drones

MDPI

Article

Thermal Image Tracking for Search and Rescue Missions with a Drone

Seokwon Yeom

Department of Artificial Intelligence, Daegu University, Gyeongsan 38453, Republic of Korea; yeom@daegu.ac.kr; Tel.: +82-53-850-6643

Abstract: Infrared thermal imaging is useful for human body recognition for search and rescue (SAR) missions. This paper discusses thermal object tracking for SAR missions with a drone. The entire process consists of object detection and multiple-target tracking. The You-Only-Look-Once (YOLO) detection model is utilized to detect people in thermal videos. Multiple-target tracking is performed via track initialization, maintenance, and termination. Position measurements in two consecutive frames initialize the track. Tracks are maintained using a Kalman filter. A bounding box gating rule is proposed for the measurement-to-track association. This proposed rule is combined with the statistically nearest neighbor association rule to assign measurements to tracks. The track-to-track association selects the fittest track for a track and fuses them. In the experiments, three videos of three hikers simulating being lost in the mountains were captured using a thermal imaging camera on a drone. Capturing was assumed under difficult conditions; the objects are close or occluded, and the drone flies arbitrarily in horizontal and vertical directions. Robust tracking results were obtained in terms of average total track life and average track purity, whereas the average mean track life was shortened in harsh searching environments.

Keywords: search and rescue missions; thermal image; object detection; target tracking; bounding box gating

check for updates

Citation: Yeom, S. Thermal Image Tracking for Search and Rescue Missions with a Drone. *Drones* **2024**, *8*, 53. https://doi.org/10.3390/drones8020053

Academic Editor: Giordano Teza

Received: 10 January 2024
Revised: 31 January 2024
Accepted: 3 February 2024
Published: 5 February 2024

1. Introduction

Multirotor drones are widely applied in a variety of fields [1]. The ability to hover and maneuver by the operator or as programmed makes them valuable tools in numerous industries. By capturing images from various angles and heights, drones can obtain cost-effective aerial views covering arbitrary areas.

Thermal imaging cameras detect infrared radiation emitted by objects [2,3]. This radiation is invisible to the human eye, but thermal imaging cameras convert it into a visible image. Thermal imaging requires no illumination or ambient light, penetrating obstacles including smoke, dust, haze, and light foliage. However, the resolution of the thermal image is low, and no texture or color information is provided.

Drones equipped with thermal imaging cameras are highly effective in locating missing people for search and rescue (SAR) missions and surveillance [4–6]. The technology has been also applied to wildlife monitoring and agricultural and industrial inspection [7,8]. People detection with thermal images captured by drones has been studied in [9–15]. Persons and animals were detected using the YOLO detection model [9]. Persons and cars were detected from different observation angles of the drone using the YOLO detection model [10]. The YOLO detection model was adopted to build a human detection system using drones [11]. The spatial gray level co-occurrence matrix estimates temperature differences [12]. In [13], a person is recognized with a two-stage hot-spot detection approach. People and fire detection were studied with high-altitude thermal images obtained by optical and thermal sensors [14]. However, the tracking of people in thermal scenes using drones has been seldom researched [15]. In [15], people tracking with a thermal imaging camera mounted on a small drone was performed for the first time. However, object detection was performed using the *k*-means algorithm, which resulted in degraded performance in complex backgrounds.

Tracking of non-living objects using thermal imaging by fixed-wing drones was studied in [16,17]. The Kalman filter was adopted to track a boat [16], and a colored-noise measurement model was utilized to track a small vessel [17]. Tracking of people and animals using fixed thermal imaging cameras has been studied in [18–24]. Foreground objects were extracted by contour-based background subtraction [18]. The pedestrian detection was performed by local adaptive thresholding [19]. People in the aerial thermal view were detected and tracked with a particle filter [20]. The weighted correlation filter tracked the various targets in the thermal image database [21]. An effective feature representation was introduced based on correlation energy in [22]. An adaptively multi-feature fusion model was proposed to integrate hand-crafted features and a convolutional neural network for thermal object tracking [23]. The Kalman filter was used to detect sports players in thermal videos for real-time tracking [24]. However, the tracking conditions were limited to fixed cameras, and a plain background. In [25], a tracking-by-detection approach was studied in the thermal spectrum. A convolutional neural network, pre-trained with visible images, was transferred to the thermal tracking [26].

A multiple-target tracking system consists of a sensing system that observes the movement of multiple objects in a dynamic environment and a target tracking algorithm that simultaneously estimates the attributes (position, velocity, acceleration, etc.) of multiple objects based on the observation. As a result, a multi-target tracking system forms a trajectory (track) for each target; so, it is essential to establish the identity of the same target while it is present, and this can be achieved by accurately matching the established track with the observation. However, due to high and various maneuvering, close or occluded objects, low object detection, or high false alarms, tracks may be missed, broken, or overlapped on the same target.

In this paper, tracking of people in the mountains is studied using thermal imaging by a drone. The overall process comprises two stages: object detection and multiple-target tracking. For object detection, YOLOv5 is adopted to generate a bounding box of objects. YOLO's deep learning framework has garnered immense popularity for its versatility, ease of use, and high performance [27]. The centroid of the bounding box is considered the measured position (observation) for target tracking.

Multiple-target tracking is performed via three stages: track initialization, maintenance, and termination [28–30]. The track maintenance consists of measurement-to-track association (measurement association), track update, and track-to-track association (track association). Figure 1 shows an entire block diagram of target tracking with infrared thermal videos acquired by drones. Bold fonts inside red bold line boxes include the newly proposed contents in this paper. A scheme of the nearest neighbor measurement association, followed by track association, track termination, and validity testing, has been developed in previous works showing robust performance in ground target tracking from visible images acquired by a drone [29,30].

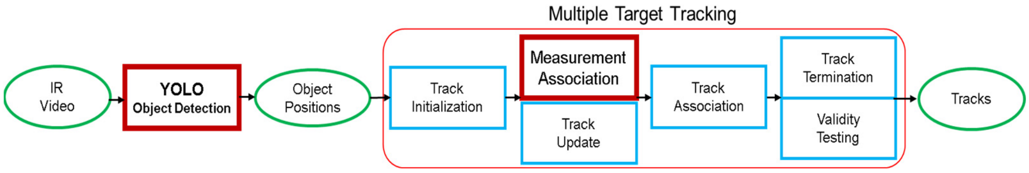


Figure 1. Block diagram of thermal image target tracking.

To the best of my knowledge, ref. [15] was the first study to track people with a thermal imaging camera mounted on a small drone. This paper has been substantially improved from [15]. The contributions of this paper are (1) the YOLO detection mode is applied to extract the position information of the thermal target. The YOLO detection model is pre-trained with visible image datasets, but it was transferred to thermal object detection for multiple-target tracking. (2) A bounding box gating scheme is proposed for the measurement association. This scheme allows track updates if the intersection over union (IoU) between the bounding box of the current frame and the bounding box of the previous frame is sufficiently high. The centroid of the current frame bounding box is the measurement that is statistically nearest to the position prediction at the current frame and the centroid of the previous frame bounding box is the measurement associated with updating the tract at the previous frame. (3) The framework combining the measurement association and the track association with the Kalman filter shows robust tracking performance of the mobility of the platform in complex backgrounds. In the presented scenarios, thermal objects are closely located and heavily occluded, and the sensing platform is allowed to move fast.

In the experiments, the drone flies at a height of 40~60 m. The video shows three hikers simulating being lost in the mountains on a winter night with no ambient light. The drone moves rapidly in horizontal and vertical directions, and the perspective of the camera is arbitrary. People are often occluded by other people or trees and leaves. Figure 2 shows three sample scenes extracted from the three videos, respectively. The experimental results show the average total track lives (TTLs) for the three videos are obtained as 0.987, 0.993, and 0.894, respectively. The corresponding average mean track lives are 0.987, 0.442, and 0.151, respectively. The average track purities (TPs) are obtained as 1, 0.999, and 0.995, respectively, for the three videos. The average MTL is reduced for two videos due to the track breakage in the harsh environments.

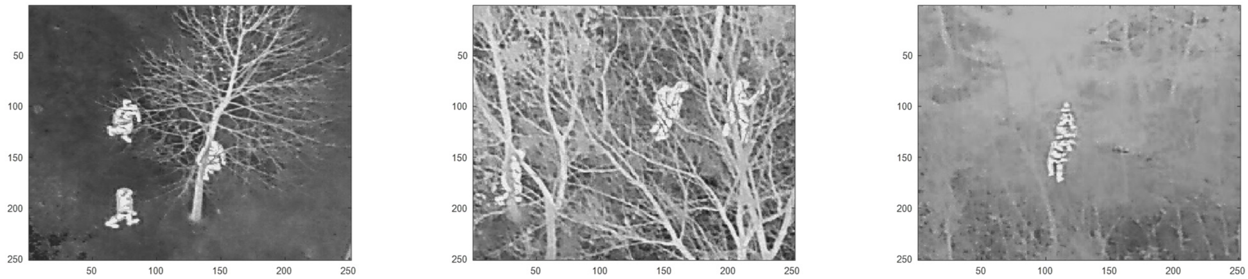


Figure 2. Sample frames of three thermal videos capturing people in distress.

The rest of the paper is organized as follows: the YOLOv5 training process is explained in Section 2. Section 3 describes each step of multiple-target tracking. The experimental results are presented in Section 4, and conclusions follow in Section 5.

2. YOLOv5 for Thermal Videos

YOLOv5 is the 5th iteration of the YOLO object detection model [27]. It is designed to provide high-speed, high-accuracy results in real-time. YOLOv5 has several pre-trained models with a visible image dataset; YOLOv5x shows the best performance among them. The output of YOLO is multiple bounding boxes surrounding the object of interest, along with the object's class name and confidence level.

The YOLO pre-trained models are trained with thermal images; the number of training images is 197 and a total of 548 rectangular boxes are manually drawn in them. The training images were obtained from a different video than the ones used for the tracking experiments. Each rectangular box indicates the instance of a person in the scene. The rectangular boxes were manually drawn using the graphical image annotation tool LabelImg [31]. During training, all instances are named with only one class, 'walking'. Three pre-trained models, YOLOv5s, YOLOv5l, and YOLOv5x, were transferred to detect thermal objects. They were trained with 100 epochs. No data augmentation was applied, and no background was used for training. Figure 3 shows five sample images, each image containing three blue rectangular boxes. Supplementary Video S1 shows a movie of all 197 images containing 548 rectangular boxes. The experimental results of detection testing will be shown in Section 4.

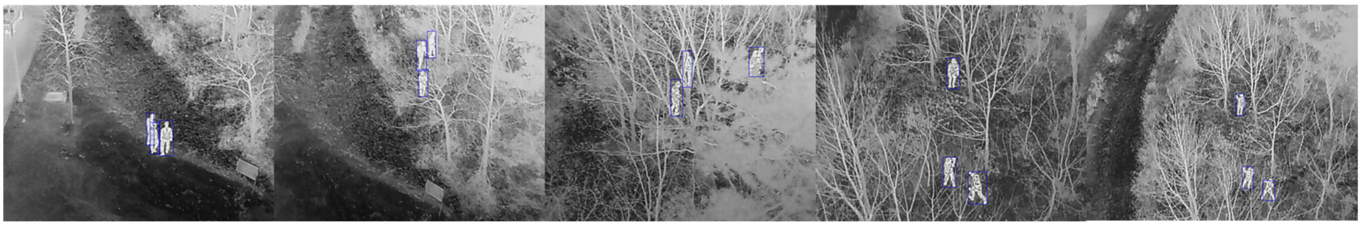


Figure 3. Sample training images with thermal object instances.

3. Multiple-Target Tracking

A block diagram of multiple-target tracking is shown in Figure 4. Tracks are initialized with two-point differencing between the measurements in consecutive frames after the initial speed gating. Tracks are maintained by track update, measurement association, and track association. Measurement association is necessary in multiple-target and clutter environments. The track association aims to handle multi-sensor environments [32], but it has been improved to fuse tracks generated with a single sensor [30]. The state of the target is estimated using the Kalman filter. The track is terminated if no measurement is available or track-fusion occurs. The validity of the track is tested with the track-life length after track termination. Each step of the block diagram is described in the following subsections.

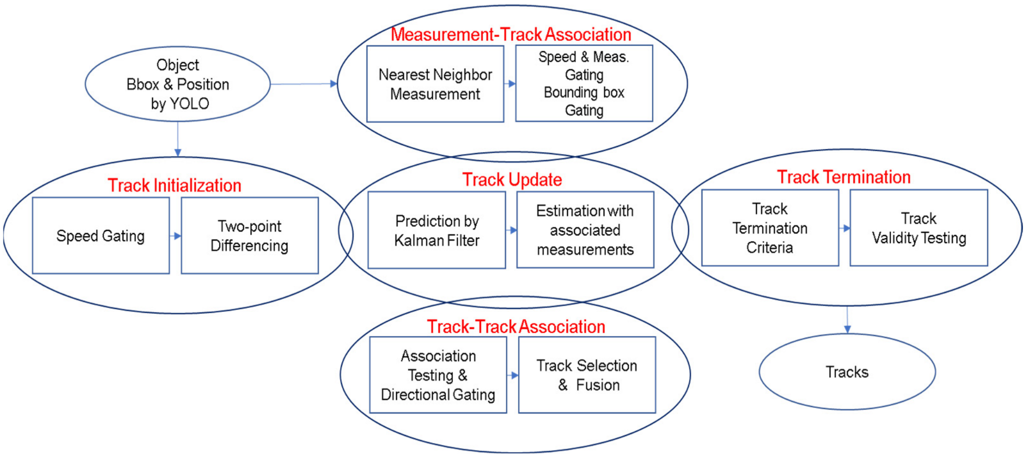

Object
Bbox & Position
by YOLO
Measurement-Track Association
Nearest Neighbor Measurement
Speed & Meas. Gating Bounding box Gating
Track Initialization
Speed Gating
Two-point Differencing
Track Update
Prediction by Kalman Filter
Estimation with associated measurements
Track Termination
Track Termination Criteria
Track Validity Testing
Track-Track Association
Association Testing & Directional Gating
Track Selection & Fusion
Tracks

Figure 4. Block diagram of multiple-target tracking.

3.1. System Modeling

For the nearly constant velocity motion, the process noise following a Gaussian distribution imposes uncertainty on the kinematic state of the target. The discrete state equation of a target is

$$\boldsymbol{x}(k+1) = F(\Delta)\boldsymbol{x}(k) + q(\Delta)\boldsymbol{v}(k), \tag{1}$$

$$F(\Delta) = \begin{bmatrix} 1\Delta00 \\ 0100 \\ 001\Delta \\ 0001 \end{bmatrix}, \quad q(\Delta) = \begin{bmatrix} \Delta^2/2 & 0 \\ \Delta & 0 \\ 0 & \Delta^2/2 \\ 0 & \Delta \end{bmatrix}, \tag{2}$$

where $\boldsymbol{x}(k) = \left[x(k)\ v_x(k)\ y(k)\ v_y(k)\right]^T$ is the state vector of a target at frame k, $x(k)$ and $y(k)$ are positions in the x and y directions, respectively, $v_x(k)$ and $v_y(k)$ are velocities in the x and y directions, respectively, T denotes the matrix transpose, Δ is the sampling time, and $\boldsymbol{v}(k)$ is a process noise vector, which is Gaussian white noise with the covariance matrix

$Q_v = diag\left(\left[\sigma_x^2\ \sigma_y^2\right]\right)$. The measurement equation representing the positions in the x and y directions of the target is

$$\boldsymbol{z}(k) = \begin{bmatrix} z_x(k) \\ z_y(k) \end{bmatrix} = H\boldsymbol{x}(k) + \boldsymbol{w}(k), \tag{3}$$

$$H = \begin{bmatrix} 1000 \\ 0010 \end{bmatrix}, \tag{4}$$

where $\boldsymbol{w}(k)$ is a measurement noise vector, which is Gaussian white noise with the covariance matrix $R = diag\left(\left[r_x^2\ r_y^2\right]\right)$.

3.2. Two-Point Initialization

A track is initialized by two positions in consecutive frames if those two positions are close enough to pass the initial speed gating. The initial state and covariance of track t at frame k are estimated as, respectively,

$$\hat{\boldsymbol{x}}_t(k|k) = \begin{bmatrix} \hat{x}_t(k|k) \\ \hat{v}_{tx}(k|k) \\ \hat{y}_t(k|k) \\ \hat{v}_{ty}(k|k) \end{bmatrix} = \begin{bmatrix} z_{tx}(k) \\ \frac{z_{tx}(k)-z_{tx}(k-1)}{\Delta} \\ z_{ty}(k) \\ \frac{z_{ty}(k)-z_{ty}(k-1)}{\Delta} \end{bmatrix},\ P_t(k|k) = \begin{bmatrix} r_x^2 & \frac{r_x^2}{\Delta} & 0 & 0 \\ \frac{r_x^2}{\Delta} & \frac{2r_x^2}{\Delta^2} & 0 & 0 \\ 0 & 0 & r_y^2 & \frac{r_y^2}{\Delta} \\ 0 & 0 & \frac{r_y^2}{\Delta} & \frac{2r_y^2}{\Delta^2} \end{bmatrix},\ t = 1, \ldots, N(k), \tag{5}$$

where $N(k)$ is the number of established tracks at frame k, which increases by 1 when a track is initialized and decreases by 1 when it terminates. The initial speed is bounded as $\sqrt{\left[\hat{v}_{tx}(k|k)\right]^2 + \left[\hat{v}_{ty}(k|k)\right]^2} \leq V_{max}$, where V_{max} is the maximum initial speed of the target.

3.3. Prediction and Filter Gain

The state and covariance predictions are iteratively computed as

$$\hat{\boldsymbol{x}}_t(k|k-1) = F\hat{\boldsymbol{x}}_t(k-1|k-1), \tag{6}$$

$$P_t(k|k-1) = FP_t(k-1|k-1)F^T + Q, \tag{7}$$

$$Q = q(\Delta)Q_v q(\Delta)^T, \tag{8}$$

where $\hat{\boldsymbol{x}}_t(k|k-1)$ and $P_t(k|k-1)$, respectively, are the state and the covariance prediction of track t at frame k. The residual covariance $S_t(k)$ and the filter gain $W_t(k)$, respectively, are obtained as

$$S_t(k) = HP_t(k|k-1)H^T + R, \tag{9}$$

$$W_t(k) = P_t(k|k-1)H^T S_t(k)^{-1}. \tag{10}$$

3.4. Measurement-to-Track Association with Bounding Box Gating

The measurement association assigns a suitable measurement to the established track in the multi-target and clutter environments. The nearest-neighbor association rule selects the m_{tk}-th measurement, which has the shortest statistical distance to the position prediction as

$$m_{tk} = \underset{m\ =1,\ldots,M(k)}{\operatorname{argmin}} \left\| \boldsymbol{\nu}_{tm}\ (k)^T \left[S_t(k)\right]^{-1} \boldsymbol{\nu}_{tm}(k) \right\|, \tag{11}$$

$$\boldsymbol{\nu}_{tm}(k) = \boldsymbol{z}_m(k) - H\hat{\boldsymbol{x}}_t(k|k-1), \tag{12}$$

where $z_m(k)$ is the m-th measurement vector at frame k, and $M(k)$ is the number of measurements at frame k. The speed and measurement gating are defined as

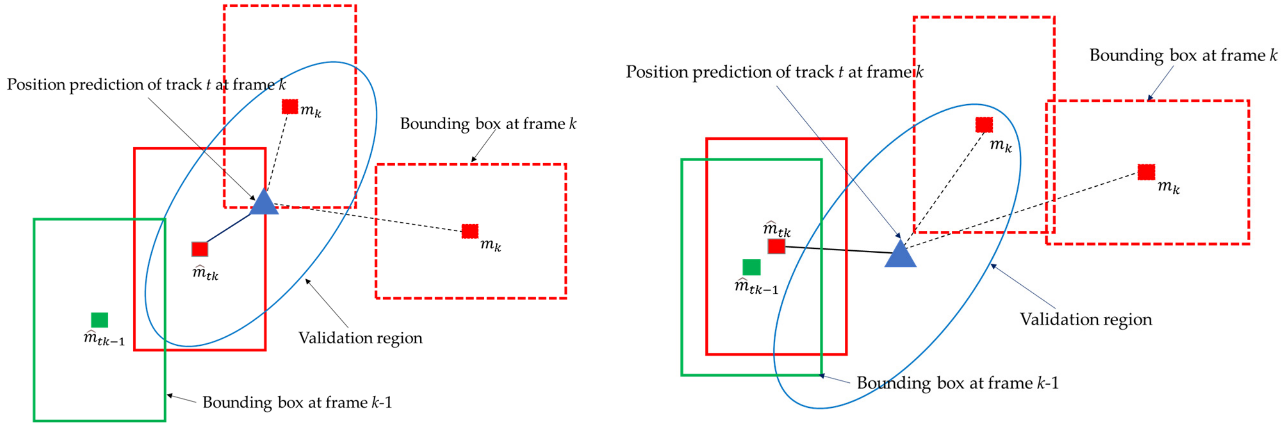
$$\frac{\|z_{m_{tk}}(k) - [\hat{x}_t(k|k) \quad \hat{y}_t(k|k)]^T\|}{\Delta} \le S_{max} \; \& \; \boldsymbol{v}_{tm_{tk}}(k)^T[S_t(k)]^{-1}\boldsymbol{v}_{tm_{tk}}(k) \le \gamma_f, \tag{13}$$

where S_{max} is the maximum target speed, γ_f is a gating size of the measurement validation region, and '&' is the AND operation. The measurement gating is the chi-square hypothesis testing on Gaussian measurement residuals [32].

The bounding box gating is proposed as

$$\text{IoU}(m_{tk}, \hat{m}_{tk-1}) = \frac{|Bbox(m_{tk}) \cap Bbox(\hat{m}_{tk-1})|}{|Bbox(m_{tk}) \cup Bbox(\hat{m}_{tk-1})|} \ge \theta_f, \tag{14}$$

where the $\hat{m}_{tk-1}$-th measurement is already associated with track t at frame $k-1$, Bbox($\cdot$) is the set of pixels in the bounding box obtained by YOLO detection, $|\cdot|$ is the operator that counts the pixels in a set, and θ_f is a threshold for the bounding box gating; if the m_{tk}-th measurement satisfies Equation (14), that is, the IoU of two bounding boxes is equal to or more than the threshold value, the m_{tk}-th measurement is associated with track t at frame k. Figure 4 illustrates the measurement gating and the bounding box gating. In Figure 5a, m_{tk} is located inside the validation region and is associated with track t, whereas in Figure 5b, it is outside the validation region, but its IoU is high enough for the measurement to be associated with track t. As a consequence, the statistically nearest measurement is associated with a track if it passes the speed-measurement gating or the bounding box gating. If the measurements are not associated with any target, they are used for the track initialization at the current or previous frame.


Position prediction of track t at frame k
m_k
Bounding box at frame k
m_k
$\hat{m}_{tk}$
$\hat{m}_{tk-1}$
Validation region
Bounding box at frame k-1
Position prediction of track t at frame k
Bounding box at frame k
m_k
m_k
$\hat{m}_{tk}$
$\hat{m}_{tk-1}$
Validation region
Bounding box at frame k-1

(a) **(b)**

Figure 5. Measurement-target association, (**a**) measurement gating, (**b**) bounding box gating.

3.5. State Estimate and Covariance Update

The state estimate and the covariance are updated after the measurement association as

$$\hat{\mathbf{x}}_t(k|k) = \hat{\mathbf{x}}_t(k|k-1) + W_t(k)\boldsymbol{v}_{t\hat{m}_{tk}}(k), \tag{15}$$

$$P_t(k|k) = P_t(k|k-1) - W_t(k)S_t(k)W_t(k)^T. \tag{16}$$

If there is no measurement associated, they are the same, with predictions of the state and covariance as

$$\hat{\mathbf{x}}_t(k|k) = \hat{\mathbf{x}}_t(k|k-1), \tag{17}$$

$$P_t(k|k) = P_t(k|k-1). \tag{18}$$

3.6. Track-to-Track Association

A track fusion method for a multi-sensor environment has been developed to associate redundant tracks [28]. The directional track association was proposed in [29]. The track association procedure was improved to search the fittest track in [30]. The fittest track for track s is

$$\hat{t} = \underset{t=1,\ldots,N(k),s\neq t}{argmin} [\hat{\boldsymbol{x}}_s(k|k) - \hat{\boldsymbol{x}}_t(k|k)]^T [T_{st}(k)]^{-1} [\hat{\boldsymbol{x}}_s(k|k) - \hat{\boldsymbol{x}}_t(k|k)], \tag{19}$$

$$T_{st}(k) = P_s(k|k) + P_t(k|k) - P_{st}(k|k) - P_{ts}(k|k), \tag{20}$$

$$P_{st}(k|k) = [I - b_s(k)W_s(k)H]\left[FP_{st}(k-1|k-1)F^T + Q\right][I - b_t(k)W_t(k)H], \tag{21}$$

where I is the identity matrix, and $b_s(k)$ and $b_t(k)$ become one when track s or t is associated with a measurement, otherwise, they are zero [32]. The fused covariance in Equation (21) is a linear recursion with the initial condition $P_{st}(0|0)$, which is the square zero matrix. The fittest track is also satisfied with the following track and directional gating as

$$\begin{gathered} [\hat{\boldsymbol{x}}_s(k|k) - \hat{\boldsymbol{x}}_{\hat{t}}(k|k)]^T \left[T_{s\hat{t}}(k)\right]^{-1} [\hat{\boldsymbol{x}}_s(k|k) - \hat{\boldsymbol{x}}_{\hat{t}}(k|k)] \leq \gamma_g \quad \& \\ max\left(\cos^{-1}\frac{|<\hat{\boldsymbol{d}}_{s\hat{t}}(k|k),\, \hat{\boldsymbol{v}}_s(k|k)>|}{\|\hat{\boldsymbol{d}}_{s\hat{t}}(k|k)\|\,\|\hat{\boldsymbol{v}}_s(k|k)\|},\ \cos^{-1}\frac{|<\hat{\boldsymbol{d}}_{s\hat{t}}(k|k),\, \hat{\boldsymbol{v}}_{\hat{t}}(k|k)>|}{\|\hat{\boldsymbol{d}}_{s\hat{t}}(k|k)\|\,\|\hat{\boldsymbol{v}}_{\hat{t}}(k|k)\|}\right) \leq \theta_g \end{gathered} \tag{22}$$

$$\hat{\boldsymbol{d}}_{s\hat{t}}(k|k) = \begin{bmatrix} \hat{x}_{\hat{t}}(k|k) - \hat{x}_s(k|k) \\ \hat{y}_{\hat{t}}(k|k) - \hat{y}_s(k|k) \end{bmatrix}, \hat{\boldsymbol{v}}_s(k|k) = \begin{bmatrix} \hat{v}_{sx}(k|k) \\ \hat{v}_{sy}(k|k) \end{bmatrix}, \tag{23}$$

where γ_g is a gating size of the track validation region, $< \cdot >$ denotes the inner product operation, and θ_g is a threshold for the directional gating. The track gating is the chi-square hypothesis testing. Tracks s and $\hat{t}$ have the error dependencies on each other if they originated from the same target [32]. The directional gating tests the maximum deviation between the displacement vector and the velocity vector.

After the fittest track is selected, the current state of track s is replaced with a fused estimate and covariance if $|P_s(k|k)| \leq |P_{\hat{t}}(k|k)|$ as

$$\hat{\boldsymbol{x}}_s(k|k) = \hat{\boldsymbol{x}}_s(k|k) + [P_s(k|k) - P_{s\hat{t}}(k|k)][P_s(k|k) + P_{\hat{t}}(k|k) - P_{s\hat{t}}(k|k) - P_{\hat{t}s}(k|k)]^{-1}[\hat{\boldsymbol{x}}_{\hat{t}}(k|k) - \hat{\boldsymbol{x}}_s(k|k)], \tag{24}$$

$$P_s(k|k) = P_s(k|k) - [P_s(k|k) - P_{s\hat{t}}(k|k)][P_s(k|k) + P_{\hat{t}}(k|k) - P_{s\hat{t}}(k|k) - P_{\hat{t}s}(k|k)]^{-1}[P_s(k|k) - P_{\hat{t}s}(k|k)]. \tag{25}$$

A more accurate track has less error covariance; thus, fusion only occurs if the determinant of the covariance matrix of track s is less than the determinant of the selected track $\hat{t}$. After track s becomes a fused track, track $\hat{t}$ becomes a potentially terminated track. The potentially terminated track is still eligible to be associated with other tracks that have not yet been fused. The potentially terminated track is terminated when no track remains for the track association.

3.7. Track Termination and Validation Testing

Tracks are terminated by two criteria. One is if a track is selected as a potentially terminated track but not fused during the track association, then the track is terminated. The other is when the number of frames not updated by the measurement exceeds a threshold, the track is terminated.

After track termination, its validity is tested through the track life length. Track life length is defined as the number of frames between the last frame and the initial frame associated with the measurement. If the track life length is less than the minimum track life length, the track is considered invalid and is removed.

3.8. Performance Evaluation

The tracking performance is evaluated in terms of the number of tracks, total track life (TTL), mean track life (MTL), and track purity (TP) [33]. The TTL, MTL, and TP are defined, respectively, as

$$\mathrm{TTL} = \frac{\text{Sum of lengths of tracks which have the same target ID}}{\text{Target life length} - 1}, \tag{26}$$

$$\mathrm{MTL} = \frac{\mathrm{TTL}}{\text{Number of tracks associated in TTL}}, \tag{27}$$

$$\mathrm{TP} = \frac{\text{Number of measurments with the same target ID in a track}}{\text{Number of measurements in the track}}, \tag{28}$$

where target life length is defined as the number of frames between the last frame and the first frame when a measurement of the target appears, and the target ID of a track is defined as the target with the most measurements associated with the track. It is noted that the track length included in the TTL is only the number of frames for which the measurement associated is the same as the target ID of the track and it also includes predictions between updates. If the track is broken or overlaps with one target, the MTL will be less than the TTL. The TP becomes one if there is only one target associated with a track.

Figure 6 illustrates the TTL, MTL, and TP, where three tracks are generated for two targets. The blue and red circles represent the measurements of Targets *t* and *s*, respectively. Mission detections are found at Frames 6 and 8 for Targets *t* and *s*, respectively. The squares, triangles, and crosses represent the initialized or updated states of Tracks 1, 2, and 3, respectively. The same color of the target and the track indicates that they are related by the measurement-track association. The empty triangles in Track 2 indicate the predictions between the updated states. The Target ID of Tracks 1 and 2 is *t*, and Target ID of Track 3 is *s*. There is a breakage between Tracks 1 and 2 for Target *t*, and an intersection between Tracks 2 and 3 for Targets *t* and *s*. The TTL of Targets *t* and *s* are 7/9 and 4/7, respectively. The MTL of Targets *t* and *s* are 3.5/9 and 4/7, respectively, since the number of tracks for Targets *t* and *s* is 2 and 1, respectively. The TP of Tracks 1, 2, and 3 are 1, 3/5, and 2/3, respectively.

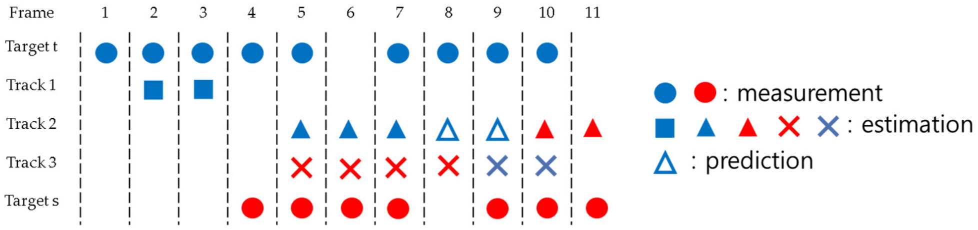


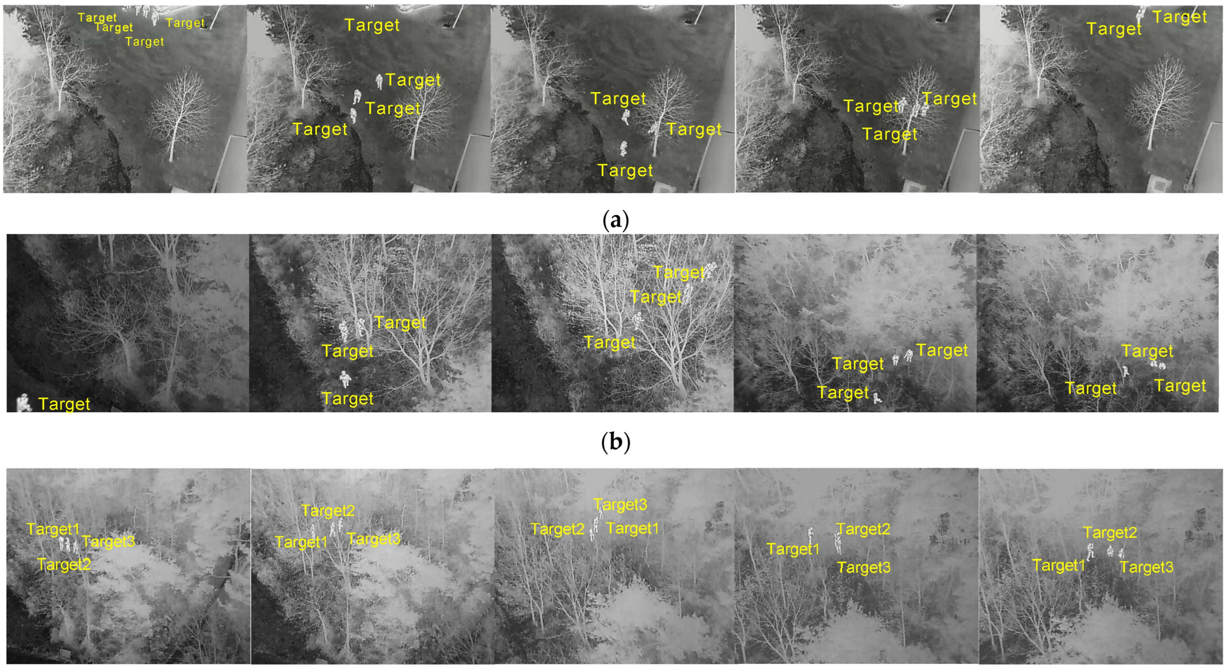
Figure 6. An illustration of three tracks generated for two targets.

4. Results

The experimental results are presented with video description, parameter settings, YOLO object detection, and multiple-target tracking.

4.1. Video Description

An infrared thermal imaging camera captured three videos (Videos 1–3). The camera, FILR Vue Pro R640 (f = 19 mm, FOV = 32° × 26°, spectrum band: 7.5–13.5 µm, pixel pitch: 17 µm), is mounted on a DJI Inspire 2 drone. The resolution of the image is 640 × 512 pixels, and the frame rate is set to 30 fps. The videos were shot in the mountains on a winter night with no ambient light. In these environments, the visible image is completely dark. The thermal video capturing was assumed to be under harsh circumstances; the drone flies in any direction and altitude slowly or swiftly, and the perspective of the drone is arbitrary. Image characteristics vary from image to image depending on weather conditions and surrounding objects. In Video 1, three hikers walk around the mountain for 60 s. In Videos 2 and 3, three hikers are simulated as missing in the mountains for 120 s. Figure 7a shows the 1st, 501th, 901th, 1301th, and 1701th frame of Video 1, and Figure 7b,c show the 1st, 901th, 1801th, 2701th, 3601th frame of Videos 2 and 3, respectively. In each frame, a target ID, numbered 1 to 3 or 4, was displayed close to the object. It is assumed that there are four people in Video 1 and one of them appears with his legs at the top of the frame.



(**a**)

(**b**)

(**c**)

Figure 7. Sample frames with Target ID's of (**a**) Video 1, (**b**) Video 2, (**c**) Video 3.

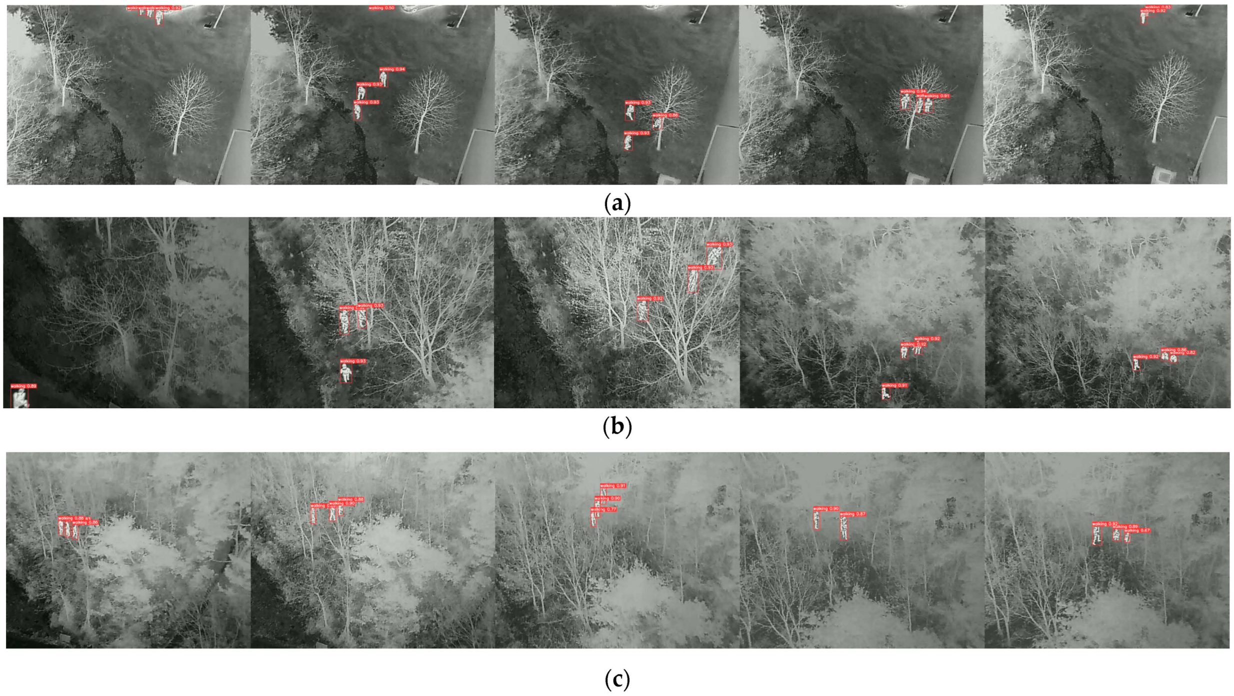
4.2. People Detection by YOLOv5

YOLOv5 detects people in every frame of Videos 1–3. The detection results are summarized in Table 1. For YOLOv5l, the number of detections in Video 1 is more than the number of true instances, mostly due to false alarms generated on the car at the top of the frame. For YOLOV5x, the number of detections in Video 2 is more than the number of instances. This is because more than one bounding box was generated for one object. The recall of YOLOv5x is calculated lower than YOLOvl for Video 1 because YOLOv5x does not always detect the small part of the legs that appear at the top of the frame.

Table 1. YOLOv5 detection results.

		Video 1	Video 2	Video 3
Descriptions	Num. of frames (duration)	1801 (1 min)	3601 (2 min)	
	Num. of objects (people)	4	3	
	Num. of instances	6000	10,652	10,800
YOLOv5s	Num. of detections	5760	10,492	9143
	Num. of detections over 0.5 conf. level	5176	10,283	8425
	Recall over 0.5 conf. level	0.863	0.965	0.780
YOLOv5l	Num. of detections	6634	10,610	9693
	Num. of detections over 0.5 conf. level	5811	10,474	9248
	Recall over 0.5 conf. level	0.969	0.983	0.856
YOLOv5x	Num. of detections	5734	10,699	9947
	Num. of detections over 0.5 conf. level	5406	10,567	9676
	Recall over 0.5 conf. level	0.901	0.992	0.862

The detection of YOLOv5x with a minimum confidence level of 0.5 was used for the target tracking as it provides the most accurate detections without generating false alarms. For YOLOv5x, the recall of Videos 1 to 3 at a minimum confidence level of 0.5 is 0.901, 0.992, and 0.862, respectively, and the precision is 1 for all videos. In Video 3, harsh conditions, such as close and occluding objects and rapid movement of the drone, degrade the detection performance. Figure 8 shows YOLOv5x detection with a minimum confidence level of 0.5 for the sample frames. The centroids of the object in all frames are shown in Figures 9 and 10 in the first frame and on a white background, respectively. Supplementary Videos S2–S4 show the YOLOv5x object detection results of Videos 1, 2, and 3, respectively.

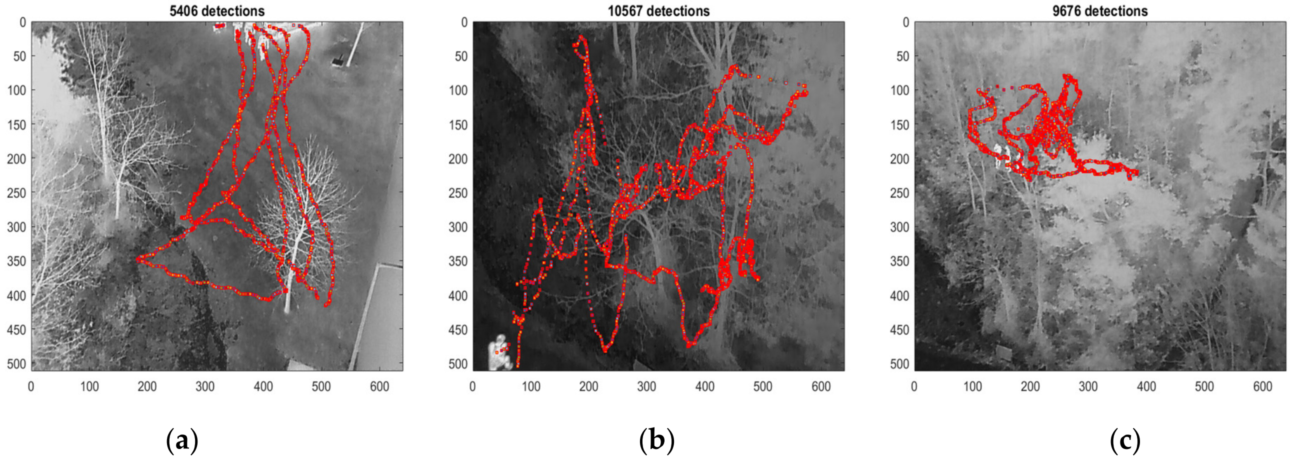


(**a**)

(**b**)

(**c**)

Figure 8. Detections of the sample frames, (**a**) Video 1, (**b**) Video 2, (**c**) Video 3.

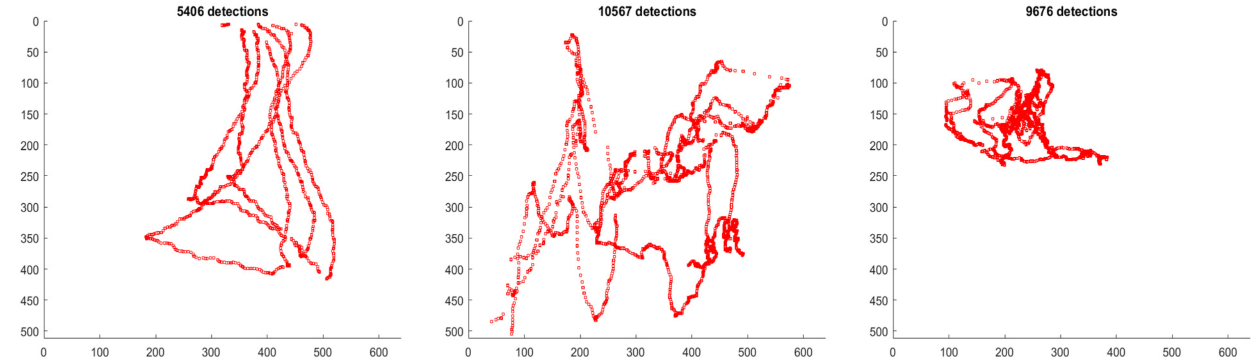


(**a**)

(**b**)

(**c**)

Figure 9. Detections in the 1st frame of (**a**) Video 1: 5406 detections, (**b**) Video 2: 10,567 detections, (**c**) Video 3: 9676 detections.



(**a**) (**b**) (**c**)

Figure 10. Detections only, (**a**) Video 1: 5406 detections, (**b**) Video 2: 10,567 detections, (**c**) Video 3: 9676 detections.

4.3. Multiple-Target Tracking

4.3.1. Parameter Set-Up

The target-tracking parameters are shown in Table 2. Video 1 was processed every two frames, thus the sampling time is 1/15 s for Video 1, whereas it is 1/30 s for Videos 2 and 3. The one-pixel coordinate is scaled to 0.03 m, 0.04 m, and 0.05 m for Videos 1, 2, and 3, respectively. The maximum initial speed of the target is set to 3 m/s. The process noise standard deviation of the Kalman filter is set to 0.5 m/s^2 for Videos 1 and 2.5 m/s^2 for Videos 2 and 3. The process noise standard deviation is determined considering the acceleration of the target and the rapid movement of the drone. The measurement noise standard deviation is set to 0.5 m for all videos. The maximum target speed for the track maintenance is set to 10 m/s for Videos 1 and 3 and 20 m/s for Video 2. The threshold for measurement gating is set to four and the minimum IoU for the bounding box gating is set to 0.6 for Videos 1 and 2 and 0.4 for Video 3. The gate threshold and angular threshold for the track association are set at 10 and 45° for Videos 1 and 3, respectively, and 20 and 60° for Video 2, respectively. They are set up when better results are produced. For the track termination, the maximum number of searches is set at 20 frames, and tracks shorter than 2 s are removed as invalid.

Table 2. Target tracking parameters.

Parameters (Unit)		**Video 1**	**Video 2**	**Video 3**
Sampling Time (second)		1/15	1/30	
Max. initial target speed, V_{max} (m/s)		3		
Process noise std.	$\sigma_x = \sigma_y$ (m/s^2)	0.5	2.5	
Measurement noise std.	$r_x = r_y$ (m)	0.5		
Measurent association	Max. target speed, S_{max} (m/s)	10	20	10
	Gate threshold, γ_m	4		
	Bbox threshold, b_m	0.6		0.4
Track association	Gate threshold, γ_t	10	20	10
	Angular threshold, θ_t (degree)	45°	60°	45°
Track termination	Maximum searching number (frame)	20		
Min. track life length for track validity (second)		2		

4.3.2. Target Tracking Results

The target tracking results for Videos 1 to 3, including the number of valid tracks, average TTL, average MTL, and average TP, are shown in Tables 3–5, respectively. Case 1, the first column of the table, is the result of applying both the bounding box gating and the track association. Case 2, the second column, is the result of applying the track association without the bounding box gating, and Case 3, the third column, is the result of applying neither the track association nor the bounding box gating. It is noted that Target 4 at the top of Video 1 was excluded from evaluating the tracking performance because it is a small part of a person's foot that does not move. For Video 1, the tracking results of Case 3 are perfect. For Videos 2 and 3, the results of Case 1 are better than others. For Case 1, the average TTL of Videos 1 to 3 are obtained as 0.987, 0.993, and 0.894, respectively. The corresponding average MTLs are 0.987, 0.442, and 0.151, respectively. The average TPs are obtained as 1, 0.999, and 0.995, respectively, for the three videos. The average MTL is reduced for Videos 2 and 3 due to the track breakage caused by the high maneuvering of the drone or object occlusion.

Table 3. Tracking results of Video 1.

	Case 1	**Case 2**	**Case 3**
Num. of Tracks	3	3	3
Avg. TTL	0.987	0.982	1
Avg. MTL	0.987	0.982	1
Avg. TP	1	0.991	1

Table 4. Tracking results of Video 2.

	Case 1	**Case 2**	**Case 3**
Num. of Tracks	7	7	18
Avg. TTL	0.993	0.945	0.943
Avg. MTL	0.442	0.417	0.193
Avg. TP	0.999	0.954	0.963

Table 5. Tracking results of Video 3.

	Case 1	**Case 2**	**Case 3**
Num. of Tracks	20	22	29
Avg. TTL	0.894	0.878	0.890
Avg. MTL	0.151	0.130	0.097
Avg. TP	0.995	0.995	0.986

Figures 11 and 12 show the tracks for all frames in random colors. The tracks are shown in the first frame in Figure 11 and on a white background in Figure 12. Supplementary Videos S5–S7 show the tracking results for Case 1 for Videos 1, 2, and 3, respectively.

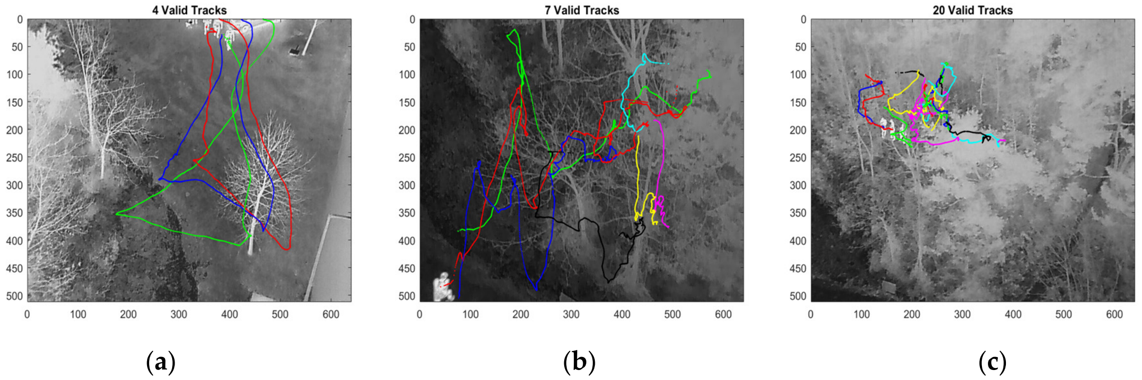


Figure 11. Tracks in the 1st frame, (**a**) Video 1: 4 valid tracks, (**b**) Video 2: 7 valid tracks, (**c**) Video 3: 20 valid tracks.

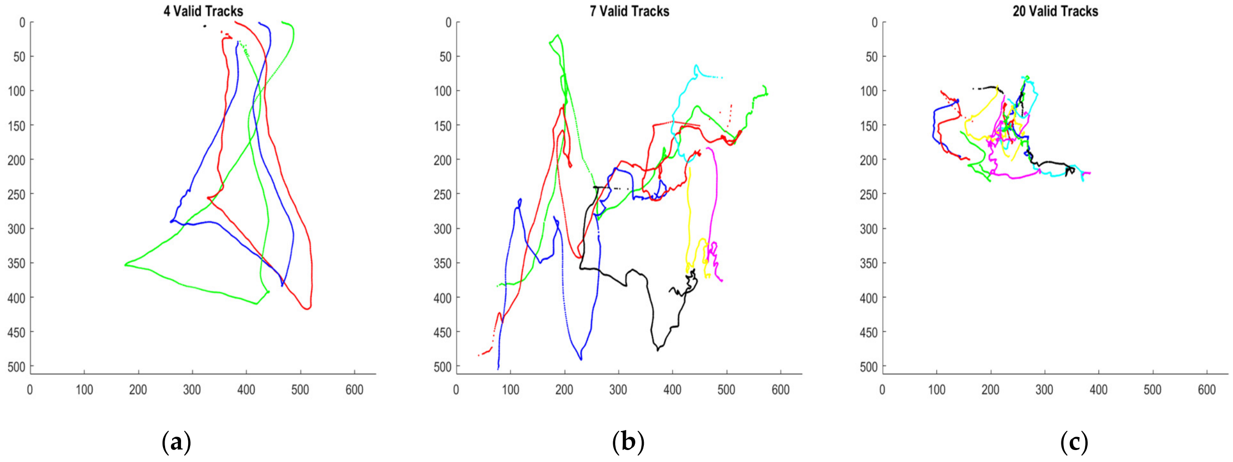


Figure 12. Tracks only, (**a**) Video 1: 4 valid tracks, (**b**) Video 2: 7 valid tracks, (**c**) Video 3: 20 valid tracks.

Seven Supplementary Multimedia Files (MP4 format) are available online. Supplementary Materials Video S1 is a video of YOLO training images. Training instances are indicated by blue rectangles. Supplementary Materials Videos S2–S4 are the YOLOv5x detection results with a minimum confidence level of 0.5 for Videos 1 to 3, respectively. The videos display bounding boxes, including the class name and confidence level. Supplementary Materials Videos S5–S7 are the tracking results from Videos 1 to 3 applying both the bounding box gating and the track association, respectively. The bounding box and its centroid of YOLOv5x are shown as red squares. Position estimates are shown as blue circles. Valid tracks were numbered in the order they were created.

5. Discussion

The thermal videos were captured under extremely challenging conditions, simulating people needing search and rescue missions in non-visible environments. The experimental scenarios included overgrown terrain, in which the warm objects were either partially visible or invisible. The objects are often occluded by other people and natural objects, such as trees, bushes, and foliage. The drone is manually operated, allowing rapid movements.

Since the more accurate the location information, the better the tracking quality, the detections of YOLOv5x with a minimum confidence level of 0.5 were utilized. The detection accuracy is expected to increase with larger training data since less than 200 images were trained in this paper.

TTL and MTL are defined on the target and TP is defined on the track. High average TTLs and average TPs are achieved for all videos. However, the average MTLs were lowered for Videos 2 and 3 as more than one valid track was generated from the target. The number of valid tracks is increased due to the track breakage caused by missing detection. The missing detection is mainly caused by the high maneuvering of the drone or the object occlusion. One track intersection occurs in both Videos 2 and 3, resulting in the slightly lower average TP. The track breakage and intersection both lower the average TTL. The bounding box gating has improved tracking performance for all metrics.

Low-resolution and gray-scaled infrared thermal frames can be transmitted to a ground station with less bandwidth. Adopting a lighter and later version of the YOLO model, such as YOLOv8 [34], is also considered to increase implementation feasibility, as well as detection performance.

6. Conclusions

In the paper, multiple-target tracking using thermal imaging was studied for the purpose of search and rescue missions with a drone. The object-detection multiple-target tracking scheme has been shown to be very powerful for tracking people in thermal videos acquired by drones. The harsh conditions of simulated search and rescue missions, including (1) no ambient lighting environment, (2) complex backgrounds, (3) closely located and heavily occluded targets, and (4) arbitrary moving platforms, can be overcome with the proposed solution.

To evaluate tracking performance, three metrics TTL, MTL, and TP were obtained. TTL and TP provide solid performance, but MTL decreases when tracks are broken. The proposed solution is direct and simple, yet quite effective. It is also suitable for security and surveillance in civil and military fields and wildlife monitoring. It can also be applied to pedestrian tracking in crowds and object tracking in sports analysis. Track segment association to increase the track continuity remains for future studies. Adopting higher iterations of the YOLO model for detection performance and feasible implementation also remains for future work.

Supplementary Materials: The following are available online at https://zenodo.org/records/10467489 accessed on 26 January 2024, Video S1: YOLOv5x training data, Video S2: Video 1 YOLOv5x detection, Video S3: Video 2 YOLOv5x detection, Video S4: Video 3 YOLOv5x detection. Videos S2–S4 are the detection results with 0.5 minimum confidence level. Video S5: Video 1 target tracking, Video S6: Video 2 target tracking, Video S7: Video 3 target tracking. Videos S5–S7 are the tracking results with the bounding box gating and the track association.

Funding: This research was supported by Daegu University Research Grant, 2023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Conflicts of Interest: The author declares no conflict of interest.

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. Vollmer, M.; Mollmann, K.-P. *Infrared Thermal Imaging: Fundamentals, Research and Applications*; Wiley-VCH: Weinheim, Germany, 2010.
3. Kirk, J.; Havens Edward, J. *Sharp, Thermal Imaging Techniques to Survey and Monitor Animals in the Wild*; Academic Press: Cambridge, MA, USA, 2016; ISBN 9780128033845. [CrossRef]
4. Rudol, P.; Doherty, P. Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; pp. 1–8. [CrossRef]
5. Jamjoum, M.; Siouf, S.; Alzubi, S.; AbdelSalam, E.; Almomani, F.; Salameh, T.; Al Swailmeen, Y. DRONA: A Novel Design of a Drone for Search and Rescue Operations. In Proceedings of the 2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 20–23 February 2023; pp. 1–5. [CrossRef]
6. Thermal Camera-Equipped UAVs Spot Hard-to-Find Subjects. Available online: https://www.photonics.com/Articles/Thermal_Camera-Equipped_UAVs_Spot_Hard-to-Find/a63435 (accessed on 30 January 2024).
7. Gonzalez, L.F.; Montes, H.G.A.; Puig, E.; Johnson, S.; Mengersen, K.; Gaston, K.J. Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors* **2016**, *16*, 97. [CrossRef] [PubMed]
8. Messina, G.; Modica, G. Applications of UAV Thermal Imagery in Precision Agriculture: State of the Art and Future Research Outlook. *Remote Sens.* **2020**, *12*, 1491. [CrossRef]
9. Krišto, M.; Ivasic-Kos, M.; Pobar, M. Thermal Object Detection in Difficult Weather Conditions Using YOLO. *IEEE Access* **2020**, *8*, 25459–125476. [CrossRef]
10. Jiang, C.; Ren, H.; Ye, X.; Zhu, J.; Zeng, H.; Nan, Y.; Sun, M.; Ren, X.; Huo, H. Object detection from UAV thermal infrared images and videos using YOLO models. *J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102912. [CrossRef]
11. Kannadaguli, P. YOLO v4 Based Human Detection System Using Aerial Thermal Imaging for UAV Based Surveillance Applications. In Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 1213–1219. [CrossRef]
12. Levin, E.; Zarnowski, A.; McCarty, J.L.; Bialas, J.; Banaszek, A.; Banaszek, S. Feasibility Study of Inexpensive Thermal Sensor and Small UAS Deployment for Living Human Detection in Rescue Missions Application Scenario. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B8, 2016 XXIII ISPRS Congress, Prague, Czech Republic, 12–19 July 2016.

13. Teutsch, M.; Mueller, T.; Huber, M.; Beyerer, J. Low Resolution Person Detection with a Moving Thermal Infrared Camera by Hot Spot Classification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Work-Shops, Columbus, OH, USA, 23–28 June 2014; pp. 209–216. [CrossRef]
14. Giitsidis, T.; Karakasis, E.G.; Gasteratos, A.; Sirakoulis, G.C. Human and Fire Detection from High Altitude UAV Images. In Proceedings of the 2015 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, Turku, Finland, 4–6 March 2015; pp. 309–315. [CrossRef]
15. Yeom, S. Moving People Tracking and False Track Removing with Infrared Thermal Imaging by a Multirotor. *Drones* **2021**, *5*, 65. [CrossRef]
16. Leira, F.S.; Helgensen, H.H.; Johansen, T.A.; Fossen, T.I. Object detection, recognition, and tracking from UAVs using a thermal camera. *J. Field Robot.* **2021**, *38*, 242–267. [CrossRef]
17. Helgesen, H.H.; Leira, F.S.; Johansen, T.A. Colored-Noise Tracking of Floating Objects using UAVs with Thermal Cameras. In Proceedings of the 2019 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 11–14 June 2019; pp. 651–660. [CrossRef]
18. Davis, J.W.; Sharma, V. Background-Subtraction in Thermal Imagery Using Contour Saliency. *Int. J. Comput. Vis.* **2007**, *71*, 161–181. [CrossRef]
19. Soundrapandiyan, R. Adaptive Pedestrian Detection in Infrared Images Using Background Subtraction and Local Thresh-olding. *Procedia Comput. Sci.* **2015**, *58*, 706–713. [CrossRef]
20. Portmann, J.; Lynen, S.; Chli, M.; Siegwart, R. People detection and tracking from aerial thermal views. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1794–1800. [CrossRef]
21. He, Y.-J.; Li, M.; Zhang, J.; Yao, J.-P. Infrared target tracking via weighted correlation filter. *Infrared Phys. Technol.* **2015**, *73*, 103–114. [CrossRef]
22. Yu, T.; Mo, B.; Liu, F.; Qi, H.; Liu, Y. Robust thermal infrared object tracking with continuous correlation filters and adaptive feature fusion. *Infrared Phys. Technol.* **2019**, *98*, 69–81. [CrossRef]
23. Yuan, D.; Shu, X.; Liu, Q.; Zhang, X.; He, Z. Robust thermal infrared tracking via an adaptively multi-feature fusion model. *Neural Comput. Appl.* **2022**, *35*, 3423–3434. [CrossRef] [PubMed]
24. Gade, R.; Moeslund, T.B. Thermal Tracking of Sports Players. *Sensors* **2014**, *14*, 13679–13691. [CrossRef] [PubMed]
25. WEl Ahmar, A.; Kolhatkar, D.; Nowruzi, F.E.; AlGhamdi, H.; Hou, J.; Laganiere, R. Multiple Object Detection and Tracking in the Thermal Spectrum. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–24 June 2022; pp. 276–284. [CrossRef]
26. Liu, Q.; Lu, X.; He, Z.; Zhang, C.; Chen, W.-S. Deep convolutional neural networks for thermal infrared object tracking. *Knowl.-Based Syst.* **2017**, *134*, 189–198. [CrossRef]
27. Available online: https://github.com/ultralytics/yolov5 (accessed on 30 January 2024).
28. Yeom, S.; Nam, D.-H. Moving Vehicle Tracking with a Moving Drone Based on Track Association. *Appl. Sci.* **2021**, *11*, 4046. [CrossRef]
29. Yeom, S. Long Distance Moving Vehicle Tracking with a Multirotor Based on IMM-Directional Track Association. *Appl. Sci.* **2021**, *11*, 11234. [CrossRef]
30. Yeom, S. Long Distance Ground Target Tracking with Aerial Image-to-Position Conversion and Improved Track Association. *Drones* **2022**, *6*, 55. [CrossRef]
31. Available online: https://github.com/HumanSignal/labelImg (accessed on 30 January 2024).
32. Bar-Shalom, Y.; Li, X.R. *Multitarget-Multisensor Tracking: Principles and Techniques*; YBS Publishing: Storrs, CT, USA, 1995.
33. 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]
34. Available online: https://github.com/ultralytics/ultralytics (accessed on 30 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.