Estimate’s Statistics in the Performance Evaluation of Extended Object Tracker

Teodor Dehelean¹, Corina Nafornita¹, Alexandru Isar¹
¹ Communications Dept. Politehnica University of Timisoara, Timisoara, Romania
cosmin.dehelean@student.upt.ro, corina.nafornita@upt.ro, alexandru.isar@upt.ro

Abstract—This paper considers the statistical parameters of all characteristic points of a target’s estimate, when assessing the performance of extended object tracker. For this, the established optimal subpattern assignment metric is modified by replacing the Euclidean base distance between the true and the estimated target state-vectors with the Mahalanobis distance. It represents the distance between a deterministic ground truth target position and its corresponding estimate, which is characterized by a multivariate Gaussian distribution.

The necessity of considering the track covariance into the metric is proven by experimental results, obtained with a Gaussian Mixture Probability Hypothesis Density Filter (GM-PHD).

Adding the track uncertainty to the OSPA metric definition for extended object tracker leads to a fast and intuitive assessment method for modern tracking algorithms, as used for example in automotive lidar sensors, radar sensors or future traffic control systems for flying taxis.

Keywords: OSPA metric, Mahalanobis distance, extended object tracker, radar

I. INTRODUCTION

To assess the performance of multiple object tracking systems, the distance between two sets of tracks, the true and the system’s estimates set must be measured. For extended object tracking algorithms, various measures of effectiveness must be merged to obtain a reduced number of key performance indicators that outline all relevant aspects. If considered individually, these measures can lead to wrong or inadequate results as some of them are correlated and intransitive. For example, waiting some time prior to opening new tracks in order to avoid a high number of false alarms typically increases the track initiation delay.

Past publications [3] consider for both the estimated and the true target state some assigned uncertainty to assess the localization error as part of the OSPA metric. Using the Hellinger distance as an alternative to the original Euclidean distance in [1] and [2] assumes that the ground truth annotation is itself ambiguous. In tool-based as well as in human tool-supported annotation of computer vision data, intensive efforts are invested to ensure the highest quality of this reference data. Very performant and in consequence very expensive reference sensor systems are used for ground truth data acquisition, supplemented with a thorough, subsequent step conducted by test and quality assurance engineers.

Lidars, radars and multisensor systems permit through their improved resolution a precise representation of each tracked object in the field of view. Relevant targets could be 2D or 3D objects and are tracked as extended objects, where multiple resolution cells are occupied by the same object. A typical car model inside an embedded automotive radar sensor consists of eight characteristic points, four on the car corners and four points on the wheel-houses, denoted as a Set of Points on a Rigid Body (SPRB) [4]. All characteristic points per extended object are present in both the estimated and the true data set, leading to a very computation-intensive processing, if the Hellinger distance is taken as basis for the OSPA metric. Again, the introduction and the selection of a covariance matrix for the true state allow a theoretical valuable, yet counterintuitive metric definition. In this paper we propose to consider the estimate’s uncertainty when computing the OSPA metric, by using instead of the original Euclidean distance the Mahalanobis distance. It represents the distance between a deterministic ground truth target position and its corresponding estimate, which is characterized by a Gaussian distribution, as in [8].

The paper is organized as follows. In Section 2, we describe the norm-based and matrix-based distances proposed in literature, as well as their disadvantages; in Section 3 we propose an enhanced method to assess extended object tracker performances based on the OSPA metric and the Mahalanobis distance; Section 4 contains simulations results and last section gives some concluding remarks and possible directions for future work.

II. BASE DISTANCES

A. Euclidean Distance

A norm is a mapping of a vector space \( V \) to \( \mathbb{R}^n \), the set of positive, real numbers:

\[
\| \cdot \| : V \rightarrow \mathbb{R}^n, \mathbb{R}^n = \{ x \in \mathbb{R} | x > 0 \}
\]

(1)

Various norms, so called p-norms, can be described through the following equation:

\[
\| X \|_p = (\sum_{i=1}^{n} x_i^p)^{1/p}, X \in \mathbb{R}^n and p \in \mathbb{R}, 1 \leq p \leq \infty \]  \nas (2)

The commonly used Euclidean distance is derived from the 2-norm, by the relation:

\[
d_2(X, Y) = \| X - Y \|_2, with p = 2 \]  \nas (3)

Note that \( X \) and \( Y \) represent n-dimensional vectors in the state space \( \mathbb{R}^n \), containing for example geometric, velocity and
acceleration components of one characteristic point of a target of interest.

### B. Mahalanobis Distance

The first field in which the Mahalanobis distance was applied is in craniometrics and anthropological studies, having nowadays various applications like medical diagnosis and statistical pattern recognition in remote object sensing. The original idea was to use the distance between a test point and several distributions to classify this point as belonging to one of the available classes.

In our application, the deterministic point is represented by the ground truth state \( x_k \) of a track at time \( t_k \), while the random variable is given by the estimate’s state. For one target, the corresponding estimated state at time \( t_k \) is a multidimensional vector \( \mathbf{y}_k \) which may include range (height, width and length), velocity \((v_H, v_W, v_L)\) and acceleration \((a_{H}, a_{W}, a_{L})\) coordinates. In the following, we will consider \( \mathbf{y}_k \) as being a multivariate normal distribution, parametrized by the mean vector \( \mathbf{\mu}_y \) and the covariance matrix \( \mathbf{\Sigma}_y \).

The Mahalanobis distance \( D_m \) is unitless and scale-invariant [5]. It emphasizes the correlations between the variables of a data set and is given by the relation:

\[
D_m^2(x_k) = (x_k - \mathbf{\mu}_y)^T \mathbf{\Sigma}_y^{-1}(x_k - \mathbf{\mu}_y) \tag{4}
\]

A radar sensor for example computes for every radar cycle the mean values and the standard deviations for all constitutive components of the state vector. This means that for every radar cycle, multiple detections are assigned to the same characteristic point. Obviously, these individual variables are correlated, so that also pairwise covariance values are also available per radar cycle. The estimate’s mean vector is (where operator \( \mathbf{\Sigma}^T \) represents the transpose):

\[
\mathbf{\mu}_y = \left( \mu_H, \mu_W, \mu_L, \mu_{a_H}, \mu_{a_W}, \mu_{a_L} \right)^T \tag{5}
\]

and the covariance matrix is:

\[
\mathbf{\Sigma}_y = \begin{pmatrix}
\text{var} (H) & \text{cov} (H, L) & \ldots & \text{cov} (H, a_L) \\
\text{cov} (L, H) & \text{var} (L) & \ldots & \text{cov} (L, a_L) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov} (a_H, H) & \text{cov} (a_L, L) & \ldots & \text{var} (a_L)
\end{pmatrix} \tag{6}
\]

As an example, let’s consider a three dimensional state-vector \( \mathbf{est} = (H, W, L)^T \) and four observation groups, originated each from a different tracking algorithm, around a target with the ground truth position \( x_{GT} = (0, 0, 0)^T \). The goal is to compare the localization precision of the four synthetic trackers with respectively without taking into account the associated uncertainties. Tracker1 is the baseline. Tracker2 has increased variances of all variables of the state vector. Tracker3 provides a higher correlation between the height \( (H) \) and the width \( (W) \) values, while tracker4 shows a very small correlation between height and width. The covariance matrices of the four tracking algorithms were selected as shown in Fig. 1. Increased parameters are marked with green, smaller ones with red. By varying the mean values of the estimates in every dimension by one respectively two digits, we obtain the results in Figure 2.

\[
\begin{align*}
\text{Sigma}_y &= [1.0 \, 0.8 \, 0.9 \, 1.0 \, 7.0 ; \, 0.8 \, 0.7] \\
\text{Sigma}_y &= [2.0 \, 0.8 \, 0.9 \, 2.0 \, 7.0 ; \, 0.8 \, 0.7] \\
\text{Sigma}_y &= [2.1 \, 0.8 \, 0.1 \, 2.0 \, 7.0 ; \, 0.8 \, 0.7] \\
\text{Sigma}_y &= [2.0 \, 0.1 \, 0.8 \, 1.0 \, 7.0 ; \, 0.8 \, 0.7]
\end{align*}
\]

Figure 1 Covariance matrices for the four analyzed trackers

First interesting remark is that the different tracking algorithms have significant performance differences in localizing targets, which cannot be observed when looking at the Euclidean distance only. The squared Euclidean distance \( D_e^2 \) considers only the localization error between the estimate’s mean vector and the ground truth reference and is the same for every tracker at a given mean vector.

\[
\begin{align*}
\text{Mean Vector (1,0,0)T} &\quad \text{Mean Vector (0,1,0)T} &\quad \text{Mean Vector (0,0,1)T} \\
\text{Dm}^2 [10e-3] &\quad \text{De}^2 [10e-3] &\quad \text{Dm}^2 [10e-3] \\
\text{Tracker 1} &\quad 6625.5 &\quad 5721.1 &\quad 2322.7 \\
\text{Tracker 2} &\quad 885.1 &\quad 676.4 &\quad 693.3 \\
\text{Tracker 3} &\quad 2543.1 &\quad 2670.5 &\quad 420.5 \\
\text{Tracker 4} &\quad 474.3 &\quad 557.9 &\quad 1110.5
\end{align*}
\]

Figure 2 Mahalanobis \((D_m)\) vs. Euclidean \((D_e)\) distances of four analyzed trackers

Secondly, jointly increased variances of the height, width and length values lead to improved, i.e. reduced squared Mahalanobis distances \( D_m^2 \) (see tracker2 vs. tracker1). It is more likely that the ground truth point belongs to the wider distribution than to a narrow one. Figure 3 illustrates why, for a one-dimensional case. The probability of a test point to belong to the Gaussian distribution 1 (blue line, with \( \mu_1 = 0 \) and \( \sigma_1 = 1 \)) is lower than the probability to belong to the Gaussian distribution 2 (red line, with \( \mu_2 = 0 \) and \( \sigma_2 = 2 \)), if the test point is outside of the interval given by the two intersection points [-1,35, 1,35]. For a test point outside this interval, the Mahalanobis distance would be bigger relative to the blue Gaussian distribution 1 than relative to the red Gaussian distribution 2.

\[
\begin{align*}
\text{Mean Vector (2,0,0)T} &\quad \text{Mean Vector (0,2,0)T} &\quad \text{Mean Vector (0,0,2)T} \\
\text{Dm}^2 [10e-3] &\quad \text{De}^2 [10e-3] &\quad \text{Dm}^2 [10e-3] \\
\text{Tracker 1} &\quad 2734.8 &\quad 2212.3 &\quad 9217.1 \\
\text{Tracker 2} &\quad 3041.2 &\quad 2630.3 &\quad 3775.2 \\
\text{Tracker 3} &\quad 10831.3 &\quad 10384 &\quad 2296.7 \\
\text{Tracker 4} &\quad 2038.7 &\quad 2439.6 &\quad 3974.4
\end{align*}
\]

Figure 3 Crossings of two normal distributions at (-1.35, 0.16) and (1.35, 0.16)
The effect of an increased respectively decreased correlation between two variables of the state-vector can be seen in the results for tracker3 and tracker4 shown in Figure 2. A strong correlation between the height and the width distributions leads to higher Mahalanobis distances, while observations that are less correlated effect the convergence of the Mahalanobis distance to the Euclidean one.

III. INTRODUCING TRACK COVARIANCE IN OSPA METRIC

A. Extended Object Tracker

In multiple object tracking problems, the target objects like cars, passengers, bicyclists, airplanes or ships have a spatial extent. Also, small-size particles like cells or atoms occupy several resolution cells, if the examination sensor is performant enough. Depending on the sensor performance, mainly its resolution, to distinguish the target’s details, a point tracking or an extended object tracking problem is to be solved. Extended objects generate multiple estimates per measurement cycle, which are spatially grouped around the same object.

In this article, we use a measurement model that considers automotive radar tracking vehicles in near and middle range and a number of L = 3 reflection points per vehicle, the most-left, the closest and the most-right point. The estimates for the different reflection points are independently of each other. Accordingly, the measurement likelihood of the Set of Points on a Rigid Body [9] is given by the formula:

\[ p(Z|x) = \sum_{\theta} \prod_{i=1}^{n} (1 - p_\theta) \prod_{\theta > 0} p_\theta p_\theta(x_\theta|x) \]  

(7)

where \( p_\theta \) is the detection probability of the l-th reflection point, depending on the point’s state, Z is the set of measurements that were caused by the extended object and \( \theta \) is an assignment variable. It is noted that the likelihood determination is done by the tracking algorithms.

B. M-OSPA Metric for extended object tracker

A decisive advantage when evaluating tracker performances relies in the knowledge of mean and covariance values for every individual estimate. Unlike the tracker algorithm itself, metrics computation schemes do not deal with unknown statistical parameters of the estimates, but get them supplied from the trackers.

As shown in [6], the scan-by-scan OSPA metric definition for to the ground truth and the estimates data sets, \( X_k \) respectively \( \mathcal{Y}_k \), is given by \( D_{p,c}(X_k, \mathcal{Y}_k) = \) 

\[ \left[ \frac{1}{n} \left( \min_{\pi \in \Pi_n} \sum_{i=1}^{n} \left( D_c(\tilde{x}_{k,i}, \tilde{y}_{k,\pi(i)}) \right)^p + (n - m) c^p \right) \right]^{1/p} \] 

(8)

\( \tilde{x}_{k,i} \equiv (l_{k,i}, x_{k,i}) \), \( \tilde{y}_{k,\pi(i)} \equiv (s_{\pi(i)}, y_{k,\pi(i)}) \) and 

• \( D_c(\tilde{x}, \tilde{y}) = \min(c, D_m(\tilde{x}, \tilde{y})) \), is the parametrizable cut-off distance, and \( D_m \) represents the Mahalanobis base distance, given in (4).

• \( \Pi_n \) is the set of permutations of length m with values taken from \{1, ..., n\}, with \( m \leq n \); \( m \) and \( n \) are the cardinalities of \( X_k \) respectively \( \mathcal{Y}_k \).

• \( 1 \leq p \leq \infty \) is the OSPA metric order parameter, we select \( p = 2 \), as in [6].

For every measurement cycle, the number of ground truth tracks is \( n \) and the number of estimated targets is \( m \). First, we compute the Mahalanobis distances between all pairs of closest points in the n-length ground truth data set, denoted in the following as \( \| \| \).

If \( \min_{\pi \in \Pi_n} \| x_\pi \| \geq f \) then \( D_{p,c,f}(X_k, \mathcal{Y}_k) = \)

\[ \left[ \frac{1}{n} \left( \min_{\pi \in \Pi_n} \sum_{i=1}^{n} \left( D_c(\tilde{x}_{k,i}, \tilde{y}_{k,\pi(i)}) \right)^p + (n - m) c^p \right) \right]^{1/p}, \quad \text{else} \]

\[ \left[ \frac{1}{n} \left( \min_{\pi \in \Pi_n'} \sum_{i=1}^{n'} \left( D_c(\tilde{x}_{k,i}, \tilde{y}_{k,\pi(i)}) \right)^p + (n' - m') c^p \right) \right]^{1/p} \] 

(9)

where \( n' = L \cdot n \); \( m' = L \cdot m \), \( L \) is the number of reference points per target and \( f \) represents the form factor of the targets of interest. Again, the base distance in the pairwise cutoff distance \( D_m(\tilde{x}_{k,i}, \tilde{y}_{k,\pi(i)}) \) is given by the Mahalanobis distance, well taking into account the estimate’s statistics (\( \mu_k, \Sigma_k \)). The parameter selection for M-OSPA follows the selection criteria in [6] and [8]. For our simulations, we set \( f = 3 \), simulating half of the diagonal length of a mid-class SUV, as well as the cutoff parameter \( c = 10m \) and \( L = 3 \) (most-left, closest and most-right point). Please note that the estimate vector (5), the covariance matrix (6) and also the M-OSPA value (9) are considered for every individual point l of the SPRB.

C. Gaussian Mixture Probability Hypothesis Density Filter

The state-of-the-art approach to model the extended object state is in form of a Random Finite Set (RFS). As established compute-friendly approximations of the classical Bayes multi-target filter, the Probability Hypothesis Density (PHD) [11] and Cardinalized PHD (CPHD) filters propagate moments and cardinality distribution. Their basic approach is to identify and maximize the multi-object likelihood function (see (7) for SPRB). For simulation purposes, we selected the multiple Gaussian Mixture implementation of the PHD filter (GMPHD) as model for the multiple extended object tracking problem [10]. Further work will aim for comparisons between multi PHD, CPHD, GMHDP and variants of multi Bernoulli filters [7], based on M-OSPA.

IV. RESULTS

Based on the same scenario as the one used in [6], we consider four targets engaged in a parking maneuver, simulated with the Matlab toolbox [12]. The goal of this simulation is to highlight the effect of various variance and covariance values of the estimate’s state on the OSPA metric. Future work will present more extensive simulation results in complex scenarios, considering an increased number of reflection points per vehicle, additional variables for the multi-object state as well as a set of underlying extended object filters.

Further we consider a 2D model (length and width) of the target respectively the estimate state, having the following [2x2] covariance matrices, corresponding to four exemplary distributions of the GMPHD filter:
\[
\Sigma_{Y_1} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}; \Sigma_{Y_2} = \begin{bmatrix} 2 & 0.9 \\ 0.9 & 2 \end{bmatrix}
\]

\[
\Sigma_{Y_3} = \begin{bmatrix} 2 & 1.8 \\ 1.8 & 2 \end{bmatrix}; \Sigma_{Y_4} = \begin{bmatrix} 2 & 0.1 \\ 0.1 & 2 \end{bmatrix}
\]

The circles in Figure 4 show the position where the targets appeared in the radar field of view, the triangles the position where they vanished.

![Figure 4 Simulation of a parking maneuver in the case of four cars. The circles show where the targets appear in the radar field of view, while the triangles show the position where the targets disappear.](image)

In conformance with the results in Figure 2, we get in Figure 5 higher variances of the x and y coordinates that exceed the crossings of the two distributions \(\Sigma_{Y_1}\) and \(\Sigma_{Y_2}\), improved i.e. reduced OSPA values. Stronger correlation between x and y leads to reduced performance while reduced correlation between the coordinates emphasis a better tracker performance, i.e. reduced OSPA metric.

![Figure 5 Top: localization error of OSPA, bottom: localization error of M-OSPA with form factor; \(\Sigma_{Y_1}\) in black, \(\Sigma_{Y_2}\) in red, \(\Sigma_{Y_3}\) in blue, \(\Sigma_{Y_4}\) in cyan, time is in ms.](image)

The benefit of introducing the form factor was discussed in [6]. By also considering the estimate’s uncertainty, it becomes possible to select the best tracker with the lowest M-OSPA (tracker4 with cyan M-OSPA curve in Figure 5).

V. CONCLUSIONS

Real-world physical values that are under the interest of automotive sensors or airspace traffic management systems are basically correlated. Positive relative accelerations lead to higher relative velocities and typically to increased values of the target’s ranges. Modern sensors with strong measurement resolution can determine the spatial extents of several objects in the field of view, where every object generates multiple measurements, spatially structured around the same target. Target state information like range, shape, or velocity is provided by sensors with corresponding statistics, i.e. mean values and variances/covariances. “For arbitrary object shapes, the determination of suitable performance metrics for the evaluation of the shape estimate is still an open research question” concludes the elaborated overview article on extended object tracking in [4].

This paper proposes an enhanced method how to process an appropriate, intuitive and compute-friendly metric to evaluate the performances of extended object tracking algorithms. For this, the established optimal subpattern assignment (OSPA) metric is modified by replacing the Euclidean base distance between the true and the estimated target state-vectors with the Mahalanobis distance. It represents the distance between a deterministic ground truth target position and its corresponding estimate, which is characterized by a multivariate Gaussian distribution. The necessity of considering the track covariance into the metric is proven by experimental results, obtained with a Gaussian Mixture Probability Hypothesis Density Filter (GM-PHD).

REFERENCES


