sensors

MDPI

Article

Detection and Tracking of Moving Targets for Thermal Infrared Video Sequences

Chenming Li and Wenguang Wang *

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China; lchm1990@163.com
* Correspondence: wwenguang@buaa.edu.cn; Tel.: +86-10-8231-7240

Received: 26 September 2018; Accepted: 12 November 2018; Published: 14 November 2018

Abstract: The joint detection and tracking of multiple targets from raw thermal infrared (TIR) image observations plays a significant role in the video surveillance field, and it has extensive applied foreground and practical value. In this paper, a novel multiple-target track-before-detect (TBD) method, which is based on background subtraction within the framework of labeled random finite sets (RFS) is presented. First, a background subtraction method based on a random selection strategy is exploited to obtain the foreground probability map from a TIR sequence. Second, in the foreground probability map, the probability of each pixel belonging to a target is calculated by non-overlapping multi-target likelihood. Finally, a δ generalized labeled multi-Bernoulli (δ-GLMB) filter is employed to produce the states of multi-target along with their labels. Unlike other RFS-based filters, the proposed approach describes the target state by a pixel set instead of a single point. To meet the requirement of factual application, some extra procedures, including pixel sampling and update, target merging and splitting, and new birth target initialization, are incorporated into the algorithm. The experimental results show that the proposed method performs better in multi-target detection than six compared methods. Also, the method is effective for the continuous tracking of multi-targets.

Keywords: joint detection and tracking of multi-target; thermal infrared (TIR) image; track-before-detect (TBD); background subtraction; labeled random finite sets (RFS); δ-GLMB filter

1. Introduction

The detection and tracking of moving targets is a challenging vision task that has attracted extensive research. Because of the comparatively lower cost, omnipresence, and 24 $\times$ 7 applicability, thermal infrared (TIR) sensors have provided new application areas [1]. Since pedestrians are the major participants in many events of interest, the joint detection and tracking of multi-targets (usually meaning pedestrians, but not exclusively in this paper) becomes one primary task borne by the TIR surveillance system [2]. The main advantages of thermal sensors are their ability to see in complete darkness, their robustness to illumination changes and shadow effects, and their comparatively lower degree of intrusion regarding privacy. Despite many superiorities, the main disadvantages of TIR imaging include low resolution, many dead pixels, lack of color information, low foreground/background contrast, and associated heavy noise [2,3]. As well as the above disadvantages, in top-down surveillance scenes, the target usually occupies fewer pixels and it is difficult to extract the effective appearance features, including textural and contour information. Moreover, the targets in surveillance scenes are highly variable in pose, size, shape, and intensity [2]. Multiple moving-target detection and tracking remain crucial objectives and are identified as the key issues in the TIR surveillance system.

Usually, detection and tracking tasks are separable in computer vision, and most multi-target tracking objectives require a detection operation to produce measurements [4]. Emerging technologies, such as proposal detection method and deep convolution neural network method, usually cannot

achieve as admirable detection performance in the TIR surveillance images as the optical images due to the above-mentioned shortcomings. The traditional method still plays a significant role in moving-target detection. In addition, the detection methods in video streams can be divided into three categories: frame difference, optical flow, and background subtraction methods [5]. Because of its low computational cost and high accuracy, the background subtraction method is more popular [6,7]. Unfortunately, the classic background subtraction method has two debilitating drawbacks. First, it achieves detection tasks based on a threshold, which results in information loss [8]. During detection, information loss can significantly degrade tracking performance, especially in obscure feature cases [4,9]. Second, the detection foreground pixels from different targets are not discriminable. To track different targets individually, some post-processes must be executed to differentiate them. Therefore, the joint processing of detection and tracking tasks is of fundamental interest to reduce information loss and simplify the process. Some typical algorithms such as dynamic programming [9,10], Bayesian existence process [11], and multi-modal distributions [12] have shown great success. Among them, the random finite set (RFS) framework approaches that jointly detect and track have attracted significant attention [4,13,14].

The RFS-based methods consist of two procedures in a Bayesian framework: prediction and update. Two methods can be used to implement them, one based on the Gaussian mixtures (GM) model and the other based on the sequential Monte Carlo (SMC) model; the latter is also known as particle implementation. The RFS-based filters can be divided into unlabeled filters and labeled filters. The probability hypothesis density (PHD), cardinalized PHD (CPHD), and multi-target multi-Bernoulli (MeMBer) filters are the typical unlabeled category [15–17]. Although these filters have been successfully applied to visual tracking [4,18,19], they provide only unidentified estimates and require additional post-processing to form tracks. The generalized labeled multi-Bernoulli (GLMB) filter and δ-GLMB filter belonging to the label class of RFS-based filters can distinguish and maintain different tracks by adding a label to each target [20,21]. Based on the labeled RFS filter, a δ-GLMB track-before-detect (TBD) approach with a separable likelihood function was introduced in a radar-tracking scenario in [22]. Subsequently, an improved GLMB-TBD algorithm, which can handle non-separable likelihood situations, was proposed in [23] for generic measurement models, although the considerable computational cost limits its application. Currently, the labeled RFS-based methods focus on non-overlapping targets and point target tracking [14,18,22].

To alleviate the above problem in multi-target detection and tracking of TIR surveillance system, a joint detection and tracking approach based on particle implementation, which combines a background subtraction method with a GLMB filter, is proposed. First, according to the ViBe algorithm [24], a random selection strategy background subtraction method without threshold detection is designed to yield a TIR foreground probability map. Second, a multi-target likelihood function is used to calculate the probability of each pixel belonging to a target in the foreground probability map. Finally, the δ-GLMB-TBD filter is applied to produce the state of the multi-target along with their labels. Most RFS-based TBD methods describe a target in the image as a rectangle or a single point [14,18], which is too rough for a target with complex contour and may degrade the detection and tracking performance. In the proposed method, the target is represented by its own pixel set. This means that the δ-GLMB-TBD filter is extended to track irregular areas of a target by benefiting from shape similarity in consecutive frames. To be practical, the algorithm also includes some extra procedures, such as pixel sampling and update, target merging and splitting, and new birth target initialization, which accommodate target deformation and overlapped and dynamic change via gathering, splitting, birth, and death.

The main contribution of this paper is the proposal of a joint multi-target detection and tracking method based on background subtraction and a δ-GLMB-TBD filter for infrared surveillance system, along with its particle implementation. As well as producing a multi-target state estimate, the proposed method can track the multi-targets successfully and individually keep their labels. Based on the proposed method, we also developed the following:

- More effective multi-target estimates which are from a pixel set instead of a rectangle or a single point;
- Several procedures to accommodate target deformation and multi-target dynamic processes, such as pixel sampling and update, target merging, splitting, and new target initialization;
- A random selection strategy background subtraction method which can be used to pre-process the images without threshold segmentation.

This paper is organized as follows. Section 2 describes the proposed algorithm in detail, including the background subtraction method, multi-target likelihood calculation, and the recursion of the δ-GLMB-TBD filter. The results and analysis of the experiments are presented in Section 3. Conclusions are drawn in Section 4.

2. Background-Subtraction-Based δ-GLMB-TBD Filter

In this section, the background-subtraction-based δ-GLMB-TBD filter is introduced in three major parts: background subtraction, multi-target likelihood function calculation, and implementation of the δ-GLMB-TBD filter [22]. A block diagram is presented in Figure 1. Background subtraction transforms the original TIR image to a foreground probability map where each pixel can be interpreted as the probability of the pixel belonging to the foreground. Then the map is used to generate new birth targets and to calculate the multi-target likelihood. Finally, the δ-GLMB-TBD is used to produce multi-target estimates. To accommodate target deformation and dynamic change, some extra procedures, such as splitting and merging, are included.

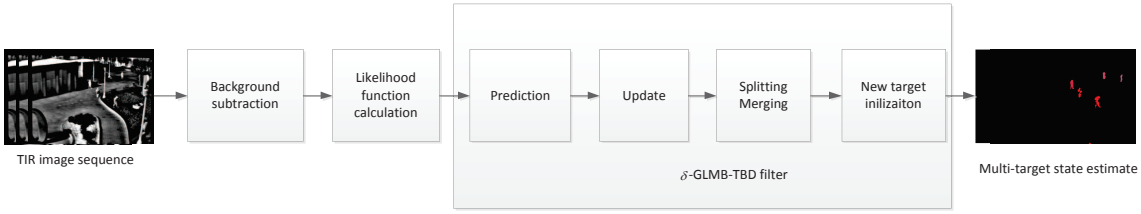


Figure 1. Schematic of the proposed method.

2.1. Background Subtraction Method for TIR

Figure 2 shows three typical TIR surveillance scenes, in which it is difficult to detect all targets because of low foreground/background contrast, fewer pixels, heavy noise, and lack of textural and contour information. One popular method is background subtraction. However, background subtraction usually produces two segmentations that denote background and foreground by threshold detection, which results in information loss and then leads to inferior tracking performance. To alleviate these problems, the GLMB-TBD filter without the need to detect targets is employed to achieve the joint detection and tracking of multi-target on a background suppression image. Many methods have the potential to subtract the background, such as ViBe [24], KDE [25], GMM [26] and so on. In this paper, we propose a random selection strategy based on ViBe algorithm [24] to subtract the background, which is because the ViBe algorithm has the characters of low complexity, high stability, and excellent background subtraction effect [27]. The new background subtraction method consists of two parts: background model initialization and background model update. The differences between the new approach and the ViBe method are as follows: (1) less missing detection without threshold segmentation; (2) pixel background model update with respect to the probability of being a foreground pixel; (3) morphology operations to eliminate scattered noise and maintain shape. The details are introduced in the following sections.

(a) (b) (c)

Figure 2. Three typical TIR images. (**a**) Campus; (**b**) Parking lot; (**c**) Community.

2.1.1. ViBe Method

Before discussing the proposed method, we will first summarize the standard ViBe method. The background subtraction is regarded as a classification problem in the ViBe method. However, as there is no way to model each background pixel as a probability density function (PDF), then, no estimation or classification result could be given directly from the PDF. Therefore, the ViBe method establishes a background pixel set for each pixel, and each element in this set can be seen as a sample obtained by the true PDF of the background [24].

The key problems of the ViBe method are: (1) how to get the background pixel set, effectively; and (2) how to classify a pixel as a background or foreground according to its given background pixel set. For the first problem, a random selection strategy-based method is employed to update the background pixel set. This can make the samples be more compliant to the pdf of the background without increasing the number of the samples and discarding the earlier samples. For the second problem, the new pixel should compare with its background pixel set. The steps are listed as follows.

Step 1: Calculate the Euclidean distances between the new sample and each sample in its background pixel set;

Step 2: Obtain the number of the Euclidean distances shorter than a given threshold;

Step 3: Compare the number with another given threshold; if the number is greater than the threshold, then the new pixel is classified as a background pixel and vice versa.

2.1.2. Background Model Initialization

Regarding the proposed method, the background model initialization will be discussed first. The proposed method only employs the first frame to initialize the background model. Let y_i denote the ith pixel value in the original TIR image I_{ori} ($1 \leq i \leq N_{img}$, where N_{img} denotes the number of all pixels in the TIR image), and $M_i(k)$ denotes the background model of pixel i at time k (in the background model initialization, set $k = 1$); all $M_i(k)$ make up the image background model $M(k)$. Each $M_i(k)$ is a collection of N background samples.

$$M_i(k) = \{m_{i,1}, m_{i,2}, ..., m_{i,N}\}. \quad (1)$$

where $m_{i,n}$ ($1 \leq n \leq N$) is a sample and initialized as follows:

$$m_{i,n} = y_i + v_{ran}(v_l, v_h) \quad (2)$$

where $v_{ran}(v_l, v_h)$ denotes a uniform random number between v_l and v_h. The main parameter in initial model is: N.

2.1.3. Background Model Update

The background model will be updated with each new frame. The update step is the core procedure used to yield accurate results over time. In this step, a conservative update policy is used.

For each pixel i in frame I_{ori} at time k ($k > 1$), its equivalent background is the average of its background model $M_i(k)$ given by

$$y_{i,equ} = \frac{1}{N}(m_{i,1} + m_{i,2} + ... + m_{i,N}). \tag{3}$$

We compare the absolute difference between its current value y_i and its equivalent $y_{i.equ}$ with a threshold th_{diff}. The comparison result s_i can be obtained by

$$s_i = \begin{cases} abs(y_i - y_{i,equ})/th_{diff} & if \;\; abs(y_i - y_{i,equ}) < th_{diff} \\ 1 & if \;\; abs(y_i - y_{i,equ}) \geq th_{diff} \end{cases} \tag{4}$$

where $abs(\cdot)$ denotes the "absolute" operation. s_i = 1 means the pixel i is classified as foreground. According to a conservative update policy, a foreground pixel should never be used to update the background. Thus, only when $s_i < 1$ can the current pixel value y_i replace one sample in the background model $M_i(k-1)$ with the probability of $(1 - s_i)$. So th_{diff} plays a significant role in updating the background model. When th_{diff} is too small, s_i is sensitive to the pixel change and noise; when th_{diff} is too large, the background model may be polluted by the pixel belonging to moving target. In contrast with the first-in-first-out strategy, the sample substituted by y_i is chosen randomly by a uniform probability density function. These operations can extend the time windows covered by the background models. The long lifespan of the background samples significantly aids in the detection of slow-moving multi-targets. The conservative update policy can make a sharp detection of a moving target without introducing the foreground.

Unfortunately, one disadvantage of the conservative update policy is that it can lead to deadlock situations and ghosts. To eliminate these influences, an improved "detection support map" method [28], which counts the number of times a pixel is classified as foreground consecutively, is employed. At frame k, the detection support map $DSM_i(k)$ is given by

$$DSM_i(k) = \begin{cases} DSM_i(k-1) + 1 & s_i = 1 \\ 0 & otherwise \end{cases} \tag{5}$$

We assume the maximum duration of a target remaining stationary is t_{sta} and that the corresponding frame number is N_{sta}. If the time of one target remaining stationary reaches t_{sta}, this target is then classified as a background target. When the $DSM_i(k)$ reaches the threshold N_{sta}, a random strategy is then used to update the background model. During the update, n_{ran} samples (n_{ran} is a uniform random positive integer between 1 and N) in the background model of pixel i are replaced by new samples. To speed up ghost elimination, when a sample has been updated, one sample from its 8-connected neighborhood pixel background model should be replaced by a new value according to a uniform law. This operation uses the spatial consistency assumption of the background, as in the ViBe method. The assumption is that the background pixel shares a similar distribution as its immediate neighbors. In addition, each new sample is obtained by (2). The detection support map is updated following (6), and the updated background model $M_i(k)$ at time k is obtained.

$$DSM_i(k) = DSM_i(k) - n_{ran} \tag{6}$$

Based on the above steps, the whole update background model $M(k)$ can be obtained by recursion with the new TIR frame I_{ori}. Then, (4) can be executed with the updated $y_{i,equ}$ calculated from $M_i(k)$; all s_i ($1 \leq i \leq N_{img}$) constitute the foreground probability map $S(k)$. Figure 3a shows the foreground probability map of Figure 2a. The whole target is separated into several small targets and there is some scattered noise in $S(k)$ in Figure 3a. Morphology erosion with small structure can be used to eliminate the scattered noise. Before normalizing the absolute difference by th_{diff} in (4), we conduct erosion and dilation operation on the absolute difference image whose pixel value is $abs\left(y_i - y_{i,equ}\right)$.

Figure 3b shows the foreground probability map with dilation following erosion. The pseudo-code for the background model update is presented in Algorithm 1. Lines 1–3 determine a pixel to be a foreground pixel ($s_i = 1$) or a background pixel ($s_i < 1$). If the pixel belongs to the background, then we update its background model as lines 4–9. If it belongs to the foreground, besides updating $M_i(k-1)$ lines 11–18 also show how to use the detection support map to eliminate deadlock situations and ghosts.

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

Figure 3. The two foreground probability maps of Figure 2a. (**a**) The original foreground probability map; (**b**) the foreground probability map with dilation following erosion.

Algorithm 1: Background model update

Input: $M(k-1)$: background model. I_{ori}: the kth TIR image. th_{diff}: normalization threshold. $DSM_i(k)$: detection support map. N_{sta}: the frame number of a target remaining stationary.

Output: $M(k)$: background model. $S(k)$: foreground probability map.

1 **for** i = *1:*N_{img} **do**
2 calculate $y_{i,equ}$ according to (3);
3 calculate s_i according to (4);
4 **if** $s_i \neq 1$ **then**
5 $DSM_i(k) \leftarrow 0$;
6 **if** $v_{ran}(0,1) > s_i$ **then**
7 $m_{i,n} \leftarrow y_i$, where n is a random positive integer between 1 and N ;
8 update $M_i(k-1)$ with $m_{i,n}$;
9 **end**
10 **else**
11 $DSM_i(k) \leftarrow DSM_i(k-1) + 1$;
12 **if** $DSM_i(k) == N_{sta}$ **then**
13 **for** j=*1:*n_{ran},*(where* n_{ran} *is a random positive integer between 1 and N)* **do**
14 update $M_i(k-1)$ according to (1) and (2);
15 update $M_q(k-1)$ according to (1) and (2), where q is a random pixel index and y_q is located at the 8-connected neighborhood of pixel i;
16 $DSM_i(k) \leftarrow DSM_i(k-1) - 1$;
17 **end**
18 **end**
19 **end**
20 **end**
21 $M(k) \leftarrow M(k-1)$;
22 do morphology operation on the absolute difference image;
23 update s_i and make up $S(k)$;

2.2. Multi-Target Likelihood Function Calculation

In the foreground probability map $S(k)$, each s_i ($1 \leq i \leq N_{img}$) lies in the interval [0, 1], which can be interpreted as the probability that pixel i should be classified as the foreground. In $S(k)$, each target x illuminates a set of pixels denoted by $T(x)$. Inspired by [4], if a pixel $i \in T(x)$, its intensity distribution follows the foreground likelihood function $g_F(x)$, and if $i \notin T(x)$, its intensity distribution follows the background likelihood function $g_B(x)$. These two likelihood functions (or probability density functions of intensity) are of the form:

$$g_F(x) = \varsigma_F \exp(x/\delta_F) \tag{7}$$

$$g_B(x) = \varsigma_B \exp(-x/\delta_B) \tag{8}$$

where δ_F and δ_B determine the spread rates of the foreground and background intensities respectively, and ς_F and ς_B are normalizing factors. In general, the ratio of δ_F to δ_B determines the detection threshold for judging whether a pixel belongs to a target or background. The greater the ratio is, the lower the associated target threshold will be. Because the TBD method allows more suspected targets to be input into the δ-GLMB-TBD filter, δ_B should be significantly smaller than δ_F. The background intensity remains constant unless it is quite close to the target, whereas the foreground intensity has a significantly more variable and spreading intensity function [4,29]. As the references state, the fluctuation of the ratio does not cause significant changes to the tracking results.

Let $\bar{s}(x)$ denote the average of all pixels in $T(x)$, i.e.,

$$\bar{s}(x) = \frac{1}{|T(x)|} \sum_{i \in T(x)} s_i \tag{9}$$

where $|\cdot|$ denotes the cardinality (the number of elements) of a set. We assume that all illumination regions of influences of the multi-target in the TIR image are not overlapped, i.e., $x \neq x' \Rightarrow T(x) \cap T(x') = \emptyset$. This assumption is reasonable, because an individual cannot recognize the identities and states of overlapped targets. For example, in the TIR image when some pedestrians overlap with each other, an individual cannot determine whether one of the pedestrians has disappeared, whether a pedestrian appears in the surveillance scene, or when or whether the group of pedestrians will separate from each other. In practical application, tracking is a dynamic process. When targets are overlapped, the merging procedure (seen in Section 2.3) will guarantee that all targets are treated as one. After separation, they will be tracked individually by the splitting procedure (seen in Section 2.3). The GLMB-TBD filter will generate the separable likelihood function based on the assumption that the targets do not overlap. Then, if given a target set X with statistically independent pixel values, the likelihood that the set X illuminates the region $\bigcup_{x \in X} T(x)$ can be expressed as $\prod_{x \in X} g_F(\bar{s}(x))$.

Let $\bar{s}_B(X)$ denote the average intensity of the map $S(k)$ after filling all target regions with the background pixel value of 0, i.e.,

$$\bar{s}_B(X) = \frac{1}{N_{img}} \left(\sum_{i=1}^{N_{img}} s_i - \sum_{x \in X} \sum_{i \in T(x)} s_i \right). \tag{10}$$

Substituting $\bar{s}_B(X)$ into (8),

$$\begin{aligned} g_B(\bar{s}_B(X)) &= \varsigma_B \exp\left(-\frac{\sum_{i=1}^{N_{img}} s_i - \sum_{x\in X}\sum_{i\in T(x)} s_i}{\delta_B N_{img}}\right) \\ &= \varsigma_B \exp\left(-\frac{\sum_{i=1}^{N_{img}} s_i}{\delta_B N_{img}}\right) \prod_{x\in X} \exp\left(\frac{\sum_{i\in T(x)} s_i}{\delta_B N_{img}}\right) \\ &= \varsigma_B \exp\left(-\frac{\sum_{i=1}^{N_{img}} s_i}{\delta_B N_{img}}\right) \prod_{x\in X} \exp\left(\frac{|T(x)|\cdot \bar{s}(x)}{\delta_B N_{img}}\right). \end{aligned} \tag{11}$$

Then, given the target set X, the multi-target likelihood of $S(k)$ is the product of the foreground and background, i.e.,

$$\begin{aligned} g(S(k)|X) &= g_B(\bar{s}_B(X)) \prod_{x\in X} g_F(\bar{s}(x)) \\ &= \underbrace{\varsigma_B \exp\left(-\frac{\sum_{i=1}^{N_{img}} s_i}{\delta_B N_{img}}\right)}_{independent\ of\ X} \prod_{x\in X} \underbrace{\exp\left(\frac{|T(x)|\cdot \bar{s}(x)}{\delta_B N_{img}}\right) g_F(\bar{s}(x))}_{dependent\ on\ x}. \end{aligned} \tag{12}$$

(12) shows that this likelihood is separable.

2.3. δ-GLMB-TBD Filter

After obtaining the multi-target likelihood for the foreground probability map, we describe how to detect and track multi-targets using the improved δ-GLMB-TBD filter with a separable likelihood function [30]. The main contribution in this part is that the improved filter can produce the appearance of the target in contrast to the center position or rectangle estimates obtained by the standard δ-GLMB-TBD filter. Based on this, we also develop the following: (1) merging and splitting procedures are employed to handle situations where multi-targets merge into one group and one group splits into several multi-targets; (2) pixel sampling and updating are used to accommodate target deformation; (3) birth target initialization procedure is to open up the applications of the filter. We discuss each of these improvements in the following subsections.

2.3.1. Basic Theory

In this subsection, the standard δ-GLMB-TBD filter, consisting of two steps, prediction and update, with a separable likelihood function, is briefly discussed [23]. Before introducing the recursion of the standard δ-GLMB-TBD filter, some notations are shown for convenience.

(1) Notation

For the remainder of the paper, let lowercase letters (e.g., x) denote single-target state and uppercase letters (e.g., X) denote multi-target states. The labeled target states are indicated by boldface letters (e.g., $\mathbf{x}$, $\mathbf{X}$). Space is represented by a letter with a tilde (e.g., $\tilde{\mathbf{X}}$ denotes the state space, $\tilde{\mathbf{L}}$ denotes the discrete label space). A labeled target can be written as $\mathbf{x} = (x, l)$, where $l \in \tilde{\mathbf{L}}$ and $l = (k, i)$, k means the target birth time, and i is a unique index to distinguish targets born at the same time. A labeled multi-Bernoulli (LMB) RFS with state space $\tilde{\mathbf{X}}$ and label space $\tilde{\mathbf{L}}$ can be written as $v = \{r^{(l)}, p^{(l)}\}_{l\in\tilde{\mathbf{L}}}$, where $r^{(l)}$ and $p^{(l)}$ mean the existence probability and the probability density of a target with label l. $\langle \alpha, \beta \rangle = \int \alpha(x)\beta(x)dx$ denotes the standard inner product, and $h^X = \prod_{x\in X} h(x)$ denotes the

multi-target exponential, where $h^{\varnothing} = 1$ by convention and $\alpha(x)$, $\beta(x)$, and $h(x)$ are real-valued functions. The generalized Kronecker delta function and inclusion function are defined as

$$\delta_Y(X) = \begin{cases} 1 & if\ X = Y \\ 0 & otherwise \end{cases}, \tag{13}$$

$$1_Y(X) = \begin{cases} 1 & if\ X \subseteq Y \\ 0 & otherwise \end{cases}. \tag{14}$$

(2) Standard δ-GLMB-TBD Filter

From [23], the multi-target posterior at time k has the following δ-GLMB form:

$$v_k(\mathbf{X}) = \Delta(\mathbf{X}) \sum_{I \in \mathcal{F}(\tilde{\mathbf{L}}_{0:k})} \delta_I(\mathcal{L}(\mathbf{X})) w_k^{(I)} [p_k^{(I)}]^{\mathbf{X}} \tag{15}$$

where $\mathbf{X}$ is the current multi-target state; $\Delta(\mathbf{X}) = \delta_{|\mathbf{X}|}(|\mathcal{L}(\mathbf{X})|)$ denotes the distinct label indicator, which means that the cardinalities of the set of labels and the set of state vectors are identical; $\tilde{\mathbf{L}}_{0:k}$ denotes the label space of targets born between time 0 and time k, where the subscript 0:k means time interval $[0, k]$; $\mathcal{F}(\cdot)$ denotes collections of all finite subsets of a given space; $\mathcal{L}$ is a projection from space $\tilde{\mathbf{X}} \times \tilde{\mathbf{L}}$ to $\tilde{\mathbf{L}}$ and hence $\mathcal{L}(\mathbf{X}) = \{\mathcal{L}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$ is the set of labels of $\mathbf{X}$; and $w_k^{(I)}$ denotes the joint existence probability of the label set I, while the multi-target exponential $[p_k^{(I)}]^{\mathbf{X}}$ denotes the joint probability density of $\mathbf{X}$, conditional on their corresponding label set I.

The new birth model covering labeled Poisson, labeled identically and independently distributed cluster and labeled multi-Bernoulli filter can be given by [20]

$$f_B = \Delta(\mathbf{Y}) w_B(\mathcal{L}(\mathbf{Y}))[p_B]^{\mathbf{Y}}. \tag{16}$$

where $\mathbf{Y}$ is the state of new birth targets, and w_B and $[p_B]^{\mathbf{Y}}$ are the joint existence probability and probability density. This model can also be written as the LMB birth model [23,31] as follows:

$$w_B(L) = \prod_{i \in \tilde{\mathbf{L}}_k} (1 - r_B^{(i)}) \prod_{l \in L} \frac{1_{\tilde{\mathbf{L}}_k}(l) r_B^{(l)}}{1 - r_B^{(l)}} \tag{17}$$

$$p_B(x, l) = p_B^{(l)}(x) \tag{18}$$

where $r_B^{(l)}$ and $p_B^{(l)}(x)$ mean the existence probability and the probability density of a birth target with label l.

Proposition 1. *If the multi-target state posterior with δ-GLMB form is given as (15) at time k, with the new birth model (16), the multi-target prediction density also has a δ-GLMB form [23]:*

$$v_{k+1|k}(\boldsymbol{X}) = \Delta(\boldsymbol{X}) \sum_{I \in \mathcal{F}(\tilde{\boldsymbol{L}}_{0:k+1})} \delta_I(\mathcal{L}(\boldsymbol{X})) w_{k+1|k}^{(I)} [p_{k+1|k}^{(I)}]^{\boldsymbol{X}} \tag{19}$$

where

$$w_{k+1|k}^{(I)} = w_S^{(I)}(I \cap \tilde{\boldsymbol{L}}_{0:k}) w_B(I \cap \tilde{\boldsymbol{L}}_{k+1}) \tag{20}$$

$$w_S^{(I)}(L) = [\eta_S^{(I)}]^L \sum_{J \in \tilde{\boldsymbol{L}}_{0:k}} 1_J(L) [1 - \eta_S^{(I)}]^{J-L} w_k^{(I)}(J) \tag{21}$$

$$p_{k+1|k}^{(I)}(x, l) = 1_{\tilde{\boldsymbol{L}}_{0:k}}(l) P_S^{(I)}(x, l) + (1 - 1_{\tilde{\boldsymbol{L}}_{0:k}}(l)) p_B(x, l) \tag{22}$$

$$P_S^{(I)}(x,l) = \frac{\left\langle p_S(\cdot,l) f_{k+1|k}(x|\cdot,l), p_k^{(I)}(\cdot,l) \right\rangle}{\eta_S^{(I)}(l)} \tag{23}$$

$$\eta_S^{(I)}(l) = \left\langle p_S(\cdot,l), p_k^{(I)}(\cdot,l) \right\rangle \tag{24}$$

$f_{k+1|k}(\cdot)$ denotes a single-target transition function. $p_S(x,l)$ denotes the survival probability of target x. $\tilde{\boldsymbol{L}}_{k+1}$ is the label space of a target born at time $k+1$.

Proposition 2. *If the multi-target prediction density has the form of δ-GLMB as (19), then, with the measurement set S and separable likelihood function $\gamma_S(\boldsymbol{x})$, the multi-target posterior density also has the same form [23]:*

$$v_{k+1}(\boldsymbol{X}|S) = \Delta(\boldsymbol{X}) \sum_{I \in \mathcal{F}(\tilde{\boldsymbol{L}}_{0:k+1})} \delta_I(\mathcal{L}(\boldsymbol{X})) w_{k+1}^{(I)}(S) [p_{k+1}^{(I)}(\cdot|S)]^{\boldsymbol{X}} \tag{25}$$

where

$$w_{k+1}^{(I)}(S) \propto w_{k+1|k}^{(I)} [\eta_S]^I \tag{26}$$

$$p_{k+1}^{(I)}(x,l|S) = p_{k+1|k}^{(I)}(x,l) \gamma_S(x,l) \Big/ \eta_S(l) \tag{27}$$

$$\eta_S(l) = \left\langle p_{k+1|k}^{(I)}(\cdot,l), \gamma_S(\cdot,l) \right\rangle. \tag{28}$$

Proposition 3. *If the multi-target state posterior with the δ-GLMB form is given as (15) at time k, the cardinality distribution $\rho_k(n)$ can be given by [20]*

$$\rho_k(n) = \sum_{I \in \mathcal{F}(\tilde{\boldsymbol{L}}_{0:k})} \delta_n(|I|) w_k^{(I)}. \tag{29}$$

The cardinality estimates can be obtained by

$$\hat{n} = \arg\max_n \rho_k(n) \tag{30}$$

The multi-target state estimate is the mean estimate of the multi-target state conditioned on the estimated cardinality $\hat{n}$ as in [20].

2.3.2. Recursion

Particle implementation is based on the standard δ-GLMB filter implementation [20,21]. Each target density $p_k^{(l)}$ with label l in $p_k^{(I)}$ is modeled as a set of weighted samples $\{(\Omega_{k,n}^{(l)}, \mathbf{x}_{k,n})\}_{n=1}^{J_k^{(l)}}$, where $J_k^{(l)}$ is the number of particles, and $\mathbf{x}_n$ denotes the state x_n and label l. Besides the prediction and update steps, we add four other parts: pruning, splitting, merging, and new birth target initialization to form the whole solution. The details are as follows.

(1) Prediction

The pixels illuminated by the target can describe the target more effectively than a single point or a rectangle. Let us assume the single-target state x have two fields: a geometric center location $x.cen$ and cover area $x.cov$, which can be obtained from $T(x)$. For convenience, let $\mathbf{x}.cen = x.cen$ and $\mathbf{x}.cov = x.cov$ in this paper. This new approach representing a single-target benefits from the similar target shape in consecutive frames. To accommodate the small amount of deformation, pixel sampling is executed in this prediction step. Each pixel in $x.cov$ has the probability of p_{pix} to be selected to survive, the value of which depends on the deformation degree. In general, when the deformation is small, p_{pix} is close to 1. It is obvious that sampling may reduce the covered area in the final estimates. Fortunately, this can be alleviated in the update step.

If at time k, the multi-target state posterior is given by (15), when the LMB new birth model $\{(r_{B,k+1}^{(l)}, p_{B,k+1}^{(l)})\}_{l\in\tilde{\mathbf{L}}_{k+1}}$ is known (in practice for the unknown birth model, the third part in this subsection describes how to initialize the new birth model), where $p_{B,k+1}^{(l)}$ is $\{(\Omega_{B,k+1,n}^{(l)}, \mathbf{x}_{B,k+1,n})\}_{n=1}^{B^{(l)}}$ and $B^{(l)}$ is the number of particles, according to proposition 1, we obtain

$$\eta_S^{(I)}(l) = \sum_{n=1}^{J_k^{(l)}} \Omega_{k,n}^{(l)} p_S(\mathbf{x}_{k,n}) \tag{31}$$

and $p_{k+1|k}^{(I)}(x,l)$ can be represented as

$$\{(1_{\tilde{\mathbf{L}}_{0:k}}(l)\tilde{\Omega}_{S,k+1|k,n}^{(l)}, \mathbf{x}_{S,k+1|k,n})\}_{i=n}^{J_k^{(l)}} \cup \{(1_{\tilde{\mathbf{L}}_{0:k}}(l)\Omega_{B,k+1,n}^{(l)}, \mathbf{x}_{B,k+1,n})\}_{i=n}^{B^{(l)}}, \tag{32}$$

where $\mathbf{x}_{S,k+1|k,n}.cen \sim q(\cdot|\mathbf{x}_{k,n}.cen)$ $(n = 1,2,...,J_k^{(l)})$, $x_{S,k+1|k,n}.cov$ is generated by a random pixel sample as described above; $\Omega_{S,k+1|k,n}^{(l)} = \frac{\Omega_{k,n}^{(l)} f_{k+1|k}(x_{S,k+1|k,n}|x_{k,n},l) p_S(x_{k,n},l)}{q(\mathbf{x}_{S,k+1|k,n}.cen|\mathbf{x}_{k,n}.cen)}$; $\tilde{\Omega}_{S,k+1|k,n}^{(l)} = \Omega_{S,k+1|k,n}^{(l)} \Big/ \sum_{n=1}^{J_k^{(l)}} \Omega_{S,k+1|k,n}^{(l)}$; and $q(\cdot|x_{k,n}.cen)$ is a proposal density. The procedure for calculating $w_{k+1|k}^{(I)}$ in (20) is bothersome and complex; however, reference [21] offers a method based on a K-shortest paths algorithm to carry it out. To understand the procedure more clearly, reference [21] is strongly recommended.

(2) Update

With the assumption that targets are not overlapped, we obtain a separable likelihood function shown as (12). After adding a distinct label to each target, (12) can be written as

$$g(S(k)|\mathbf{X}) \propto \prod_{\mathbf{x}\in\mathbf{X}} \gamma_S(\mathbf{x}) \tag{33}$$

where $\gamma_S(\mathbf{x}) = g_F(\bar{s}(\mathbf{x}))$.

According to proposition 2, if each single-target density $p_{k+1|k}^{(l)}$ is modeled by a particle set $\{(\Omega_{k+1|k,n}^{(l)}, \mathbf{x}_{k+1|k,n})\}_{n=1}^{J_{k+1|k}^{(l)}}$. Then,

$$\eta_S(l) = \sum_{n=1}^{J_{k+1|k}(l)} w_{k+1|k,n}(l)\gamma_S(\mathbf{x}_{k+1|k,n}) \tag{34}$$

and $p_{k+1}^{(I)}(x,l|S)$ can be represented as

$$\{(\Omega_{k+1,n}^{(l)}, \mathbf{x}_{k+1,n})\}_{n=1}^{J_{k+1}^{(l)}}, \tag{35}$$

where $w_{k+1,n}(l) = w_{k+1|k,n}(l)\gamma_S(x_{k+1|k,n}) \Big/ \sum_{n=1}^{J_{k+1|k}(l)} w_{k+1|k,n}(l)\gamma_S(x_{k+1|k,n})$; $\mathbf{x}_{k+1,n} = \mathbf{x}_{k+1|k,n}$; and $J_{k+1}^{(l)} = J_{k+1|k}^{(l)}$. Substituting (34) into (26), we obtain $w_{k+1}^{(I)}(S)$.

Analogous to the standard particle filter, resampling each target density $\{(\Omega_{k+1,n}^{(l)}, \mathbf{x}_{k+1,n})\}_{n=1}^{J_{k+1}^{(l)}}$ must be executed to reduce the degeneracy. For simplicity, in this paper, multinomial resampling is used for numerical studies that would otherwise be carried out with other multi-Bernoulli filters [17].

To eliminate the influence of pixel sampling in the prediction step, a pixel set update procedure is used to correct the shape of the target. Taking target x as an example, the procedure is described as

follows. First, for all $J_{k+1}^{(l)}$ particles representing target x, count the pixel i ($i \in \bigcup_{j=1}^{J_{k+1}^{(l)}} T(x_{k+1,j})$) occupied times $OT_i^{J_{k+1}^{(l)}}$. The occupied times are initialized to 0, i.e., $OT_i^0 = 0$, and for the jth ($1 \le j \le J_{k+1}^{(l)}$) particle state $x_{k+1,j}$, OT_i^j is given by

$$OT_i^j = \begin{cases} OT_i^{j-1}+1 & if\, i \in T(x_{k+1,j}) \\ OT_i^{j-1} & otherwise \end{cases} . \tag{36}$$

Second, compare the $OT_i^{J_{k+1}^{(l)}}$ with a threshold $p_{pix}J_{k+1}^{(l)}/2$. When $OT_i^{J_{k+1}^{(l)}}$ reaches the threshold, the pixel i is classified as an occupied pixel. Third, if the occupied pixel i is greater than an extremely low value in the foreground probability map, then it is considered to be a pixel in the updated pixel set for target x. The updated pixel set is the estimated cover area of target x. The pseudo-code for the pixel set update procedure is presented in Algorithm 2. Algorithm 2 shows us that for each pixel belonging to the covered areas occupied by all particles representing x, the occupied time is calculated (line 3) and then it will be used to determine the true cover area of the target (lines 5–9). The filter will propagate a pixel set as the target state in the recursion and will directly produce the pixel set as its output. This operation can facilitate target recognition and extraction and subsequent processing in computer vision applications.

Algorithm 2: Pixel set update

Input: $S(k)$: foreground probability map. x: single-target state. $J_{k+1}^{(l)}$: the number of particles describing target x. p_{pix}: the pixel sampling rate in the prediction step. $x_{k+1,j}$: the jth particle state ($1 \le j \le J_{k+1}^{(l)}$).

Output: x: the update target state.

```
1  for i ∈ ⋃_{j=1}^{J_{k+1}^{(l)}} T(x_{k+1,j}) do
2      for j = 1 : J_{k+1}^{(l)} do
3          calculate OT_i^j according to (36);
4      end
5      if OT_i^{J_{k+1}^{(l)}} > p_pix J_{k+1}^{(l)} / 2 then
6          if the pixel i in S(k) is greater than an extremely low threshold then
7              pixel i is classified as an updated pixel of target x;
8          end
9      end
10 end
11 output the updated target state x ;
```

(3) Pruning, Splitting, Merging and Birth Target Initialization

To reduce computational complexity, besides the truncation [20,21], the pruning procedure is also needed. The multi-target set **X** should be discarded if its weight $w_{k+1}^{(I)}$ is less than a threshold th_{pr}. In practical application, the dynamic processes of multi-targets are complex. To accommodate these processes, we propose splitting and merging, which are described as follows. For each target **x**, clustering is executed to check whether the target **x** should be split into several small targets. If yes, the small target with the most pixels inherits the identity (label) of target **x**, while the others are labeled as new birth targets. For two targets in **X** with substantial overlap, if the overlap ratio of the pixel intersection area to the smaller target area is higher than a threshold th_{me}, these two targets should be merged; the merged label is the same as the earlier born target.

Birth target initialization has two pre-processing steps. Step 1: clustering the foreground probability map to obtain current cluster target $\mathbf{x}_{c,i}$ ($0 \leq i \leq n$, n denotes the number of all current cluster targets). Step 2: for each cluster target $\mathbf{x}_{c,i}$, calculate the overlap ratio between the $\mathbf{x}_{c,i}$ and all existing targets. If the maximum overlap ratio is higher than a threshold th_{bir}, then remove this $\mathbf{x}_{c,i}$. Subsequently, we use the remaining $\mathbf{x}_{c,i}$ to initiate the new birth target LMB modeled as $v_{B,k+1} = \{(r^{(l)}_{B,k+1}, p^{(l)}_{B,k+1})\}_{l \in \tilde{\mathbb{L}}_{k+1}}$. For simplicity, the $r^{(l)}_{B,k+1}$ is initialized as a constant and the target density $p^{(l)}_{B,k+1}$ is modeled as $\{(\Omega^{(l)}_{B,k+1,n}, \mathbf{x}_{B,k+1,n})\}^{B^{(l)}}_{n=1}$, where all $\Omega^{(l)}_{B,k+1,n}$ have equal weight; and $\mathbf{x}_{B,k+1,n}.cen \sim \mathcal{N}(u^{(l)}_B, Q^{(l)}_B)$, where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution, $u^{(l)}_B$ is the geometric center location of one remaining $\mathbf{x}_{c,i}$, and $Q^{(l)}_B$ is the variance. $\mathbf{x}_{B,k+1,n}$.cov can be obtained from $T(\mathbf{x}_{c,i})$ using random sampling by probability p_{pix}.

These procedures are designed to broaden the applications of the filter. Without splitting and merging, this filter would produce a cardinality estimation error. Missing the new target birth process would invalidate the δ-GLMB-TBD filter. Using proposition 3 can produce a multi-target estimate.

3. Experimental Results

In this section, the optimal parameters of the background-subtraction-based δ-GLMB-TBD filter are determined. Based on the optimal parameters, detection performance is compared to several state-of-the-art techniques. Then, tracking performance is tested. Based on experimental results, the advantages and disadvantages of the proposed algorithm are discussed. In these experiments, the kinematic transition is modeled as the random walk model; the survival probability of a target is 0.98; the maximum number of frames in which the target remains stationary N_{sta} is 60; the morphology erosion and dilation operators are executed once with the same flat disk-shaped structuring element whose radius is 1. The particle number for each target is 200; the pruning threshold th_{pr} is 0.01; the DBSCAN method is chosen for clustering [32]; the threshold th_{bir} in birth targets initialization is 0.5; the merging threshold th_{me} is 0.7; the existence probability $r^{(l)}_{B,k+1}$ is 0.3; and the variance $Q^{(l)}_B$ is [1,0;0,1]. Theoretically, the same detection result and data association can achieve the same tracking performance. Therefore, the detection performance is the main evaluation criteria for the joint detection and tracking method. This novel TBD filter emphasizes detection by exploiting the trajectory information. The measure metrics of recall, precision, and F-measure (FM) are used to evaluate performance, and they are defined as:

$$recall = \frac{tp}{tp + fn}, precision = \frac{tp}{tp + fp}, FM = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{37}$$

where tp denotes the number of true positives, and fp and fn represent the number of false positives and false negatives.

3.1. Determination of Parameters

From previous discussions, the parameters to be determined are as follows: (1) the threshold th_{diff} in (4), (2) the number of samples N, (3) δ_F and δ_B, and (4) pixel sampling rate p_{pix}. The tested sequence is OTCBVS Dataset 03 2a [33]. The fluctuation of the δ_F to δ_B ratio does not cause obvious changes in the tracking result, where $\delta_F = 0.2$ and $\delta_B = 0.05$ are suitable for our experiments.

The proposed method can be divided into two parts: detection and tracking, which are connected by the foreground probability map $S(k)$. The two parts can be considered as conditional independence (CI) based on $S(k)$. Therefore, the parameters also can be divided into two groups: (1) th_{diff} and N (before obtaining $S(k)$); (2) p_{pix} (after obtaining $S(k)$). Let us use the "trial-and-error" method to determine the two groups of parameters respectively as follows.

To choose the optimal value for th_{diff} and N, the performance metrics are calculated for th_{diff} as: 20, 40, 60, 80, and 100 while N as: 10, 20, 30, 40, and 50. The other parameters are fixed at δ_F = 0.2,

δ_B = 0.05 and p_{pix} = 0.8. The results are shown in Tables 1–3. From Table 1, we can see that the recall decreases with increasing th_{diff}. When th_{diff} is set 20, the recall can achieve the highest value. The number of samples N has less influence on recall; in general, with the increase of N, the recall tends to become higher. From Table 2, the result of precision is opposite to the recall metric. Table 3 shows that the balance metric FM has the highest value at th_{diff} = 60. In this TBD method, recall plays a more important role than precision. The range of th_{diff} can be set between 40 and 60, N is between 20 and 40.

Table 1. Average recall with different th_{diff} and N.

th_{diff} \ N	10	20	30	40	50
20	0.9315	0.9219	0.9057	0.9082	0.9241
40	0.8560	0.8532	0.8549	0.8756	0.8654
60	0.7311	0.7880	0.8031	0.8121	0.8213
80	0.6343	0.6996	0.7146	0.7108	0.7288
100	0.5083	0.5804	0.6307	0.6316	0.6423

Table 2. Average precision with different th_{diff} and N.

th_{diff} \ N	10	20	30	40	50
20	0.4710	0.3937	0.3557	0.3361	0.3158
40	0.7279	0.6752	0.6461	0.6206	0.6039
60	0.8311	0.8296	0.8146	0.8056	0.8049
80	0.8910	0.8949	0.8931	0.8889	0.8853
100	0.9316	0.9304	0.9374	0.9349	0.9338

Table 3. Average FM with different th_{diff} and N.

th_{diff} \ N	10	20	30	40	50
20	0.6221	0.5482	0.5061	0.4847	0.4665
40	0.7848	0.7479	0.7303	0.7228	0.7077
60	0.7709	0.8052	0.8050	0.8048	0.8031
80	0.7358	0.7796	0.7909	0.7849	0.7964
100	0.6480	0.7062	0.7498	0.7477	0.7552

Now, let us determine the pixel sampling rate p_{pix}. When we set th_{diff} to 60, δ_F to 0.2, δ_B to 0.05, and N to 20, Table 4 shows the experimental results. From Table 4, the evaluation metrics do not change significantly, because the p_{pix} mainly affects the detection and tracking of dim targets that occupy few pixels and may be deformed in consecutive frames. A p_{pix} closer to 1 is of minimal help in detecting and tracking deformation targets, and a smaller p_{pix} may cause a dim target to be undetected. In the experiment, the p_{pix} is set to 0.8.

Table 4. Average evaluation metrics of different p_{pix}.

p_{pix}	0.6	0.7	0.8	0.9	1
recall	0.9204	0.9157	0.9234	0.9296	0.9299
precision	0.8085	0.8047	0.8020	0.8037	0.8020
FM	0.8532	0.8431	0.8510	0.8545	0.8539

3.2. Comparison with Other Techniques

To the best of our knowledge, there is no common dataset for TIR multiple moving-target detection and tracking in the surveillance scene. We collected 15 sequences containing about 14,753 images in

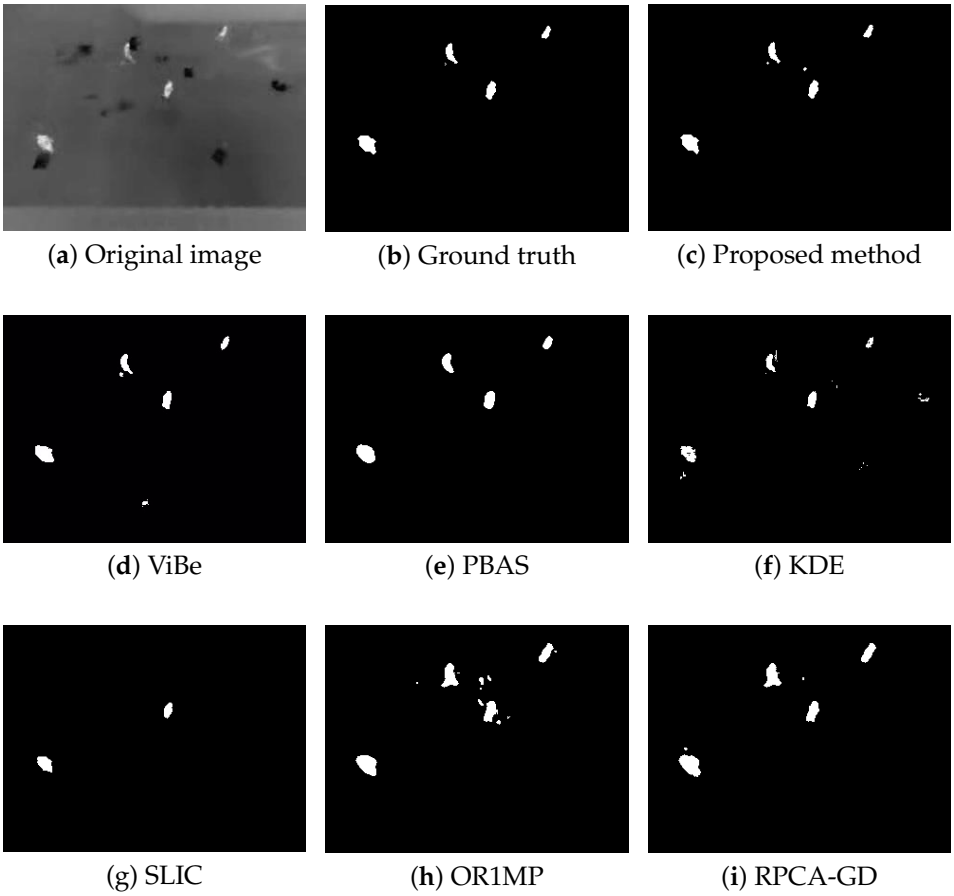
7 different locations at different times from OTCBVS [34] (IEEE OTCBVS WS Series Bench and Roland Miezianko, Terravic Reasearch Infrared Database) and PTB-TIR [35]. The moving targets are mainly pedestrians, but also include pets, cars, and trucks. In this subsection, the performance of the proposed method is compared with 6 methods: ViBe [24], pixel-based adaptive segmenter (PBAS) [36], and kernel density estimator (KDE) [25], SLIC-based method [37], orthogonal rank-one matrix pursuit method (OR1MP) [38], and robust PCA via Gradient Descent (RPCA-GD) [39]. The ViBe, PBAS, and KDE are classic state-of-the-art methods. The SLIC, OR1MP, and RPCA-GD are 3 new methods. Among them, SLIC is a proposal method designed for infrared target detection; OR1MP and RPCA-GD can be recursive and unable to give instantaneous detection because of the non-causality. The programs of the ViBe, KDE, OR1MP, and RPCA-GD methods are provided by their authors, and the parameters used in the experiments are suggested by the authors. The PBAS method is available in the BGSlibrary [40]. SLIC is implemented by us using MATLAB; we define $th_d = means + 4stds$ as the detection threshold, where means and stds are the mean and the standard deviation of all saliency scores, respectively.

First, we use four typical sequences to validate the effectiveness of the proposed method. They cover many typical situations such as targets crossing, entering/leaving, gathering/separation, occlusion, stopping /restarting, and irregular motion. The sequences are divided into two categories: sparse target scenarios and dense target scenarios. The results of the four sequences are well represented for examining the performance of the proposed method. The parameters used in our model are th_{diff} = 60, $N = 20$, $\delta_F = 0.2$, $\delta_B = 0.05$, and $p_{pix} = 0.8$. For the proposed method, the output results are the estimates of the filter, while for other methods, the results are the foreground detections.

Figures 4–7 shows examples of moving-target detection for four typical frames chosen from four different sequences. Table 5 presents the average evaluation metrics of the seven methods. For all four sequences, the proposed method can detect all moving targets, despite some false alarms. Compared with our method, ViBe detects more false alarms and PBAS produces missing detections. The KDE produces obvious worse detection results than the proposed method. Because the proposal-based method only uses the features of the target to detect and pay no attention to the information in inter frame, SLIC can only detect the highlight areas in the images. In general, other algorithms that do not use information between consecutive frames, such as the SLIC method, will obtain similar results. OR1MP and RPCA-GD can also obtain good detection performance on Seq. 2; this is because their non-causality allows them to use the follow-up frames to build the current time model to eliminate the ghost. The non-causality will limit their application. However, for Seq.1, OR1MP and RPCA-GD detection results occupy more pixels than the ground truth; for Seq. 3 and Seq. 4 they produce obvious missing detection. These results are validated by the metric scores in Table 5. The good performance of the proposed method benefits from the detection support map, the neighbor pixel update processing, and the use of trajectory information. Figures 4–7 and Table 5 indicate that the proposed method outperforms the other six methods obviously.

Table 5. Average evaluation metrics of different methods.

Metrics		Proposed Method	ViBe	PBAS	KDE	SLIC	OR1MP	RPCA-GD
Seq. 1	recall	0.9996	0.9875	0.9434	0.8477	0.6090	0.9992	0.9996
	precision	0.9487	0.8489	0.7743	0.8266	0.8411	0.5716	0.6175
	FM	0.9587	0.8952	0.8269	0.8295	0.6216	0.7169	0.7533
Seq. 2	recall	0.9625	0.7127	0.6132	0.3660	0.0679	0.9598	0.9621
	precision	0.8231	0.8487	0.8523	0.7512	0.0633	0.67851	0.8031
	FM	0.8663	0.7893	0.7150	0.4547	0.0611	0.8076	0.8651
Seq. 3	recall	0.9656	0.9632	0.9007	0.5276	0.3869	0.5238	0.5778
	precision	0.8477	0.4931	0.6529	0.5408	0.1362	0.7826	0.7743
	FM	0.8764	0.6240	0.7096	0.5107	0.1874	0.5618	0.6020
Seq. 4	recall	0.8847	0.8311	0.6150	0.7524	0.0210	0.7477	0.7786
	precision	0.7909	0.7142	0.8731	0.6306	0.0257	0.8103	0.7776
	FM	0.8244	0.7626	0.7035	0.6784	0.0173	0.7716	0.7669

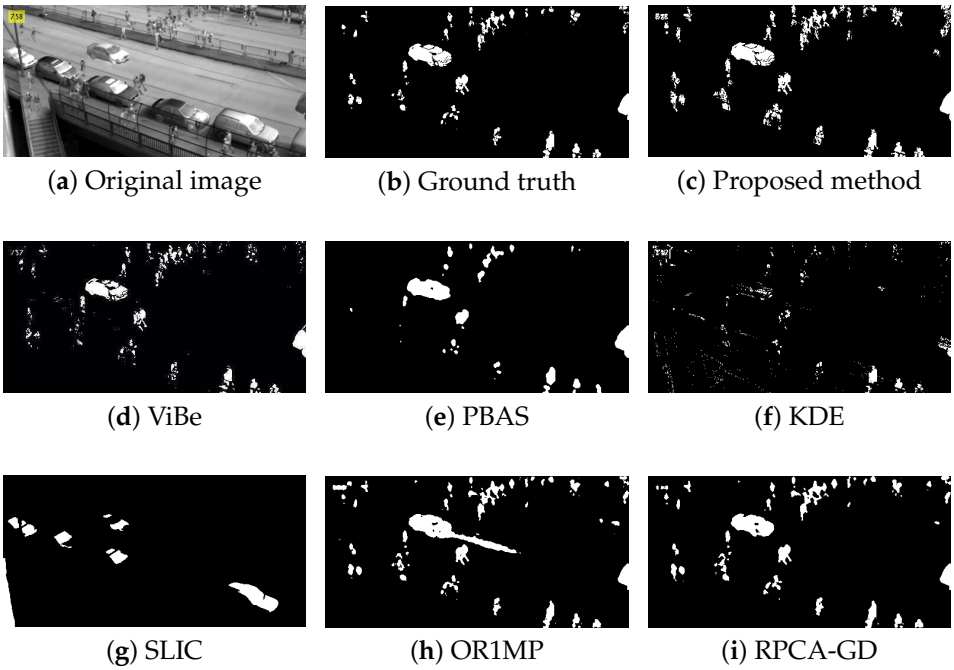


(**a**) Original image (**b**) Ground truth (**c**) Proposed method

(**d**) ViBe (**e**) PBAS (**f**) KDE

(g) SLIC (**h**) OR1MP (**i**) RPCA-GD

Figure 4. The detection results of the 190th frame in Seq 1.

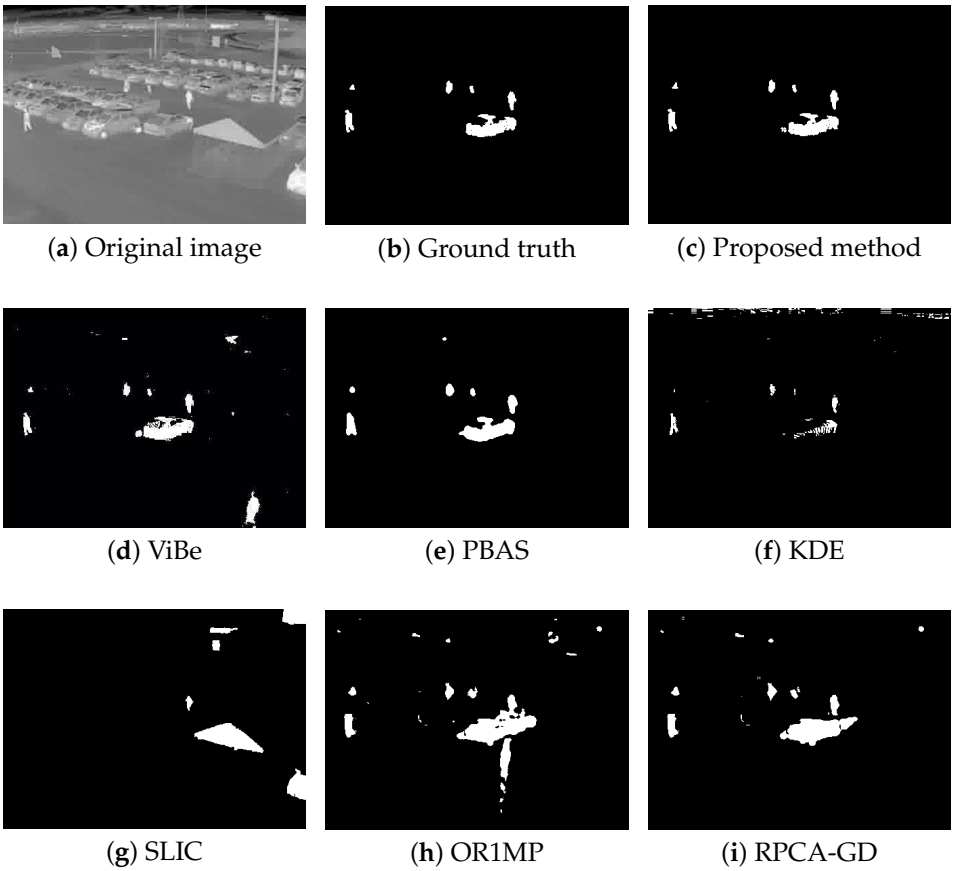


(**a**) Original image (**b**) Ground truth (**c**) Proposed method

(**d**) ViBe (**e**) PBAS (**f**) KDE

(**g**) SLIC (**h**) OR1MP (**i**) RPCA-GD

Figure 5. The detection results of the 758th frame in Seq. 2.

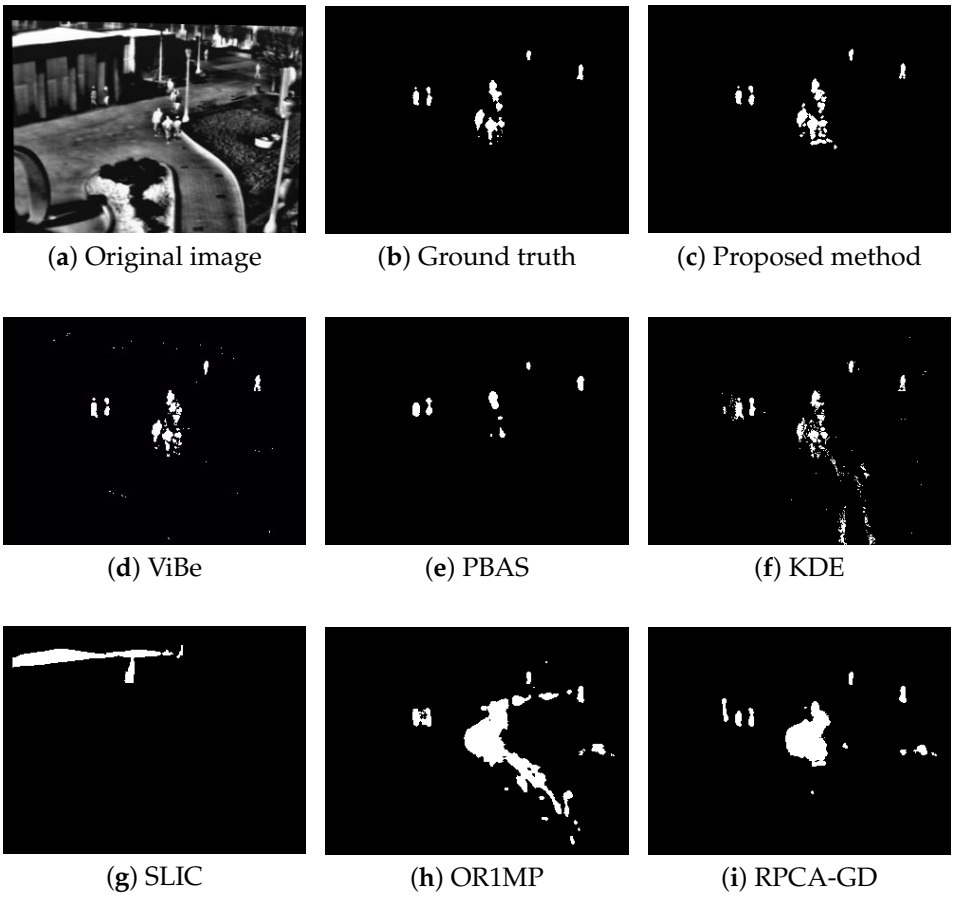


(**a**) Original image (**b**) Ground truth (**c**) Proposed method

(**d**) ViBe (**e**) PBAS (**f**) KDE

(**g**) SLIC (**h**) OR1MP (**i**) RPCA-GD

Figure 6. The detection results of the 173th frame in Seq. 3.



(**a**) Original image (**b**) Ground truth (**c**) Proposed method

(**d**) ViBe (**e**) PBAS (**f**) KDE

(**g**) SLIC (**h**) OR1MP (**i**) RPCA-GD

Figure 7. The detection results of the 741rd frame in Seq. 4.

In addition, Table 6 shows average evaluation metrics and runtime per frame for all 15 sequences. From Table 6, we can see that the proposed method can get the highest recall, precision, and FM metrics even compared with the non-causal method. Also, Table 6 validates that the proposed method has the best detection performance among the 7 methods.

Table 6. Average evaluation metrics for all sequences.

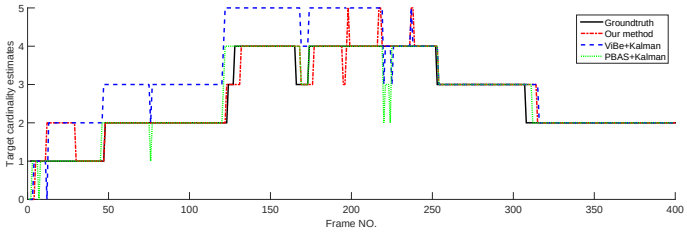
	Proposed Method	**ViBe**	**PBAS**	**KDE**	**SLIC**	**OR1MP**	**RPCA-GD**
recall	0.9839	0.9332	0.8619	0.7682	0.4435	0.8449	0.8714
precision	0.8732	0.6116	0.8121	0.5993	0.3082	0.7871	0.7935
FM	0.9094	0.6438	0.8001	0.5600	0.1441	0.7543	0.7775

3.3. Discussion of Tracking Performance

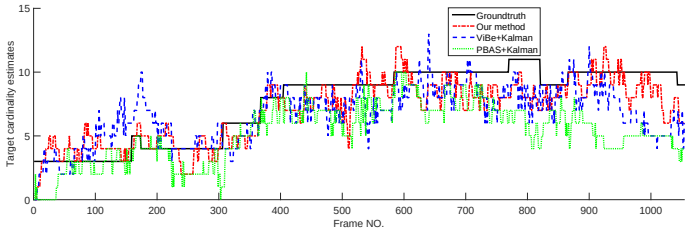
The background-subtraction-based δ-GLMB-TBD method can also produce target trajectories without additional post-processing, and accommodate multi-target dynamic process and measurement process, although its main task is target detection. In this section, the optimal sub-pattern assignment (OSPA) metric [41] interpreted as per-target tracking error containing cardinality error and state estimation error with parameters p = 1 and c = 50 pixel will be employed as the main performance metric.

In this section, the tracking performance of the proposed method is compared with the standard Kalman filter whose detections are provided by ViBe and PBAS methods. The models and parameters used in Kalman filter are the same as the proposed method, but in Kalman filters, the target state is represented by its centroid. The cardinality estimates and OSPA curves tested on Seq. 1 and Seq. 4 are

shown in Figures 8 and 9. From Figure 8; we can see that without overlapped targets, our method and PBAS + Kalman method can almost produce the correct cardinality; and with complex multi-target motion and occlusion, the 3 filters produce obvious cardinality estimation error, but they still can reflect the trend of cardinality change. From Figure 9, we can see, in both simple and complex scenarios, the proposed method can obtain the lowest OSPA value, this means the proposed method has the best multi-target tracking performance in the 3 methods. This is because the proposed method can yield accurate appearance and centroid estimation. The average OSPA values of all 15 sequences is shown in Table 7. Also, Table 7 indicates the proposed method can obtain the best multi-target tracking performance.

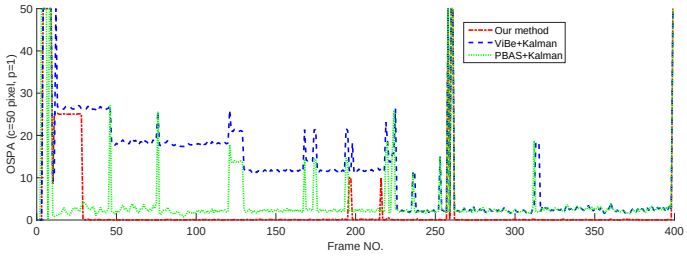


(**a**) Seq. 1

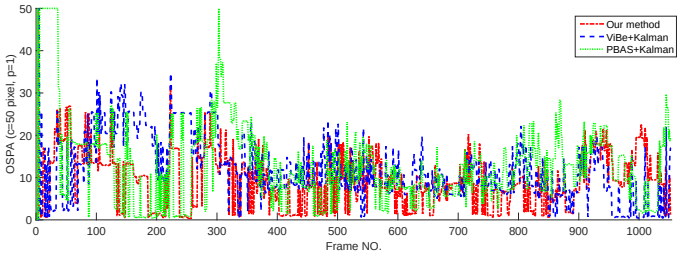


(**b**) Seq. 4

Figure 8. The cardinality estimates of different trackers.



(**a**) Seq. 1



(**b**) Seq. 4

Figure 9. The OSPA of different trackers.

Table 7. Average OSPA values for all sequences.

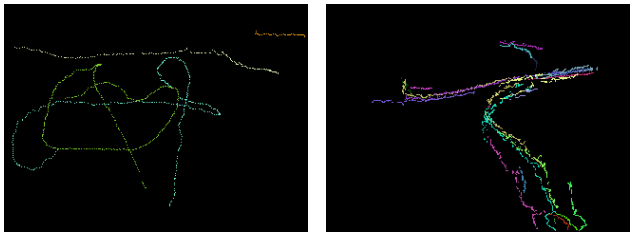
Proposed Method	ViBe + Kalman	PBAS + Kalman
5.00	11.13	9.16

Figure 10 shows two typical frames with estimation error. The proposed method outputs targets that are filled with different colors. From Figure 10, we can see the "ghost" (seen in Figure 10b), similarity between the targets and background (red rectangles with number 1, 2, 6, 7 in Figure 10b) can cause the over estimation; occlusions between targets or between target and background (red rectangles with number 6, 9 in Figure 10b) is the main reason to produce low estimation.

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

Figure 10. The two typical results of the proposed method. (**a**) The 20th frame in Seq. 1; (**b**) the proposed method estimates of (**a**); (**c**) the 486th frame in Seq. 4; (**b**) the proposed method estimates of (**c**).

As described above, this new filter can add a label to the target to maintain the track. Figure 11 shows the estimated trajectories of moving targets whose duration exceeds 30 frames. The different colors in Figure 11 denote different trajectories. According to Figure 11, the filter can track the targets successfully along with their labels without overlap or occlusion. When a target is separated from the crowd, this new filter can continue tracking it as a new target. The estimates from the labeled filter can facilitate tracking, recognition, and other subsequent processing in computer vision. In future, the trace association could be used to merge the tracks before and after occlusion to form a long-time trajectory.



(**a**) Seq. 1 (**b**) Seq. 4

Figure 11. The trajectory estimate results.

4. Conclusions

A novel method for moving-target detection and tracking directly from the TIR sequence in surveillance scenes was proposed based on background-subtraction and the δ-GLMB-TBD filter. First, a background subtraction method using a random selection strategy was used to produce the foreground probability map. Separable non-overlapped multi-target likelihood was exploited to obtain the probability of the pixels belonging to the foreground. Then, the δ-GLMB-TBD filter was used to provide estimates. Unlike other RFS-based filters, the proposed method used the pixel set, which was the target projection in the image, to describe the target instead of a rectangle or a single point. This means the δ-GLMB-TBD filter directly produced a continuous trajectory as well as accurate multi-target shape estimates. In implementation, several procedures including pixel sampling and

update, target merging and splitting, and new birth target initialization were combined in the method to accommodate target deformation and multi-target dynamic change: gathering, splitting, birth, and death. After describing the method, the optimal parameters were determined by experiments. Then, the performance of the novel method was compared with six existing methods. According to the experimental results, the proposed TBD method obtained the highest FM scores, meaning that it outperformed the other six methods in the detection of moving targets. The experiments also show that the proposed method can achieve better tracking performance than the Kalman filters with different detections. In future, the proposed method could be extended to the detection and tracking of moving targets without non-overlap assumptions.

Author Contributions: Methodology, C.L. and W.W.; Validation, C.L. and W.W.; Formal Analysis, C.L.; Writing—Original Draft Preparation, C.L. and W.W.; Writing—Review & Editing, W.W. and C.L.; Project Administration, W.W.; Funding Acquisition, W.W.

Funding: This research was funded by the National Natural Science Foundation of China [61771028].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gade, R.; Moeslund, T.B. Thermal cameras and applications: A survey. *Mach. Vis. Appl.* **2014**, *25*, 245–262. [CrossRef]
2. Tan, Y.; Guo, Y.; Gao, C.; Tan, Y.; Guo, Y.; Gao, C. Background subtraction based level sets for human segmentation in thermal infrared surveillance systems. *Infrared Phys. Technol.* **2013**, *61*, 230–240. [CrossRef]
3. Berg, A.; Ahlberg, J.; Felsberg, M. Channel Coded Distribution Field Tracking for Thermal Infrared Imagery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1248–1256.
4. Hoseinnezhad, R.; Vo, B.N.; Vo, B.T. Visual Tracking in Background Subtracted Image Sequences via Multi-Bernoulli Filtering. *IEEE Trans. Signal Process.* **2013**, *61*, 392–397. [CrossRef]
5. Shaikh, S.H.; Saeed, K.; Chaki, N. *Moving Object Detection Using Background Subtraction*; Springer International Publishing: Cham, Switzerland, 2014; pp. 15–23.
6. St-Charles, P.L.; Bilodeau, G.A.; Bergevin, R. Universal Background Subtraction Using Word Consensus Models. *IEEE Trans. Image Process.* **2016**, *25*, 4768–4781. [CrossRef]
7. Gemignani, G.; Rozza, A. A Robust Approach for the Background Subtraction Based on Multi-Layered Self-Organizing Maps. *IEEE Trans. Image Process.* **2016**, *25*, 5239–5251. [CrossRef] [PubMed]
8. Hurney, P.; Waldron, P.; Morgan, F.; Jones, E.; Glavin, M. Review of pedestrian detection techniques in automotive far-infrared video. *Iet Intell. Transp. Syst.* **2015**, *9*, 824–832. [CrossRef]
9. Yi, W.; Morelande, M.R.; Kong, L.; Yang, J. An Efficient Multi-Frame Track-Before-Detect Algorithm for Multi-Target Tracking. *J. Sel. Top. Signal Process.* **2013**, *7*, 421–434. [CrossRef]
10. Jiang, H.; Yi, W.; Kirubarajan, T.; Kong, L.; Yang, X. Track-Before-Detect Strategies for Radar Detection in G0-distributed Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 2516–2533.
11. Vermaak, J.; Maskell, S.; Briers, M.; Perez, P. Bayesian visual tracking with existence process. In Proceedings of the IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. I-721–I-724.
12. Czyz, J.; Ristic, B.; Macq, B. A particle filter for joint detection and tracking of color objects. *Image Vis. Comput.* **2007**, *25*, 1271–1281. [CrossRef]
13. Punithakumar, K.; Kirubarajan, T. A sequential Monte Carlo probability hypothesis density algorithm for multitarget track-before-detect. In Proceedings of the SPIE—The International Society for Optical Engineering, Boston, MA, USA, 24–25 October 2005; pp. 59131S–59131S–8.
14. Li, M.; Li, J.; Zhou, Y. Labeled RFS-Based Track-Before-Detect for Multiple Maneuvering Targets in the Infrared Focal Plane Array. *Sensors* **2015**, *15*, 30839–30855. [CrossRef] [PubMed]
15. Mahler, R.P.S. Multitarget Bayes filtering via first-order multitarget moments. *IEEE Trans. Aerosp. Electron. Syst.* **2004**, *39*, 1152–1178. [CrossRef]
16. Mahler, R. PHD filters of higher order in target number. *IEEE Trans. Aerosp. Electron. Syst.* **2008**, *43*, 1523–1543. [CrossRef]

17. Vo, B.T.; Vo, B.N.; Cantoni, A. The Cardinality Balanced Multi-Target Multi-Bernoulli Filter and Its Implementations. *IEEE Trans. Signal Process.* **2009**, *2*, 409–423.
18. Hoseinnezhad, R.; Vo, B.N.; Vo, B.T.; Suter, D. Visual tracking of numerous targets via multi-Bernoulli filtering of image data. *Pattern Recognit.* **2012**, *45*, 3625–3635. [CrossRef]
19. Zhou, X.; Li, Y.F.; He, B. Entropy distribution and coverage rate-based birth intensity estimation in GM-PHD filter for multi-target visual tracking. *Signal Process.* **2014**, *94*, 650–660. [CrossRef]
20. Vo, B.T.; Vo, B.N. Labeled Random Finite Sets and Multi-Object Conjugate Priors. *IEEE Trans. Signal Process.* **2013**, *61*, 3460–3475. [CrossRef]
21. Vo, B.N.; Vo, B.T.; Phung, D. Labeled Random Finite Sets and the Bayes Multi-Target Tracking Filter. *IEEE Trans. Signal Process.* **2014**, *62*, 6554–6567. [CrossRef]
22. Papi, F.; Vo, B.T.; Bocquel, M.; Vo, B.N. Multi-target Track-Before-Detect using labeled random finite set. In Proceedings of the 2013 International Conference on Control, Automation and Information Sciences (ICCAIS), Nha Trang, Vietnam, 25–28 November 2013; pp. 116–121.
23. Papi, F.; Vo, B.N.; Vo, B.T.; Fantacci, C.; Beard, M. Generalized Labeled Multi-Bernoulli Approximation of Multi-Object Densities. *IEEE Trans. Signal Process.* **2015**, *63*, 5487–5497. [CrossRef]
24. Barnich, O.; Droogenbroeck, M.V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. *IEEE Trans. Image Process.* **2011**, *20*, 1709. [CrossRef] [PubMed]
25. Elgammal, A.; Duraiswami, R.; Harwood, D.; Davis, L.S. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proc. IEEE* **2002**, *90*, 1151–1163. [CrossRef]
26. Zivkovic, Z. Improved adaptive Gaussian mixture model for background subtraction. In Proceedings of the 17th International Conference on Pattern Recogntion, Cambridge, UK, 23–26 August 2004; Volume 2, pp. 28–31.
27. Xu, Y.; Dong, J.; Zhang, B.; Xu, D. Background modeling methods in video analysis: A review and comparative evaluation. *CAAI Trans. Intell. Technol.* **2016**, *1*, 43–60. [CrossRef]
28. Haritaoglu, I.; Harwood, D.; David, L.S. *W4: Real-Time Surveillance of People and Their Activities*; IEEE Computer Society: Washington, DC, USA, 2000; pp. 809–830.
29. Punchihewa, Y.; Papi, F.; Hoseinnezhad, R. Multiple target tracking in video data using labeled random finite set. In Proceedings of the International Conference on Control, Automation and Information Sciences, Gwangju, South Korea, 2–5 December 2014; pp. 13–18.
30. Reuter, S.; Vo, B.T.; Vo, B.N.; Dietmayer, K. The Labeled Multi-Bernoulli Filter. *IEEE Trans. Signal Process.* **2014**, *62*, 3246–3260.
31. Papi, F.; Gostar, A.K. Bayesian Track-Before-Detect for closely spaced targets. In Proceedings of the IEEE 2010 International Conference on Signal Processing Conference, Beijing, China, 24–28 October 2010, pp. 1979–1983.
32. Ester, M.; Kriegel, H.P.; Xu, X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
33. Davis, J.W.; Sharma, V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182. [CrossRef]
34. Wu, Z.; Fuller, N.; Theriault, D.; Betke, M. A thermal infrared video benchmark for visual analysis. In Proceedings of the 10th IEEE Workshop on Perception Beyond the Visible Spectrum, Columbus, OH, USA, 23–28 June 2014.
35. Liu, Q.; He, Z. PTB-TIR: A Thermal Infrared Pedestrian Tracking Benchmark. *arXiv* **2018**, arXiv:1801.05944.
36. Hofmann, M.; Tiefenbacher, P.; Rigoll, G. Background segmentation with feedback: The Pixel-Based Adaptive Segmenter. In Proceedings of the Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; pp. 38–43.
37. Sun, N.; Jiang, F.; Yan, H.; Liu, J.; Han, G. Proposal generation method for object detection in infrared image. *Infrared Phys. Technol.* **2017**, *81*, 117–127. [CrossRef]
38. Wang, Z.; Lai, M.J.; Lu, Z.; Fan, W.; Davulcu, H.; Ye, J. Orthogonal Rank-One Matrix Pursuit for Low Rank Matrix Completion. *SIAM J. Sci. Comput.* **2014**, *37*, A488–A514. [CrossRef]
39. Yi, X.; Park, D.; Chen, Y.; Caramanis, C. *Fast Algorithms for Robust PCA via Gradient Descent*; Neural Information Processing Systems Foundation Inc.: La Jolla, CA, USA, 2016.

40. Sobral, A. BGSLibrary: An OpenCV C++ Background Subtraction Library. In Proceedings of the IX Workshop De Visio Computacional, Rio de Janeiro, Brazil, 3–5 June 2013.
41. Schuhmacher, D.; Vo, B.T.; Vo, B.N. A consistent metric for performance evaluation of multi-object filters. *IEEE Trans. Signal Process.* **2008**, *56*, 3447–3457. [CrossRef]