

# Integration of shape constraints in data association filters

Giambattista Gennari, Alessandro Chiuso, Fabio Cuzzolin, Ruggero Frezza

*Abstract*—Many algorithms have been proposed in literature to deal with the tracking and data association problem. A common assumption made in the proposed algorithms is that the targets are independent. There are however many interesting applications in which targets exhibit some sort of coordination, they satisfy shape constraints. In the current work a general and well formalized method which allows to embed such constraints into data association filters is proposed. The resulting algorithm performs robustly in challenging scenarios.

## I. INTRODUCTION

Tracking and data association are key problems in many interesting applications such as satellite surveillance systems, air defence, air traffic control, visual tracking and motion capture. All these applications deal with a set of points (or targets) moving in space and generating measurements of their positions. Tracking the points means estimating their state recursively over time starting from the observations. If all points were equal and they were not producing an identifying label when they generate the measurement of their position, the problem is associating measurements to trajectories. The data association problem is complicated further by false detections and missing data. There is an extensive literature on this problem and standard methods can be found for instance in [1]. These includes the popular JPDAF (joint probabilistic data association filter). In [7] the MHT (multi hypothesis trackers) have been introduced. In such techniques a target is described through a mixture of gaussian densities. Monte Carlo approaches have been recently adopted [13], [14] as well. Adaptive algorithms [12] have been proposed in order to cope with particular uncertainties such as the unknown inputs which typifies maneuvering targets. In multiple models based techniques [8] each target is described by a finite number of state space models.

In the approaches proposed in literature targets are modelled independently. In some interesting applications there might be some sort of coordination between targets; for instance one might have a set of targets rigidly connected.

This work has been supported by ASI, EU and MIUR

Giambattista Gennari is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, Via Gradenigo 6/A, 35131 Padova, Italy, [simpson@dei.unipd.it](mailto:simpson@dei.unipd.it)

Alessandro Chiuso is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova Via Gradenigo 6/A, 35131 Padova, Italy, [chiuso@dei.unipd.it](mailto:chiuso@dei.unipd.it)

Fabio Cuzzolin is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova Via Gradenigo 6/A, 35131 Padova, Italy, [@dei.unipd.it](mailto:@dei.unipd.it)

Ruggero Frezza is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova Via Gradenigo 6/A, 35131 Padova, Italy, [frezza@dei.unipd.it](mailto:frezza@dei.unipd.it)

Taking into account information about the coordinated motion would be of great help. Consider a group of coordinated points and assume that some of them are occluded or generate measurements which are similar to those generated by other points or to false detections. Their state can be better estimated exploiting the coordination and the measurements generated by the other points forming the given group.

The motion of a rigid pattern is only one example of coordination. Think for instance to formations of robots or to points placed on articulated bodies or to a network of systems which operate to achieve a common task. Coordination is a natural concept for humans. Deriving a general model for arbitrary forms of coordination would allow to develop algorithms capable of recognizing and tracking reliably arbitrary coordinated groups. Unfortunately, obtaining such a model does not seem to be so simple.

To simplify the problem we note that, in many cases, the measurements generated by points which would be intuitively classified as coordinated satisfy constraints which are invariant with respect to the motion. Start from a simple case of coordination, the rigid motion. Rigidly linked points exhibit constant distances and angles. The latter are a prior constraint on the positions of points which are characterized by such distances or angles in each instant of time.

Similar considerations hold for non rigid sets of targets as well. Targets moving in formation are coordinated and intuitively compact independently from the motion. Formally, each point could be represented by a node on a graph. An edge would be placed between two nodes if the corresponding points were spatially close. The connectivity of the graph could formalize the instinctive notion of compactness. Such connectivity holds independently from the trajectories that such points follow in the space or equivalently it is a prior constraint on the points positions (from which the graph can be constructed at each instant of time).

Finally, think to coordinated points forming a circle which spreads and contracts periodically. The law describing the distances between the points and the center of the configuration are prior constraints on the points positions.

The prior constraints can be conveniently described by a probabilistic model. The latter allows for instance to describe non perfectly rigid configurations of points whose pairwise distances can be assumed to be distributed according to appropriate density functions.

Motivated by all considerations above, we assume that a prior density function of the positions of points is given.

A general method to embed such prior knowledge into standard tracking algorithms is presented. An overall observation model which takes into account also the past history of the measurements is obtained. Moreover a general procedure which allows to learn the prior density from appropriate data sequences is introduced. Such density embeds prior knowledge and hence its validity extends to arbitrary motions of the points which can be even quite different from those observed during the learning phase. In the sequel the motion invariant knowledge is called shape. The resulting algorithm copes robustly with quite unpredictable motions performed by points in presence of high numbers of false detections and of occlusions lasting several consecutive time steps.

The current work is strictly related to [4], [2] where a constrained version of the JPDAF has been proposed for visual tasks. The main contribution of our approach is the general and well formalized way adopted to extend the standard tracking and data association techniques. Our approach differs substantially from [4], [2] for what concerns the occlusions as well. Moreover, a general learning methodology is introduced. The solution to the problem of missing data is also the main innovation brought to our previous work [3]. Some analogies with this work can be found in motion capture literature [9], [10]. However the focus is on specific and technical aspects of the human body tracking rather than on the development of a general and comprehensive methodology for tracking and data association problems capable to deal with general form of coordination. In [6] the coordination of point features placed on a human body is described by a convenient probability density function of positions and velocities. However the authors assume that the trajectories are provided by an appropriate tracking algorithm and such model is used for detection and labelling (recognition) of human parts. In [5] closely spaced targets are not considered independently in order to avoid track coalescence. However the problem solved and the proposed approach differ completely from those presented in this work.

The paper is structured as follows: in section II a brief introduction to the tracking and data association problem and to the standard solutions is presented. An alternative expression for the likelihood of the observation is derived in order to deal with missing data. In section III the shape constraints are introduced. A particular expression for the prior density embedding the motion invariant knowledge well suited to deal with rigid and articulated bodies is presented. A general method which allows to learn such density from data is described. In section IV an overall observation model taking into account targets dynamics and shape constraints is introduced. In section V such overall model is employed in the computation of the associations probabilities. The computation of such probabilities in presence of missing data is treated in some

details. A Monte Carlo approach is adopted in order to estimate the positions of occluded points. In section VI the state estimation process is presented. A general method which allows to embed the Monte Carlo estimates into the Kalman filters in order to obtain the overall state estimates of occluded points is explained in some details. Finally, in section VII our methodology is tested on real data. The experiments performed demonstrate the validity of our approach. The developed algorithm performs robustly in presence of many false detections and occlusions lasting several time steps.

## II. TRACKING AND DATA ASSOCIATION

Let us consider the problem of tracking a set of points (or targets) moving in space generating unlabelled<sup>1</sup> measurements of their positions. We assume the  $i$ -th point can be described by a linear state space model of the form:

$$\begin{cases} x_i(k+1) = F_i \cdot x_i(k) + v_i(k) \\ y_i(k) = H_i \cdot x_i(k) + w(k) \end{cases} \quad (1)$$

where  $x_i$  and  $y_i$  are respectively the state (for instance position and velocity) and measurable position of target  $i$ . The model and measurement noises  $v_i$  and  $w$  are assumed to be white and zero mean gaussian distributed with covariances  $Q$  and  $R$  respectively.

In order to estimate the state of the point  $i$  over time a Kalman filter is associated to the model (1). Assume a state estimate of each point is available at time  $k-1$ . The state predictions for the next time step  $k$  can be computed using the standard Kalman formulas. At time  $k$ , a set of  $M_k$  unlabelled measurements  $\mathbf{z}(k) = [z_1(k), \dots, z_{M_k}(k)]$  is received. In what follows the time index  $k$  is dropped whenever unnecessary and  $\hat{x}_i^-$  and  $P_i^-$  denote the prediction of state  $x_i(k)$  and its error covariance respectively given the measurements up to time  $k-1$ . If the measurements were labelled, i.e. we knew which measurement is associated to which target, they could be used to update in the standard way the state predictions. To deal with unlabelled measurements it is customary [1] to introduce the *association events* (or hypothesis)  $\theta \in \Theta$ .  $\Theta$  is the association space, i.e. the set of all possible ways to associate measurements to targets. According to an association  $\theta$ , each measurement is either associated to a target or it is considered a false detection<sup>2</sup>. The underlying assumption is that each target either generates one (and only one) measurement or it is occluded, i.e. it is not guaranteed that every target generates one measurement.

Given an association  $\theta$ ,  $j(i, \theta)$  will denote the measurement associated to the point  $i$  or equivalently:

$$y_i = z_{j(i, \theta)}$$

<sup>1</sup>Unlabelled means that it is unknown from which point a given measurement has been generated.

<sup>2</sup>In this case one usually says that it comes from *clutter*.

Conditionally on an association  $\theta$  the estimate  $\hat{x}_{i,\theta}$  of the state  $x_i$  and its error covariance  $P_{i,\theta}$  are given by standard Kalman recursions associated to the model (1) driven by measurements  $z_{j(i,\theta)}$ . For convenience of notation, we shall denote with  $K_x(z_{j(i,\theta)}, R)$  and  $K_P(R)$  respectively the state and covariance update<sup>3</sup> given measurement  $y_i$  ( $R$  is the covariance of the measurement noise  $w$  in (1)), i.e.  $\hat{x}_{i,\theta} = K_x(z_{j(i,\theta)}, R)$  will be the estimate of the state at time  $k$  given measurements up to time  $k$  and  $P_{i,\theta} = K_P(R)$  its error covariance for all hypothesis  $\theta$  which associates a measurement to the point  $i$ .

When the point  $i$  generates no measurement according to the hypothesis  $\theta$ , its conditional state estimate will coincide with the prediction given by the Kalman filter, i.e.  $\hat{x}_{i,\theta} = \hat{x}_i^-$  and  $P_{i,\theta} = P_i^-$  for all  $\theta$  which assign no measurement to the point  $i$ .

Finally, in order to iterate the procedure described above, one needs to compute an overall state estimate for each point. This can be done starting from the conditional state estimates in the following way. Each conditional state estimate is given a degree of reliability which is the posterior probability of the corresponding association event. Let us denote with  $Z^k = \{\mathbf{z}(0), \dots, \mathbf{z}(k)\}$ ; in a Bayesian setting, i.e. when a prior  $p(\theta)$  is given, each hypothesis  $\theta$  has associated a posterior probability  $p(\theta | Z^k)$  which can be computed according to [1]:

$$\begin{aligned} p(\theta | Z^k) &= c \cdot p(\mathbf{z}(k) | \theta, Z^{k-1}) \cdot p(\theta | Z^{k-1}) \\ &= c \cdot p(\mathbf{z}(k) | \theta, Z^{k-1}) \cdot p(\theta) \end{aligned} \quad (2)$$

where  $c$  is a normalization constant. In what follows  $Z$  and  $Z^-$  indicate  $Z^k$  and  $Z^{k-1}$  respectively.

An overall estimate  $\hat{x}_i$  of the state  $x_i$  and its error covariance  $P_i$  can be obtained as follows:

$$\begin{cases} \hat{x}_i = \sum_{\theta} \hat{x}_{i,\theta} \cdot p(\theta | Z) \\ P_i = \sum_{\theta} P_{i,\theta} \cdot p(\theta | Z) \end{cases} \quad (3)$$

Expressions for the prior probability  $p(\theta)$  and for the likelihood of the measurements  $p(\mathbf{z} | \theta, Z^-)$  are given in [1]. We introduce an alternative expression for  $p(\mathbf{z} | \theta, Z^-)$  which will be very useful to deal with missing data in section V.

The output variables  $y_1, \dots, y_N$  are described by an appropriate density  $p(y_1, \dots, y_N | Z^-)$  which depends on the past history  $Z^-$  of the observations. For instance, when the points are described independently by (1), the density  $p(y_1, \dots, y_N | Z^-)$  is given by:

$$p(y_1, \dots, y_N | Z^-) = \prod_{i=1}^N p(y_i | Z_i^-) \quad (4)$$

where  $p(\cdot | Z_i^-)$  is a Gaussian density of mean  $H \cdot \hat{x}_i^-$ , the predicted position of point  $i$  provided by the  $i$ -th Kalman filter, and covariance  $H \cdot P_i^- \cdot H^T + R$ . In the

<sup>3</sup>In filtering form.

remaining part of this section a generic observation model  $p(y_1, \dots, y_N | Z^-)$  is considered and a general expression for the computation of the likelihood of measurements is derived. This will be adopted in section V whereby  $p(y_1, \dots, y_N | Z^-)$  has a more general form of (4) since relations among points will be taken into account.

The likelihood  $p(\mathbf{z} | \theta, Z^-)$  can be factorized as follows:

$$p(\mathbf{z} | \theta, Z^-) = p(\mathbf{z}_F | \theta, Z^-) \cdot p(\mathbf{z}_T | \theta, Z^-) \quad (5)$$

where  $\mathbf{z}_T$  and  $\mathbf{z}_F$  denote the true (generated by points) and false measurements according to the given association  $\theta$  respectively which are assumed to be independent.  $p(\mathbf{z}_F | \theta, Z^-)$  is usually taken uniform in the observed space. In order to compute  $p(\mathbf{z}_T | \theta, Z^-)$  we proceed in the following manner.  $\mathbf{z}_T = \{z_{j(i,\theta)}, i \in D_{\theta}\}$  where  $D_{\theta}$  is the set of indices of all detected points (which have generated a measurement according to  $\theta$ ) and  $j(i, \theta)$  is the measurement associated to point  $i$  in the current hypothesis  $\theta$ , i.e.  $y_i = z_{j(i,\theta)}$ . Denote with  $\mathbf{y}_{D_{\theta}} = \mathbf{z}_T$  the following set of equations:

$$\mathbf{y}_{D_{\theta}} = \mathbf{z}_T : y_i = z_{j(i,\theta)} \quad i \in D_{\theta}$$

When the all  $N$  points are detected (i.e.  $D_{\theta} = \{1, \dots, N\}$ ),  $p(\mathbf{z}_T | \theta, Z^-)$  can be computed through  $p(y_1, \dots, y_N | Z^-)$  as follows:

$$p(\mathbf{z}_T | \theta, Z^-) = p(\mathbf{y}_{D_{\theta}} = \mathbf{z}_T | Z^-) \quad (6)$$

Some complications arise when the generic point  $i$  is occluded and the corresponding output variable  $y_i$  is unknown according to a given association  $\theta$ . In order to deal with this problem the following solution is proposed. Denote with  $M_{\theta}$  the set of occluded points according to a given  $\theta$  and with  $\mathbf{y}_{M_{\theta}}$  their unknown positions:

$$\mathbf{y}_{M_{\theta}} = \{y_i, i \in M_{\theta}\}$$

The likelihood  $p(\mathbf{z}_T | \theta, Z^-)$  can be computed as in (6) marginalizing over the positions  $\mathbf{y}_{M_{\theta}}$  of the missing points:

$$p(\mathbf{z}_T | \theta, Z^-) = \int p(\mathbf{y}_{D_{\theta}} = \mathbf{z}_T, \mathbf{y}_{M_{\theta}} | Z^-) d\mathbf{y}_{M_{\theta}} \quad (7)$$

When the observation model is of the form (4)

$$\begin{aligned} p(\mathbf{y}_{D_{\theta}} = \mathbf{z}_T, \mathbf{y}_{M_{\theta}} | Z^-) &= \prod_{i \in D_{\theta}} p(z_{j(i,\theta)} | Z_i^-) \\ &\cdot \prod_{i \in M_{\theta}} p(y_i | Z_i^-) \end{aligned} \quad (8)$$

and the integration (7) yields:

$$p(\mathbf{z}_T | \theta, Z^-) = \prod_{i \in D_{\theta}} p(z_{j(i,\theta)} | Z_i^-)$$

which is the standard expression used in [1]. Equation (7) will be very useful in section V where a more complicated observation model is considered.

In conclusion, from (7), (5) and (2), the posterior probability  $p(\theta | Z)$  of a given association  $\theta$  is given by:

$$p(\theta | Z) = c \cdot p(\theta) \cdot p(\mathbf{z}_F | \theta, Z^-) \cdot \int p(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} | Z^-) d\mathbf{y}_{M_\theta} \quad (9)$$

Standard expressions for  $p(\theta)$  and  $p(\mathbf{z}_F | \theta, Z^-)$  can be found in [1].

### III. SHAPE CONSTRAINTS

As anticipated in the introduction there might be some sort of coordination between targets; their spacial configuration may be conveniently described by an a priori probabilistic model of the form

$$\bar{p}(y_1(k), \dots, y_N(k)) \quad (10)$$

which depends only on the relative positions and not on motion of the single targets. We say that (10) describes the shape constraints of the points.

In the next sections a general methodology which allows to embed an arbitrary  $\bar{p}(y_1, \dots, y_N)$  into a tracking algorithm is proposed. In the remaining part of this section a particular expression for  $\bar{p}(y_1, \dots, y_N)$  which is useful to deal with rigid and articulated bodies is derived. Such expression will be used in the experiments presented in section VII.

In order to construct the model  $\bar{p}(y_1, \dots, y_N)$  we start by choosing properties of the positions of points which can be reasonably described by prior constraints. If the points form rigid or articulated patterns pairwise distances between rigidly connected points are a reasonable choice. In real scenarios the configurations of points may not be perfectly rigid and therefore a probabilistic model is more suited. We assume the distance between point  $i$  and  $j$  can be modeled by a gaussian density with mean  $\mu_{ij}$  and variance  $\sigma_{ij}^2$ . It is convenient to introduce the set  $\mathcal{R}$  of pairs of points whose distances are characterized by variances which are sufficiently small (the variances would be equal to zero if pairs were perfectly rigid). Let us define

$$\mathcal{N}(d | \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(d-\mu)^2}{\sigma^2}}$$

The density  $\bar{p}(y_1(k), \dots, y_N(k))$  is chosen of the following form<sup>4</sup>:

$$\bar{p}(y_1(k), \dots, y_N(k)) = c \cdot \prod_{(i,j) \in \mathcal{R}} \mathcal{N}(\|y_i(k) - y_j(k)\| | \mu_{ij}, \sigma_{ij}^2) \quad (11)$$

where  $c$  is a normalization constant and  $\|\cdot\|$  denotes the  $L_2$  norm.

The means  $\mu_{ij}$  and the variances  $\sigma_{ij}^2$  can be learned from data in the following manner. A standard tracking

<sup>4</sup>Even though  $\|y_i(k) - y_j(k)\|$  is always positive we still assume that it can be approximately described by a gaussian density.

algorithm (based on independent dynamics (1)) provides the trajectories of the points from which all the possible pairwise distances are computed. At a given instant of time, all past values of a given distance can be used to compute its mean and variance. Pairs of points whose distances exhibit a variance greater than a convenient threshold are no longer considered. At the end of the learning phase all pairs which have not been discarded form the set  $\mathcal{R}$  and the corresponding means and variances are available. Meaningful results are obtained if the trajectories are sufficiently and persistently exciting [11].

### IV. OBSERVATION MODEL

In this section we introduce an overall observation model which combines (10) describing mutual configuration and (4) modeling separately target dynamics.

We assume the overall observation model  $p(y_1(k), \dots, y_N(k) | Z^-)$  can be factored in two terms embedding (10) and (4) in the following form

$$p(y_1(k), \dots, y_N(k) | Z^-) = c \cdot \prod_{i=1}^N p(y_i(k) | Z_i^-) \cdot \bar{p}(y_1(k), \dots, y_N(k)) \quad (12)$$

The observation model (12) takes into account all the available information and it is employed in the computation of the associations probabilities as explained in the next section.

### V. ASSOCIATIONS PROBABILITIES

In order to compute the posterior probability (9) of a given association  $\theta$  the likelihood (7) of the true measurements  $\mathbf{z}_T$  needs to be computed. Similarly to (8), the overall observation model (12) can be written as follows:

$$p(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} | Z^-) = c \cdot \prod_{i \in D_\theta} p(z_{j(i,\theta)} | Z_i^-) \cdot \prod_{i \in M_\theta} p(y_i | Z_i^-) \cdot \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta}) \quad (13)$$

which substituted into (7) yields:

$$p(\mathbf{z}_T | \theta, Z^-) = c \cdot \prod_{i \in \mathcal{D}_\theta} p(z_{j(i,\theta)} | Z_i^-) \cdot \int \prod_{i \in \mathcal{M}_\theta} p(y_i | Z_i^-) \cdot \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta}) d\mathbf{y}_{M_\theta} \quad (14)$$

The integral in (14) is solved through 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 as discussed in some details in the next section.

Following Monte Carlo approach one draws an appropriate number  $N_s$  of independent and identically distributed samples:

$$\mathbf{y}_{M_\theta}^{(n)} \triangleq \{y_i^{(n)}, i \in \mathcal{M}_\theta\} \sim \prod_{i \in \mathcal{M}_\theta} p(\cdot | Z_i^-) \quad n = 1, \dots, N_s \quad (15)$$

and computes the  $n$ -th weight through the following expression:

$$\pi^{(n)} = \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} = \mathbf{y}_{M_\theta}^{(n)}) \quad (16)$$

Finally, the integral is computed as follows:

$$\int \prod_{i \in \mathcal{M}_\theta} p(y_i | Z_i^-) \cdot \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta}) \, d\mathbf{y}_{M_\theta} \propto \sum_{n=1}^{N_s} \pi^{(n)} \quad (17)$$

and substituted in (14) yields  $p(\mathbf{z}_T | \theta, Z^-)$ .

## VI. STATE ESTIMATION

A Kalman estimator based on the model (1) is associated to each target. Consider a target  $i$  which is detected according to a given association  $\theta$ . The corresponding measurement  $z_{j(i,\theta)}$  is used to update the  $i$ -th Kalman estimator as in standard methods (see section II). Nevertheless, the probability of the hypothesis  $\theta$  is computed in a different way with respect to the standard approaches, i.e. taking into account the shape information encoded in  $\bar{p}(y_1, \dots, y_N)$  as explained in the previous section. The associations probabilities contribute to the computation of the overall state estimates (3) which hence depend on the shape information as well.

Consider now a point  $i$  which generates no measurement according to the association  $\theta$ . In the standard approaches the corresponding Kalman estimator cannot be updated and the conditional state estimates of point  $i$  coincide with its state predictions as explained in section II.

When an appropriate prior density  $\bar{p}(y_1, \dots, y_N)$  is available, a better solution can be adopted. The conditional density  $p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-)$  of the occluded points positions  $\mathbf{y}_{M_\theta}$  given the detected ones  $\mathbf{y}_{D_\theta}$  and the past history  $Z^-$  of the observations can be written as follows:

$$p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-) = \frac{p(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} | Z^-)}{p(\mathbf{y}_{D_\theta} = \mathbf{z}_T | Z^-)}$$

The probability  $p(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta} | Z^-)$  is given by (13) and  $p(\mathbf{y}_{D_\theta} = \mathbf{z}_T | Z^-)$  does not depend on  $\mathbf{y}_{M_\theta}$ , hence the following expression holds:

$$p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-) \propto \prod_{i \in \mathcal{D}_\theta} p(z_{j(i,\theta)} | Z_i^-) \cdot \prod_{i \in \mathcal{M}_\theta} p(y_i | Z_i^-) \cdot \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta})$$

and noting that  $\prod_{i \in \mathcal{D}_\theta} p(z_{j(i,\theta)} | Z_i^-)$  does not depend on  $\mathbf{y}_{M_\theta}$ , the following expression is finally derived:

$$p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-) \propto \prod_{i \in \mathcal{M}_\theta} p(y_i | Z_i^-) \cdot \bar{p}(\mathbf{y}_{D_\theta} = \mathbf{z}_T, \mathbf{y}_{M_\theta}) \quad (18)$$

The density  $p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-)$  is approximated (see (15) and (16) and recall factored sampling techniques) by the following set:

$$\{(\mathbf{y}_{M_\theta}^{(n)}, \pi^{(n)}), n = 1, \dots, N_s\} \quad (19)$$

Such set can be used to compute a gaussian approximation of the density  $p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-)$  which will be useful in order to compute the state estimates of occluded targets:

$$p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-) \simeq \prod_{i \in \mathcal{M}_\theta} \mathcal{N}(y_i | \hat{y}_{i,\theta}, \hat{\Sigma}_{y_i,\theta})$$

$\hat{y}_{i,\theta}$  and  $\hat{\Sigma}_{y_i,\theta}$  are the weighted sample means and covariances respectively computed starting from (19). The posterior density  $p(\mathbf{y}_{M_\theta} | \mathbf{y}_{D_\theta} = \mathbf{z}_T, Z^-)$  embeds all the available information: the predictions  $Z_i^-$  provided by the Kalman filters, the shape constraints  $\bar{p}(\cdot)$  and the measurements  $\mathbf{z}_T$  of the detected points (see equation (18)). Hence, it is reasonable to say that  $\hat{y}_{i,\theta}$  and  $\hat{\Sigma}_{y_i,\theta}$  are the estimates and the error covariances of the positions of occluded points according to a given hypothesis  $\theta$ :

$$\begin{cases} \hat{y}_{i,\theta} = E[H \cdot x_i | \theta, \mathbf{z}_T, \bar{p}, Z_i^-] \\ \hat{\Sigma}_{y_i,\theta} = \text{Var}[H \cdot x_i | \theta, \mathbf{z}_T, \bar{p}, Z_i^-] \end{cases} \quad (20)$$

where  $x_i$  is the state of point  $i$  and  $H$  is the output matrix in the state model (1).  $\bar{p}$  and  $Z_i^-$  indicate explicitly that the prior constraints and the predictions provided by the Kalman estimators are used to obtain  $\hat{y}_{i,\theta}$  and  $\hat{\Sigma}_{y_i,\theta}$ . When  $H$  is the identity matrix  $\hat{y}_{i,\theta}$  and  $\hat{\Sigma}_{y_i,\theta}$  are the conditional state estimate (according to  $\theta$ ) and the error covariance respectively. When  $H$  is not the identity matrix and it is not invertible (for instance in those cases in which both the position and velocity of a point are embedded in the state vector) the following solution can be adopted. Associate to the occluded point  $i$  an unknown fictitious measurement  $z = H \cdot x_i + w$  where  $w \sim \mathcal{N}(0, R)$  and  $R$  is unknown. Such measurement is supposed to be equivalent to the information given by  $\bar{p}$  and by the measurements  $\mathbf{z}_T$  of the detected points in the given association  $\theta$  or more formally:

$$\begin{cases} E[H \cdot x_i | \theta, \mathbf{z}_T, \bar{p}, Z_i^-] = E[H \cdot x_i | \theta, z, Z_i^-] \\ \text{Var}[H \cdot x_i | \theta, \mathbf{z}_T, \bar{p}, Z_i^-] = \text{Var}[H \cdot x_i | \theta, z, Z_i^-] \end{cases} \quad (21)$$

The conditional estimates of the state given the measurement  $z$  and the past history  $Z_i^-$  are given by the Kalman estimator:

$$\begin{cases} E[H \cdot x_i | \theta, z, Z_i^-] = H \cdot K_x(z, R) \\ \text{Var}[H \cdot x_i | \theta, z, Z_i^-] = H \cdot K_P(R) \cdot H^T \end{cases} \quad (22)$$

where  $K_x(z, R)$  and  $K_P(R)$  indicate the state estimate and the relative error covariance provided by the estimator (see section II) whose explicit expressions are the following:

$$\begin{cases} K_x(z, R) = \hat{x}^- + P^- H^T \Lambda^{-1} \cdot e \\ K_P(R) = P^- - P^- H^T \Lambda^{-1} H P^- \end{cases} \quad (23)$$

$\hat{x}^-$  and  $P^-$  are the state prediction and the corresponding error covariance,  $e = z - Hx^-$  is the innovation and  $\Lambda$  is the innovation covariance which is given by the following expression:

$$\Lambda = HP^-H^T + R$$

In order to obtain the state estimate  $K_x(z, R)$  and the corresponding error covariance  $K_P(R)$  the unknowns quantities  $\Lambda^{-1}$  and  $e$  need to be computed. From equations (20), (21) and (22) follows that

$$\begin{cases} \hat{y} = H \cdot K_x(z, R) \\ \hat{\Sigma}_y = H \cdot K_P(R) \cdot H^T \end{cases} \quad (24)$$

where the subscripts  $i$  and  $\theta$  are dropped for convenience of notation. Using (23) in (24) yields

$$\begin{cases} \hat{y} = H\hat{x}^- + S\Lambda^{-1}e \\ \hat{\Sigma}_y = S - S\Lambda^{-1}S \end{cases} \quad (25)$$

where  $S = HP^-H^T$ . From the second equation of (25) follows that

$$\Lambda^{-1} = S^{-1}(I - \hat{\Sigma}_y S^{-1}) \quad (26)$$

( $I$  is the identity matrix) which is meaningful under the following condition:

$$\hat{\Sigma}_y < HP^-H^T$$

Substituting (26) in the first equation of (25) yields the expression for the innovation:

$$e = (I - \hat{\Sigma}_y S^{-1})^{-1}(\hat{y} - H\hat{x}^-) \quad (27)$$

Using (26) and (27) in (23) yields the expression of the state estimate

$$\begin{aligned} K_x(z, R) &= \hat{x}^- + P^-H^T S^{-1}(I - \hat{\Sigma}_y S^{-1})(I - \hat{\Sigma}_y S^{-1})^{-1} \\ &\quad \cdot (\hat{y} - H\hat{x}^-) \\ &= \hat{x}^- + P^-H^T(HP^-H^T)^{-1}(\hat{y} - H\hat{x}^-) \\ &= K_x(\hat{y}, R = 0) \end{aligned}$$

and of the estimate error covariance:

$$\begin{aligned} K_P(R) &= P^- - P^-H^T(HP^-H^T)^{-1}[I - \hat{\Sigma}_y(HP^-H^T)^{-1}] \\ &\quad \cdot HP^- \\ &= P^- - P^-H^T(HP^-H^T)^{-1}HP^- + P^-H^T S^{-1} \\ &\quad \cdot \hat{\Sigma}_y S^{-1}HP^- \\ &= K_P(R = 0) + P^-H^T S^{-1}\hat{\Sigma}_y S^{-1}HP^- \end{aligned}$$

## VII. EXPERIMENTS

The algorithm is tested on real data obtained through a motion capture system. Points on a rigid body are moving in space and generating measurements of their positions. Many other observations are generated by other points (see figure (1)). Moreover some of the points of interest can be occluded for several frames.

The dynamics of targets are described by (1) where the state of each point is formed by its position and velocity. The shape model is of the form (11) and it is learned from data

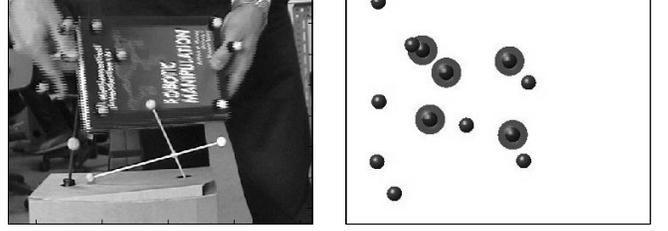


Fig. 1. On the left hand side a picture of the scene is reported. Five points are placed on the cover of a book. Many points are located on other objects. On the right hand side the overall set of observations at a given instant of time is depicted (the measurements of the five points on the book are highlighted).

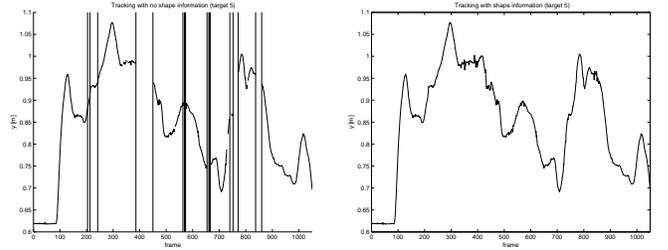


Fig. 2. The trajectory of the y-coordinate of a point of interest. At first, shape is not embedded into the tracker. The point is given many labels because the tracker is distracted by the measurements generated by other points. Moreover, an occluded point cannot be tracked reliably for many frames and it is given a new label when it is detected again. The segments of the point trajectory identified by many different labels are manually merged together and reported on the left hand side. Vertical lines indicate the time instants in which the point is given a new label. Note that in certain intervals of time the trajectory is not available since the point is occluded. When shape information is considered the tracker reconstructs reliably the trajectory of the point which is reported on the right hand side.

as explained in section III. The algorithm performs robustly. The state estimates are not affected by the many observations close to those generated by the points of interest. The trajectories of occluded points are reconstructed reliably. The tracking algorithm does not exhibit such robustness when shape knowledge is not considered (see figure (2)). In absence of shape one could improve performances adopting for instance multi hypothesis techniques. However, many measurements which are not generated by targets of interest are persistently present in the observed space. This implies in general many and persistent hypothesis (i.e. high ambiguity) on the state of a given point of interest. Moreover, when shape is not available, it is not possible to reconstruct the trajectories of occluded targets even using sophisticated tracking algorithms.

Also computational aspects have been examined. The bottleneck of the algorithm could be considered the Monte Carlo computations performed to deal with missing points. In order to reduce the number of times numerical integration has to be solved hypothesis that a certain target is occluded are considered only if there are no other associations where it is matched to some measurement which have a reasonably high likelihood.

## VIII. CONCLUSIONS AND FUTURE WORK

A general and well formalized methodology which allows the integration of prior probabilistic shape constraints into data association filters has been presented. Such methodology is flexible since the shape prior can be of arbitrary form. Shape constraints can be learned from data and their validity extends to arbitrary motions. A general and comprehensive solution to the problem of missing data is given. The resulting algorithm performs robustly in presence of many false detections and occlusions lasting several consecutive time steps.

Short term future work concerns the on line learning of shape constraints which is done embedding the learning phase into the tracking process. This would eliminate the off line learning which requires sufficiently and persistently exciting data sequences without heavy clutter and long occlusions. Clearly, prior information implies great robustness and hence it should be used whenever possible.

A long term goal is the derivation of a general description of coordination which would lead to the development of powerful tracking algorithms capable of recognizing and tracking groups of targets exhibiting arbitrary forms of coordination.

## REFERENCES

- [1] Y. Bar-Shalom and T.E. Fortmann "Tracking and data association", New York, Academic Press, 1988.
- [2] C. Rasmussen, G.D. Hager, Probabilistic Data Association Methods for Tracking Complex Visual Objects, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No.6, June 2001, pp. 560-576.
- [3] G. Gennari, A. Chiuso, F. Cuzzolin, R. Frezza, "Integrating probabilistic shape and dynamic models in data association and tracking", *Proc. IEEE Int. Conf. on Decision and Control*, Las Vegas, 2002.
- [4] C. Rasmussen G. D. Hager, "Joint Probabilistic Techniques for tracking multi-part objects", *Proc. Computer Vision and Pattern Recognition*, 1998, pp. 16-21.
- [5] Henk A. P. Blom and Edwin A. Bloem, Probabilistic Data Association Avoiding Track Coalescence, *IEEE Trans. on Automatic Control*, Vol. 45, No. 2, February 2000, pp. 247-259.
- [6] Yang Song, Luis Goncalves and Pietro Perona, "Monocular perception of biological motion-clutter and partial occlusion", *Proc. of European Conference of Computer Vision*, June 2000.
- [7] Donald B. Reid, An algorithm for tracking multiple targets, *IEEE Trans. on Automatic Control*, Vol. 24, No. 6, pp. 843-854, 1979.
- [8] E. Mazor, A. Averbuch, Y. Bar Shalom, J. Dayan, Interacting Multiple Model Methods in Target Tracking: A survey, *IEEE Trans. on Aerospace and Electronic systems*, Vol. 34, No. 1, January 1998, pp. 103-123.
- [9] M. Ringer and J. Lasenby, "Multiple Hypothesis Tracking for Automatic Optical Motion Capture", *European Conf. on Computer Vision*, 2002, pp. 524-536.
- [10] L. Herda, P. Fua, R. Plankers, R. Boulic and D. Thalmann, Skeleton-Based Motion Capture for Robust Reconstruction of Human Motion, *Computer Animation*, IEEE press, May 2000.
- [11] L.Ljung, *System identification : theory for the user*, Prentice Hall, 1997.
- [12] Y. Bar-Shalom and X.R. Li "Multitarget-Multisensor Tracking: Principles and techniques". YBS Publishing, 1995.
- [13] Carine Hue, Jean-Pierre Le Cadre, Patrick Perez, Sequential Monte Carlo Methods for Multiple Target Tracking and Data Fusion, *IEEE Trans. on Signal Processing*, Vol. 50, No.2, February 2002, pp. 309-325.
- [14] A. Doucet, B-N. Vo, C. Andrieu, and M. Davy, "Particle filtering for multi-target tracking and sensor management", *International Conference on Information Fusion*, Vol. 1, pp. 474-481, 2002.