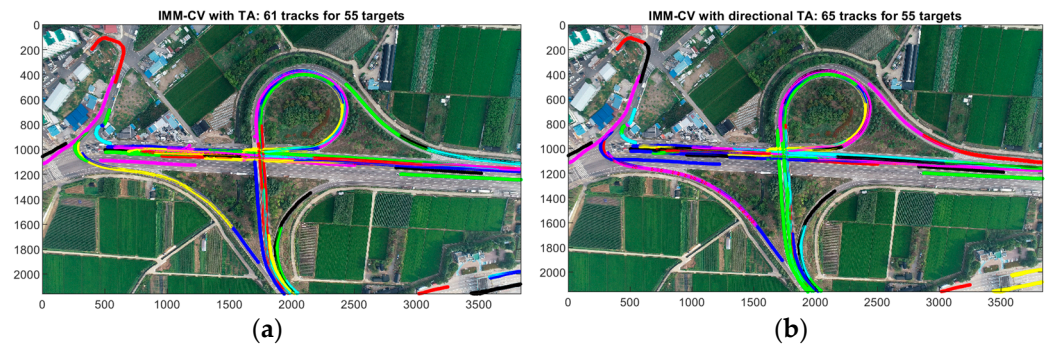
Figure 10. IMM-CV: (**a**) track association and (**b**) directional track association.
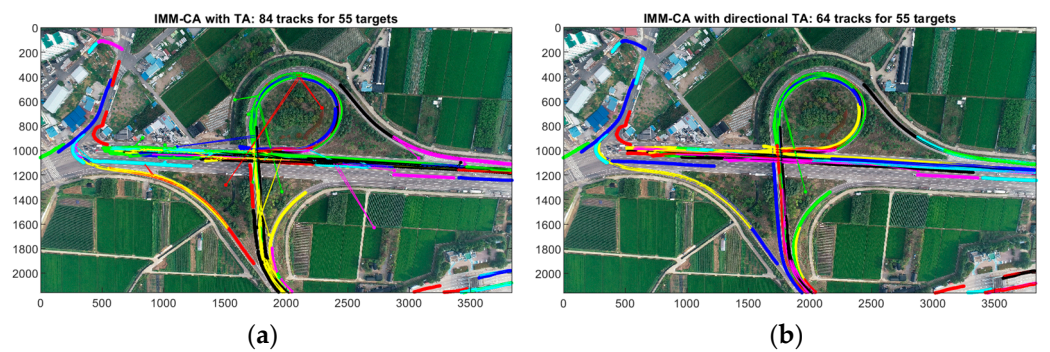

Figure 11. IMM-CA: (**a**) track association and (**b**) directional track association.

Figures 12 and 13 show the TTL and MTL of IMM-CV and IMM-CA, respectively. The TTL and MTL are the same if only one track is generated for one target. There are track breakages on 8, 9, 19, and 9 targets in Figure 12a,b and Figure 13a,b, respectively. The TTL and MTL become zero if the target is missing. Two targets (targets 1 and 11) are missing for all cases. In addition, no track is established on targets 13, 33, and 54 in Figure 12a, targets 33 and 54 in Figure 12b, and target 13 in Figure 13a.
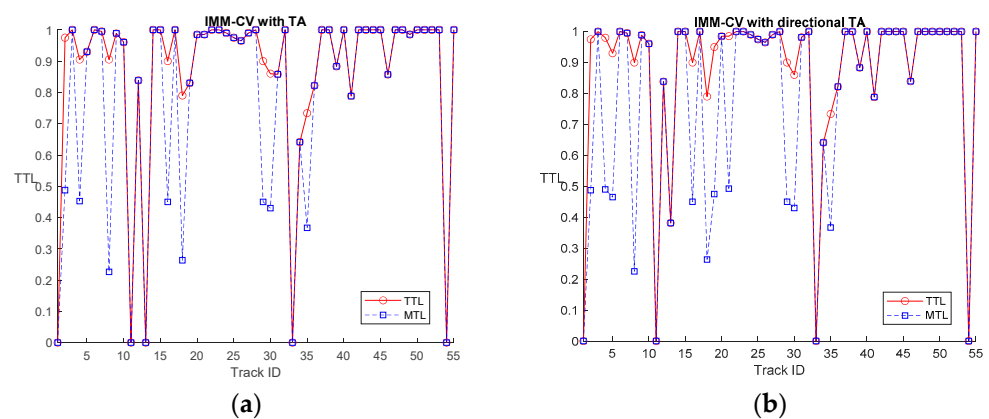

Figure 12. TTL and MTL of IMM-CV: (**a**) track association and (**b**) directional track association.
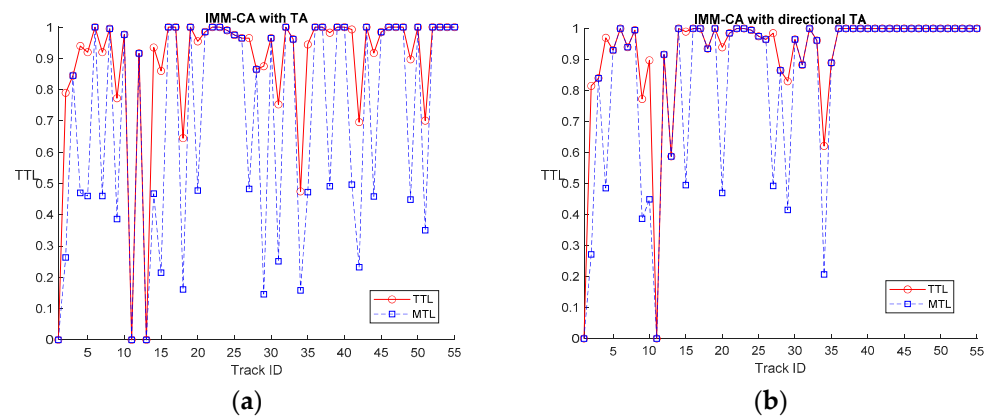
Figure 13. TTL and MTL of IMM-CA: (**a**) track association and (**b**) directional track association.

Table 2 shows the overall tracking performance, number of tracks, average TTL and MTL, number of targets with broken tracks, and number of missing targets. No false tracks were generated for all cases. The IMM-CA with the directional track association outperforms other methods showing that the average TTL is 0.917 and the average MTL is 0.842.

Table 2. Overall tracking performance.

	IMM-CV		IMM-CA	
	Track Association	Directional TA	Track Association	Directional TA
Number of tracks	61	65	84	64
Avg. TTL	0.859	0.871	0.879	0.917
Avg. MTL	0.789	0.775	0.705	0.842
Number of targets with broken tracks	8	9	19	9
Number of missing targets	5	4	3	2

5. Discussion

In the paper, three target tracking associations have been successfully developed for a real-world scenario: measurement-to-measurement (initialization), measurement-to-track, and track-to-track associations. The video of a highway interchange and its surroundings features a variety of moving vehicles, including vehicles driving straight and on curves, occluded, intersected on and under the bridge, and stopping at the traffic signal. Since a minimum number of frames is required for a track to be valid, only targets with a life length over 20 frames (2 s) were considered. There can be incorrect results in the tracking during the last 20 frames because the track is invalid if it is initialized during the last 20 frames.

The tracks slower than the minimum speed are terminated, which is set at 1 m/s in the paper. It is noted that very slow or stopped targets are not of interest as they do not pose an imminent threat to security and surveillance.

In all cases, the proposed target tracking scheme does not generate false tracks in a heavy clutter environment. The IMM scheme was able to handle multiple targets with various maneuvers. The directional track association is particularly effective when two vehicles are moving very closely side by side. The average TTL and MTL are 0.917 and 0.842, respectively, for the IMM-CA with the directional track association; the MTL is less than the TTL on 9 targets. They were caused by track breakages as can be seen in the Supplementary Material Video S4. Therefore, the IMM-CA equipped with the directional track association was able to fuse all tracks overlapping on the target. The track breakages were caused by complex reasons including missing detection, wrong association, and high maneuvering.

Two targets (Targets 1 and 11) were missed for the IMM-CA with directional track association because the initialization failed; a valid initialization requires two consecutive detections in the direction of the target. The low detection is mainly caused by the low-resolution object. The surveillance area is more than 0.23 km^2 (=640 m × 360 m) in the paper. Higher drone altitudes will increase the surveillance area but generate lower-resolution object images and lower detection rates, which remains an issue for future research.

6. Conclusions

In the paper, two strategies were newly developed for multi-target tracking by a multirotor. One is the IMM estimator equipped with track association. The other is the directional track association. The IMM-CA scheme was compared with the IMM-CV when applying track association or directional track association. It was found that the IMM-CA using the directional track association outperformed other methods.

Comprising two stages of moving object detection and multi-target tracking, the overall process is effective for computing as it does not require high-resolution video streaming or storage. However, the object recognition is not included in the process.

This system is suitable for vehicle tracking and traffic control or vehicle counting. It can be extended to various security, surveillance, and other applications including a large number of people and animal tracking over long distances. Target tracking with a dynamic drone with various perspectives remains for future study.

References

1. Alzahrani, B.; Oubbati, O.S.; Barnawi, A.; Atiquzzaman, M.; Alghazzawi, D. UAV assistance paradigm: State-of-the-art in applications and challenges. *J. Netw. Comput. Appl.* **2020**, *166*, 102706. [CrossRef]
2. Zaheer, Z.; Usmani, A.; Khan, E.; Qadeer, M.A. Aerial surveillance system using UAV. In Proceedings of the 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN), Hyderabad, India, 21–23 July 2016; pp. 1–7.
3. Theys, B.; Schutter, J.D. Forward flight tests of a quadcopter unmanned aerial vehicle with various spherical body diameters. *Int. J. Micro Air Veh.* **2020**, *12*, 1–8. [CrossRef]
4. Li, S.; Yeung, D.-Y. Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models. In Proceeding of the Thirty-Frist AAAI conference on Artificial Intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017; pp. 4140–4146.
5. Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. In *Computer Vision—ECCV 2018*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; pp. 375–391. [CrossRef]
6. Zhang, H.; Lei, Z.; Wang, G.; Hwang, J. Eye in the Sky: Drone-Based Object Tracking and 3D Localization. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 899–907. [CrossRef]
7. Zhang, S.; Zhuo, L.; Zhang, H.; Li, J. Object Tracking in Unmanned Aerial Vehicle Videos via Multifeature Discrimination and Instance-Aware Attention Network. *Remote Sens.* **2020**, *12*, 2646. [CrossRef]
8. Kouris, A.; Kyrkou, C.; Bouganis, C.-S. Informed Region Selection for Efficient UAV-Based Object Detectors: Altitude-Aware Vehicle Detection with Cycar Dataset. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 4–9 November 2019; pp. 51–58.
9. Kamate, S.; Yilmazer, N. Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles. *Procedia Comput. Sci.* **2015**, *61*, 436–441. [CrossRef]
10. Fang, P.; Lu, J.; Tian, Y.; Miao, Z. An Improved Object Tracking Method in UAV Videos. *Procedia Eng.* **2011**, *15*, 634–638. [CrossRef]

11. Jianfang, L.; Hao, Z.; Jingli, G. A novel fast target tracking method for UAV aerial image. *Open Physics* **2017**, *15*, 420–426. [CrossRef]

12. Sinha, A.; Kirubarajan, T.; Bar-Shalom, Y. Autonomous Ground Target Tracking by Multiple Cooperative UAVs. In Proceedings of the 2005 IEEE Aerospace Conference, Big Sky, MT, USA, 5–12 March 2005; pp. 1–9.

13. Guido, G.; Gallelli, V.; Rogano, D.; Vitale, A. Evaluating the accuracy of vehicle tracking data obtained from Unmanned Aerial Vehicles. *Int. J. Transp. Sci. Technol.* **2016**, *5*, 136–151. [CrossRef]

14. Rajasekaran, R.K.; Ahmed, N.; Frew, E. Bayesian Fusion of Unlabeled Vision and RF Data for Aerial Tracking of Ground Targets. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 1629–1636.

15. Stone, L.D.; Streit, R.L.; Corwin, T.L.; Bell, K.L. *Bayesian Multiple Target Tracking*, 2nd ed.; Artech House: Boston, MA, USA, 2014.

16. Blom, H.A.P.; Bar-shalom, Y. The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Trans. Autom. Control* **1988**, *33*, 780–783. [CrossRef]

17. Bar-Shalom, Y.; Fortmann, T.E. *Tracking and Data Association*; Academic Press Inc.: Orlando, FL, USA, 1989.

18. Bar-Shalom, Y.; Li, X.R. *Multitarget-Multisensor Tracking: Principles and Techniques*; YBS Publishing: Storrs, CT, USA, 1995.

19. Lee, M.-H.; Yeom, S. Multiple target detection and tracking on urban roads with a drone. *J. Intell. Fuzzy Syst.* **2018**, *35*, 6071–6078. [CrossRef]

20. Yeom, S.; Cho, I.-J. Detection and Tracking of Moving Pedestrians with a Small Unmanned Aerial Vehicle. *Appl. Sci.* **2019**, *9*, 3359. [CrossRef]

21. Yeom, S.; Nam, D.-H. Moving Vehicle Tracking with a Moving Drone Based on Track Association. *Appl. Sci.* **2021**, *11*, 4046. [CrossRef]

22. Yeom, S. Moving People Tracking and False Track Removing with Infrared Thermal Imaging by a Multirotor. *Drones* **2021**, *5*, 65. [CrossRef]

23. Houles, A.; Bar-Shalom, Y. Multisensor Tracking of a Maneuvering Target in Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1989**, *25*, 176–189. [CrossRef]

24. Peng, Z.; Li, F.; Liu, J.; Ju, Z. A Symplectic Instantaneous Optimal Control for Robot Trajectory Tracking with Differential-Algebraic Equation Models. *IEEE Trans. Ind. Electron.* **2020**, *67*, 3819–3829. [CrossRef]

25. Li, X.R.; Jilkov, V.P. Survey of maneuvering target tracking, part I: Dynamic models. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 1333–1364.

26. Moore, J.R.; Blair, W.D. *Multitarget-Multisensor Tracking: Applications and Advances*; Bar-Shalom, Y., Blair, W.D., Eds.; Practical Aspects of Multisensor Tracking, Chap. 1; Artech House: Boston, MA, USA, 2000; Volume III.

27. Yeom, S.-W.; Kirubarajan, T.; Bar-Shalom, Y. Track segment association, fine-step IMM and initialization with doppler for improved track performance. *IEEE Trans. Aerosp. Electron. Syst.* **2004**, *40*, 293–309. [CrossRef]