A TRACK-BEFORE-DETECT ALGORITHM USING JOINT PROBABILISTIC DATA ASSOCIATION FILTER AND INTERACTING MULTIPLE MODELS

Andrea Mazù Simone Chiappino Lucio Marcenaro Carlo S. Regazzoni

DITEN, University of Genoa
Via Opera Pia 11A 16145 Genoa – Italy
{andrea.mazu, s.chiappino, mlucio, carlo} @ginevra.dibe.unige.it

ABSTRACT

Detection of dim moving point targets in cluttered background can have a great impact on the tracking performances. This may become a crucial problem, especially in low-SNR environments, where target characteristics are highly susceptible to corruption. In this paper, an extended target model, namely Interacting Multiple Model (IMM), applied to Track-Before-Detect (TBD) based detection algorithm, for far objects, in infrared (IR) sequences is presented. The approach can automatically adapts the kinematic parameter estimations, such as position and velocity, in accordance with the predictions as dimensions of the target change. A sub-par sensor can cause tracking problems. In particular, for a single object, noisy observations (i.e. fragmented measures) could be associated to different tracks. In order to avoid this problem, presented framework introduces a cooperative mechanism between Joint Probabilistic Data Association Filter (JPDAF) and IMM. The experimental results on real and simulated sequences demonstrate effectiveness of the proposed approach.

Index Terms— IR sequences, Track-Before-Detect, JPDAF, extended objects, IMM.

1. INTRODUCTION

In the past decades, infrared (IR) target detection and tracking have received considerable attention in civil and military applications, such as infrared precise guidance, early warning, video surveillance, search and tracking [15], etc. The increasing interest in IR sensors is related to their capability of reconstructing a scene from thermal radiation, even in poorly illuminated scenes [1].

The infrared search and track (IRST) systems are mostly used to detect long-range maritime and airborne threats. Because of its typically long distance, the target in IR images is small, weak and embedded in strong clutter [13]. This presents a motivation for the development of the track-before-detect (TBD) framework. TBD algorithms use information of more frames to make the detection take advantage of the observation that the trajectories of the targets are continuous and consistent, while those of noise are not.

Several approaches related to the TBD paradigm are proposed in the literature. Zhang et al. proposed in [5] a TBD method, which first reduces the 3D spatio-temporal scanning to 2D spatial hunting; and then used the Constant False Alarm Rate (CFAR) method to find the consistent trajectories. In [16], an auxiliary particle filter with a modified measurement likelihood function is used. Gaussian particle filter based Quasi-Monte Carlo sampling has been proposed in [3], where the detection phase relies on the difference between the estimation covariance in the step k-n and k.

Also, in [9] and [8], a dynamic programming approach was proposed for the TBD paradigm, to detect and track a unique target, while in [12] an extension to the multi-targets case was addressed.

Wei in [7] used a band-stop filter and an adaptive Hough transform-based processing to search lines that may have been produced by the targets.

Boers et al. in [14] dealt with extended objects. They concluded that the accuracy in the position and velocity estimates improves if the dimension of the object is also considered, due to the virtual accelerations introduced by the change of size.

In this work, a TBD algorithm for the detection and tracking of moving point in a 3D space is presented. An Interacting Multiple Model (IMM) is used to distinguish the movements in the image plane from the movements towards the sensor and two independent Joint Probabilistic Data Association Filters (JPDAFs) are employed. In the first one, a constant size is assumed while in the second a growing dimension is considered. The filter that produces the lowest error is chosen for the state estimation. The proposed approach performances can also improve in the case of a non-ideal sensor. Specifically, in case of noisy measures a single object could be fragmented. Therefore, different observations of the same object should be associated to different tracks (i.e. generating false targets).

The remainder of this paper is organized as follows: in section 2, the theoretical background for the proposed algorithm is briefly discussed. In section 3, the proposed TBD algorithm with the use of JPDAF and IMM is presented in detail, while Section 4 presents and discusses
the experimental results of the proposed method. Conclusions are given in section 5.

2. BACKGROUND

2.1. Joint Probabilistic Data Association Filter (JPDAF)

The JPDAF provides an optimal solution for the data association problem when more than one target is present in the scene [6].

The main idea is to calculate jointly the update step of multiple trackers.

Consider the problem of tracking T objects.

Let \( X^k = \{x^k_1, \ldots, x^k_T \} \) denote the state of these objects at time \( k \). Furthermore, let \( Z(k) = \{z_1(k), \ldots, z_m(k) \} \) denotes a set of measurements (i.e. observations) at time \( k \), where \( z_j(k) \) is one feature of such a measurement. \( Z_k \) is the sequence of all measurements up to time \( k \).

Two sets of conditional independence relations are assumed: \( z(k) \) is independent of all other observations and states given \( X^k \); \( X^k \) is independent of \( X^1, X^2, \ldots, X^{k-1} \) given \( X^k \).

In the JPDAF framework, a joint association event \( \theta \) is a set of pairs \( (j,i) \in \{0, \ldots, m \} \times \{1, \ldots, T \} \). Each \( \theta \) uniquely determines which feature is assigned to which object. Let \( \Theta_{ji} \) denote the set of all valid joint association events which assign feature \( j \) to the object \( i \). At time \( k \), the JPDAF computes the posterior probability that feature \( j \) is caused by object \( i \) according to

\[
\beta_{ij} = \sum_{\theta \in \Theta_{ji}} P(\theta | Z^k).
\]

In the framework of Bayesian filtering, the update equation for the prediction of the new state of an object is

\[
p(x^k_t | Z^{k-1}) = \int p(x^k_t | x^k_{t-1}, t) p(x^k_{t-1} | Z^{k-1}) \, dx^k_{t-1},
\]

where \( t \) denotes the time expired since the previous update. Whenever new sensory input arrives, the state update is performed according to

\[
p(x^k_t | Z^k) = \alpha p(Z(k) | x^k_t) p(x^k_t | Z^{k-1}),
\]

where \( \alpha \) is a normalization factor. Since we do not know which of the features in \( Z(k) \) is caused by \( i \)-th object, we integrate the single features according to the assignment probabilities \( \beta_{ji} \) [4]

\[
p(x^k_t | Z^k) = \alpha \sum_{j=0}^{m_k} \beta_{ji} p(z_j(k) | x^k_t) p(x^k_t | Z^{k-1}).
\]

2.2. Interacting Multiple Model (IMM)

The tracking of a single target is based on the choice of a model to describe the movement of an object. So, the filter runs according to this model. In this way, since the choice is made offline, can be present a mismatch between the target mode and the filter model.

So one can use more than one model to operate the filtering and to reduce the error in the state estimation.

The flowchart of this kind of strategy can be summarized in these three points:

1. To assume a set of models as possible candidates;
2. To run a bank of filters, each based on a unique model in the set;
3. To generate the overall estimates by means of the results of the elemental filters.

Using IMM, the Bayesian Network can be represented as shown in Fig. 1.

How can be noticed by the Bayesian Network, the variable \( \beta \) can switch the model from one to another by choosing that one that better approximates the object behavior.

![Fig 1: Graphical model of IMM.](image)

For this work, two dynamical models are used: in the first one, it is assumed that the dimensions of the objects do not change on time (i.e. so called constant size filter). While in the second one, a growing size has been designed by adding a constant value to the present dimension or to a multiple of this (i.e. so called growing size filter).

The state vector is defined as:

\[
x^k_t = [x^k_x, \hat{x}^k_x, y^k_y, \hat{y}^k_y, d^x_k, d^y_k, 1]^T,
\]

where \( x \) and \( y \) lie parallel to the image plane and \( d_x \) and \( d_y \) are the maximal extension of the object on the \( x \) and \( y \) axis respectively. Let us suppose the motion of a point target follows the model

\[
x^{k+1}_i = A_i x^k_i + n^{k+1}_i
\]

where the state transition matrix models the object dynamics and \( n^{k+1} \) is a zero mean Gaussian vector with covariance matrix \( Q \). Since there are two models, there will be two transition matrices, one for the growing objects (\( A_1 \)), and one for constant size objects (\( A_2 \)):

\[
A_1 =
\begin{bmatrix}
1 & \Delta T & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & \Delta T & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & \epsilon_1 \\
0 & 0 & 0 & 0 & 0 & 1 & \epsilon_2 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
3. PROPOSED METHOD

The proposed method with IMM approach is able to correctly detect and track small object moving in image plane or in 3D space.

The method performs in two steps. Firstly, an image preprocessing is performed to find the brightest pixels in the scene and then, the system attempts to track this points for a number of frames. When a large number of measures are associated, the detection is done; else, they are considered as noise.

The image preprocessing consists of limiting the atmosphere non-linearity and suppressing the complex background.

The pixels intensity in the IR images are strictly related to the temperature of the objects, but for the background the temperature depends on the ground height. Therefore, pixels in the same row will have roughly the same temperature and the same intensity [11]. To remove the non-linear distribution of the background due to the atmosphere temperature, the mean of the row can be subtracted from each pixel belonging to that row [5].

Considering an M x N image, the new intensity value of each pixel is

\[
P_{ij} = P_{ij} - \frac{1}{N} \sum_{j=1}^{N} P_{ij}
\]

After this operation, let us consider the temperature non-linear distribution effectively suppressed. Now, to extract the brightest points in the frame, a CFAR method is introduced.

Since the objects are very small, the biggest of the image will be background. Therefore, if a Gaussian distribution of the background is assumed, it can also be assumed that the parameters of the Gaussian are equal to the mean and the standard deviation of the image.

By fixing a false alarm probability, this equals to

\[
P_{FA} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = Q(\lambda)
\]

Therefore, the threshold can be calculated as follows:

\[
T = \mu_b + \lambda \sigma_b
\]

where \( \mu_b \) and \( \sigma_b \) represent mean and standard deviation of the image, and \( \lambda = Q^{-1}(P_{FA}) \).

After these operations, we have a binary image that includes all the potential targets.

In order to recover the real dimension and shape of the object, an object refinement method is used.

In IR images, objects with different levels of IR radiation appear as different gray-levels [15]. A good segmentation algorithm can generate a labeled image that contains information about position and size of the objects.

To this purpose, Simple Linear Iterative Clustering (SLIC) is used [10]. This method allows of finding different groups of pixels, i.e. super-pixels.

After this first part, two JPDAFs run to track the potential targets. The data association step is influenced by the proximity and dimension of the measures.

At each step, the variance of the track of the covariance matrices is considered.

Lower values of the variance of the position, lower will be the error between real and the estimated position thus increasing reliability of the filter. Therefore, the update with lower track of the covariance matrix is chosen and the state of the other filter is re-initialized with this information.

![Fig 2: Result of the detection for “Plane Motion and Tracking” sequence collection 1. Top: original sequence (frame 145, 151 and 157). Down: results of the detection and tracking.](image1)

![Fig 3: Result of the detection for [18]. Top: original sequence (frame 861, 870 and 888). Down: results of the detection and tracking.](image2)

4. EXPERIMENTAL RESULTS

In order to test the performance of the system, real and synthetic sequences have been used. The real image sequences, from OTCBVS Benchmark Dataset Collection [17] and [18], contain small objects under strong cluster.

Fig 2 and Fig 3 show the results of the detection and tracking of objects for the “Plane Motion and Tracking” sequence collection 1 [17] and for the sequence of [18].

In the first sequence, the plane size does not change and the filter that applies the dynamic model \( A_2 \) (i.e. constant size) performs better than \( A_1 \) (i.e. growing size).
In Fig 4 the covariance matrices tracks are plotted. In this case, the constant size filter presents a lower covariance matrices curve trend than the growing size filter. In the second sequence, the object size changes. In this case, the $A_1$ collaborates with the $A_2$ computing an optimal estimation of the object state. A comparison between the tracks of the covariance matrices are shown in Fig 5. The proposed method is able to adapt its parameter in according to the object movements. Also in this case the combined approach, of constant and growing size filters, presents a covariance matrix track curve lower than the single filters. In this part of the experiment, we want investigate the fragmentation problem due to not ideal sensors. To this end, we have simulated four image sequences formed by 150 frames 25 [fps]. Each sequence has been characterize by a different fragmentation percentage of the object: 0%, 30%, 50% and 80% of the total number of target pixel. In such sequences a moving object is changing its size, e.g. while it is approaching to the sensor. When the object size changes, this fragmentation can generate a greater number of false measures and a single constant size filter fails the tracking. Fig. 6 and Fig. 7 show a comparison between the proposed method and the constant size filter. Specifically, Fig. 6 presents the differences between the tracks of the covariance matrices, while in Fig. 8 the differences between the position errors are drawn. The results show how the proposed method compute a more reliable object tracking, which presents a lower error between observed and estimated state vector.

5. CONCLUSIONS

In this paper, an Interacting Multiple Model (IMM) is used to distinguish the movements in the image plane from the movements towards the sensor and two independent Joint Probabilistic Data Association Filters (JPDAFs) are employed. The proposed method has been tested on IR and visible image sequences. The proposed IMM is able to optimize the state estimation compared with the single constant and growing size filters. However, we point out that cooperative approach, between constant and growing size filters, presents a high level of improvements in the case of fragmentation problems. Future developments of this work will include a Bayesian representation for dynamical modelling the switching between different filters.
REFERENCES


