A metric for sets of trajectories that is practical and mathematically consistent

José Bento jose.bento@bc.edu
Jia Jie Zhu zhuuv@bc.edu

***Abstract*—Metrics on the space of sets of trajectories are important for scientists in the field of computer vision, machine learning, robotics, and general artificial intelligence. However, existing notions of closeness between sets of trajectories are either mathematically inconsistent or of limited practical use. In this paper, we outline the limitations in the current mathematically-consistent metrics, which are based on OSPA [1]; and the inconsistencies in the heuristic notions of closeness used in practice, whose main ideas are common to the CLEAR MOT measures [2] widely used in computer vision. In two steps, we then propose a new intuitive metric between sets of trajectories and address these limitations. First, we explain a solution that leads to a metric that is hard to compute. Then we modify this formulation to obtain a metric that is easy to compute while keeping the useful properties of the previous metric. Our notion of closeness is the first demonstrating the following three features: the metric 1) can be quickly computed, 2) incorporates confusion of trajectories' identity in an optimal way, and 3) is a metric in the mathematical sense.**

I. Introduction

Similarity measures for sets of trajectories are very important. In computer vision, they are used to evaluate the performance of multi-object tracking algorithms. If GT and O are the ground-truth and output trajectories of a tracker, a similarity measure $\mathcal{D}$ can be used to distinguish a good tracker from a bad one, i.e. $\mathcal{D}(GT, O) = small$ implies O is a good tracker. In machine learning, algorithms such as [3], [4] and [5], can only cluster, classify and do a nearest neighbor search on sets of trajectories, if we have a similarity measure.

Given their importance, one would expect that existing widely-used measures would be easy to compute and would not produce counter-intuitive results. Surprisingly, this is not the case, leaving a critical problem unsolved.

The main limitation of most similarity measures is that they are not a mathematical metric. For example, the CLEAR MOT measures, widely used to evaluate the performance of trackers, are based on a heuristic and are not a metric. If a measure does not satisfy, for example, the triangle inequality (cf. Section III), then we cannot guarantee that two good trackers (good according to this measure) produce similar outputs, which is counter-intuitive. In addition to this inconsistency, the CLEAR MOT also produce other counter-intuitive results (cf. Section VI). Furthermore, these results indirectly affect many other similarity measures that internally use the CLEAR MOT, or similar heuristics, to precompute an association between two sets of trajectories (e.g. [6]–[9]).

However, even existing measures that are a metric, and that, we believe, are all variants of OSPA, can produce unreasonable results in simple scenarios (cf. Section II). We use OSPA-ST to refer to OSPA applied to sets of trajectories since it was not originally defined for this setting.

On the positive side, both the CLEAR MOT and the OSPA-based metrics are easy to define and have a computation time that scales well with the number and length of trajectories. These two attributes are part of the reason why the CLEAR MOT are popular.

In this paper we introduce the first measures that are simultaneously a mathematical metric, do not replicate the counter-intuitive results of OSPA-ST or CLEAR MOT and can also be quickly computed (cf. Section VII).

In the next section, we use an example to introduce the primary technical challenge in defining a measure for sets of trajectories and how this is addressed by OSPA-ST, the CLEAR MOT, and our metrics. We focus on OSPA-ST and the CLEAR MOT because they give a simple but powerful overview of the main ideas leading up to our work. We include a detailed list of related work in Section V.

II. Main challenge: the association problem

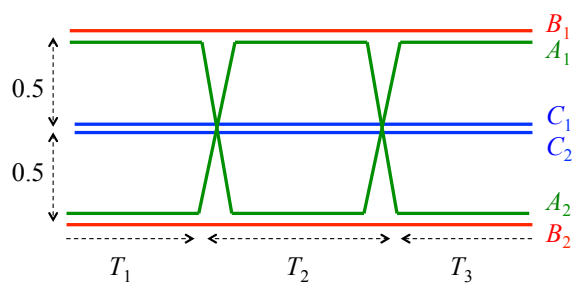

Fig. 1. Two people, A_1 and A_2, move along a line and are followed by two trackers that generate outputs B_1, B_2, C_1 and C_2. For visualization purposes, trajectories that are close are actually on top of each other.

Figure 1 shows three sets of trajectories A, B and C. Set A has the ground truth trajectories of two people that start at a distance 1.0 from each other and, twice, quickly exchange positions; B and C are the trajectories that two different tracking algorithms output. We represent time on the x-axis (left to right) in frames and space on the y-axis (top to bottom). For convenience, T_1, T_2 and T_3 are normalized to sum to 1.0. We seek a distance measure $\mathcal{D}$ for sets of trajectories to determine if B (and/or C) is a good tracker.

If A and B only contained one trajectory, say A_1 and B_1, we could define $\mathcal{D}(A, B)$ by computing the distance between A_1 and B_1 at each frame and returning the average distance over frames. Since this is not the case, the next idea is to define $\mathcal{D}$ in two steps. *Step 1*: determine, for each frame, if we should compute the distance between A_1 and B_1 and between A_2 and B_2 *or* if we should compute the

distance between A_1 and B_2 and between A_2 and B_1. We refer to step 1 as *finding an association* between A and B. If, at a certain frame, we decide to compute the distance between A_1 and B_2 and between A_2 and B_1, then we say that, at that frame, *we associate* A_1 *with* B_2 *and* A_2 *with* B_1. Later, we formally define associations using permutations (See Section III). *Step 2*: given this association, compute the average distance between associated points.

The main problem in defining $\mathcal{D}$ is to establish an association, i.e. step 1 above. Loosely speaking, typical approaches involve choosing an association that makes the average distance between associated points small. We can divide different approaches in three cases. The association between A and B (i) can change freely from frame to frame; (ii) cannot change; and (iii) can change from frame to frame but we pay a (smaller/higher) cost for (smaller/bigger) changes. We call this last cost a *switching cost*. We name the total (or average) distance between associated points the *distance cost*.

In (i), a "good tracker" tracks position accurately even if it does not track peoples' identities (ID) correctly. At each frame, we associate A to B such that the sum of the distance between associated points is minimal. For Fig 1, we obtain an average distance between A and B (over frames and objects) that is $\mathcal{D}(A, B) = 0$ and an average distance between A and C that is $\mathcal{D}(A, C) = 0.5$. Tracker B is better than C. Notice that from T_1 to T_2 track B_1 changes the person it is tracking. We call this as an *identity switch*.

In (ii), the tracker's output B_1 must either be associated with A_1 for all frames or associated with A_2 for all frames. B_2 is associated with the other trajectory in A. Between these two possibilities, we again choose the one that minimizes the average distance (over frame and objects). Now $\mathcal{D}(A, B) = \min\{T_2, 1 - T_2\}$ and $\mathcal{D}(A, C) = 0.5$. Tracker C can be as good as tracker B (for $T_2 = 1/2$), which is counter-intuitive, because C's output is never close to the ground-truth.

In (iii), a "good tracker" must trade off some position accuracy (distance cost) for some ID accuracy (switching cost). The switching cost penalizes identity switches. It is also evident that category (iii) includes (i) and (iii) as particular cases of extreme tradeoffs.

The most widely used measures fall in category (iii). In computer vision, the prototypical example is MOTP, one of the CLEAR MOT measures. At each frame, MOTP heuristically tries to maintain the association between A and B as close as possible to the association made in the previous frame. It makes corrections to the association only if the association created in the previous frame applied to the current frame produces distances between matched points that are larger than a certain threshold thr_{MOT}. In the first frame, MOTP uses the association that minimizes the total distance. (cf. Definition 7 in Appendix A for a formal definition). By changing this threshold one can have different results. We always have $\mathcal{D}(A, C) = 0.5$. If thr_{MOT} is small, then $\mathcal{D}(A, B) = 0$, because we switch association twice. However, if thr_{MOT} is large, $\mathcal{D}(A, B) = T_2$, because we never change from the association made in the first frame. Unfortunately, this heuristic leads to MOTP not being a metric and to other counter-intuitive results (cf. Section VI). In addition, all metrics (based on OSPA) fall in category (ii) because their associations are fix in time. Thus, the applications where they produce intuitive results are limited. We give another example of how OSPA-ST can produce counter-intuitive results in Appendix D.

Like the CLEAR MOT, our measures fall in category (iii). However, ours do not compute the association between A and B heuristically or sequentially. Rather, we solve a *global optimization problem* such that the sum of the distance cost and the switching cost over all time frames is minimized. Thus we avoid counter-intuitive results and obtain a metric.

To appreciate the difference between optimizing associations globally or sequentially, notice that, for Figure 1, MOTP either does not change the association across frames or changes it twice, from T_1 to T_2 and from T_2 to T_3. Regardless of T_1, T_2 and T_3, MOTP never considers changing association just once. This happens because MOTP uses a fixed threshold to decide when to change associations. Hence, it fails to explore possibly better ways in which to compare A and B.

III. Setup and notation

We denote the collection of all finite sets of finite trajectories by S. We reserve the letters A, B and C to represent finite sets of finite trajectories. A_i is the ith trajectory in A. Each trajectory A_i is a finite set of time-state pairs (t, x) with time $t \in \mathbb{N}$ and state $x \in \mathbb{R}^p$. We focus on $t \in \mathbb{N}$ but it is possible to generalize our results to continuous time. We use $A_i(t)$ to represent the state of the ith trajectory in A at time t.

When A and B are defined for all time instants and have the same number of trajectories, we can express $\mathcal{D}(A, B)$ using very simple notation. However, $A_i(t)$ might not be defined for all values of t. In addition, A and B might have a different number of trajectories. Thus, for mathematical convenience, we define a symbol $*$ and the following extension procedure.

Given $A = \{A_i\}_{i=1}^{m_1}$ and $B = \{B_i\}_{i=1}^{m_2}$, we let $m = m_1 + m_2$ and define $A^+ = \{A_i^+\}_{i=1}^{m}$ and $B^+ = \{B_i^+\}_{i=1}^{m}$ as follows. Let T be the largest time index for which either A or B have trajectories with defined states. If $i \leq m_1$, then, for all t such that $A_i(t)$ is defined, we set $A_i^+(t) = A_i(t)$. Otherwise, $A_i^+(t) = *$. If $m_1 < i \leq m$ then $A_i^+(t) = *$ for all $t \in \{1, ..., T\}$. We call these A_i^+, $*$-only trajectories. We define B^+ in the same way. Now A^+ and B^+ both have m trajectories and their states are defined in the extended set $\mathbb{R}^p \cup \{*\}$ for all $t \in \{1, ..., T\}$. We call A^+ and B^+ extended sets of trajectories. Figure 2-(b) exemplifies this procedure. For example, A_1^+ agrees with A_1 for all t except for $t = 3$ for which A_1 is not defined and $A_1^+ = *$.

The meaning of an instant t for which $A_i(t)$ is not defined, i.e. $A_i^+(t) = *$, depends on the application, e.g. it might mean an occlusion. Other interpretations are possible, e.g., an object has yet to come into existence. We refrain for adhering to a particular interpretation of $*$. Our use of $*$ resembles [10], where a null symbol $\emptyset$ is used.

We use $\mathcal{D} : S \times S \mapsto \mathbb{R}_0^+$ to represent distance measures on S. If $A, B \in S$, then $\mathcal{D}(A, B)$ measures the distance between A and B. The main goal of this work is to introduce mathematical metrics that are also easy to compute and produce intuitive results. Recall that, for any $A, B, C \in S$, a metric $\mathcal{D}$ must satisfy the properties (i) **(non-negativity)**

$\mathcal{D}(A,B) \geq 0$, (ii) **(coincidence)** $\mathcal{D}(A,B) = 0$ iff $A = B$, (iii) **(symmetry)** $\mathcal{D}(A,B) = \mathcal{D}(B,A)$ and (iv) **(sub-additivity)** $\mathcal{D}(A,C) \leq \mathcal{D}(A,B) + \mathcal{D}(B,C)$.

Our definition of $\mathcal{D}$ in the following sections needs two ingredients: an association between extended sets of trajectories and a distance between extended trajectories' states.

We formally define an association between two sets of m elements using permutations. A permutation $\sigma : i \mapsto \sigma_i$ is a bijective map from $\{1, ..., m\}$ to itself. Π denotes the set of all permutations. We define the composition of $\sigma, \sigma' \in \Pi$ as $\sigma' \circ \sigma : i \mapsto \sigma'_{\sigma_i}$. σ^{-1} is the inverse map of the bijection σ.

An association $\sigma \in \Pi$ between A and B for which $\sigma_i = j$ tells us that, to compute $\mathcal{D}(A,B)$, we will use distances between the states of A_i^+ and the states of B_j^+ (See e.g. Def. 4). In this case, we say that σ associates trajectory A_i^+ to trajectory B_j^+. Because computing $\mathcal{D}(A,B)$ might involve computing distances between different pairs of trajectories A_i, B_j at different points in time, we extend the terminology "association" to also mean a sequence of permutations. We define $\Pi^T = \{\Sigma : \Sigma = (\Sigma(1), \Sigma(2), ..., \Sigma(T)), \Sigma(t) \in \Pi \ \forall t\}$ as the set of all length-T sequences of associations. $\Sigma_i(t)$ is the image of i by the map $\Sigma(t) \in \Pi$. We define $\Sigma^{-1} \triangleq (\Sigma(1)^{-1}, ..., \Sigma(T)^{-1}) \in \Pi^T$ and $\Sigma' \circ \Sigma \triangleq (\Sigma'(1) \circ \Sigma(1), ..., \Sigma'(T) \circ \Sigma(T)) \in \Pi^T$. An association sequence $\Sigma \in \Pi^T$ between A and B for which $\Sigma_i(t) = j$ tells us that, to compute $\mathcal{D}(A,B)$, we will use the distance between $A_i^+(t)$ and $B_j^+(t)$ (See e.g. Def. 3). In this case, we say that Σ associates the state $A_i^+(t)$ to the state $B_j^+(t)$ at time t.

We use $d : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}_0^+$ to indicate a distance between state elements. We define the extended distance $d^+ : \mathbb{R}^p \cup \{*\} \times \mathbb{R}^p \cup \{*\} \mapsto \mathbb{R}_0^+$ such that for every $x, y \in \mathbb{R}^p$ we have (i) $d^+(x,y) = \min\{2M, d(x,y)\}$, (ii) $d^+(x,*) = d^+(*,x) = M > 0$ and (iii) $d^+(*,*) = 0$. Property (iii) guarantees that placeholders by themselves do not change the distance between A and B. Property (ii) defines a cost M for comparing two states, one of which is not defined. For example, a larger M might indicate a higher penalty for a missed or false track. Property (i) of d^+ is a technicality that makes d^+ be a metric. We define $D^{AB}(t) \in \mathbb{R}^{m \times m}$ with element (ij) equal to $d^+(A_i^+(t), B_j^+(t))$.

In the context of computer vision tracking, having m_2 $*$-only trajectories in A^+, i.e. trajectories will all states equal to $*$, and m_1 $*$-only trajectories in B^+, allows two important types of associations. In computing $\mathcal{D}(A,B)$, we might not want to use any distance between a ground-truth trajectory in A, say A_1, and any reconstructed trajectories in B. We get this by associating A_1^+ to a $*$-only trajectory in B^+. The points in A_1 are *miss detections*. A reconstructed trajectory, say B_1, might not be related to any ground truth trajectories in A. We represent this by associating B_1^+ to a $*$-only trajectory in A^+. B_1 is a *spurious trajectory*. To create the possibility that all ground truth trajectories might be missed and all tracker trajectories might be spurious, we need at least m_2 $*$-only trajectories in A^+ and m_1 $*$-only trajectories in B^+.

With a particular application in mind, we might want to distinguish occlusions and no-target or penalize missed and false tracks differently. In this case, we might introduce multiple different placeholder symbols during the extension procedure, say $*$ and $\#$, and extend the definition of d^+ to penalize different symbol comparisons differently, e.g. $d^+(*,\#) = M_1$, $d^+(*,x) = M_2$, $d^+(\#,x) = M_3$, $d^+(*,*) = d^+(\#,\#) = 0$.

Symbols	Meaning
A, A^+	Set of trajectories and set of extended trajectories
$A_i(t)$, $A_i^+(t)$	State of ith trajectory in A, and A^+, at time t
$*$	Symbol to mark states outside Euclidean space
S	Set of all possible trajectories
$\mathcal{D}$	Distance between trajectory sets
d, d^+	Distance between states and between extended states
M	Cost between state $*$ and state in Euclidean space
$D_{ij}^{AB}(t)$	Distance d^+ between extended states $A_i^+(t)$ and $B_j^+(t)$
Π	Set of all possible permutations (perms.)
σ, σ^{-1}, σ_i	Permutation, its inverse, the image of i by the map σ
Σ, Σ^{-1}	Sequence of permutations, sequence of their inverses
$\Sigma_i(t)$	The image of i by the map $\Sigma(t)$ in the sequence Σ
Π^T	Set of all possible sequences of perms. of length T
$\mathcal{K}(\Sigma)$	Switching cost of permutation sequence Σ
w, W	Doubly stochastic matrix (d.s.m.), sequence of d.s.m.
$W_{ij}(t)$	Entry (ij) of the tth d.s.m. of the sequence W
$\mathcal{P}$	Set of all possible d.s.m.
$\mathcal{P}^T$	Set of all possible sequences of d.s.m. of length T
$\|.\|$, $\dagger$, **tr**	Operators for matrix norm, transpose, trace

Not all interpretations and uses of these symbols make sense. Let B be the output of a tracker and A the ground truth. If B^+ has one $*$-only trajectory, a miss detection, and A has one $*$-only trajectory, an always occluded object, then $\mathcal{D}(A,B) = 0$, even tough there is a cardinality error. This occurs because the meaning of $*$ in A and B is inconsistent. We could have used a different symbol for miss detections, $*$, and occluded ground truth, $\#$. Furthermore, it is unreasonable to let always-hidden objects be part of the ground truth, i.e. let A have one $*$-only trajectory. Otherwise, a tracker is only good if it can find the number of objects behind a curtain.

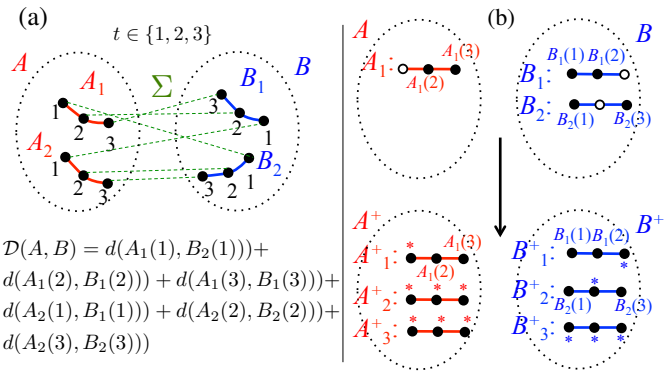

Fig. 2. (a) Role of associations in computing $\mathcal{D}$; (b) Extension procedure.

Figure 2-(a) illustrates how we can use an association Σ between A and B and distance d^+ between state elements to define a distance $\mathcal{D}(A,B)$. In Fig. 2-(a) we have dropped $^+$ for clarity. In this figure we have $\Sigma = (\Sigma(1), \Sigma(2), \Sigma(3))$ where $\Sigma(1) = (2,1)$, $\Sigma(2) = (1,2)$, $\Sigma(3) = (1,2)$.

In this paper we denote the set of doubly stochastic matrices as $\mathcal{P} = \{w \in \mathbb{R}^{m \times m} : w^\dagger \mathbf{1} = \mathbf{1}, w\mathbf{1} = \mathbf{1}, w \geq 0\}$. We also define $\mathcal{P}^T = \{W : W = (W(1), ..., W(T)), W(t) \in \mathcal{P} \ \forall t\}$ as the set of all length-T sequences of doubly stochastic matrices.

Without specification, $\|\cdot\|$ denotes the Euclidean norm, $\|\cdot\|_1$ denotes the 1-norm, $\dagger$ means transpose and **tr** matrix trace.

The above table gathers the most important notation used.

IV. Proposed metrics

Here we introduce two novel families of metrics. We study their properties in Section VII.

Consider a map $\mathcal{K}$:$\Pi^T \mapsto \mathbb{R}_0^+$ that gives a score, a switching cost, to sequences of associations and a map $d : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}_0^+$ that computes distances between trajectories' states. Recall that $D_{ij}^{AB}(t) = d^+(A_i^+(t), B_j^+(t))$ depends on d's definition.

Definition 1. *The* **natural distance** *induced by $\mathcal{K}$ and d between two sets of trajectories is a map $\mathcal{D}_{nat} : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}_0^+$ such that for any $A, B \in \mathcal{S}$*

$$\mathcal{D}_{nat}(A,B) = \min_{\Sigma \in \Pi^T} \Big\{ \mathcal{K}(\Sigma) + \sum_{t=1}^{T} \sum_{i=1}^{m} D_{i\,\Sigma_i(t)}^{AB}(t) \Big\}. \quad (1)$$

A natural choice for d is the Euclidian metric, that is, $d(x, x') = \|x - x'\|_2$. One intuitive choice for $\mathcal{K}$ is $\mathcal{K}_{count}(\Sigma) = \alpha \sum_{t=1}^{T-1} \mathbb{I}(\Sigma(t+1) \neq \Sigma(t))$, $\alpha > 0$, that basically counts the number of times we change association. We give other definitions for $\mathcal{K}$ in Section VII.

Generally, computing $\mathcal{D}_{nat}$ is a combinatorially hard problem so we introduce a new metric that uses doubly stochastic matrices instead of permutations as associations between A and B [1]. Consider any norm $\|.\|$ on the space of matrices.

Definition 2. *The* **natural computable distance** *induced by $\|.\|$ and d between two sets of trajectories is a map $\mathcal{D}_{comp} : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}_0^+$ such that for any $A, B \in \mathcal{S}$*

$$\mathcal{D}_{comp}(A,B) = \min_{W \in \mathcal{P}^T} \Big\{ \sum_{t=1}^{T-1} \|W(t+1) - W(t)\| + \sum_{t=1}^{T} \sum_{i,j=1}^{m} W_{ij}(t)\, D_{ij}^{AB}(t) \Big\}. \quad (2)$$

A few important observations follow. First, notice that both measures minimize the sum of a distance cost plus a switching cost. Second, notice that we are defining a family of measures, not just two measures. Since d, $\mathcal{K}$ and $\|.\|$ are generic, *our definitions can easily incorporate a scaling factor in front of each term of equations* (1) *and* (2). Therefore, we can control the relative importance of the switching and distance costs.

Third, $\mathcal{P}$ is a convex set and any norm $\|.\|$ is convex in $\mathcal{P}$, so computing $\mathcal{D}_{comp}$ amounts to solving a convex optimization problem, for which there are easy-to-use and efficient packages, e.g. CVX [11]. Later we show one choice of $\|.\|$, among several, that reduces (2) to a linear program (LP). From this LP, we can approximate $\mathcal{D}_{nat}$ by forcing the variables to be in $\{0, 1\}$ and then use techniques such as branch-and-bound, e.g. available in MATLAB and CPLEX.

Both $\mathcal{D}_{comp}$ and $\mathcal{D}_{nat}$ solve a global optimization problem to find the best tradeoff between distance cost and switching cost. As explained in Section II, this is in contrast to most measures used in practice such as MOTP, as explained below.

[1] By the Birkhoff–von Neumann theorem we know that $\mathcal{P}$ is smallest convex set that contains all permutations (represented as permutation matrices).

Definition 3. *The CLEAR MOT's distance measure for evaluating tracking precision is called MOTP and is defined as*

$$\mathcal{D}_{MOTP}(A,B) = \sum_{t=1}^{T} \sum_{i=1}^{m} D_{i\,\Sigma_{MOT_i}(t)}^{AB}(t). \quad (3)$$

In the definition above, $\Sigma_{\text{MOT}} \in \Pi^T$ is the sequence of associations that CLEAR MOT builds heuristically (explained in Section II and formally in Appendix A).

We also know that $\mathcal{D}_{nat}$ and $\mathcal{D}_{comp}$ are more flexible than OSPA-ST, where associations cannot change with time. To see this, compare (1) and (2) with the definition of OSPA-ST below.

Definition 4. *The OSPA-ST metric is defined as*

$$\mathcal{D}_{OSPA-ST}(A,B) = \min_{\sigma \in \Pi} \sum_{t=1}^{T} \sum_{i=1}^{m} D_{i\,\sigma_i}^{AB}(t). \quad (4)$$

Our metrics include OSPA-ST as a particular case. Indeed, we recover OSPA-ST metric if we choose $\mathcal{K}$ as: $\mathcal{K}_{OSPA-ST}(\Sigma) = 0$ if $\Sigma = (\sigma, \sigma, ..., \sigma)$, for some $\sigma \in \Pi$, and $\mathcal{K}_{OSPA-ST}(\Sigma) = \infty$ otherwise. In this case, we can compute $\mathcal{D}_{comp}$ in polynomial time, just like for OSPA-ST, using e.g. the Hungarian algorithm [12] (cf. [1]).

The doubly stochastic matrices $W(t)$ of $\mathcal{D}_{comp}$, can be transformed into permutations using a projection procedure that amounts to solving a LP. However, we cannot obtain doubly stochastic matrices from $\mathcal{D}_{nat}$. We can use the associations from $\mathcal{D}_{nat}$ and $\mathcal{D}_{comp}$ to improve the computation of other heuristic measures of multi-target tracking performance, e.g. the number of ID switches or track purity.

Since $d^+(*, *) = 0$, the $*$-only trajectories add zero distance cost when associated with each other and so our metrics do not produce irrelevant switches between $*$-only trajectories. In addition, although $\mathcal{K}$ is a function of Σ but not of A and B, when a particular application is in mind, we can have $\mathcal{K}$ treat switches between different indices differently. We might, for instance, impose that confusing the ID of tracks 1 and 20 (e.g. different-team players) should be heavily penalized but confusing the ID of tracks 1 and 4 (e.g. same-team players) should not. Also, we can impose that a switch between a detected person and a $*$, a non-detected person, should have a different cost than a switch between two detected persons.

Finally, using duality, and *from now on using more compact matrix notation*, we can equivalently define $\mathcal{D}_{comp}$ (cf. (2)) as

$$\mathcal{D}_{comp}(A,B) = \min_{W \in \mathcal{P}^T} \sum_{t=1}^{T} \mathbf{tr}\left(W(t)^{\dagger} D^{AB}(t)\right) \quad \text{subject to} \quad \sum_{t=1}^{T-1} \|W(t+1) - W(t)\| \leq \alpha$$

for some α. Acting by analogy, $\mathcal{D}_{nat}$ can also be re-defined in a similar way. In other words, we do not need to worry about the sum of distance cost and switching cost.

V. Related work

Except for our work, we found no similarity measure for sets of trajectories that is both mathematically consistent (a metric) and, at the same time, is useful and can deal with

identity switches, i.e. it allows time-dependent associations. We focus our discussion around these characteristics. However, several of the ideas we review can be incorporated into our metrics to define new variants that compete with past work in settings not discussed, given the scope of this paper.

Our work is related to the general problem of defining a distance between two sets, however, this is a topic too vast to review in this paper alone. In [13], the reader can find many of those distances. In this article, the two sets A and B are sets of trajectories, and this limits the scope of our discussion. In the simplest case, when trajectories have only one vector, typical definitions compute an average or sum distance between all pairs of elements from A and B or just from a few pairs, e.g. [14]. Knowing which pairs to use when computing distances requires a procedure that matches elements of A with elements of B, e.g. [1], [15]. In general, however, each trajectory is composed of a set of vectors indexed by time, which limits our discussion even more.

To the best of our knowledge, the most rigorous works on metrics for sets of trajectories are based on [1]. In [1] the authors propose a distance for sets of vectors, the optimal sub-pattern assignment metric (OSPA), and explain its advantages over other distances in the context of multi-object filters. The OSPA metric has better sensitivity than the Hausdorff metric to cardinality differences between sets and does not lead to complicated interpretations as does the optimal mass transfer metric (OMAT) [16], proposed to address the limitations of the Hausdorff metric. OSPA, a metric for sets of vectors, can be defined from any distance between two vectors.

All the spin-offs of [1] focus on designing a metric for two sets of trajectories for the purpose of evaluating the performance of tracking algorithms. We recall however that, apart from tracking, many applications in machine learning and AI benefit if we work with a metric rather than a similarity measure that is not a metric. For example, the algorithms mentioned in the introduction [3]–[5] can only cluster, classify and find nearest neighbors if $\mathcal{D}$ is a metric. None of the spin-offs of [1] allow associations that change with time and hence suffer from the same restricted applicability as OSPA (cf. Section II-(ii)).

The OSPA-T metric [10] for sets of trajectories is defined in two steps. First, we optimally match full tracks in A to full tracks in B while considering that tracks have different lengths and can be incomplete. Second, we assign labels to each track based on this match and compute the OSPA metric for each time using a new metric d for pairs of vectors. d uses both the vectors' components and their labels. Finally, we sum the OSPA values across all time instants. Although the first step alone defines a metric, the authors in [17] point out that the full two-step procedure that describes OSPA-T is not a metric.

[17] defines the OSPAMT to be a metric and be more reliable than OSPA-T. The OSPAMT metric also optimally matches complete trajectories in A to complete trajectories in B but unlike OSPA-T allows us to match one trajectory in A to multiple trajectories in B (and vice-versa). The authors make this design choice to not penalize a tracker when it outputs only one track for two objects that move closely together.

Some extensions to OSPA incorporate measurement uncertainty. The Q-OSPA metric [18] weighs the distance between pairs of points by the product of their certainty and by adding a new term that is proportional to the product of the uncertainties. The H-OSPA metric [19] uses OSAP with distributions as states instead of vectors and uses the Hellinger distance between distributions instead of the Euclidean distance between vectors. Both works only allow sets that contain points, not trajectories. However, it is easy to combine them with [10] or [17] and get a metric for sets of trajectories.

The papers above are relatively recent, as is the search for mathematical metrics for sets of trajectories. However, researchers studying multi-object tracking have been using similarity measures for sets of trajectories to rank tracking algorithms for a while. Reviewing all the work done in this area is impossible. Especially because evaluating tracking performance has many challenges other than the problem of defining a similarity measure. See [20], [21] for some examples. Nonetheless, we mention various works that relate to our problem. We emphasize that none of the following works defines a metric mathematically.

Most of our paper is about finding an association between the elements in A and B (cf. Section II). In [22], which is expanded in [23], the authors are one of the first to identify this as the central problem in defining $\mathcal{D}$, although some of their ideas draw from the much earlier Ph.D. thesis of one of the authors [24]. They propose an one-to-one association between A and B but this association is optimally computed independently at every time instant. Also, there is no discussion about the number of association changes that might occur.

The CLEAR MOT are used widely in part because they allow control over the number of times that the association between A and B changes. It appears that [25] is one of the first that allow this. Like the CLEAR MOT, [25] uses a sequential matching procedure that tries to keep the association from the previous time instant if possible. The association that [25] uses at each time t is not one-to-one optimal like in [22] or in the CLEAR MOT. Rather, the authors use a simple thresholding rule to decide which elements to associate. The idea of using a simple threshold rule to associate A and B has survived until relatively recently. For example, [26] match a full trajectory in A to a whole trajectory in B if they are close in space for a sufficiently long time interval. The authors in [27] use a similar thresholding rule.

Shortly after [22] and [23], some of the same authors discuss again the problem of associating A to B. [28] proposes four different methods for this problem. First: A and B are optimally matched independently at each time instant, possibly generating different associations at every time instant. Second: the association between A and B cannot change with time. Like in the OSPA-based metrics, this restricts the measure's applicability (cf. Section II). Third: allow the association between A and B to change in time but only in special circumstances. Unfortunately, as they point out, this leads to an NP-hard problem. Fourth: use heuristics to minimize the number of changes in association. Some of these ideas resemble those used in the CLEAR MOT.

Measuring trackers' performance is not just about defining similarity measures for sets of trajectories, or only concerned

with establishing associations between A and B. If we have already matched the elements of A and B, which quantity should we compute from this match? A typical quantity that researchers calculate is the average or total distance between matched points. Our metrics are a combination of this quantity with the switching cost in the A-B association. Distance-related quantities apart, researchers also compute quality measures for the match itself, e.g. the number of trajectories in A that do not match to any trajectory in B (spurious tracks and missed tracks) or the number of times the match between A and B changes. Quantities like track purity and coalescence are often also computed once a match is established. [6]–[8] and [9] list many other measures of similarity between A and B and measures of quality for individual tracks.

It is worth mentioning a few non-mainstream works. [29] defines a distance based on comparing the occurrence of special discrete events in A and B. [30] proposes an information theoretic similarity measure. Finally, [31] proposes a similarity measure based on hidden Markov models to allow unequal temporal sampling rates in the trajectories.

VI. Limitations of the CLEAR MOT

In the context of visual tracking, the next lemma shows that there are situations in which the average tracking distance cost over time gets arbitrarily close to zero with time while the CLEAR MOT produces an association Σ_{MOT} that fails to see that. In other words, the CLEAR MOT can be counter-intuitive. Since none of the definitions of Section IV deal with time averages, in this section we use the following two definitions.

For any $\Sigma \in \Pi^T$, $A, B \in \mathcal{S}$ we let

$$swi(\Sigma) = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}\{\Sigma(t) \neq \Sigma(t+1)\} \tag{5}$$

denote the empirical frequency of identity switches, i.e. association changes, where $\mathbb{I}\{\cdot\}$ is the indicator function. Let

$$dist(\Sigma, A, B) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m} D^{AB}_{i\,\Sigma_i(t)}(t) \tag{6}$$

denote the average distance between A and B under $\Sigma = (\Sigma(1), ..., \Sigma(T))$. Note that, for example, $\mathcal{D}_{MOTP} = T \times dist(\Sigma_{MOTP}, A, B)$.

Lema 1. *For any threshold $thr_{MOT}>0$, there exists trajectory sets $A, B \in \mathcal{S}$ and an association $\Sigma \in \Pi^T$ such that*

$$dist(\Sigma, A, B) < \mathcal{O}(1/T),$$
$$swi(\Sigma) = 0 \text{ and},$$
$$dist(\Sigma_{MOTP}, A, B) > m\,thr - \mathcal{O}(1/T).$$

The following lemma shows that, for any $thr_{MOT} > 0$, the MOTP measure is mathematically inconsistent.

Lema 2. *MOTP is not a metric for any threshold $thr_{MOT}>0$.*

The proofs of these two lemmas are in Appendix A.

It is also possible to prove that MOTA, the other widely used measure in the CLEAR MOT, does not define a metric. See www.jbento.info/papers/metriccompanion.pdf.

VII. Properties of our metrics

The metrics we introduce do not produce the counter-intuitive results of OSPA-ST or MOTP because they control the tradeoff between switches and distances in a globally optimal way. In addition, we also show that they are a metric under mild conditions on $\mathcal{K}$, $\|.\|$ and d.

Definition 5. *A map $\mathcal{K} : \Pi^T \mapsto \mathbb{R}_0^+$ is a* **permutation measure** *if it satisfies the following three properties (i) $\mathcal{K}(\Sigma) = 0$ if and only if Σ is constant $\Sigma = (\sigma, \sigma, ..., \sigma)$ for some $\sigma \in \Pi$, (ii) $\mathcal{K}(\Sigma^{-1}) = \mathcal{K}(\Sigma)$ and (iii) $\mathcal{K}(\Sigma \circ \Sigma') \leq \mathcal{K}(\Sigma) + \mathcal{K}(\Sigma')$.*

A few observations are in order. Clearly, if $\mathcal{K}$ is a permutation measure so is $\alpha\mathcal{K}, \alpha > 0$. Also, property (iii) requires that all permutation matrices have the same dimensions, otherwise we cannot talk about their composition. In practice, this can be naturally achieved if e.g. all trajectories have a common maximum observation time T, sets of trajectories have a common maximum number of trajectories M, and permutation matrices have dimensions $2M \times 2M$. Finally, Def. 5 implies that $\mathcal{K}$ is invariant under reindexing. Indeed, by property (iii) followed by property (i) $\mathcal{K}(\Sigma \circ (\sigma, \sigma, \ldots, \sigma)) \leq \mathcal{K}(\Sigma) + \mathcal{K}((\sigma, \sigma, \ldots, \sigma)) = \mathcal{K}(\Sigma)$. On the other hand, by property (iii) followed by property (i) followed by property (ii) $\mathcal{K}(\Sigma) = \mathcal{K}(\Sigma \circ (\sigma, \sigma, \ldots, \sigma) \circ (\sigma, \sigma, \ldots, \sigma)^{-1}) \leq \mathcal{K}(\Sigma \circ (\sigma, \sigma, \ldots, \sigma)) + \mathcal{K}((\sigma, \sigma, \ldots, \sigma)^{-1}) = \mathcal{K}(\Sigma \circ (\sigma, \sigma, \ldots, \sigma))$. Therefore, $\mathcal{K}(\Sigma \circ (\sigma, \sigma, \ldots, \sigma)) = \mathcal{K}(\Sigma)$. A similar argument shows that $\mathcal{K}((\sigma, \sigma, \ldots, \sigma) \circ \Sigma) = \mathcal{K}(\Sigma)$. This invariance is part of the reason why we can later show that, under certain conditions, $\mathcal{D}_{nat}$ and $\mathcal{D}_{comp}$ are metrics.

One straightforward choice for $\mathcal{K}$ is to count the number of times that the association between A and B changes.

Theorem 1. *Let $\mathcal{K}_{count}(\Sigma) = \sum_{t=1}^{T-1} \mathbb{I}(\Sigma(t+1) \neq \Sigma(t))$. $\mathcal{K}_{count}$ is a permutation measure.*

Another two desirable choices for $\mathcal{K}$ are (a) the function that sums the minimum number of transpositions to go from one permutation to the next and (b) the function that sums the number of adjacent transpositions to go from one permutation to the next. In what follows $k_{Cayley}(\sigma)$ gives the minimum number of transpositions to obtain the identity permutation from $\sigma \in \Pi$ and $k_{Kendall}(\sigma)$ gives the number of adjacent transposition that the Bubble-Sort algorithm performs when sorting σ to obtain the identity permutation. The Cayley distance goes back to [32] and the Kendall distance to [33].

Theorem 2. *Let $\mathcal{K}_{trans}(\Sigma) = \sum_{t=1}^{T-1} k_{Cayley}(\Sigma(t+1) \circ \Sigma(t)^{-1})$. $\mathcal{K}_{trans}$ is a permutation measure.*

The proof of Theorems 1 and 2 is in Appendix B.

Although many intuitive choices for $\mathcal{K}$ satisfy properties (i), (ii) and (iii) of Definition 5, some natural ones do not. For example, given a $\beta \geq 1$ we might want to define $\mathcal{K}_{maxcount}$ as $\mathcal{K}_{maxcount}(\Sigma) = \mathcal{K}_{count}(\Sigma)$ if $\mathcal{K}_{count}(\Sigma) \leq \beta$ and $\mathcal{K}_{maxcount}(\Sigma) = \infty$ if $\mathcal{K}_{count}(\Sigma) > \beta$ (we can replace ∞ by some very large number if we want to be technical about the range of $\mathcal{K}$ being $\mathbb{R}_0^+$). This $\mathcal{K}$ forces us not to create an association between A and B more intricate than a certain amount β, something natural to desire. The following is proved in Appendix B.

Theorem 3. *$\mathcal{K}_{maxcount}$ does not satisfy (iii) in Definition 5. Thus, it is not a permutation measure.*

The following theorem is similar to Theorem 3 and its proof is omitted.

Theorem 4. *Let $\mathcal{K}_{adjtrans}(\Sigma) = \sum_{t=1}^{T-1} k_{Kendall}(\Sigma(t+1) \circ \Sigma(t)^{-1})$. $\mathcal{K}_{adjtrans}$ does not satisfy (iii) in Definition 5. Thus, it is not a permutation measure.*

We now state our first main result.

Theorem 5. *If $\mathcal{K}$ is a permutation measure and d is a metric, then the map $\mathcal{D}_{nat}$ induced by them is a metric on $\mathcal{S}$.*

Even if a function $\mathcal{K}$ violates some of the properties in Definition 5, it could still induce a $\mathcal{D}_{nat}$ that is a metric. However, we often find that, from a set of associations Σ and Σ' that violate Definition 5, we can build three sets of trajectories A, B and C that violate some of the properties required of a metric. This is the case, for example, for $\mathcal{D}_{nat}$ induced by $\mathcal{K}_{maxcount}$ and d equal to the Euclidean metric.

Theorem 6. *The map $\mathcal{D}_{nat}$, induced by $\mathcal{K} = \mathcal{K}_{maxcount}$ and the Euclidean distance d, is not a metric.*

The proof of this theorem is in Appendix B. In this sense, Definition 5 is required for $\mathcal{D}_{nat}$ to be a family of metrics.

We now focus on $\mathcal{D}_{comp}$.

Definition 6. *A matrix norm $\|.\|$ is a* **switching norm** *if for any four matrices $w_1, w_2, w_1', w_2' \in \mathcal{P}$*

$$\|w_2' w_2 - w_1' w_1\| \leq \|w_2' - w_1'\| + \|w_2 - w_1\|. \quad (7)$$

The following Lemma gives sufficient conditions for a metric $\|.\|$ to satisfy property (7). See Appendix C for the proof.

Lema 3. *If $\|.\|$ is a sub-multiplicative norm and $\|W\| \leq 1$ for all $W \in \mathcal{P}$ then $\|.\|$ is a switching norm.*

This lemma implies, for example, that the 1-norm, ∞-norm and spectral norm for matrices are valid choices for $\|.\|$.

We now state our second most important result.

Theorem 7. *If $\|.\|$ is a switching norm and d is a metric then the map $\mathcal{D}_{comp}$ induced by $\|.\|$ and d is a metric on $\mathcal{S}$.*

The proof of this theorem is in Appendix C.

The use of $\|.\|_1$ (matrix norm) in $\mathcal{D}_{comp}$ is extremely useful because it induces the changes of association to be sparse and, as the next theorem shows, reduces $\mathcal{D}_{comp}$ to solving a linear program. Recall that all LPs can be solved in polynomial time [34]. The theorem's proof is in Appendix C.

Theorem 8. *For any $A, B \in \mathcal{S}$, the metric $\mathcal{D}_{comp}(A, B)$ induced by the matrix 1-norm and any d can be computed (in polynomial time) by solving a linear program.*

This LP can be a made sparse if we impose that for every (i, j, t) such that $A_i(t)$ and $B_j(t)$ are distant we must have $W_{ij}(t) = 0$, i.e. we cannot associate distant points. Sparsity allows us to reduce the effective number of optimization variables in (2) and speedup computations further.

To end this section, let us explicitly include a scaling factor $\alpha > 0$ in the definition of $\mathcal{D}_{comp}$. To be specific,

$$\mathcal{D}_{comp}(A, B) = \min_{W \in \mathcal{P}^T} \quad \alpha \times swi(W) + dist(W, A, B), \quad (8)$$

where, changing the definition of (5) and (6) in Section VI,

$$swi(W) = \sum_{t=1}^{T-1} \|W(t+1) - W(t)\| \text{ and} \quad (9)$$

$$dist(W, A, B) = \sum_{t=1}^{T} \mathbf{tr}(W(t)^{\dagger} D^{AB}(t)). \quad (10)$$

If we compute $\mathcal{D}_{comp}$ for different α's we obtain different pairs of values $dist$ and swi. We can obtain more pairs of values if we compute $dist$ and swi using as W the permutation matrices representing the association produced by the CLEAR MOT or OSPA-ST. Even using MOTP alone, we can compute different values for swi and $dist$ by changing thr_{MOT}.

If we fix A and B, we can display all these different pairs on a 2D scatter plot where each point is a pair $(dist, swi)$ evaluated on a different $W \in \mathcal{P}^T$. Using such a scatter plot is a great way to assess the performance of a tracker B on ground-truth A under different similarity measures or when we do not know which α or thr_{MOT} to use to compute $\mathcal{D}_{comp}$ and $\mathcal{D}_{MOTP}$. For a fixed A and B, a good measure $\mathcal{D}$ generates pairs closer to the origin. A very good measure $\mathcal{D}$ generates pairs only on the Pareto frontier of this scatter plot. Not surprisingly, the pairs $(dist, swi)$ that $\mathcal{D}_{comp}$ produces for different values of α generate this Pareto frontier. This follows from a well-known result in convex optimization theory that we state in Theorem 9 (see [35] for a proof). In this context, $\mathcal{D}_{comp}$ is the best metric we can hope. In particular, none of the examples where MOTP or OSPA-ST produce counter-intuitive results, like the counter example behind the proof of Lemma 1, lead to $\mathcal{D}_{comp}$ giving counter-intuitive results. The reader unfamiliar with convex optimization can jump to Section VIII.

Let Ω be a convex set, e.g. $\mathcal{P}^T$, and let f and g be two convex functions in Ω, e.g. $swi(.)$ and $dist(., A, B)$. Let $\mathcal{R} = \{(s, t) \in \mathbb{R}^2 : s \geq f(x) \text{ and } t \geq g(x), x \in \Omega\}$. This set could be, for example, the points in our scatter plot plus points with worse costs. Let $\partial\mathcal{R}$ be the boundary of $\mathcal{R}$. Since $\mathcal{R}$ is convex, $\partial\mathcal{R}$ is its Pareto frontier. For instance, $\partial\mathcal{R}$ could be the Pareto frontier we discussed above. Let $p(\alpha) \in \mathbb{R}^2$ be the curve of (f, g) pairs generated by solving $\min_{x \in \Omega} f(x) + \alpha g(x)$ for different values of $\alpha > 0$. This could be a tradeoff curve of $(swi, dist)$ pairs generated by $\mathcal{D}_{comp}$ for different α's.

Theorem 9. *If $(s, t) \in \partial\mathcal{R}$ then either $(s, t) = p(\alpha_0)$ for some $\alpha_0 > 0$ or (s, t) is a convex combination of $p(\alpha_0)$ and $p(\alpha_1)$ for some $\alpha_0, \alpha_1 > 0$.*

VIII. Numerical results: $\mathcal{D}_{comp}$ computation time

In practice, there are many ways to compute $\mathcal{D}_{comp}$, like when solving a convex optimization problem. To illustrate how easy it is to get code working, we have a simple Matlab code for $\mathcal{D}_{comp}$ in www.jbento.info/papers/metriccompanion.pdf. In the same document, we also have a simple Matlab code to estimate $\mathcal{D}_{nat}$. However, to explore the limits of practical performance, we coded a non-trivial solver in C using

$parADMM$, an implementation of the Alternating Direction Method of Multipliers (ADMM) that is available at github.com/parADMM/engine. $parADMM$ has good performance in practice [36]. The ADMM is known to scale well, and its modular nature makes it easy to research future variants of $\mathcal{D}_{comp}$ without having to re-write much code. Furthermore, for strongly convex problems, its optimally-tuned convergence rate is as fast as that of the fastest first-order method [37].

[38] is a good introduction to the ADMM and its applications.

In Figure 3 we plot run-time in computing $\mathcal{D}_{comp}$ against the total duration T of the input data for a different number of free association variables per time instant t. We use the term "free association variables" because, as explained after Thm. 8, a few variables $W_{ij}(t)$ can be set to zero to sparsify the problem and save computation time. We choose the euclidean distance for d and the one-norm for matrix norm $\|.\|$ in the switching cost. Similar run-times hold for other metrics. The run-time is computed for randomly generated A and B on a single core of a 1.4GHz Intel Core i5 MacBook Air. ADMM always converged to 1% accuracy in less than 150 iterations.

To interpret the plots, imagine we want to evaluate the quality of a tracker when tracking 22 objects. Imagine that the tracker operates at 30 frames per second and also that our tracker is noisy so that it produces a few false tracks that create approximately 10 extra points per instant. To compute $\mathcal{D}_{comp}$, and after we extend the ground-truth and hypothesis sets from A and B to A^+ and B^+, we are dealing with distance matrices $D^{AB}(t)$ with about $((22+10)\times 2)^2 = 4096$ variables per instant t. Using Fig. 3 we see that it take us about 40 seconds to evaluate the accuracy of $800/30 = 26.6$ seconds of data. If we reduce the number of free variables per frame to half, e.g. by setting $W_{ij}(t) = 0$ if $D^{AB}_{ij}(t)$ is larger than a given threshold, then we can reduce the time to process 26.6 secs. of data to about 20 secs.

IX. Numerical results: Optimal tradeoff curves

To the best of our knowledge, $\mathcal{D}_{comp}$ is the first measure that is a metric, can be computed in polynomial time and, in addition, deals with the tradeoff between distance cost and switching cost optimally. In this sense, $\mathcal{D}_{comp}$ (and future variants) is the best metric we can hope for (cf. Section VII).

To illustrate this optimal tradeoff, we build and compare tradeoff curves obtained from $\mathcal{D}_{comp}$ and MOTP. A *tradeoff curve* is a set of points $(swi, dist)$ where swi and $dist$ are computed using (9) and (10). For $\mathcal{D}_{comp}$, we obtain the different points along the curve by changing $\alpha > 0$ in (8). For MOTP we generate the tradeoff curve by changing thr_{MOT}. We do this for both synthetic and real data. We use the Euclidean metric for d^+ and the component-wise 1-norm for the matrix norm $\|.\|$.

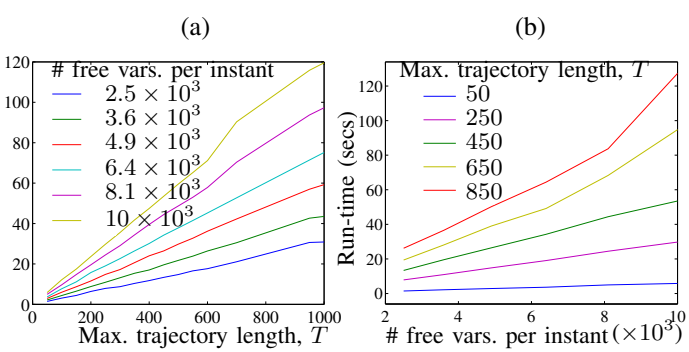

Fig. 3. Time to compute $\mathcal{D}_{comp}$ (within 1% accuracy) as a function of (a) the length of trajectories; and (b) the number of association variables per instant.

The direct interpretation of our results is that $\mathcal{D}_{comp}$ is better than MOTP. However, the important underlying fact is that $\mathcal{D}_{comp}$ builds better associations than (i) the heuristics widely used in the literature, e.g. the CLEAR MOT, and (ii) the optimal associations before our work that do not allow switches, e.g. OSPA-ST. Therefore, although we restrict the comparison to MOTP, we can show similar improvement over many measures that first establish a heuristic association between A and B. For example, Multiple Object Tracking Accuracy (MOTA), False Alarms per Frame, Ratio of Mostly Tracked trajectories, Ratio of Mostly Lost trajectories, the number of False Positives, number of False Negatives, number of ID Switches, the number of tracks Fragmentation and many of the measures listed in [7] and [9].

A. Real trackers and real data

In Figure 4-(a) we show the performance of the trackers in [39] and [40] on the AVG-TownCentre data set. We call these trackers $Tracker09$ and $Tracker11$ respectively. The data set is part of the Multiple Object Tracking Benchmark [41] and is widely used in computer vision. It comes from a pedestrian street filmed from an elevated point for 3 minutes and 45 seconds and can be downloaded from http://www.robots.ox.ac.uk/ActiveVision/Research/Projects/2009bbenfold_headpose/project.html#datasets. In Figure 4-(b) we show the performance of the trackers in [42] and [43] on the PETS2009 data set, also part of the Multiple Object Tracking Benchmark. We call these trackers $Tracker12$ and $Tracker15$ respectively. Its duration is 1 minute and 54 seconds, and it can be downloaded from http://www.cvg.reading.ac.uk/PETS2009/a.html. More exact knowledge of these data sets and trackers is outside the scope of this paper. The point we want to make is in regard to comparing similarity measures. Not about comparing trackers' performance on different data sets. We produced all the plots using the exact same output that each tracker produced in its respective paper, thanks to the authors who provided us with their trackers' output. When coding d^+, we set $M = 20$ for AVG-TownCenter and $M = 50$ for PETS2009.

As expected, the tradeoff curves from MOTP understate the performance of the trackers: all trackers are achieving a substantially smaller number of switches without incurring larger distance costs than what MOTP reports. Interestingly, for these trackers and data sets, $\mathcal{D}_{comp}$ keeps the same relative ordering of performance as MOTP. It is conceivable that there are situations in which a tracker 1 is better than a tracker 2 according to MOTP but not according to $\mathcal{D}_{comp}$. It would be interesting to find such examples in future work.

B. Random ensemble of trajectories

Above, A and B are relatively close to each other because all the trackers are good trackers. In this section, we aim

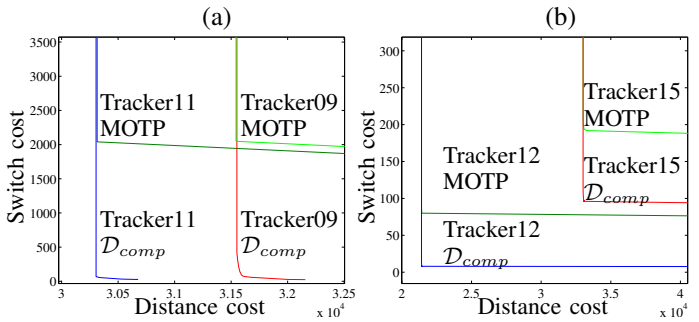

Fig. 4. (a) Tradeoff plot for $Tracker09$ and $Tracker11$ on the AVG-TownCenter data set; and (b) tradeoff plot for $Tracker12$ and $Tracker15$ on the PETS2009 data set. Smaller values (in either axis) is better.

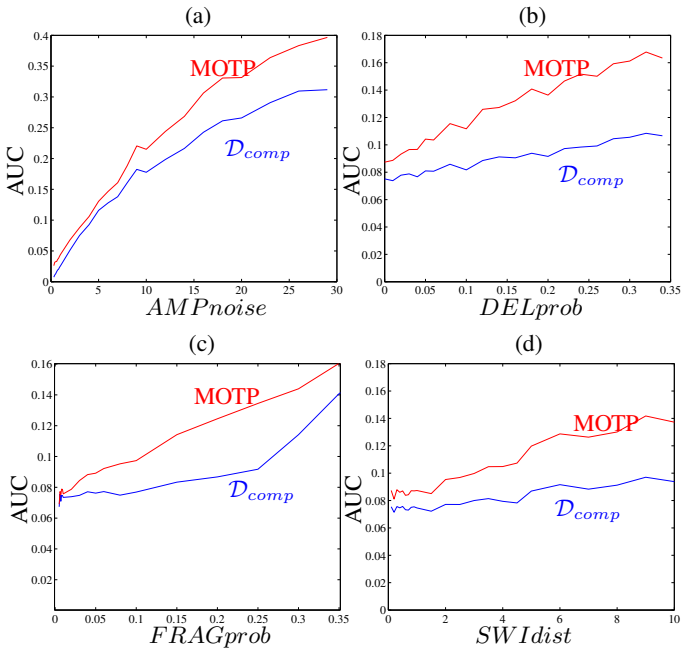

Fig. 5. AUC (normalized) versus (a) noise amplitude; (b) point deletion probability; (c) fragmenting probability; and (d) switching distance. Smaller AUC is better.

to study A and B that are more different. Hence, we use synthetic data to control the distance between A and B. We randomly generate A with 25 trajectories and make B a distorted version of the trajectories in A. The trajectories in A have random start and end times and the objects in each trajectory randomly change their velocity's direction along the way. The trajectories in B are generated by randomly (i) fragmenting the trajectories in A, (ii) removing some of the resulting trajectories, (iii) adding noise to all trajectories and (iv) flipping or not the ID of two trajectories if they pass by each other close enough. In the end, B might have more or less than 25 trajectories. In total, we have four knobs to increase or reduce the distance between A and B. These knobs are, (i) the amplitude of noise, $AMPnoise$; (ii) the probability of fragmenting a track at each point in time, $FRAGprob$; (iii) the probability of deleting a points in the track, $DELprob$; and (iv) the threshold distance after which we allow to tracks ID to be switched or not randomly, $SWIdist$.

Here, trajectories are far more diverse and complex than those in Section IX-A and in most publicly available real data sets. Real objects, like people, have relatively simple trajectories. In addition, we do not just test two data sets, like in Section IX-A. We generate data for about 20 different configurations for each of the four knobs described above and for each of these configurations we generate 30 random sets A and B. Hence, and in the context of computer vision tracking, it is as if we test 2400 different data sets of ground-truth and output trajectories.

We study the similarity between A and B using tradeoff plots with distance cost on the y-axis and switching cost on the x-axis. The smaller the area under the curve (AUC), the closer A and B are. In Fig. 5 we show the average AUC for A and B under different knob settings. Each AUC is normalized by the largest AUC possible. The largest AUC is the product of the largest distance cost possible with the largest switching cost possible. Each point in the plots is an average over 30 random pairs A and B with the same knobs' setting. In each plot we keep all but one knob constant.

As expected from Theorem 9, the AUC of $\mathcal{D}_{comp}$ is smaller than the AUC of MOTP. Note that it is incorrect to interpret these results as saying that MOTP is a scaled version of $\mathcal{D}_{comp}$. We are computing exactly the same quantities for both measures: $dist$ and swi according to (10) and (9). We just use a different W for each measure. The curves of Figure 5 have a much deeper significance: they say that the conclusions that hold for real data hold in great generality, now in 2400 different tests and not just for two real data sets and four tackers as above. If we interpret A and B as the ground-truth and output of a tracker respectively, MOTP almost always says that the tracker is worse than what it really is. $\mathcal{D}_{comp}$ can see similarities between A and B that MOTP cannot.

X. Conclusion

The problem of defining a similarity measure for sets of trajectories is crucial for computer vision, machine learning, and general AI. An essential aspect of this problem is finding a meaningful association between the elements of the sets. Existing measures that define useful associations fail to be a metric mathematically speaking, e.g., the CLEAR MOT. The ones that are a metric, only consider restrictive associations, i.e. associations that cannot change in time, e.g. OSPA-based metrics. $\mathcal{D}_{comp}$ is the first that simultaneously (1) is a metric, (2) allows time-dependent associations and hence can associate parts of trajectories to parts of trajectories, (3) allows controlling the complexity of this association in a globally optimal way and (4) has polynomial computation time.

The general idea of defining mathematical metrics for sets of trajectories using convex programs is our greatest overarching contribution. We are currently exploring variants of our metric that allow incorporating uncertainty, as well as richer comparisons between A and B without losing its useful properties. We are also exploring the use of this metric in a machine learning application for fast classification and retrieval of human actions/activities. We will present these extensions in future work.

References

[1] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *Signal Processing, IEEE Trans. on*, vol. 56, no. 8, pp. 3447–3457, 2008.

[2] B. Keni and S. Rainer, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, 2008.

[3] V. Ganti, R. Ramakrishnan, J. Gehrke, A. Powell, and J. French, "Clustering large datasets in arbitrary metric spaces," in *Data Engineering, 1999. Proceedings., 15th Intern. Conf. on*. IEEE, 1999, pp. 502–511.

[4] J. Kleinberg and E. Tardos, "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields," *Journal of the ACM (JACM)*, vol. 49, 2002.

[5] P. N. Yianilos, "Data structures and algorithms for nearest neighbor search in general metric spaces," in *Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1993, pp. 311–321.

[6] S. S. Blackman, "Multiple-target tracking with radar applications," *Dedham, MA, Artech House, Inc., 1986, 463 p.*, vol. 1, 1986.

[7] R. L. Rothrock and O. E. Drummond, "Performance metrics for multiple-sensor multiple-target tracking," in *AeroSense 2000*. Intern. Society for Optics and Photonics, 2000, pp. 521–531.

[8] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[9] A. A. Gorji, R. Tharmarasa, and T. Kirubarajan, "Performance measures for multiple target tracking problems," in *Information Fusion (FUSION), 2011 Proceedings of the 14th Intern. Conf. on*. IEEE, 2011, pp. 1–8.

[10] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *Signal Processing, IEEE Trans. on*, vol. 59, no. 7, pp. 3452–3457, 2011.

[11] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[12] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[13] M. M. Deza and E. Deza, *Encyclopedia of distances*. Springer, 2009.

[14] O. Fujita, "Metrics based on average distance between sets," *Japan Journal of Industrial and Applied Mathematics*, vol. 30, 2013.

[15] A. Gardner, J. Kanno, C. A. Duncan, and R. Selmic, "Measuring distance between unordered sets of different sizes," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conf. on*. IEEE, 2014.

[16] J. R. Hoffman and R. P. Mahler, "Multitarget miss distance via optimal assignment," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Trans. on*, vol. 34, no. 3, pp. 327–336, 2004.

[17] T. Vu and R. Evans, "A new performance metric for multiple target tracking based on optimal subpattern assignment," in *Information Fusion (FUSION), 2014 17th Intern. Conf. on*. IEEE, 2014, pp. 1–8.

[18] H. Xiaofan, R. Tharmarasa, T. Kirubarajan, and T. Thayaparan, "A track quality based metric for evaluating performance of multitarget filters," *Aerospace and Electronic Systems, IEEE Trans. on*, vol. 49, 2013.

[19] S. Nagappa, D. E. Clark, and R. Mahler, "Incorporating track uncertainty into the ospa metric," in *Information Fusion (FUSION), 2011 Proceedings of the 14th Intern. Conf. on*. IEEE, 2011, pp. 1–8.

[20] T. Ellis, "Performance metrics and methods for tracking in surveillance," in *Proceedings of the 3rd IEEE Intern. Workshop on Performance Evaluation of Tracking and Surveillance (PETS02)*. Citeseer, 2002.

[21] A. Milan, K. Schindler, and S. Roth, "Challenges of ground truth evaluation of multi-target tracking," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conf. on*. IEEE, 2013.

[22] B. E. Fridling and O. E. Drummond, "Performance evaluation methods for multiple-target-tracking algorithms," in *Orlando'91, Orlando, FL*. Intern. Society for Optics and Photonics, 1991, pp. 371–383.

[23] O. E. Drummond and B. E. Fridling, "Ambiguities in evaluating performance of multiple target tracking algorithms," in *Aerospace Sensing*. Intern. Society for Optics and Photonics, 1992, pp. 326–337.

[24] O. E. Drummond, "Multiple-object estimation," 1975.

[25] S. B. Colegrove, L. Davis, and S. J. Davey, "Performance assessment of tracking systems," in *Signal Processing and Its Applications, 1996. ISSPA 96., Fourth Intern. Symposium on*. IEEE, 1996.

[26] F. Yin, D. Makris, and S. A. Velastin, "Performance evaluation of object tracking algorithms," in *10th IEEE Intern. Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007)*, 2007.

[27] F. Bashir and F. Porikli, "Performance evaluation of object detection and tracking systems," in *IEEE Intern. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, vol. 5, 2006.

[28] O. E. Drummond, "Methodologies for performance evaluation of multitarget multisensor tracking," in *SPIE's Intern. Symposium on Optical Science, Engineering, and Instrumentation*. Intern. Society for Optics and Photonics, 1999, pp. 355–369.

[29] S. Pingali and J. Segen, "Performance evaluation of people tracking systems," in *Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on*. IEEE, 1996, pp. 33–38.

[30] K. K. Edward, P. D. Matthew, and B. H. Michael, "An information theoretic approach for tracker performance evaluation," in *Computer Vision, 2009 IEEE 12th Intern. Conf. on*. IEEE, 2009, pp. 1523–1529.

[31] F. Porikli, "Trajectory distance metric using hidden markov model based representation," in *IEEE European Conf. on Computer Vision, PETS Workshop*, vol. 3, 2004.

[32] A. Cayley, "Lxxvii. note on the theory of permutations," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 34, no. 232, pp. 527–529, 1849.

[33] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, pp. 81–93, 1938.

[34] L. Khachiian, "Polynomial algorithm in linear programming," in *Akademiia Nauk SSSR, Doklady*, vol. 244, 1979, pp. 1093–1096.

[35] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[36] N. Hao, A. Oghbaee, M. Rostami, N. Derbinsky, and J. Bento, "Testing fine-grained parallelism for the admm on a factor-graph," *arXiv preprint arXiv:1603.02526*, 2016.

[37] G. França and J. Bento, "An explicit rate bound for over-relaxed admm," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 2104–2108.

[38] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[39] B. Benfold and I. Reid, "Guiding visual surveillance by tracking human attention." in *BMVC*, 2009, pp. 1–11.

[40] ——, "Stable multi-target tracking in real-time surveillance video," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conf. on*. IEEE, 2011, pp. 3457–3464.

[41] "Multiple object tracking benchmark," http://www.motchallenge.net, accessed: 2015-03-01.

[42] B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conf. on*. IEEE, 2012, pp. 1918–1925.

[43] F. Poiesi and A. Cavallaro, "Tracking multiple high-density homogeneous targets," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 25, no. 4, pp. 623–637, 2015.

Appendix A
Theory on the limitations of the CLEAR MOT

Let us formally describe the heuristic used by the CLEAR MOT.

Definition 7. *The CLEAR MOT matching heuristic defines Σ_{MOT} sequentially as follows.*

1) *Initialize $\Sigma_{MOT}(1)$ such that $\sum_i d^+(A_i^+(1), B^+_{\Sigma_{MOT_i}(1)}(1))$ is minimal;*
2) *For each $t > 1$ do: for all $i, j \in \{1, ..., m\}$ such that $\Sigma_{MOT_i}(t-1) = j$ and $d^+(A_i^+(t), B_j^+(t)) < thr_{MOT}$ fix $\Sigma_{MOT_i}(t) = j$. We call such matches as* **anchored**. *Set the non-fixed components of $\Sigma_{MOT}(t)$ such that $\sum_i d^+(A_i^+(t), B^+_{\Sigma_{MOT_i}(t)}(t))$ is minimal.*

Proof of Lemma 1. We construct a validating example for any $1 < thr_{MOT} < 2$, with $A = \{A_1, A_2\}$ and $B = \{B_1, B_2\}$, i.e. $m = 2$, and where A_i and B_i are 1D trajectories. We generalize this example to any thr_{MOT} and m at the end.

Consider the two sets of one-dimensional trajectories $A = \{A_1, A_2\}$ and $B = \{B_1, B_2\}$ defined in Figure 6. Time is on the x-axis (left to right) and space on the y-axis (bottom to top). T_1 and T_2 are fixed. T_3 grows with T.

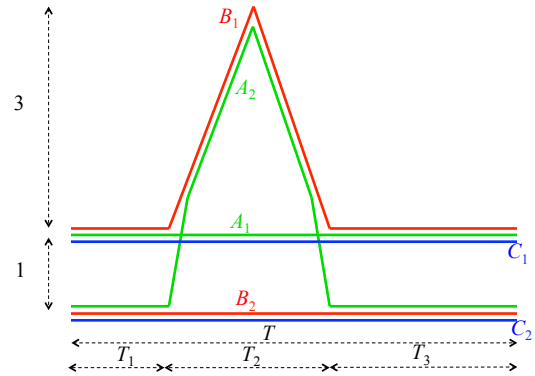

Fig. 6. Example that shows that (a) the CLEAR MOT heuristic is bad and (b) MOTP is not a metric. For visualization purposes, trajectories that are close are actually on top of each other.

Since we assume $1 < thr_{MOT} < 2$, the CLEAR MOT builds an association sequence Σ_{MOT} that initially matches A_1 to B_1 but some time after T_1 matches A_1 to B_2 to minimize the distance between matched points. After $T_1 + T_2$, this last association is anchored given that $1 < thr_{MOT}$. Hence,

$$\Sigma_{\text{MOT}} = \{(1,2), ..., (1,2), (2,1), (2,1), ...\}.$$

The number of times that $\Sigma_{\text{MOT}}(t) \neq \Sigma_{\text{MOT}}(t+1)$ is 1, thus $swi(\Sigma_{\text{MOT}}) = \frac{1}{T-1} = \mathcal{O}(1/T)$. For t after $T_1 + T_2$, $\sum_i \|A_i(t) - B_{\Sigma_{\text{MOT}_i}(t)}(t)\| = 2$. Thus $dist(\Sigma_{\text{MOT}}, A, B) > \frac{2(T-T_1-T_2)}{T} = 2 - \mathcal{O}(1/T)$. However, if we choose $\Sigma = \{(1,2), ..., (1,2)\}$, we have $swi(\Sigma) = 0$ and $dist(\Sigma, A, B) < \frac{T_2 \times 4}{T} = \mathcal{O}(1/T)$. The result of the lemma follows.

To see that the proof holds for any m, we extend A and B as follows. Without loss of generality, we assume m is even. If the 1D trajectory A_i above is equal to $(A_i(1), ..., A_i(T))$, define the 2D trajectory $A_i^{(k)}$ such that

$$A_i^{(k)}(t) = [A_i(t); Ck] \in \mathbb{R}^2,$$

where C is a constant large enough such that trajectories for different k's are not close to each other under the d^+ measure. We define $B_i^{(k)}$ similarly. Define

$$A = \{A_1^{(0)}, A_2^{(0)}, A_1^{(1)}, A_2^{(1)}, ..., A_1^{(m-1)}, A_2^{(m-1)}\},$$

and similarly B. In this setting, the bounds previously computed on $swi(\Sigma_{\text{MOT}})$ and $swi(\Sigma)$ change by a factor of m, while the bounds on $dist(\Sigma_{\text{MOT}}, A, B)$ and $dist(\Sigma, A, B)$ change by a factor of $m/2$. Thus the statement of the lemma still holds.

To extend the proof for an odd m, append to A and B two equal trajectories far away from all other trajectories such that they are matched to each other. They do not contribute to either swi or $dist$.

Finally, to see that the proof also holds for any thr_{MOT}, we rescale space-axis in Figure 6. This changes the bounds on $dist$ by a factor of thr_{MOT} and leaves the bounds on swi unchanged, thus obtains the result. □

Proof of lemma 5. We construct $A, B, C \in \mathcal{S}$ for which the triangle inequality is violated. Specifically, $\mathcal{D}(A,B) > \mathcal{D}(A,C) + \mathcal{D}(C,B)$.

Without loss of generality, we assume $1 < thr_{MOT} < 2$. To see that the result holds for all $thr > 0$, we employ the space-rescaling trick in the proof of Lemma 1.

Consider the sets $A = \{A_1, A_2\}$, $B = \{B_1, B_2\}$ and $C = \{C_1, C_2\}$ as in Fig. 6, where the trajectories extend to some large T. T_1 and T_2 are fixed. T_3 grows with T. To make calculations simpler, we work with $\mathcal{D}$ divided by T in Def. 3.

Let us compute $\mathcal{D}(A,B)$ first. The association Σ_{MOT} for this distance is $\{(1,2), .., (1,2), (2,1), ...\}$ because (i) we start with the association $\{A_1 \leftrightarrow B_1, A_2 \leftrightarrow B_2\}$, (ii) at some point after T_1 we need to change the association to $\{A_1 \leftrightarrow B_2, A_2 \leftrightarrow B_1\}$ because the initial association exceeds $thr_{MOTP} < 2$ and (iii) for times after the $T_1 + T_2$ the association $\{A_1 \leftrightarrow B_2, A_2 \leftrightarrow B_1\}$ is anchored because $thr_{MOT} > 1$. Therefore, $\mathcal{D}(A,B) > \frac{2(T-T_2)}{T}$.

Now we compute $\mathcal{D}(A,C)$. The association for $\mathcal{D}(A,C)$ is $\Sigma_{\text{MOT}} = \{(1,2), ..., (1,2)\}$ because (i) we start with $\{A_1 \leftrightarrow C_1, A_2 \leftrightarrow C_2\}$, (ii) the association $A_1 \leftrightarrow C_1$ is always anchored because the distance between A_1 and C_1 is always zero and thus always smaller than $thr_{MOT} > 1$ and (iii) after some point, when the distance between A_2 and C_2 exceeds $thr_{MOT} < 2$, MOTP still keeps the association $A_2 \leftrightarrow C_2$ because A_1 and C_1 are already anchored. Under this association, numerical computation leads to $\mathcal{D}(A,C) < \frac{4T_2}{T}$. Similarly, $\mathcal{D}(C,B) < \frac{4T_2}{T}$.

Therefore, for T large enough we have $\mathcal{D}(A,B) > \frac{2(T-T_2)}{T} > \frac{4T_2}{T} + \frac{4T_2}{T} > \mathcal{D}(A,C) + \mathcal{D}(C,B)$. □

Appendix B
Properties of our metrics: $\mathcal{D}_{nat}$

To prove Theorem 5, we need the following lemma.

Lema 4. *The map d^+ is a metric on $\mathbb{R}^p \cup \{*\}$.*

Proof. We verify that d^+ satisfies the four conditions of a metric. Let $x'', x', x \in \mathbb{R}^p \cup \{*\}$.

Non-negativity and symmetry are obvious.

To verify the coincidence property, observe that $d^+(x,x') = 0$ implies either $x = x' = *$ or, since $M > 0$, $d^+(x,x') = d(x,x') = 0$. Because d is a metric, this in turn implies that $x = x' = 0$. In other words, $d^+(x,x') = 0 \Leftrightarrow x = x'$.

To verify the subadditivity property, we need to consider eight different cases of the triangle inequality. We first consider the non-trivial cases. If $x, x', x'' \in \mathbb{R}^p$, then

$$\begin{aligned} d^+(x,x'') &= \min\{2M, d(x,x'')\} \\ &\leq \min\{2M, d(x,x') + d(x',x'')\} \\ &\leq \min\{2M, d(x,x')\} + \min\{2M, d(x',x'')\} \\ &= d^+(x,x') + d^+(x',x''). \end{aligned}$$

If $x'{=}*$ and $x, x'' \in \mathbb{R}^p$, then $M{=}d^+(x',x''){=}d^+(x,x')$ and

$$d^+(x,x'') = \min\{2M, d(x,x'')\} \leq d^+(x,x') + d^+(x',x'').$$

It is easy to check the other six cases. □

Proof of Theorem 5. Let A, B, C be three elements in $\mathcal{S}$. We verify the four conditions of a metric for $\mathcal{D}_{nat}$.

Coincidence property: We show that $\mathcal{D}_{nat}(A,B) = 0$ if and only if $A = B$. Recall that A and B are unordered

sets of trajectories. Hence $A = B$ implies that there is an isomorphism between A and B. In other words, they are equal apart from a relabeling of the elements. If $A = B$ and we set $\Sigma = (\sigma, \sigma, ..., \sigma)$ then the function to minimize on the right-hand-side of equation (1) is equal to zero. This σ is an isomorphism between A and B. Since the minimum of (1) must always be non-negative, we conclude that $A = B \Rightarrow \mathcal{D}_{nat}(A, B) = 0$. Conversely, assume that $\mathcal{D}_{nat}(A, B) = 0$ and let $\Sigma^* = (\Sigma^*(1), ..., \Sigma^*(T))$ be a minimizer in (1). $\mathcal{D}_{nat}(A, B) = 0$ implies that $\mathcal{K}(\Sigma^*) = 0$. Therefore $\Sigma_i^*(t) = \Sigma_i^*(1)$, for all t and i. Since the labeling of the trajectories does not affect the computation of $\mathcal{D}_{nat}$, we assume without loss of generality that their labeling is such that we can write $\Sigma_i^*(t) = i$. $\mathcal{D}_{nat}(A, B) = 0$ also implies that, for all t and i, we have $d^+(A_i^+(t), B^+_{\Sigma_i^*(t)}(t)) = d^+(A_i^+(t), B_i^+(t)) = 0$. Since d^+ is a metric, this in turn implies that $A_i^+(t) = B_i^+(t)$ for all i and t, which is the same as saying that $A^+ = B^+$. Hence $A = B$. To be more specific, A is equal to B apart from a relabeling of its trajectories.

Symmetry property: Since $\mathcal{D}_{nat}$ only depends on A and B through d^+ we have $\mathcal{D}_{nat}(A, B) = \mathcal{D}_{nat}(B, A)$.

Subadditivity property: We prove $\mathcal{D}_{nat}(A, C) \leq \mathcal{D}_{nat}(A, B) + \mathcal{D}_{nat}(B, C)$. First, notice that we can add any extra number of $*$-only trajectories to A, B or C without changing $\mathcal{D}_{nat}$. Recall that m is number of trajectories in A^+, B^+ and C^+. In this part of the proof, m should be the sum of the cardinalities of the two sets of highest cardinality among A, B and C. In Section III, m was just the sum of the cardinalities of A and B. T is the maximum time index observed in A, B and C.

Since d^+ is a metric, we can write that

$$d^+(A_i^+(t), C^+_{\Sigma_i(t)}(t)) \leq d^+(A_i^+(t), B^+_{\Sigma'_{\Sigma_i(t)}(t)}(t)) + d^+(B^+_{\Sigma'_{\Sigma_i(t)}(t)}(t), C^+_{\Sigma_i(t)}(t)) \tag{11}$$

for any $\Sigma' = (\Sigma'(1), ..., \Sigma'(T)) \in \Pi^T$ and for all i and t. Now notice that

$$\sum_{i=1}^m d^+(B^+_{\Sigma'_{\Sigma_i(t)}(t)}(t), C^+_{\Sigma_i(t)}(t)) = \sum_{i=1}^m d^+(B^+_{\Sigma'_i(t)}(t), C_i^+(t)).$$

Using this together with (11), we can write

$$\sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), C^+_{\Sigma_i(t)}(t)) \leq \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), B^+_{\Sigma'_{\Sigma_i(t)}(t)}(t)) + d^+(B^+_{\Sigma'_i(t)}(t), C_i^+(t)).$$

Let us define $\Sigma'' = (\Sigma''(1), ..., \Sigma''(T)) \in \Pi^T$ where $\Sigma''_i(t) = \Sigma'_{\Sigma_i(t)}(t)$. This means $\Sigma'' = \Sigma' \circ \Sigma$. We can use Σ' to rewrite the expression above as

$$\sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), C^+_{\Sigma_i(t)}(t)) \leq \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), B^+_{\Sigma''_i(t)}(t)) + \sum_{t=1}^T \sum_{i=1}^m d^+(B^+_{\Sigma'_i(t)}(t), C_i^+(t)). \tag{12}$$

Using Definition 5, we can write

$$\mathcal{K}(\Sigma) = \mathcal{K}(\Sigma'^{-1} \circ \Sigma' \circ \Sigma) \leq \mathcal{K}(\Sigma'^{-1}) + \mathcal{K}(\Sigma' \circ \Sigma) = \mathcal{K}(\Sigma') + \mathcal{K}(\Sigma''). \tag{13}$$

Now we add both sides of (12) and (13) and obtain $\mathcal{K}(\Sigma) + \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), C^+_{\Sigma_i(t)}(t)) \leq \mathcal{K}(\Sigma'') + \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), B^+_{\Sigma''_i(t)}(t)) + \mathcal{K}(\Sigma') + \sum_{t=1}^T \sum_{i=1}^m d^+(B^+_{\Sigma'_i(t)}(t), C_i^+(t))$.

Finally, we find the minimum of both sides of the inequality over all pairs of Σ and Σ'. Recall the relationship $\Sigma'' = \Sigma' \circ \Sigma$ and the fact that we can choose Σ' independently of Σ. Consequently,

$$\begin{aligned}
&\min_{\Sigma \in \Pi^T} \mathcal{K}(\Sigma) + \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), C^+_{\Sigma_i(t)}(t)) \\
&\leq \min_{\Sigma'' \in \Pi^T} \mathcal{K}(\Sigma'') + \sum_{t=1}^T \sum_{i=1}^m d^+(A_i^+(t), B^+_{\Sigma''_i(t)}(t)) \\
&+ \min_{\Sigma' \in \Pi^T} \mathcal{K}(\Sigma') + \sum_{t=1}^T \sum_{i=1}^m d^+(B^+_{\Sigma'_i(t)}(t), C_i^+(t)).
\end{aligned}$$

Hence, $\mathcal{D}_{nat}(A, C) \leq \mathcal{D}_{nat}(A, B) + \mathcal{D}_{nat}(B, C)$. □

Proof of Theorem 1. The first two properties of the permutation measure are trivial to verify. To verify the third property, it is sufficient to prove that

$$\begin{aligned}
&\mathbb{I}\left((\Sigma'(t+1) \circ \Sigma(t+1)) \circ (\Sigma'(t) \circ \Sigma(t))^{-1} \neq I\right) \\
&\leq \mathbb{I}(\Sigma(t+1) \circ \Sigma(t)^{-1} \neq I) + \mathbb{I}(\Sigma'(t+1) \circ \Sigma'(t)^{-1} \neq I).
\end{aligned}$$

Since the left-hand-side is at most 1, we only need to consider the case in which the right-hand-side is less than 1, i.e.,

$$\mathbb{I}(\Sigma(t+1) \circ \Sigma(t)^{-1} \neq I) + \mathbb{I}(\Sigma'(t+1) \circ \Sigma'(t)^{-1} \neq I) = 0.$$

In this case, it follows that $\Sigma(t+1) = \Sigma(t)$ and $\Sigma'(t+1) = \Sigma'(t)$. Consequently, the left-hand-side equals 0 as well in this case. Hence the property is verified. □

Proof of Theorem 2. The first two properties of the permutation measure are immediate to check from the definition of $\mathcal{K}_{trans}$. To prove the third property it suffices to show that

$$\begin{aligned}
&k_{Cayley}\left((\Sigma'(t+1) \circ \Sigma(t+1)) \circ (\Sigma'(t) \circ \Sigma(t))^{-1}\right) \\
&\leq k_{Cayley}(\Sigma(t+1)) \circ \Sigma(t)^{-1}) + k_{Cayley}(\Sigma'(t+1) \circ \Sigma'(t))^{-1}).
\end{aligned}$$

To prove this we use of the following two facts.

Fact 1: if $\sigma, \sigma' \in \Pi$ then $k_{Cayley}(\sigma \circ \sigma') \leq k_{Cayley}(\sigma) + k_{Cayley}(\sigma')$. To see why this is true, notice that this inequality can be rewritten as $k_{Cayley}(a \circ c^{-1}) \leq k_{Cayley}(a \circ b^{-1}) + k_{Cayley}(b \circ c^{-1})$ where $a = \sigma \circ \sigma'$, $b = \sigma'$ and $c = I$, the identity permutation. Then, it is basically saying that shortest set of transpositions that take σ_1 to σ_3 is smaller than any set of transpositions that first take σ_1 to σ_2 and then σ_2 to σ_3.

Fact 2: if $\sigma, \sigma' \in \Pi$ then $k_{Cayley}(\sigma \circ \sigma' \circ \sigma^{-1}) = k_{Cayley}(\sigma')$. This is true because $\sigma \circ \sigma' \circ \sigma^{-1}$ is just a relabeling of the permutation σ' and k_{Cayley} is invariant to relabeling.

Let

$\sigma_A = \Sigma(t) \circ \Sigma(t+1)^{-1}, \quad \sigma_B = \Sigma'(t) \circ \Sigma'(t+1)^{-1},$
and $\sigma_C = \Sigma'(t) \circ \sigma_A \circ \Sigma'(t)^{-1}$.

Observe that the permutation $\sigma_C \circ \sigma_B$ satisfies

$$\sigma_C \circ \sigma_B \circ \left((\Sigma'(t+1) \circ \Sigma(t+1)) \circ (\Sigma'(t) \circ \Sigma(t))^{-1}\right) = I.$$

Finally, we can apply Fact 2 followed by Fact 1, and obtain (dropping $_{Cayley}$ for clarity).

$$\begin{aligned} &k\left((\Sigma'(t+1) \circ \Sigma(t+1)) \circ (\Sigma'(t) \circ \Sigma(t))^{-1}\right) \\ &= k(\sigma_B^{-1} \circ \sigma_C^{-1}) \leq k(\sigma_B^{-1}) + k(\sigma_C^{-1}) = k(\sigma_B^{-1}) + k(\sigma_A^{-1}) \\ &= k(\Sigma(t+1) \circ \Sigma(t)^{-1}) + k(\Sigma'(t+1) \circ \Sigma'(t)^{-1}). \end{aligned}$$

□

Proof of Theorem 3. We give a counter example that violates property (iii) for $\beta = 1$. It is also easy to come up with similar counter examples that violate property (iii) for any value of $\beta \geq 1$. Let $I = (1, 2)$ be the identity permutation and let $\sigma = (2, 1)$ be the permutation that swaps 1 and 2. Let $\Sigma = (I, \sigma, \sigma)$ and let $\Sigma' = (I, I, \sigma)$. We have $\mathcal{K}_{maxcount}(\Sigma) = 1$ and $\mathcal{K}_{maxcount}(\Sigma') = 1$ but $\mathcal{K}_{maxcount}(\Sigma' \circ \Sigma) = \infty > \mathcal{K}_{maxcount}(\Sigma') + \mathcal{K}_{maxcount}(\Sigma)$. □

Proof of Theorem 6. We provide a special case where the triangle inequality is violated. Let A, B and C be three sets of trajectories. We assume $\beta = 1$, but it is easy to change A, B and C such that the proof holds for any β,[2]. Let $I = (1, 2)$ be the identity permutation and let $\sigma^0 = (2, 1)$ be the permutation that swaps 1 and 2. Let $A = \{A_1, A_2\}$, $B = \{B_1, B_2\}$, $C = \{C_1, C_2\}$, where

$$\begin{aligned} &A_1 = (2, -2, -2), \quad B_1 = (2, 2, 2), \quad C_1 = (2, 2, -2), \\ &A_2 = (-2, 2, 2), \quad B_2 = (-2, -2, -2), \quad C_2 = (-2, -2, 2). \end{aligned}$$

Now consider equation (1) with $\mathcal{K} = \mathcal{K}_{maxcount}$ and $\beta = 1$. The optimization problem for $\mathcal{D}_{nat}(A, B)$,

$$\min_{\Sigma \in \Pi^T} \left\{ \mathcal{K}(\Sigma) + \sum_{t=1}^{3} \sum_{i=1}^{2} d^+(A_i^+(t), B^+_{\Sigma_i(t)}(t)) \right\}$$

has a minimum of 1 at $\Sigma = (I, \sigma^0, \sigma^0)$. We can see this because $\mathcal{K}$ forces us to either do one switch only or no switch at all. Doing no switch at all makes us incur a cost larger than 1 in the distance term. Similarly, the optimization problem for $\mathcal{D}_{nat}(B, C)$ has a minimum of 1 at $\Sigma = (I, I, \sigma^0)$. When we solve the optimization problem for $\mathcal{D}_{nat}(A, C)$, we are only allowed to perform one change in the association between A and C, otherwise the term $\mathcal{K}_{maxcount}$ makes us pay a very large cost. With only one change in the association, we incur a distance of 4 for $t = 1$ or $t = 2$ or $t = 3$. Hence, $\mathcal{D}_{nat}(A, C) \geq 4 > 1 + 1 = \mathcal{D}_{nat}(A, B) + \mathcal{D}_{nat}(B, C)$. □

[2]It is easy to see that we do not in fact need d to be the Euclidean distance for the proof to hold.

Appendix C
Properties of our metrics: $\mathcal{D}_{comp}$

Proof of Theorem 7. Let A, B and C be any elements of $\mathcal{S}$.
Coincidence property: If $\mathcal{D}_{comp}(A, B) = 0$, then $W(t) = W(1)$ for all t. Hence, $D_{ij}^{AB}(t) = 0$ for all i, j such that $W_{ij}(1) > 0$. Recall that if W_{ij} is a doubly stochastic matrix, then there exists a permutation $\sigma = (\sigma_1, \ldots, \sigma_m) \in \Pi$, such that $W_{i\sigma_i} > 0$ for all i. Therefore, $D_{i\sigma_i}^{AB}(t) = 0$ for all t and i. In other words, A and B are identical apart from a relabeling of their elements.
Symmetry property: Using the properties of **trace**, we have

$$\begin{aligned} &\mathbf{tr}(W^\dagger(t) D^{AB}(t)) = \mathbf{tr}((W^\dagger(t) D^{AB}(t))^\dagger) \\ &= \mathbf{tr}(D^{AB}(t)^\dagger W(t)) = \mathbf{tr}(D^{BA}(t) W(t)) = \mathbf{tr}(W(t) D^{BA}(t)). \end{aligned}$$

Minimizing with respect to $\{W(t)\}$ is the same as minimizing with respect to $\{W(t)^\dagger\}$. Thus $\mathcal{D}_{comp}(A, B) = \mathcal{D}_{comp}(B, A)$.
Subadditivity property: We prove that $\mathcal{D}_{comp}(A, C) \leq \mathcal{D}_{comp}(A, B) + \mathcal{D}_{comp}(B, C)$.

First, notice that we can add any extra number of $*$-only trajectories to A, B or C without changing $\mathcal{D}_{comp}$. Recall that m is number of trajectories in A^+, B^+ and C^+. In this part of the proof, m should be the sum of the cardinalities of the two sets of highest cardinality among A, B and C. In Section III, m was just the sum of the cardinalities of A and B. T is the maximum time index observed in A, B and C.

First note that, since d^+ is a metric we have that $D^{AC}(t)_{ij} \leq D_{ik}^{AB}(t) + D_{kj}^{BC}(t)$ for any k. Let $W_1 = (W_1(1), ..., W_1(T)) \in \mathcal{P}^T$ and $W_2 = (W_2(1), ..., W_2(T)) \in \mathcal{P}^T$. We multiply both sides of the previous inequality by $W_1(t)_{ik} W_2(t)_{kj}$ and sum over i, j, k to obtain

$$\sum_{i,j,k=1}^{m} W_{1ik} W_{2kj} {D^{AC}}_{ij} \leq \sum_{i,j,k=1}^{m} W_{1ik} W_{2kj} D_{ik}^{AB} + W_{1ik} W_{2kj} D_{kj}^{BC},$$

where we omitted t in W and D. Since $W_1(t), W_2(t) \in \mathcal{P}$, we have $\sum_{j=1}^{m} W_2(t)_{kj} = 1$ and $\sum_{i=1}^{m} W_1(t)_{ik} = 1$. To simplify this expression, and again omitting t in W and D, we re-write the previous inequality in matrix notation as

$$\mathbf{tr}\left((W_1 W_2)^\dagger D^{AC}\right) \leq \mathbf{tr}({W_1}^\dagger D^{AB}) + \mathbf{tr}({W_2}^\dagger D^{BC}).$$

From our assumption on $\|.\|$ (c.f. Property (7)), we have

$$\|W_1(t+1) W_2(t+1) - W_1(t) W_2(t)\| \leq \|W_1(t+1) - W_1(t)\| + \|W_2(t+1) - W_2(t)\|.$$

Adding the last two inequalities, summing over t and minimizing over $W_1, W_2 \in \mathcal{P}^T$ we have

$$\min_{W_1,W_2\in\mathcal{P}^T} \sum_{t=1}^{T-1} \|W_1(t+1)W_2(t+1) - W_1(t)W_2(t)\| + \sum_{t=1}^{T} \mathbf{tr}\left((W_1(t)W_2(t))^\dagger D^{AC}(t)\right)$$
$$\leq \min_{W_1\in\mathcal{P}^T} \sum_{t=1}^{T-1} \|W_1(t+1) - W_1(t)\| + \sum_{t=1}^{T} \mathbf{tr}(W_1(t)^\dagger D^{AB}(t))$$
$$+ \min_{W_2\in\mathcal{P}^T} \sum_{t=1}^{T-1} \|W_2(t+1) - W_2(t)\| + \sum_{t=1}^{T} \mathbf{tr}(W_2(t)^\dagger D^{BC}(t)).$$

Note that for any $W_1(t), W_2(t) \in \mathcal{P}$, we have $W_1(t)W_2(t) \in \mathcal{P}$. Therefore, in the minimization performed on the left hand side, we can replace $W_1(t)W_2(t)$ by a single $W(t)$. The subadditivity property follows. □

Proof of Lemma 3. Let $w_1, w_2, w_1', w_2' \in \mathcal{P}$, whose norms are less than 1.

$$\|w_2'w_2 - w_1'w_1\| \leq \|w_2'(w_2 - w_1 + w_1) - w_1'w_1\|$$
$$\leq \|w_2'(w_2 - w_1)\| + \|w_2'w_1 - w_1'w_1\| \leq \|w_2'\|\|(w_2 - w_1)\|$$
$$+ \|(w_2' - w_1')\|\|w_1\| \leq \|(w_2 - w_1)\| + \|(w_2' - w_1')\|.$$

The third inequality above is obtained using sub-multiplicity. □

Proof of Theorem 8. The constraints which define $\mathcal{P}$ are a set of linear constraints. In addition, the first term in the objective is a linear function of $\{W(t)\}$. Notice also that we can replace the term $\sum_{t=1}^{T-1} \|W(t+1) - W(t)\|$ in the objective by $\sum_{t=1}^{T-1} e_t$, if we add the additional constraints that $\|W(t+1) - W(t)\| \leq e_t$ for all t. We can also represent these additional constraints as linear constraints. Specifically, since

$$\|W(t+1) - W(t)\| = \max_j \sum_i |W_{ij}(t+1) - W_{ij}(t)|,$$

each of these constraints can be replaced by $\sum_i h_{ij}(t) \leq e_t$ for all t and j if we add the additional constraints

$$|W_{ij}(t+1) - W_{ij}(t)| \leq h_{ij}(t)$$

for all i, j and t. Each of these constraints can be replaced two linear constraints, namely

$$W_{ij}(t+1) - W_{ij}(t)) \leq h_{ij}(t),$$
$$-W_{ij}(t+1) + W_{ij}(t)) \leq h_{ij}(t)$$

□

Appendix D
Some extra observations

In this section make some extra observations that are not essential to understand what was presented above. We use the same setup and notation as explained in Section III of the main paper.

Another example of OSPA-ST producing counter-intuitive results

Consider the setup in Figure 7 where two people, A_1 and A_2, pass by each other and two trackers, B and C, produce output trajectories B_1, B_2, C_1 and C_2. Space is measure on the y-axis (bottom to top) and time on the x-axis (left to right). To aid visualization, close lines should be considered on top of each other.

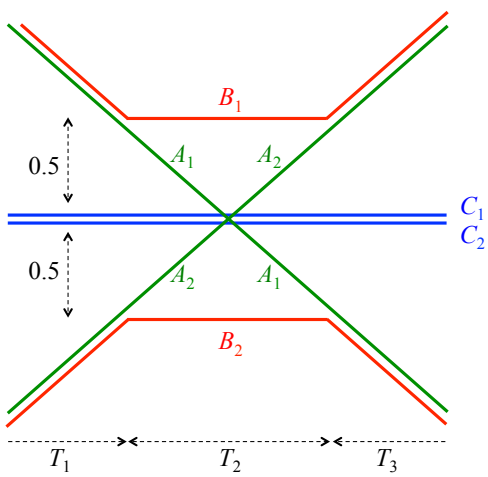

Fig. 7. Counter example that shows that $\mathcal{D}_{MOTA}$ is not a metric.

Let $T_1 = T_3$. By symmetry and the fact that OSPA-ST does not allow associations that change with time we can assume without loss of generality that in computing $\mathcal{D}_{OSPA-ST}(A, B)$ OSPA-ST associates A_1 to B_1 and A_2 to B_2 and that in computing $\mathcal{D}_{OSPA-ST}(A, C)$ OSPA-ST associates A_1 to C_1 and A_2 to C_2.

If $T_2 << T_1 = T_3$, the average distance between A and B can be arbitrarily close to the average distance between A and C. More formally, $\mathcal{D}_{OSPA-ST}(A, C)/\mathcal{D}_{OSPA-ST}(A, B) \to 1$ as $T_1 = T_3 \to \infty$. In other words, OSPA-ST says that tracker C is as good as tracker B while our intuition says that B is the better tracker. B produces good tracks for most of the time and simply makes an identity switch during internval T_2. On the other hand, C is a trivial tracker that always outputs 0 regardless of the input.

MOTA does not define a metric

First recall how to compute MOTA between two sets of trajectories A and B. First compute the CLEAR MOT association Σ_{MOT} between A and B as described in Appendix A. Second, compute a positive linear combination, μ, of three quantities ν_1, ν_2 and ν_3 where: ν_1 is the number of changes of association between trajectories; ν_2 is the number of points of trajectories in A unassociated to any point in B; and ν_3 is the number of points of trajectories in B unassociated to any point in A. MOTA is defined as $1 - \mu$ but since distances must decrease as A and B get similar we define $\mathcal{D}_{MOTA}(A, B) = \mu$.

Lema 5. *$\mathcal{D}_{MOTA}$ is not a metric for any threshold $thr_{MOT} > 0$ or positive linear combination used to define μ.*

Proof. We given an example of three sets A, B and C such $\mathcal{D}_{MOTA}(A, B) > 0$ while $\mathcal{D}_{MOTA}(A, C) = \mathcal{D}_{MOTA}(C, B) = 0$, hence the triangle inequality is violated. Our counter example works for $0.5 < thr_{MOT} < 1$ but by scaling space we can prove the lemma for any thr_{MOT}.

Consider A, B and C as in Figure 1 in the main paper, which we reproduce here for convenience. To aid visualization, close lines should be considered on top of each other and transitions in A almost instantaneous. Time is on the x-axis (left to right) and space is on the y-axis (bottom to top).

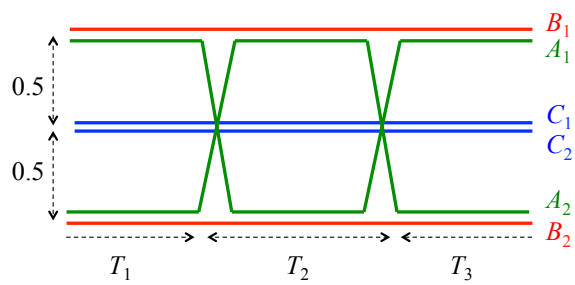

Fig. 8. Counter example that shows that $\mathcal{D}_{MOTA}$ is not a metric.

When we compute $\mathcal{D}_{MOTA}(A, C)$, Σ_{MOT} associates A_1 to C_1 and A_2 to C_2 for all times t. (we can interchange C_1 and C_2 since they are equal). There are no changes in association because the distances computed are always less than 0.5 and $0.5 < thr_{MOT}$, thus $\nu_1 = 0$. In addition, there are no unallocated tracks so $\nu_2 = \nu_3 = 0$. Therefore $\mathcal{D}_{MOTA}(A, C) = \mu = 0$ regardless of the coefficients in the linear combination μ. Similarly we get $\mathcal{D}_{MOTA}(C, B) = 0$.

When we compute $\mathcal{D}_{MOTA}(A, B)$ we find the following: before T_1, Σ_{MOT} associates A_1 to B_1; after T_1 and before $T_1 + T_2$, the distance between A_1 and B_1 becomes $1 > thr_{MOT}$ and so Σ_{MOT} changes association and matches A_1 to B_2. After $T_1 + T_2$, the distance between A_1 and B_2 becomes $1 > thr_{MOT}$ and so Σ_{MOT} changes association back to A_1 matches to B_1. We change association twice and so $\nu_1 > 0$ We just focused on who A_1 matches to. The remaining tracks are matched to each other so $\nu_2 = \nu_3 = 0$. In short $\mathcal{D}_{MOTA}(A, B) = \mu > 0$.

□