

# A Metric for Performance Evaluation of Multi-Target Tracking Algorithms

Branko Ristic<sup>\*</sup>, Ba-Ngu Vo<sup>†</sup>, Daniel Clark<sup>‡</sup>, Ba-Tuong Vo<sup>§</sup>

## Abstract

Performance evaluation of multi-target tracking algorithms is of great practical importance in the design, parameter optimisation and comparison of tracking systems. The goal of performance evaluation is to measure the distance between two sets of tracks: the ground truth tracks and the set of estimated tracks. This paper proposes a mathematically rigorous metric for this purpose. The basis of the proposed distance measure is the recently formulated consistent metric for performance evaluation of multi-target *filters*, referred to as the OSPA metric. Multi-target filters sequentially estimate the number of targets and their position in the state-space. The OSPA metric is therefore defined on the space of finite sets of vectors. The distinction between *filtering* and *tracking* is that tracking algorithms output *tracks*, and a track represents a labeled temporal sequence of state estimates, associated with the same target. The metric proposed in this paper is therefore defined on the space of finite sets of tracks. Numerical examples demonstrate that the proposed metric behaves in a manner consistent with our expectations.

## Index Terms

Tracking, performance evaluation, estimation

Correspondence submitted to IEEE Trans. Signal Processing

February 22, 2011

<sup>\*</sup>Defence Science and Technology Organisation, ISR Division, Bld 94, M2.30, 506 Lorimer Street, Fishermans Bend, VIC 3207, Australia; Tel: (+61 3) 9626 8226; Fax: +61 3 9626 8341; email: branko.ristic@dsto.defence.gov.au

<sup>†</sup>School of EECE, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia; Tel: (+61 8) 6488 1767; fax: +61 8 6488 1065; email: ba-ngu.vo@uwa.edu.au

<sup>‡</sup>Joint Research Institute in Signal & Image Processing, School of EPS, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, UK; Tel: +44 131 449 5111; email: D. E. Clark@hw.ac.uk

<sup>§</sup>School of EECE, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia; Tel: (+61 8) 6488 1767; fax: +61 8 6488 1065; email: ba-tuong.vo@uwa.edu.au

## I. INTRODUCTION

Multi-target tracking refers to the sequential estimation of the number of targets and their states (positions, velocities, etc.) tagged by a unique label. Hence the output of a tracking algorithm are *tracks*, where a track represents a labeled temporal sequence of state estimates, associated with the same target.

In evaluating the performance of a multi-target tracking algorithm the goal is to measure the distance between two sets of tracks: the set of ground truth tracks and the set of estimated tracks, produced by the tracking algorithm under evaluation. Performance evaluation of multi-target tracking algorithms is of great practical importance, with applications in tracking system design, parameter tuning and tracker comparisons (e.g. in tender evaluations). Consequently the topic has been studied extensively, see for example [1], [2], [3, Ch.13], [4], [5].

The general performance evaluation methodology includes three steps: (1) the creation of a scenario of a certain level of difficulty; (2) running of the multi-target tracking algorithm under assessment; (3) assignment of a performance measure (score) as a measure of the distance between the ground truth and the tracker output. In order to estimate the expected performance, the score is typically averaged over independent Monte Carlo runs. The randomness in each run is caused by the measurements which feed the tracker (a consequence of measurement noise and false or missed detections).

In order to carry out the third step above (assessment), the standard approach to tracker performance evaluation is to first assign the tracker output to the ground truth [1]. Once the assignment is made, a list of operator defined measures of effectiveness (MoEs) are computed. These MoEs describe various aspects of tracking performance, such as [3, Ch.13]: *timeliness* (e.g. track initiation delay, track overshoot), *track accuracy* (e.g. positional, heading, velocity error), *continuity* (e.g. track fragmentation, track labeling swaps), *false tracks* (their count and duration), to name a few.

The problem a practitioner faces is twofold. The first is how to choose the relevant MoEs and the second is how to combine them into a single score (which is typically required for parameter tuning or in tender evaluations). The choice of relevant MoEs is far from clear, since the traditional MoEs listed above are fairly arbitrary with various authors arguing in favour of different ones. Combination of MoEs is even more questionable from the theoretical point of view, since the adopted MoEs can be correlated. To illustrate this point, note that a reduction in the track initiation delay typically increases the number of false tracks (conversely, the reduction of the number of false tracks leads to an increase in the track initiation delay). Another important issue is the “transitive” property of the tracker score resulting from the combination of MoEs. This property can be explained as follows. Consider two

trackers, A and B, and suppose that tracker A output is “close” (as indicated by the combined MoE) to the ground truth. If tracker B output is “close” to that of tracker A, can we then make the conclusion that tracker B output is also “close” to the ground truth? The answer is positive only if the combined MoE possesses the “transitive” property. It is far from clear if any combination of traditional MoEs would satisfy this property.

In order to overcome the aforementioned shortcomings of traditional MoEs, it is necessary to define a mathematically rigorous metric for measuring the distance between two sets of tracks. First of all, such a metric would naturally satisfy the transitive property discussed above, by the virtue of the *triangle inequality* axiom. Furthermore, a metric also has the necessary property that the estimated state is the same as the true state, if and only if the distance between them is zero. These properties are not guaranteed by other MoEs. The development of a metric is also important if we wish to determine the underlying theoretical performance of an algorithm, such as its asymptotic convergence and consistency properties, i.e. whether in the long run the algorithm does converge to the true state. Without a metric, these notions have no meaning.

There are a few known metrics for measuring the distance between any two sets. The starting point is the Hausdorff metric, whose performance was found to be unsatisfactory due to its insensitivity to the difference in the number of elements in the set (cardinality). Hoffman and Mahler [6] then proposed a new metric for sets, based on Wasserstein distance [7], which partly fixed the undesirable cardinality behaviour of the Hausdorff metric. Schuhmacher *et al.* [8], however, subsequently demonstrated a number of shortcomings of the Hoffman-Mahler metric and proposed a new consistent metric for sets, referred to as the *optimal subpattern assignment* (OSPA) metric. The OSPA metric is shown to eliminate all of the shortcomings of the Hoffman-Mahler metric. Schuhmacher *et al.* [8] demonstrated the OSPA metric in the context of performance evaluation of multi-target *filtering algorithms*. These algorithms sequentially estimate the number of currently existing targets and their position in the state-space. The OSPA metric, being a metric between two sets of points in the state space is an appropriate metric for this application: it incorporates both the cardinality error and the spacial distance of points.

For target tracking, however, we require a metric on the space of finite sets of tracks, where a track has been defined as a labeled temporal sequence. The tracks are typically of unequal length in time. The paper proposes an adaptation of the standard OSPA metric for this purpose. The resulting metric measures the distance on the joint state-space—target-label domain. In addition, for the purpose of assigning the tracker output to the ground truth, the global OSPA distance between the pairs of tracks is minimised. The resulting metric is mathematically rigorous and consistent. Preliminary formulation of the OSPA metric for tracks have been announced in [9].

The paper is organised as follows. Section II formally introduces the problem and proposes the conceptual solution based on the OPSA metric. Section III discusses the base distance to be used in OSPA metric for tracks. Numerical examples are presented in Section IV, while the contribution is summarised in Section V.

## II. PROBLEM AND ITS CONCEPTUAL SOLUTION

### A. Problem specification

Our goal is to define a metric on the space of finite sets of objects called *tracks*, defined over discrete-time support points  $\tau = (t_1, t_2, \dots, t_K)$ . A track  $X$  on  $\tau$  is a labeled sequence of length  $K$ :

$$X = (X_1, X_2, \dots, X_K) \quad (1)$$

where  $X_k$ ,  $k = 1, \dots, K$ , is either an empty set or a singleton whose element is  $(\ell, \mathbf{x}_k)$ . Here  $\ell \in \mathbb{N}$  is the track label, which does not change with time, while  $\mathbf{x}_k$  is the time evolving state vector, a point in an  $N$ -dimensional *state space*  $\mathbf{W}$ :  $\mathbf{x}_k \in \mathbf{W} \subseteq \mathbb{R}^N$ . The state space typically involves target position and velocity in the Cartesian coordinates, but may also include other target features, such as amplitude, size, shape, or similar. For convenience let us introduce an indicator  $e_k$ , which takes the value of one, if the track exists at time  $t_k$ , and zero otherwise.

Then we can write:

$$X_k = \begin{cases} \emptyset, & \text{if } e_k = 0, \\ \{(\ell, \mathbf{x}_k)\}, & \text{if } e_k = 1. \end{cases} \quad (2)$$

Our aim is to define a metric at one particular time instant  $t_k$ ,  $k = 1, \dots, K$ . Let the set of all tracks on  $\tau$  at specific time  $t_k$  be denoted  $\mathbb{X}_k$  (i.e.  $X_k \in \mathbb{X}_k$ ). Furthermore, let the set of finite subsets of  $\mathbb{X}_k$  be denoted by  $\mathcal{X}_k$ . We need to define a metric space  $(\mathcal{X}_k, D)$ , where function  $D : \mathcal{X}_k \times \mathcal{X}_k \rightarrow \mathbb{R}_+ = [0, \infty)$  is a metric which satisfies the following three axioms for all  $\mathfrak{X}_k, \mathfrak{Y}_k, \mathfrak{Z}_k \in \mathcal{X}_k$  [10]:

- *identity*:  $D(\mathfrak{X}_k, \mathfrak{Y}_k) = 0$  if and only if  $\mathfrak{X}_k = \mathfrak{Y}_k$ ,
- *symmetry*:  $D(\mathfrak{X}_k, \mathfrak{Y}_k) = D(\mathfrak{Y}_k, \mathfrak{X}_k)$ ,
- *triangle inequality*:  $D(\mathfrak{X}_k, \mathfrak{Y}_k) \leq D(\mathfrak{X}_k, \mathfrak{Z}_k) + D(\mathfrak{Z}_k, \mathfrak{Y}_k)$ .

If, as a matter of convention, we denote by  $\mathfrak{X}_k \in \mathcal{X}_k$  the set of tracks representing the ground truth at  $t_k$ , and by  $\mathfrak{Y}_k \in \mathcal{X}_k$  the set of estimated tracks produced at  $t_k$  by the algorithm under assessment, metric  $D(\mathfrak{X}, \mathfrak{Y}_k)$  should quantify the overall estimation error of the tracking algorithm at time  $t_k$ . In a mathematically rigorous manner, it should combine various aspects of tracking performance (e.g. timeliness, track accuracy, continuity, data association, false tracks, etc) into a single metric.

## B. Conceptual solution

The conceptual solution for proposed the multi-target tracking metric is based on the OSPA metric. The OSPA metric on  $\mathcal{X}_k$  [8] is a distance between any two sets of objects. In our case the objects are tracks at  $t_k$ , and the two sets are:

$$\mathfrak{X}_k = \{(\ell_1, \mathbf{x}_{k,1}), \dots, (\ell_m, \mathbf{x}_{k,m})\} \quad (3)$$

$$\mathfrak{Y}_k = \{(s_1, \mathbf{y}_{k,1}), \dots, (s_n, \mathbf{y}_{k,n})\}. \quad (4)$$

According to the convention,  $\mathfrak{X}_k \in \mathcal{X}_k$  and  $\mathfrak{Y}_k \in \mathcal{X}_k$  are the existing ground truth tracks and tracker produced tracks at  $t_k$ , respectively. The cardinalities of sets  $\mathfrak{X}_k$  and  $\mathfrak{Y}_k$ ,  $m$  and  $n$  respectively, depend on  $k$  due to (2). For  $m \leq n$ , the OSPA distance between  $\mathfrak{X}_k$  and  $\mathfrak{Y}_k$  is defined as:

$$D_{p,c}(\mathfrak{X}_k, \mathfrak{Y}_k) = \left[ \frac{1}{n} \left( \min_{\pi \in \Pi_n} \sum_{i=1}^m \left( d_c(\tilde{\mathbf{x}}_{k,i}, \tilde{\mathbf{y}}_{k,\pi(i)}) \right)^p + (n-m) \cdot c^p \right) \right]^{1/p} \quad (5)$$

where  $\tilde{\mathbf{x}}_{k,i} \equiv (\ell_i, \mathbf{x}_{k,i})$ ,  $\tilde{\mathbf{y}}_{k,\pi(i)} \equiv (s_{\pi(i)}, \mathbf{y}_{k,\pi(i)})$  and

- $d_c(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \min(c, d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}))$  is the *cut-off distance* between two tracks at  $t_k$ , with  $c > 0$  being the cut-off parameter;
- $d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  is the *base distance* between two tracks at  $t_k$ ;
- $\Pi_n$  represents the set of permutations of length  $m$  with elements taken from  $\{1, 2, \dots, n\}$ ;
- $1 \leq p < \infty$  is the OSPA metric order parameter.

For the case  $m > n$ , the definition is  $D_{p,c}(\mathfrak{X}, \mathfrak{Y}) = D_{p,c}(\mathfrak{Y}_k, \mathfrak{X}_k)$ . If both  $\mathfrak{X}_k$  and  $\mathfrak{Y}_k$  are empty sets (i.e.  $m = n = 0$ ), the distance is zero. The proof that OSPA distance is indeed a metric, including a number of arguments that it complies with intuition, is given in [8].

The choice of parameters  $c$  and  $p$  follows the guidelines of the original OSPA metric [8]. The cut-off parameter  $c$  determines the relative weighting given to the cardinality error component against the base distance error component. Larger values of  $c$  tend to emphasize cardinality errors and vice versa. Note that the cardinality error is a consequence of an unequal number of tracks in  $\mathfrak{X}$  and  $\mathfrak{Y}$ , meaning that  $c$  can be interpreted as a measure of penalty assigned to missed or false tracks. The order parameter,  $p$ , controls the penalty assigned to ‘‘outlier’’ estimates (that are not close to any of the ground truth tracks). A higher value of  $p$  increases sensitivity to outliers.

The OSPA metric is increasingly being used as a performance measure in estimation and has already made an impact in the development of new filters and estimators, such as the Set JPDA filter [11] and the minimum mean OSPA estimator [12].

In order to apply the OSPA metric for tracks described above, we need to define the base distance  $d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  between tracks  $\tilde{\mathbf{x}} \equiv (\ell, \mathbf{x})$  and  $\tilde{\mathbf{y}} \equiv (s, \mathbf{y})$ . The base distance must be a proper metric (i.e. it needs to satisfy the three axioms) and will be discussed next.

### III. THE BASE DISTANCE AND LABELING OF ESTIMATED TRACKS

#### A. The base distance between two labeled vectors

The base distance  $d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  is a metric the on space  $\mathbb{N} \times \mathbb{R}^N$  and is defined as:

$$d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \left( d(\mathbf{x}, \mathbf{y})^{p'} + d(\ell, s)^{p'} \right)^{1/p'}, \quad (6)$$

where

- $1 \leq p' < \infty$  is the base distance order parameter;
- $d(\mathbf{x}, \mathbf{y})$  is the localisation base distance, a metric on  $\mathbb{R}^N$ , typically adopted as the  $p'$ -norm:  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{p'}$ ;
- $d(\ell, s)$  is the labeling error, a metric on  $\mathbb{N}$ , adopted as follows:

$$d(s, t) = \alpha \bar{\delta}[s, t] \quad (7)$$

where  $\bar{\delta}[i, j]$  is the complement of the Kroneker delta, that is  $\bar{\delta}[i, j] = 0$  if  $i = j$ , and  $\bar{\delta}[i, j] = 1$  otherwise.

Parameter  $\alpha \in [0, c]$  in (7) controls the penalty assigned to the labeling error  $d(s, t)$  interpreted relative to the localisation distance  $d(\mathbf{x}, \mathbf{y})$ . The case  $\alpha = 0$  assigns no penalty, and  $\alpha = c$  assigns the maximum penalty.

It is straightforward to verify that the base distance  $d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  defined by (6) satisfies the axioms of a metric. To prove identity and symmetry is trivial. To prove the triangular inequality, let  $\tilde{\mathbf{x}} = (\ell, \mathbf{x})$ ,  $\tilde{\mathbf{y}} = (s, \mathbf{y})$ ,  $\tilde{\mathbf{z}} = (u, \mathbf{z})$  be any three labeled vectors in  $\mathbb{N} \times \mathbb{R}^N$ . Then we have:

$$\begin{aligned} d(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})^{p'} &= \|\mathbf{x} - \mathbf{y}\|_{p'}^{p'} + \alpha^{p'} \bar{\delta}[\ell, s] \\ &\leq \|\mathbf{x} - \mathbf{z}\|_{p'}^{p'} + \|\mathbf{z} - \mathbf{y}\|_{p'}^{p'} + \alpha^{p'} (\bar{\delta}[\ell, u] + \bar{\delta}[u, s]) \\ &= d(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})^{p'} + d(\tilde{\mathbf{z}}, \tilde{\mathbf{y}})^{p'} \end{aligned}$$

Having defined the base distance (6), we are still not in the position to apply the OSPA metric for tracks. What remains is to determine the labels of estimated tracks.

#### B. Labeling of estimated tracks

In order to explain why this step is necessary, consider an illustrative example shown in Fig.1. There are two true tracks, labeled as  $\ell_1$  and  $\ell_2$ , (shown in thin black lines) and four estimated tracks labeled as  $s_1 \dots, s_4$  (thick orange

lines). In order to apply the OSPA metric in the framework described above, it is necessary to assign the labels of true tracks to some of the estimated tracks (in this example  $\ell_1$  and  $\ell_2$ ) in a globally optimum manner. For the case shown in Fig.1 it would be desirable to assign  $\ell_1$  to track  $s_1$ , because estimated track  $s_1$  is the best approximation of  $\ell_1$ . For the same reason we should assign  $\ell_2$  to  $s_3$ , and also make sure that the other two estimated tracks are assigned labels which are different from both  $\ell_1$  and  $\ell_2$ . Recall from the introduction that the assignment of the tracker output to the ground truth is required in all standard tracking evaluation schemes [1].

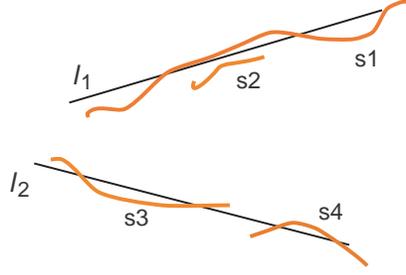


Figure 1. Example with two true tracks labeled as  $\ell_1$  and  $\ell_2$  (thin black lines) and four estimated tracks, labeled as  $s_1, \dots, s_4$ .

A simple way to carry out the assignment task is by using one of the existing two-dimensional assignment algorithms (Munkres, JVC, auction, etc) [3, p.342]. Suppose the sets of ground truth tracks and estimated tracks are  $\{X^{(1)}, \dots, X^{(L)}\}$  and  $\{Y^{(1)}, \dots, Y^{(R)}\}$ , respectively. Using the existence indicators for tracks, a track  $X^{(\ell)}$  at time  $k$ , denoted  $X_k^{(\ell)}$ , is defined as in (2). An optimal global assignment  $\lambda^*$  can be determined as follows ( $L \leq R$ ):

$$\lambda^* = \arg \min_{\lambda \in \Lambda_R} \sum_{\ell=1}^L \sum_{k=1}^K \left[ e_k^\ell e_k^{\lambda(\ell)} \min(\Delta, \| \mathbf{x}_k^\ell - \mathbf{y}_k^{\lambda(\ell)} \|_2) + (1 - e_k^\ell) e_k^{\lambda(\ell)} \Delta + e_k^\ell (1 - e_k^{\lambda(\ell)}) \Delta \right] \quad (8)$$

where

- $\Lambda_R$  is the set of permutations of length  $L$  with elements taken from  $\{1, 2, \dots, R\}$ ;
- $\Delta$  is the penalty assigned to instantaneous miss or false track;
- $e_k^\ell$  and  $e_k^{\lambda(\ell)}$  are existence indicators for true and  $\lambda$ -assigned estimated tracks.

The case  $L > R$  is a trivial modification of (8).

A close inspection of (8) reveals that the term in the square brackets is in fact the OSPA distance between  $X_k^{(\ell)}$  and  $Y_k^{\lambda(\ell)}$ , with  $p = 2$  and  $c = \Delta$  (recall from (2) that these two sets are either empty or singletons, hence the form in (8)). Thus the assignment of estimated tracks to ground-truth tracks is carried out by minimising the global OSPA distance between the pairs of tracks, accumulated over the discrete-time points  $\tau$ . Parameter  $\Delta$  controls the penalty assigned when at time  $t_k$  one of the tracks exists and the other does not. A high value of  $\Delta$  will favour

the assignment of longer duration estimated tracks to the true tracks. If an estimated track has been assigned by  $\lambda^*$  to a true track with label  $\ell$ , then its label is set to  $\ell$  too. Estimated tracks which remain unassigned according to  $\lambda^*$ , are given labels different from all the true track labels.

The basic steps in the computation of the proposed OSPA based metric for tracks are given in Algorithm 1.

---

**Algorithm 1** Computation Steps of the Proposed Metric

---

```

1: function OSPA-T( $\{X^{(1)}, \dots, X^{(L)}\}, \{Y^{(1)}, \dots, Y^{(R)}\}$ )
2:   % Label the estimated tracks
3:   For  $j = 1, \dots, R$ , Label[ $Y^{(j)}$ ] =  $I$  (initial value, different from all true track labels)
4:   Find  $\lambda^*$ , the globally optimum assignment of tracks  $\{X^{(1)}, \dots, X^{(L)}\}$  to  $\{Y^{(1)}, \dots, Y^{(R)}\}$ 
5:   For  $i = 1, \dots, L$ , Label[ $Y^{(\lambda^*(i))}$ ] = Label[ $X^{(i)}$ ]
6:   % Compute the distance
7:   For  $k = 1, \dots, K$ 
8:     Form the labeled sets  $\mathfrak{X}_k$  and  $\mathfrak{Y}_k$  at  $t_k$ , see eqs. (3) and (4)
9:     Compute the OSPA distance at  $t_k$  using (5) with the base distance (6)
10: end function

```

---

#### IV. NUMERICAL EXAMPLES

The proposed OSPA based metric for tracks is demonstrated using a scenario consisting of five crossing targets, shown in Fig. 2. The targets are moving in the  $x - y$  plane, from left to right. The total observation interval is  $K = 120$  time-steps long. The targets cross (effectively are on top of each other) at  $k = 71$ . Their duration varies: the start-time indices are  $k = 3, 5, 7, 9, 10$ ; the termination-time indices are  $k = 101, 103, 105, 108, 109$ . The sensor is placed at  $(200, 200)$ m and measures the range and bearing, with zero-mean white Gaussian noise. The measurement noise standard deviations are  $\sigma_r = 3$ m and  $\sigma_\theta = 1^\circ$ . The multi-target tracking algorithm used in demonstration is based on the Bernoulli/JoTT filter [13] and includes the linear multi-target (LM) data association [14] (details omitted, being irrelevant for the demonstration of the metric).

The numerical simulations have been carried out for four cases, listed below. Unless otherwise stated, the parameters of the proposed metric were set to:  $\Delta = 100$ m,  $\alpha = 25$ m,  $c = 25$ m and  $p = p' = 1$ . It was found that the metric is insensitive to the global assignment parameter  $\Delta$ , as long as  $\Delta > 10$ m. The choice of  $\alpha$  will be discussed in the fourth case. The state space consisted of target positions in the  $x - y$  plane, only. In all the figures shown below, the curves of the proposed metric were obtained by averaging over 100 Monte Carlo runs.

*a) Variation of the crossing angle:* Fig. 3 shows the proposed metric for two different scenarios: the solid line corresponds to the crossing angle of  $\beta = 30^\circ$ ; the dashed line to the crossing angle of  $\beta = 15^\circ$ . In both cases

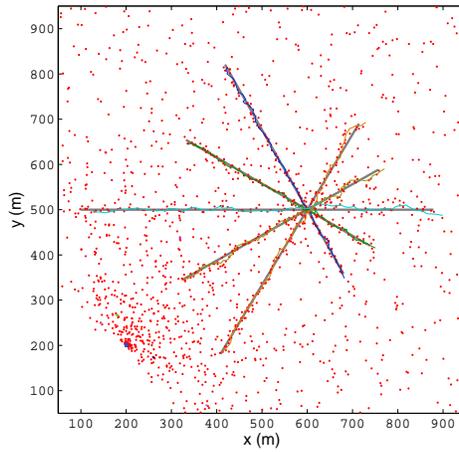


Figure 2. The testing scenario consisting of five crossing targets (crossing angle between the tracks is  $30^\circ$ ).

the probability of detection was  $P_d = 0.9$  and the average number of clutter points (false detection) per scan was  $\mu = 10$ . The general trend in the OSPA distance for tracks in this scenario is that it has high values initially and at the end of the scenario. This is due to track initiation delay and the track termination delay. Furthermore, we observe that the OSPA metric for the two different scenarios does not change until the target crossing event (after  $k = 71$ ). After this point of time, the distance between the ground truth and the tracker output becomes much higher for the smaller crossing angle  $\beta$  scenario. The result is in accordance with our intuition. There is no reason for the difference in the metric before the crossing, as the tracks are well separated. After the crossing, when some tracks are swapped or broken due to the difficulties in measurement-to-track association, the small  $\beta$  angle scenario results in the higher OSPA distance for tracks: this is to be expected as small  $\beta$  case is much more difficult for tracking.

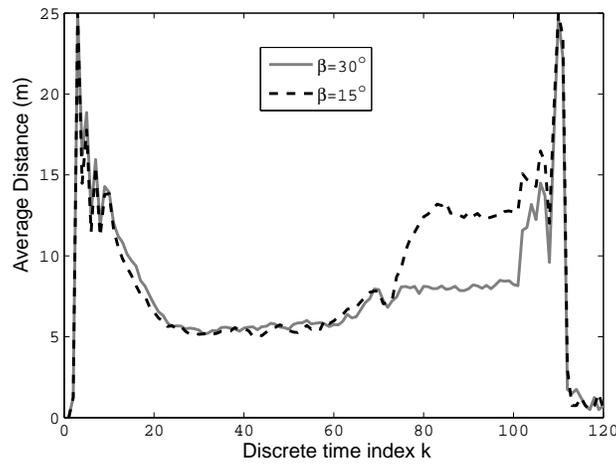


Figure 3. Average OSPA distance for tracks: different crossing angles  $\beta$  case.

*b) Variation of the probability of detection:* Fig. 4 shows the proposed metric for  $P_d = 0.9$  (thin solid line with square markers),  $P_d = 0.75$  (solid thick line) and  $P_d = 0.6$  (dashed line). In all cases  $\beta = 30^\circ$  and  $\mu = 10$ . We observe that the higher value of  $P_d$  results in the smaller value of OSPA distance between the ground truth tracks and estimated tracks, during the entire observation interval. Again this is in accordance with our intuition. The track initiation and track termination are delayed at lower  $P_d$  (as reflected in the results). In addition, the track estimation accuracy is worse at lower  $P_d$  throughout the observation interval.

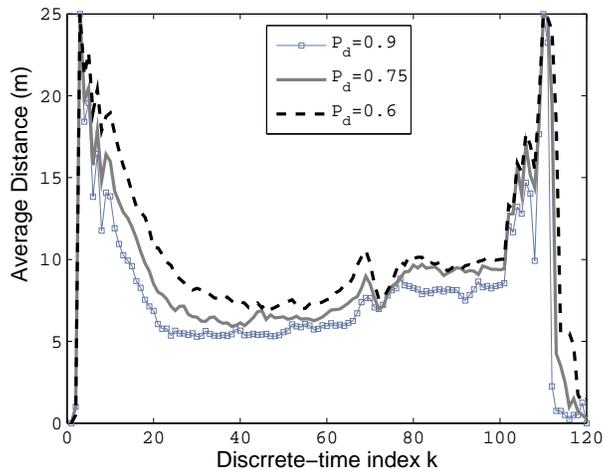


Figure 4. Average OSPA distance for tracks: different values of probability of detection  $P_d$ .

*c) Variation of the density of clutter:* Fig. 5 compares the proposed metric at  $\mu = 1$  (solid line) and  $\mu = 50$  (dashed line). In both cases  $P_d = 0.9$  and  $\beta = 30^\circ$ . In accordance with intuition, the higher the value of  $\mu$ , the greater the OSPA distance between the ground truth and estimated tracks. This is due to the increased probability of false tracks and more difficult data association at higher values of  $\mu$ .

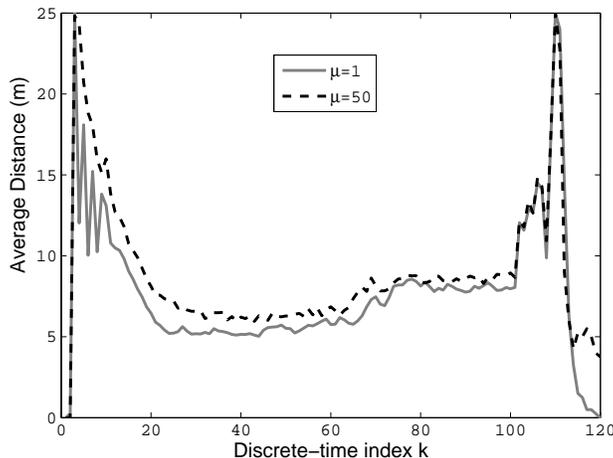


Figure 5. Average OSPA distance for tracks: different values of the average number of false detections per scan  $\mu$ .

*d) Variation of parameter  $\alpha$ :* The parameters of the simulation setup were  $P_d = 0.9$ ,  $\beta = 15^\circ$  and  $\mu = 10$ . Fig. 6 shows the proposed metric using three different values of the metric parameter  $\alpha$ :  $\alpha = 0$ ,  $\alpha = c/2 = 12.5\text{m}$  and  $\alpha = c = 12\text{m}$ . Note that for  $\alpha = 0$ , the proposed metric for tracks is equivalent to the standard OSPA metric [8]. In the considered tracking scenario, the labeling error is expected to appear after the crossing of targets, when track labels are likely to swap. Fig.6 demonstrates this effect and the importance of including the labeling error in the metric.

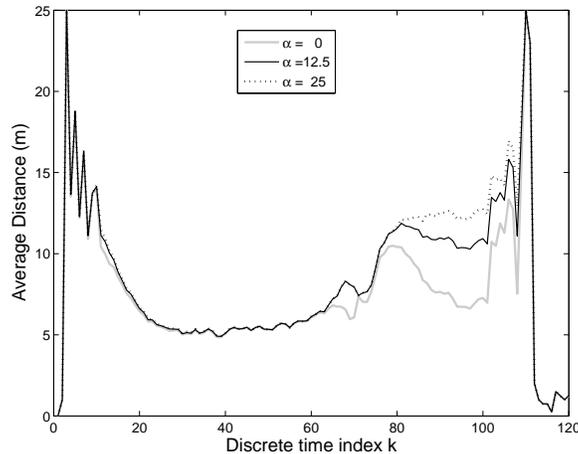


Figure 6. Average OSPA distance for tracks: the effect of  $\alpha$  parameter which penalizes the labeling error.

## V. SUMMARY

The paper introduced a mathematically rigorous metric for measuring the distance between two sets of tracks. Typically one set of tracks is the ground truth, while the other is produced by a tracker. The proposed metric is developed for the purpose of performance evaluation of tracking algorithms, with important implications to tracker design, parameter optimisation and comparison.

The proposed metric captures all aspects of the difference between the two sets of tracks: the cardinality error, the localisation error and the labeling error. The relative weighting of these aspects can be modified by the appropriate selection of the metric parameters. Numerical analysis demonstrates that the OSPA metric for tracks behaves in agreement with our intuition, in the manner consistent with our expectations. The MATLAB source code for the computation of OSPA metric for tracks is available at <http://randomsets.eps.hw.ac.uk/index.html>.

Future work will investigate alternative choices for the localisation base distance introduced in (6), based on probability density functions.

## REFERENCES

- [1] B. E. Fridling and O. E. Drummond, "Performance evaluation methods for multiple target tracking algorithms," in *Proc. SPIE, Signal and Data Processing of Small Targets*, vol. 1481, 1991, pp. 371–383.
- [2] B. Ristic, "A Tool for Track-While-Scan Algorithm Evaluation," in *Proc. Information, Decision, Control (IDC99)*, Adelaide, Australia, Feb. 1999, pp. 105–110.
- [3] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [4] S. B. Colegrove, L. M. Davis, and S. J. Davey, "Performance assessment of tracking systems," in *Proc. Intern. Symp. Signal Proc. and Applic. (ISSPA)*, Gold Coast, Australia, Aug. 1996, pp. 188–191.
- [5] R. L. Rothrock and O. E. Drummond, "Performance metrics for multi-sensor, multi-target tracking," in *Proc. SPIE, Signal and Data Processing of Small Targets*, vol. 4048, 2000, pp. 521–531.
- [6] J. R. Hoffman and R. P. S. Mahler, "Multitarget miss distance via optimal assignment," *IEEE Trans. Systems, Man and Cybernetics - Part A*, vol. 34, no. 3, pp. 327–336, May 2004.
- [7] L. Rueshendorff, "Wasserstein (Vasarshtein) metric," in *Encyclopaedia of Mathematics*, M. Hazewinkel, Ed. Springer, 2001.
- [8] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.
- [9] B. Ristic, B.-N. Vo, and D. Clark, "Performance evaluation of multi-target tracking using the OSPA metric," in *Proc. 13th Int. Conf. Information Fusion*. Edinburgh, UK: ISIF, July 2010.
- [10] E. Kreyszig, *Introductory functional analysis with applications*. Wiley, 1989.
- [11] L. Svensson, D. Svensson, and P. Willett, "Set JPDA algorithm for tracking unordered sets of targets," in *Proc. 12th Int. Conf. Information Fusion*. Seattle, USA: ISIF, July 2009.
- [12] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting two birds with two bullets: how to find minimum mean OSPA estimates," in *Proc. 13th Int. Conf. Information Fusion*. Edinburgh, UK: ISIF, July 2010.
- [13] R. Mahler, *Statistical Multisource Multitarget Information Fusion*. Artech House, 2007.
- [14] D. Musicki and B. L. Scala, "Multi-target tracking in clutter without measurement assignment," *IEEE Trans. Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 877–896, July 2008.