

Learning and exploiting invariants for multi-target tracking and data association

Ruggero Frezza[†] and Alessandro Chiuso

Department of Information Engineering, University of Padova
Via Gradenigo 6/b, 35131 Padova, Italy

ruggero.frezza@unipd.it, chiuso@dei.unipd.it

Abstract—Methods for solving multi target tracking and data association problems in presence of clutter and occlusions are based on models that describe the target dynamics and the measurements statistics. Most often the dynamics of the targets are assumed to be independent from each other. In many applications, however, the motion of the targets may be coordinated. We introduce a statistical concept of shape, or coordination, in terms of invariants w.r.t. the motion of the targets. Assuming that the rules of coordination may slowly change over time, we study the interplay among the shape and the target dynamics.

I. INTRODUCTION

In many applications from air traffic control, to tracking features on the image plane of a camera, there is the problem of tracking a multitude of targets while they are moving in space. Each target is seen by one or more sensors which generate measurements of its position in space at every time step. The problem is difficult because the dynamics of the targets are uncertain and the sensors generate unlabelled and similar measurements of their positions. It is, therefore, not trivial to associate measurements to targets. The problem is often rendered even more difficult by clutter which consists of the presence of false measurements and occlusions which happen when some targets are hidden from the sensors and do not generate measurements for some time intervals.

In the literature, the problem is known with the name of multi-target tracking and data association. There is a vast number of papers addressing it from many different facets since the late '60ies early '70ies. Standard methods are described, for instance, in [1].

Most algorithms exploit two forms of information: a dynamic model describing the motion of the targets and a statistical model of the measurements. The data association is solved by jointly examining the positions of all the received measurements w.r.t. their predictions on the basis of the dynamic model. Probably, the most well known algorithm is the JPDA (Joint Probability Data Association) which evaluates the probabilities of all possible associations and combines them accordingly in order to compute the updated estimate of the targets position. The MHT (Multiple Hypothesis Tracking) [2] is more powerful than the JPDA because it evaluates the probability of the associations on a whole time interval. It finally selects the most likely association to update the state estimate. The complexity of the MHT is, however,

much higher and a latency in the generation of the estimate is necessarily introduced. The MMF (Multiple Model Filter) is applied when the targets go through aggressive maneuvers. Multiple dynamic models are used to better describe the different phases of the maneuvers and improve the state predictions. It is clear that the more accurate the dynamic model is, the easier is the data association problem.

In recent years, in the scientific literature on computer vision, a number of papers on tracking have appeared. In vision the multi-target tracking problem is fundamental to reconstruct the trajectories of features or objects in the image plane. In [8] particle filters and the condensation algorithm have been proposed in order to estimate non-Gaussian and multi-modal posterior probability densities which arise in case of ambiguous data associations. In [9] the problem of proper re-sampling of the densities is addressed by integrating the tracker with information on the measurements which eases the data association problem. In [6], [12], [13] statistical learning techniques have been applied to the problem. The authors of [12], [13], in particular, have proposed a method for learning the joint probability density of the position of the targets in space. The approach is complementary to the standard methods such as the JPDA. There is no local information on the trajectories and, consequently, no assumptions on their regularity. All the information is collected in the joint probability density and the association events are independent in time. The computational complexity of learning the joint probability density is exponential with the size of the maximum clique of the graph describing the conditional dependencies among targets. In order to make the problem manageable, a triangulate structure is assumed. This means that the graph is composed by cliques of order three or less. The advantage of this approach is that it models statistical dependencies among targets which, with the standard algorithms such as the JPDA, are neglected in favor of local coherence of trajectories described by independent dynamic models. The goal of our research is to combine the advantages of both approaches. The main drawback of the statistical learning methods is that they require the acquisition and the labeling of a large training set for each action of interest such as a walking person. The learning set may be used for other instances of the same action, but cannot be extrapolated to other subjects or scenes.

The classical techniques based on model based filters and assuming independence are more general in this sense as they do not require particular training, but, since they do not model coordination among targets, they might produce highly segmented tracks.

In [10], [11], [3], [4] an effort to include information on shape or coordination among targets in the JPDA has been done with some success. The advantage of these approaches is that coordination may be learned during those segments of the targets trajectories which are successfully labeled by the JPDA and then it can be integrated in the scheme in order to solve the data association problem better.

In this paper we continue along the line of research introduced in [3], [4]. We assume the existence of symmetries of motion among some of the targets. Think at, for example, airplanes flying in formation. The shape of the formation is invariant w.r.t. the motion of each aircraft and it is an important feature for solving the data association problem. We model these symmetries as some functions f_i for $i = 1, \dots, p$ of the dynamical state of the targets which are constant in time or slowly varying. We, then, apply standard parameter estimation methods for estimating their statistics and, by properly adjusting the forgetting factors, we can adapt the scheme to slowly varying parameters.

II. THE MULTI-TARGET TRACKING AND DATA ASSOCIATION PROBLEM

With the purpose of introducing some notation which will be used throughout the paper, we provide a short introduction to the probabilistic approach to data association. We refer the reader to the classic book [1] for a thorough description.

Consider the problem of tracking over time a set of N_T moving targets; $y_i(k)$ shall denote the position of target $i \in [1, \dots, N_T]$ at time instant¹ $k \in \mathbb{Z}$. In many real-world scenarios one has available, at each time instant k , a set of “unlabelled” measurements $\{z_i(k)\}$, $i = 1, \dots, M_k$; this means that in general no knowledge is available regarding (a) which measurement has been originated by which target², (b) which target has generated no measurement (in which case we shall say that the target is *occluded*) and (c) which measurements are “false detections” in the sense that they have not been generated by any of the targets; in the literature these measurements are said to be originated from *clutter*. Note that at each time k the number of measurements M_k is in general different from the number of targets N_T .

Dealing with (a), (b) and (c) is the *data association* problem. Traditionally, the solution is based on assuming that the position of each target y_i is a hidden Markov process whose dynamical state is x_i defined by its probability density at the initial time $p(x_i(0))$ and by its transition density $p(x_i(t+1)|x_i(t))$. The observations are the positions of the targets. They are modeled by the conditional density

¹Without loss of generality we assume that measurements are gathered uniformly in time with unit sample time and that only the positions of targets at those time instant are observed.

²A standard assumption is that each target can originate at most one measurement.

$p(y_i|x_i)$. The filtering problem consists in the computation of the following Bayesian recursion

$$\begin{aligned} p(x(k)|Y_k) &= \frac{p(y(k)|x(k))p(x(k)|Y_{k-1})}{p(y(k)|Y_{k-1})} \\ &= p(y(k)|x(k)) \frac{\int p(y(k)|x(k-1))p(x(k-1)|Y_{k-1})dx}{\int p(y(k)|x(k))p(x(k)|Y_{k-1})dx} \end{aligned}$$

where with the capital letter Y_k we mean the σ algebra generated by all the position measurements from the initial time to time k . In the linear Gaussian setting, the above recursion is solved by the Kalman filter. The problem is that at each time instant k , a set of *unlabeled measurements* \mathbf{z} arrives and the association to the targets is unknown. It is customary [1] to denote with θ_k an “association event” (or hypothesis) at time³ k and with Θ_k the set of all possible association events at time k . Under the hypothesis θ_k , $j(i, \theta_k)$ shall denote the index of the measurement associated to target i , i.e.

$$y_i(k) = z_{j(i, \theta_k)}(k). \quad (\text{II.1})$$

If the target i is occluded then

$$j(i, \theta_k) = 0. \quad (\text{II.2})$$

It is sometimes useful to use the index 0 to denote clutter and therefore $j(0, \theta_k)$ will denote the *set*⁴ of false measurements under θ_k .

Some approaches are based on *hard decisions*, where at each time only the most likely possible association is considered; unfortunately these fail in the presence of strong clutter and occlusions.

The *Joint Probabilistic Data Association Filter* (JPDAF hereafter) (see [1] for a thorough description) is a probabilistic method which integrates (1) a dynamical model for the motion of targets, (2) a model for the clutter (false detections) and (3) the probability of occlusions.

A key observation is that under an association hypothesis θ_k , estimating the position⁵ of each target is the standard filtering problem described above. The main idea behind the JPDAF is to fuse the information from (1),(2) and (3) above in order to attach a weight (a “posterior probability”) to the possible associations. Then an estimate of the position of the targets can be obtained by conditionally weighting the state estimates on all possible⁶ associations.

We shall denote with $\mathbf{z}(k) := [z_1(k), \dots, z_{M_k}(k)]$ the set of measurements available at time k . The symbol Z_k will denote the set of measurements up to time k (included), i.e. $Z_k := \{\mathbf{z}(s), s \in [0, k]\}$.

In the gaussian linear case, the position $y_i(k)$ of the i -th target are described by a linear state space model of the

³Note that, since the number of measurements may change over time, also the set of possible associations changes over time.

⁴As explained above under each θ_k , $j(i, \theta_k)$ is a well defined function of $i \in [1, N_T]$ while $j(0, \theta_k)$ is in general a set.

⁵Or, more generally, the state of a dynamical system describing its motion.

⁶It is customary to consider only a subset of association which corresponds to “large enough” weights (posterior probabilities). This is usually done employing suitable “validation regions” for each target; see [1] for details.

form:

$$\begin{cases} x_i(k+1) &= F_i \cdot x_i(k) + v_i(k) \\ y_i(k) &= H_i \cdot x_i(k) + w_i(k) \end{cases} \quad (\text{II.3})$$

The model and measurement noises v_i and w_i are assumed to be white, zero mean and uncorrelated gaussian distributed with covariances Q_i and R_i respectively.

The generalization to nonlinear independent dynamics implies the use of nonlinear filtering methods like, for example, the Extended Kalman Filter or particle filters like the condensation algorithm [8]. Our interest, however, is focused on the data association problem which is unaffected by the fact that the dynamics are linear or not. We assume, therefore, that at time $k-1$ the conditional density of the state $x_i(k-1)$ given all measurements up to time $k-1$ is Gaussian with mean $\hat{x}_i(k-1)$ and covariance matrix $P_i(k-1)$, i.e. $p(x_i(k-1)|Z_{k-1}) \sim \mathcal{N}(\hat{x}_i(k-1), P_i(k-1))$.

Once measurements at time k become available these estimates can be updated, conditionally on an association event θ , by using standard Kalman filter formulas with measurements $^7 y_i = z_{j(i,\theta_k)}$. We denote with $\hat{x}_{i,\theta_k}(k)$ and $P_{i,\theta_k}(k)$ the updated mean and covariance conditionally on the association θ_k , initialized from initial conditions (at time $k-1$) $\hat{x}_i(k-1)$ and $P_i(k-1)$; since the dynamics are described by a Gauss-Markov model (II.3), the conditional density $p(x_i(k)|\theta_k, Z_k)$ is Gaussian with mean $\hat{x}_{i,\theta_k}(k)$ and covariance matrix $P_{i,\theta_k}(k)$.

A simple application of the Total Probability Theorem provides the conditional probability density function

$$p(x_i(k)|Z_k) = \sum_{\theta_k \in \Theta_k} p(x_i(k)|\theta_k, Z_k) p(\theta_k|Z_k) \quad (\text{II.4})$$

which turns out to be a mixture of Gaussian densities.

In order to make the computation tractable this Gaussian mixture is approximated (in the Kullback-Leibler sense for instance) by a Gaussian density with mean $\hat{x}_i(k)$ and covariance $P_i(k)$ according to

$$\begin{cases} \hat{x}_i(k) &= \sum_{\theta_k} \hat{x}_{i,\theta_k}(k) \cdot p(\theta_k | Z_k) \\ P_i(k) &= \sum_{\theta_k} P_{i,\theta_k}(k) \cdot p(\theta_k | Z_k) \\ &+ \sum_{\theta_k} (\hat{x}_{i,\theta_k}(k) - \hat{x}_i(k)) \cdot (\hat{x}_{i,\theta_k}(k) - \hat{x}_i(k))' \cdot p(\theta_k | Z_k) \end{cases} \quad (\text{II.5})$$

This allows to start again at time k with a Gaussian posterior for each target and iterate the procedure just described. This last approximation step is implicit in the classical description of JPDAF (see [1]) where only second order moments $\hat{x}_i(k)$, $P_i(k)$ are considered.

The only point left is to compute the posterior association probabilities $p(\theta_k|Z_k)$. Assume that a prior $p(\theta_k)$ on the association events is available⁸; the posterior $p(\theta_k|Z_k)$ can

⁷If under association θ_k no measurement is associated to target i then only the prediction will be computed.

⁸We shall not discuss this choice in the paper. We refer the reader to [1] for details. Suffices it to say that $p(\theta_k)$ usually depends on the probability that each target is detected, on the number of detected targets and on the number of false measurements.

be computed using Bayes' formula as follows:

$$\begin{aligned} p(\theta_k|Z_k) &= c p(\mathbf{z}(k)|\theta_k, Z_{k-1}) p(\theta_k | Z_{k-1}) \\ &= c p(\mathbf{z}(k)|\theta_k, Z_{k-1}) p(\theta_k) \end{aligned} \quad (\text{II.6})$$

where the last equality holds because associations at time k are conditionally independent upon measurements up to time $k-1$. The constant c is a normalization factor which does not play a role.

From now on, we shall omit the time index k unless needed; according to the notation introduced above, Z shall denote the set of past and present measurements, while Z^- only the past. Similarly \hat{x}_i^- will denote the prediction of the state at time k given Z^- and P_i^- its conditional error covariance.

In order to evaluate $p(\mathbf{z}|\theta, Z^-)$, it is convenient to introduce the set⁹ D_θ containing the indices of all the detected targets¹⁰; consequently we shall denote with $\mathbf{z}_{T,\theta} := \{z_{j(i,\theta)}, i \in D_\theta\}$ the set of "true" measurements, i.e. measurements which have been associated to some target and with $\mathbf{z}_{F,\theta}$ the complementary set of "false" measurements attributed to clutter. Similarly we define the set of "occluded" target indexes¹¹ as M_θ . Postulating (conditional) independence $p(\mathbf{z}|\theta, Z^-)$ can be factored in the form

$$p(\mathbf{z}|\theta, Z^-) = p(\mathbf{z}_{T,\theta}|\theta, Z^-) p(\mathbf{z}_{F,\theta}|\theta, Z^-).$$

The term $p(\mathbf{z}_{F,\theta}|\theta, Z^-)$ describing clutter is usually taken to be uniform over the volume V of interest, i.e.

$$p(\mathbf{z}_{F,\theta}|\theta, Z^-) = \left(\frac{1}{V}\right)^{N_F(\theta)} \quad (\text{II.7})$$

where $N_F(\theta)$ is the number of false measurements under hypothesis θ .

As the term $p(\mathbf{z}_{T,\theta}|\theta, Z^-)$ is concerned, it is sufficient to recall that $\mathbf{z}_{T,\theta} = \{z_{j(i,\theta)}, i \in D_\theta\} = \{y_i, i \in D_\theta\}$. Let us define the vectors $\mathbf{y}_{D_\theta} := \{y_i, i \in D_\theta\}$ and $\mathbf{y}_{M_\theta} := \{y_i, i \in M_\theta\}$ containing respectively the detected and occluded targets.

Therefore

$$p(\mathbf{z}_{T,\theta}|\theta, Z^-) = [p(\mathbf{y}_{D_\theta}|Z^-)]_{|y_i=z_{j(i,\theta)}, i \in D_\theta} \quad (\text{II.8})$$

which is the marginal of $p(y_1, \dots, y_{N_T}|Z^-) = p(\mathbf{y}_{D_\theta}, \mathbf{y}_{M_\theta}|Z^-)$ with respect to \mathbf{y}_{M_θ} :

$$p(\mathbf{y}_{D_\theta}|Z^-) = \int p(\mathbf{y}_{D_\theta}, \mathbf{y}_{M_\theta}|Z^-) d\mathbf{y}_{M_\theta}. \quad (\text{II.9})$$

Under the assumption that the target positions are conditionally independent we have that

$$p(\mathbf{z}_{T,\theta}|\theta, Z^-) = \prod_{i \in D_\theta} [p(y_i|Z^-)]_{|y_i=z_{j(i,\theta)}} \quad (\text{II.10})$$

The density $p(y_i|Z^-)$ describes the prediction of the position of target i given past measurements. From the

⁹Remind that since θ depends on time k we should use the notation D_{θ_k}
¹⁰I.e. targets to which a measurements has been associated under hypothesis θ

¹¹ M stands for "missing". Note that $D_\theta \cup M_\theta = [1, \dots, N_T]$.

assumption that x_i conditionally on Z^- is Gaussian with mean \hat{x}_i^- and covariance P_i^- it follows from (II.3) that $p(y_i|Z^-)$ is a Gaussian density with mean $\hat{y}_i = H_i \hat{x}_i^-$ and covariance matrix $\Sigma_i^- = H_i P_i^- H_i^\top + R_i$.

Remark 2.1: Equation (II.10) is fundamental in computing the association probabilities. It relies on the fact that targets are assumed to be conditionally independent given past measurements, which is not certainly true when there is coordination between targets. In [4], a shape model has been integrated into the JPDAF algorithm. When the posterior density of the positions includes the shape model, the marginalization (II.9) is not trivial. We applied Monte Carlo, or equivalently, particle methods to implement this step.

III. MOTION SYMMETRIES

During their motion, the targets may be coordinated, where by coordination we mean the existence of some statistical dependence among them. Quantifying statistical dependence is a difficult problem. The correlation coefficient, for example, can only be applied to pairs of random variables and only measures linear dependence. In the literature, there exist a number of measures of statistical dependency. In [5], for example, a generalization of Pearson's ϕ^2 measure is proposed

$$\phi^2 = \int_{\mathbf{x} \in \mathcal{X}^N} \frac{p(\mathbf{x})}{\prod_{i=1}^N p(x_i)} \mathbf{x} - 1$$

as well as a measure based on the Kullback-Leibler pseudo distance among the joint distribution and the product of the marginals. Both measures, if the targets are statistical independent, are null, otherwise they can even be unbounded. The problem with these kinds of measures is that they cannot be easily computed from the experimental observations. It would be necessary to estimate the joint density $p(\mathbf{x})$, but this is a formidable task unless particular structures of the conditional dependence graph are assumed. It is customary to assume either a tree or a triangular structure. The former implies that the maximum cliques are of order two while the latter implies that they are of order three. They are both manageable computationally, but the triangular structure is more robust w.r.t. occlusions. A single occlusion, in fact, cuts a tree structured graph. This is the reason why a triangular structure was chosen in [12], [13].

We take a different approach. We do not try to estimate the joint density of the ensemble of targets, but we search for invariants in the motion of the ensemble of targets. When a number of aircrafts are flying in formation or a set of markers are attached to a rigid body, the position of three non collinear targets fully determines the position of all the others. The motion of the whole ensemble of targets can, in this case, be factored in the motion of any triple of points and on a local, invariant, representation of the others w.r.t. the first three. In other words, the motion of the ensemble can be factored in a rigid body motion and an invariant description of the shape of the ensemble. In terms of statistical dependency among targets, the conditional density of the position of the targets given a triple is degenerate and,

in absence of noise, it is deterministic. In presence of noise it can be modeled reasonably well by a Gaussian mixture. Besides rigid motion, other interesting coordinated motions generate symmetries or, equivalently, invariants. Any kind of link or joint among articulated bodies can be described by some invariants. In general, we can model holonomic and non-holonomic constraints with invariants. If, for example, the targets are all moving along straight lines, the directions of motion are invariants, or, if the targets are orbiting about a fixed point then its coordinates are the sought invariants. Our purpose is to exploit these symmetries in order to solve the data association problem.

This paper is, in particular, on the estimate of the statistics of the motion invariants and how to combine this information with that provided by the independent dynamical models (II.3) in order to solve the data association problem. For consistency, the independent dynamical models are assumed to describe the transition density of each target. They describe, therefore, the marginal probability density of each target.

In general, let us assume that there exist some features f_i for $i = 1, \dots, p$ that are functions of the dynamical state of the targets

$$f_i = f_i(x_1, \dots, x_n)$$

which are invariant w.r.t. the motion of the targets, so that

$$\frac{df_i}{dt}(t) = \nabla f_i \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = 0.$$

Two important problems arise: (1) find or identify from the data the invariants f_i ; (2) estimate the statistics of the invariants f_i in presence of noise and uncertainties.

The first problem is also very difficult and it is not within the scope of this paper even if it is one of the main objectives of our research. We assume a list of possible invariants and, while tracking, we check if there exist group of targets that satisfy them.

An example of the definition of an invariant description is the procedure proposed by Kendall [7] to represent the shape of an ensemble of N points. It consists of the following steps:

- 1) determine the center of mass of the points $y_{cm} = \sum_{i=1}^N y_i$ and move the origin of the reference frame in y_{cm} ;
- 2) rotate by R the reference frame so that the N dimensional vector $[0, \dots, 0, 1]^T$ becomes the right kernel of the matrix

$$\bar{S} = [(y_1 - y_{cm}) \quad \dots \quad (y_n - y_{cm})] R$$

- 3) eliminate the last row of matrix \bar{S} , the $3 \times N - 1$ matrix S obtained in this fashion is a representation of shape and it is an invariant of rigid motion.

The invariant proposed by Kendall is, unfortunately, not robust w.r.t. occlusions, so, in the list of possible invariants, we also include distances among pairs of targets and angles between target velocities. We assume, therefore, that the form

of the invariants f_i is known and we address the problems of verifying their existence in the motion of the ensemble of the targets and of estimating their statistics. The existence of a specific invariant is formulated as an hypothesis test, the null hypothesis being that the invariant is true. The statistic on which the test is performed is the variance of f_i .

When an invariant is declared true, we update the estimate of its mean and covariance with a recursive estimator which is described in the following section, and on the other we add links between the correspondent targets to a coordination graph which describes the dependencies among targets.

It often happens that the invariants of motion last only for short time intervals or show slow dynamics. Can they still be exploited for the data association problem? We assume that the statistics of the invariants may slowly drift in time. This wants to be an initial model describing the dynamics of shape. While, shape, in fact, might be evanescent and persist in time briefly, it can still carry a lot of information helpful for solving the data association problem. We, *de-facto*, implement an adaptive scheme that adjusts to changing coordination strategies among targets.

The coordination graph collects information on the structure of the dependency among targets. If the targets are, for example, all attached to a rigid body the coordination graph will be complete. It is used for two reasons: book keeping and as a topological support to the data association. We have, in the past, implemented a graph matching algorithm to solve the data association problem, but it only helps when the structure of the graph is complex with more than one clique of relatively high order. This is usually the situation when the scene is composed by articulated bodies. In other situations, like a single rigid body, the graph is complete, i.e. it is a clique and the matching algorithm does not provide any information for labeling the measurements.

IV. ESTIMATING THE STATISTICS OF SLOWLY VARYING INVARIANTS

In estimation theory, it is customary to model slowly varying parameters with a linear model

$$\begin{aligned}\xi(k+1) &= A(k)\xi(k) + \nu(k) \\ f(k) &= H(k)\xi(k) + w(k)\end{aligned}\quad (\text{IV.1})$$

where the matrix $A(k)$ is close to the identity, while ν and w are independent gaussian white noises. We approximate the matrix $A(k)$ assuming that it is the identity and we assume that the observation matrix $H(k)$ is constant. The covariances of the process and measurement noises are assumed unknown.

Under our hypothesis, the extended forgetting factor recursive least squares estimator (EFRLS) [14] becomes

$$\begin{aligned}\xi(k) &= \xi(k-1) + L(k)(f(k) - H\xi(k-1)) \\ L(k) &= P(k-1)H^T(\lambda I + HP(k-1)H^T)^{-1} \\ P(k) &= \frac{1}{\lambda}(I - L(k)H)P(k-1)\end{aligned}\quad (\text{IV.2})$$

The choice of the forgetting factor λ is based on the following considerations: λ should be large and close to one whenever the process noise covariance or when the

measurement noise covariance is large. In the first case, past measurements contain information, in the second, averaging over more samples in time reduces the covariance of the estimate. λ , however, should not be too large or we loose adaptability to slow drifts of the mean of f .

An hypothesis test is then performed, based on the estimated covariance $HP(k)H^T$ of the invariant. If the norm of the estimated covariance is larger than an appropriate threshold then the alternative hypothesis is considered true and the invariant is considered not true.

V. INTEGRATION OF SHAPE IN THE JPDA

In order to compute the posterior probability of a given association event θ we need to compute the likelihood of the true measurements \mathbf{z}_T . The overall observation model can be written as follows:

$$\begin{aligned}p(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} | Z^-) &= c \cdot \prod_{i \in D_\theta} p(z_{j(i,\theta)} | Z_i^-) \cdot \\ &\cdot \prod_{i \in M_\theta} p(y_i | Z_i^-) \cdot \prod_{i=1}^p p(f_i(\mathbf{y}_{D_\theta}, \mathbf{y}_{M_\theta}))\end{aligned}\quad (\text{V.1})$$

which yields:

$$\begin{aligned}p(\mathbf{z}_T | \theta, Z^-) &= c \cdot \prod_{i \in D_\theta} p(z_{j(i,\theta)} | Z_i^-) \cdot \\ &\cdot \int \prod_{i \in M_\theta} p(y_i | Z_i^-) \cdot \prod_{i=1}^p p(f_i(\mathbf{y}_{D_\theta}, \mathbf{y}_{M_\theta})) d\mathbf{y}_{M_\theta}\end{aligned}\quad (\text{V.2})$$

As in [4], we solve the integral in (V.2) by a Monte Carlo approach. The reason is twofold. First, it is simple and consistent. Second, as a byproduct, it yields for free a set of fair samples from the posterior distribution of the occluded points positions. This allows to compute mean and covariance and hence provides a natural gaussian approximation of the more complicated posterior. We draw an appropriate number N_s of independent and identically distributed samples:

$$\mathbf{y}_{M_\theta}^{(n)} \triangleq \{y_i^{(n)}, i \in M_\theta\} \sim \prod_{i \in M_\theta} p(\cdot | Z_i^-) \quad n = 1, \dots, N_s \quad (\text{V.3})$$

and compute the n -th weight through the following expression:

$$\pi^{(n)} = \prod_{i=1}^p p(f_i(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} = \mathbf{y}_{M_\theta}^{(n)})) \quad (\text{V.4})$$

Finally, the integral is computed as follows:

$$\begin{aligned}\int \prod_{i \in M_\theta} p(y_i | Z_i^-) \cdot \\ \cdot \prod_{i=1}^p p(f_i(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta})) d\mathbf{y}_{M_\theta} \propto \sum_{n=1}^{N_s} \pi^{(n)}\end{aligned}\quad (\text{V.5})$$

which, substituted in (V.2), yields $p(\mathbf{z}_T | \theta, Z^-)$.

The conditional state estimates of a *detected point* are computed combining the Kalman updates on the basis of all the feasible associations as in the JPDA. The fundamental difference being that the symmetries f_i are instrumental in computing the likelihood of the association events. The conditional state estimates of an *occluded point* are, instead, computed exploiting the shape information starting from

the measurements generated by the detected points. It is fundamentally different than in standard approaches where it is taken equal to the state predictions.

VI. RESULTS

The multi-target tracking algorithm proposed in this paper is currently being implemented on a optical motion capture system. We implemented the algorithm in matlab and we tested it on data previously acquired with the motion capture system. Twentytwo markers have been attached to a human subject, three on the head, two on the shoulders, two on each arm, five on the torso and four on each leg. A rigid object with six markers on it was held by the subject in his hand during the acquisition of motion. The total of 28 markers was tracked by a six 50 Hz camera motion capture system for approximately two minutes, i.e. for about 6000 frames. The first 2000 frames were used to determine and initialize the estimate of the invariants of motion. All the markers attached to the rigid body held by the subject in his hand satisfy the mutual distance invariants and even Kendall's invariant when there are no occlusions. The coordination graph clearly exhibits the articulation structure of the human body. All the markers on the torso, for example, belong to the same clique of maximum order equal to five. The markers on the feet all belong to a complete subgraph. This is because the subject was asked not to move his feet in order to check adaptability to changes in the coordination and the effect of the forgetting factor.

As an example the statistics of some invariants are described in the following table.

Invariant: distance among two targets	Mean	Std	Persistence interval max (frames)
Targets 1 and 3 both on the head	16.8cm	0.12cm	All
Targets 12 and 13 on the left arm	29.7cm	0.57cm	786

The total number of trajectories segments has been taken as a performance index of the data association algorithm. Ideally, the number of trajectories should have been equal to the total number of markers i.e. 28.

An implementation of the JPDA alone generated 112 segments. The number of trajectory segments is, furthermore, highly dependent on the choice of noise covariances in the Kalman filters. If the covariances are set too small, the measurements do not fall within the validation gates and are associated to clutter. If the covariance is set too large, the data association becomes very difficult because the number of possible associations increases. After a few trials, we found a choice that led to the best result of 112 segments.

The shape integrated JPDA generated 36 segments, where most of the wrongly labeled segments were produced because of the incorrect invariants detected between the feet of the subject.

Tuning the forgetting factor λ for the invariants is important to obtain significant results. A small λ leads to the creation of invariants which persist in time very briefly. A large λ renders the scheme rigid and not adaptable so that wrong invariants declared as such because of not sufficiently exciting dynamics lead to wrong data association.

VII. CONCLUSIONS

This paper continues along the research line presented in [4]. The spirit is to include information due to the statistical dependence among the targets in standard algorithms multi target tracking algorithms that otherwise treat targets as independent. This information is of great help in solving the data association problem. The proposed schemes should also improve on the techniques proposed in the computer vision literature based on statistical learning methods which do not imply any local coherence in time of the targets trajectories.

Coordination among targets has been models by the means of motion symmetries or invariants. The shape description proposed by Kendall is used as an invariant, but, since this is not robust w.r.t. occlusions, it has been integrated with pairwise distances among targets and angles between target velocities.

The possibility of slow drifts in time of the invariants is dealt with by introducing forgetting factors in the estimate of their statistics.

In experiments with a motion capture system segmentation of the tracks has been substantially reduced compared to the standard JPDA assuming the possibility of learning the invariants on a sufficiently long time interval with persistently exciting dynamics.

REFERENCES

- [1] Y. Bar-Shalom and T. Fortman, *Tracking and data association*, Academic Press, 1988.
- [2] D. B. Reid, *An algorithm for tracking multiple targets*, IEEE Trans. on Automatic Control, **25**, No. 6, pp. 843-854, 1979.
- [3] G. Gennari, A. Chiuso, F. Cuzzolin, and R. Frezza, *Integrating shape and dynamic probabilistic models for data association and tracking*, IEEE Conference on Decision and Control, 2002.
- [4] G. Gennari, A. Chiuso, F. Cuzzolin, and R. Frezza, *Integration of shape constraints in data association filter*, IEEE Conference on Decision and Control, 2004.
- [5] I. N. Goodman and D. H. Jonson, *Orthogonal decompositions of multivariate statistical dependence measures*, Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2004), Montreal, CA, (May 2004)
- [6] I. Gordon and D. G. Lowe, *Scene modelling, recognition and tracking with invariant image features*, International Symposium on Mixed and Augmented Reality (ISMAR), Arlington, VA (Nov. 2004), pp. 110-119.
- [7] D. G. Kendall, *A survey of the statistical theory of shape (with discussion)*, Statist. Sci., **4**, 1989, pp. 87-120.
- [8] M. Isard and A. Blake, *Condensation – conditional density propagation for visual tracking*, Int. J. Computer Vision, 1998.
- [9] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little and D. G. Lowe, *A Boosted Particle Filter: Multitarget Detection and Tracking*, European Conference on Computer Vision (ECCV), Prague (May 2004), pp. 28-39.
- [10] C. Rasmussen and G.D. Hager, *Joint probabilistic techniques for tracking multi-part objects*, Int. Conf. on Computer Vision and Pattern Recognition, 1998.
- [11] C. Rasmussen and G.D. Hager, *Probabilistic data association methods for tracking complex visual objects*, IEEE Transaction on Patter Analysis and Machine Intelligence **23** (2001), 560–576.
- [12] Y. Song, L. Goncaves, E. Di Bernardo, and P. Perona, *Monocular perception of biological motion - detection and labelling*, Int. Conf. on Computer Vision, 1999, pp. 805–812.
- [13] Y. Song, L. Goncaves and P. Perona, *Unsupervised learning of human motion*, IEEE Transaction on Patter Analysis and Machine Intelligence **25** (2003), 1–14.
- [14] Y. Zhu, *Efficient Recursive State Estimator for Dynamic Systems with-out Knowledge of Noise Covariances*, IEEE Transaction on Aerospace and Electronic Systems **35** (1999), 102–114.