

Probabilistic Data Association Methods in Visual Tracking of Groups

G. Gennari

Information Engineering
University of Padova
Padova, Italy 35131

G. D. Hager

Computer Science
Johns Hopkins University
Baltimore, MD 21218

Abstract

Data association is a fundamental problem when tracking large numbers of moving targets. Most commonly employed methods of data association such as the JPDA estimator are combinatorial and therefore do not scale well to large numbers of targets. However, in many cases large numbers of targets form natural groups which can be efficiently tracked. We describe a method for defining groups based on the position and velocity of targets. This definition introduces a natural set of merging and splitting rules that are embedded into a Kalman filtering framework for tracking multiple groups. In cases where groups of different velocities cross, a general methodology for matching measurements to groups is introduced. This algorithm is based on a modified version of the PDA estimator. It is well suited to handle a high number of measurements and extends naturally to additional grouping constraints such as color or shape.

1. Introduction

Consider the problem of tracking crowds of clumped people walking in a station or flocks of birds flying together or colonies of ants. At one level of granularity, each of these phenomena consists of individual motions. However, taken together, the dominant characteristic is a gestalt group motion. Furthermore, in many cases, it is the group motion that is of interest, rather than the motions of each individual.

Typically, the problem of tracking large numbers of targets has been considered using data association techniques such as the JPDAF and its variants [1, 2]. Extensions of these ideas to computer vision problems can be found in [6]. In general, data association techniques attack the problem of distinguishing among targets so that individual tracks can be established, but

do not address the notion of abstraction to a larger group. Indeed, one of the major limitations of these methods are the combinatorics that result when scaling to large numbers of targets.

In situations like these, it is often difficult to distinguish the individual objects of interest because they are quite similar to the others and may be partially or totally occluded. Indeed, in such cases, it often makes sense to track the overall ensemble of objects as a group using a model of the group dynamics.

In this paper, a general class of partitioning functions is introduced which allows us to formally define a group. Each group is completely described by its state which takes into account information about spatial location, cardinality and velocity. We introduce rules for splitting and merging groups based on state spaces. In general this leads to situations in which groups sharing visual measurements are not merged together because they are moving with different velocities. To deal with such situations we propose a data association methodology that is a PDA estimator properly modified to embed the notion of a group. One of the unique aspects of our approach is that we first associate each *single* measurement to a group on the basis of the state predictions and then we further partition the resulting subsets of measurements into connected components. In traditional approaches [3, 4, 5, 7] the overall set of measurements is usually first partitioned into connected components *without reference to existing targets*, and only then the resulting sets of measurements are matched to targets.

2. Description of a group

In what follows we first present a formal definition of a group. Then we focus on the group tracking problem introducing an appropriate state model.

2.1. Definition of group

As we mentioned in the introduction, our goal is to differentiate between collections of clumped objects that can be treated jointly as a group, and subsets of objects that can be reliably differentiated from one another. We will formally define groups in terms of a partitioning of the set of available measurements, and will say a collection of objects forms a group when it cannot be segmented in finer partitions.

To this end consider a set $A = \{a_1, \dots, a_N\}$ and a grouping (or partitioning) function \mathcal{G} . The partitions obtained applying \mathcal{G} to the set A are indicated through the following notation:

$$\mathcal{G}(A) = \{\{a_i\}_{i \in I_1}, \dots, \{a_i\}_{i \in I_M}\}$$

where I_i form a partition of $\{1, \dots, N\}$. The cardinality of the partition, M , will be denoted $|\mathcal{G}(A)|$ in the sequel. Note that each element of the partition is an invariant set of the grouping function \mathcal{G} —that is for a given I_j above,

$$\mathcal{G}(\{a_i\}_{i \in I_j}) = \{a_i\}_{i \in I_j}$$

We say an arbitrary set A is a *group* with respect to a grouping function \mathcal{G} if it is an invariant set of \mathcal{G} , i.e. $\mathcal{G}(A) = \{A\}$ or $|\mathcal{G}(A)| = 1$ and A is maximal. Equivalently we say that A satisfies the *group constraint* with respect to \mathcal{G} . The invariance of partition elements under \mathcal{G} is the analog of saying that a group cannot be segmented into smaller groups. For instance, a process that extracts connected components of a graph is a grouping function and each connected component is an invariant set or a group.

2.2. State model

Consider now a set G of N objects moving in the scene with measured positions¹ $C_y = \{y_1, \dots, y_N\}$, and suppose that the set C_y satisfies the group constraint with respect to an appropriate grouping function (we defer the discussion of the measurement grouping function until section 3.3). In this case, the N objects of G are treated as a unique object which can be described completely through a state vector that contains the spatial location, the velocity and the cardinality of the overall ensemble:

$$G : (\xi, v, \Sigma, N)$$

¹ In general additional features such as the color can be considered as well as position; for the sake of simplicity we just use position.

where ξ and v are the position and velocity of the centroid of the group respectively. In this model, the position of the i -th object in the group is expressed as follows:

$$y_i = \xi + s_i \quad (2.1)$$

where $s_i \sim \mathcal{N}(0, \Sigma)$. In other words Σ is a covariance matrix which defines the spatial extension of the group; the positions C_y are expected to be in the following *search region*

$$SR_\xi = \{y | (y - \xi)^T \cdot \Sigma^{-1} (y - \xi) < \gamma\} \quad (2.2)$$

with a probability higher than a threshold which depends on the choice of γ . Finally N is the number of points forming the group.

Summing the equations (2.1) over i and introducing a dynamical model for the position and velocity of the group centroid we obtain the following model:

$$\begin{cases} x(k+1) = Fx(k) + n_x(k) \\ y_c(k) = Hx(k) + n_y(k) \end{cases} \quad (2.3)$$

where $x^T = (\xi \ v)^T$ and y_c is the measured position of the centroid. Note that $n_y(k)$ includes both the uncertainty in the position of a group member, and observation noise. In order to propagate x over time a standard Kalman filter is applied to the model (2.3). In addition, the group covariance Σ and the group cardinality N are estimated at time k as

$$\begin{aligned} \hat{\Sigma}(k) &= \alpha_s \cdot \hat{\Sigma}(k-1) + (1 - \alpha_s) \cdot \text{cov}(C_y) \\ \hat{N}(k) &= \alpha_N \cdot \hat{N}(k-1) + (1 - \alpha_N) \cdot |C_y| \end{aligned} \quad (2.4)$$

where $\text{cov}(C_y)$ and $|C_y|$ indicate the sample covariance and the cardinality of the set C_y of measurements generated by the group while α_s and α_N are appropriately chosen constants (positive and smaller than one).

One role of the Kalman filter is to produce an estimate for the velocity of the group, as well as its covariance matrix. By analogy with SR_ξ , we indicate with SR_v the search region defined by the predicted velocity and its corresponding error covariance matrix.

At this point, we have defined a group and, given a method for computing groups, we have defined a state updating mechanism. We must now incorporate mechanisms for managing groups (subject to a given partitioning method), and we must define a data association scheme for cases where the association of a measurement to a group is ambiguous.

3. Merging and splitting groups

In many situations groups interact, forming larger groups, or disperse, forming smaller groups. We address the formation of groups by creating connected

components of measurements that are similar in both position and velocity. When these measurements belong to other groups, the overall ensemble of groups is treated as a single group by replacing the collection of individual states by a unique merged state. A group splits when it no longer satisfies the group constraint under an appropriate partitioning function. In section 3.3 we introduce the specific group constraint adopted in our algorithm.

3.1. State space grouping

Two groups are usually considered in spatial proximity when their SR_ξ search regions overlap, and have similar velocities when their SR_v search regions overlap. We define the relation \mathcal{R}_x among groups through the following expression:

$$G_i \mathcal{R}_x G_j \iff \begin{cases} SR_{\xi_i} \cap SR_{\xi_j} \neq \emptyset \\ SR_{v_i} \cap SR_{v_j} \neq \emptyset \end{cases}$$

Thus, \mathcal{R}_x induces an undirected graph among groups. Let \mathcal{G}_x be a partitioning function that computes the connected components of this graph. That is, given existing groups $C_G = \{G_1, \dots, G_N\}$, \mathcal{G}_x computes a partition

$$\mathcal{G}_x(C_G) = \{\{G_i\}_{i \in I_1}, \dots, \{G_i\}_{i \in I_M}\}$$

3.2. Merging operator

If any of the I_i above has cardinality greater than 1, we consider the group constraint to be violated and merge the associated groups. Let (x_i, Σ_i, N_i) be the state of the i -th group G_i . The merging operator applied to the collection of groups $\{G_i\}_{i \in I_j}$ yields a new merged group $G_{m,j}$

$$\begin{cases} x_{m,j} = \sum_{i \in I_j} \bar{N}_{i,j} \cdot x_i \\ \Sigma_{m,j} = \sum_{i \in I_j} \bar{N}_{i,j} \cdot [\Sigma_i + (\xi_i - \xi_{m,j}) \cdot (\xi_i - \xi_{m,j})^T] \\ N_{m,j} = \sum_{i \in I_j} N_i \end{cases} \quad (3.1)$$

where $\bar{N}_{i,j} = \frac{N_i}{N_{m,j}}$.

3.3. Group constraint

The measurements of a given group must satisfy a specific group constraint. Generally, the group constraint is tested as estimation is performed, however there are two cases where we must group measurements with no prior state information. First, a given group

can in general be associated to several sets of measurements (more details in section 4). The group constraint will be tested for each possible configuration of measurements and when none of them satisfies the group constraint, the group is *split* into new groups initialized starting from the measurements in its search region. Second, measurements that are unclaimed by any group must themselves be grouped. The extreme case is the first frame of video, when no group yet exists.

In this section we describe the initialization process based on a group constraint for a given set of measurements. The group constraint is defined in such a way that it is guaranteed that the new groups cannot be merged back into a single group. We proceed in two stages. First, since we have no prior state information, we begin by clustering based on spatial proximity (essentially an analog to connected components). In particular, a relation \mathcal{R}_0 in the measurement space is defined as follows:

$$y_i \mathcal{R}_0 y_j \iff SR_{y_i,0} \cap SR_{y_j,0} \neq \emptyset$$

where $SR_{y,0}$ indicates a region centered in y with an appropriate radius. As before, given a collection of measurements $C_y = \{y_1, \dots, y_N\}$, each measurement y_i becomes a node in a graph whose edges are determined by \mathcal{R}_0 . C_y can then be partitioned into connected components:

$$\mathcal{G}_0(C_y) = \{\{y_i\}_{i \in I_{1,0}}, \dots, \{y_i\}_{i \in I_{M,0}}\}$$

To each set of grouped measurements $\{y_i\}_{i \in I_{j,0}}$ is associated a group $G_{j,0}$ whose initial state is obtained by computing the cardinality, the sample mean and covariance of $\{y_i\}_{i \in I_{j,0}}$. The initial velocity is assumed to be null. The final result is a collection of groups $C_{G,0} = \{G_{1,0}, \dots, G_{M,0}\}$.

To these groups, the state grouping operator $\mathcal{G}_x(\cdot)$ is applied to partition them into new groups:

$$\mathcal{G}_x(C_{G,0}) = \{\{G_{j,0}\}_{j \in I_1}, \dots, \{G_{j,0}\}_{j \in I_L}\}$$

Let \mathcal{G}_y denote the composition of these two grouping processes. That is \mathcal{G}_y is a grouping function which operates on sets of measurements and produces groups as follows:

$$\mathcal{G}_y(C_y) = \{\{\{y_i\}_{i \in I_{j,0}}\}_{j \in I_1}, \dots, \{\{y_i\}_{i \in I_{j,0}}\}_{j \in I_L}\}$$

A given set of measurements C_y satisfies the group constraint if and only if $|\mathcal{G}_y(C_y)| = 1$.

4. Data association

Data association techniques are crucial to deal with the ambiguity in the origin of measurements shared

among several groups. This is the case in our approach since groups sharing measurements (pixel coordinates) are not merged together when they are moving with different velocities. We propose a PDAF-like methodology to deal with the problem of assigning (large) sets of measurements to each group. In order to keep the computational load acceptable only an appropriate subset of all the associations is considered. The group constraint is properly embedded in the data association process.

4.1. Feasible associations

In what follows we do not consider false measurements and missed detections, focusing instead on the problem of associating measurements shared among several groups. To this end, consider a group G and let Y_G represent the set of measurements which are in the corresponding search region (2.2). Let $Y_{G,ns}$ and $Y_{G,s}$ denote the subset of measurements which G does not share with other groups and the remaining shared measurements respectively. The shared measurements $Y_{G,s}$ are sorted according to their Mahalanobis distance from the centroid of G in ascending order. $Y_{G,s}^j$ indicates the first j sorted measurements.

We now form a set of associations where the j -th association, θ_j , associates the set $Y_{G,\theta_j} = Y_{G,ns} \cup Y_{G,s}^j$ to the group G . The remaining measurements in Y_G are assumed to be generated by the other groups sharing measurements with G .

4.2. Probability of associations

Each association θ_j is assigned a probability mass which takes into account both the dynamic information given by the model (2.3) and the group constraint. We also assume prior knowledge on the range of variation over time of the cardinality N of a generic group. The mass which takes into account the dynamic information is chosen proportional to the gaussian density of the measured position of the centroid:

$$m_d(\theta_j) = k_d \cdot p(\xi(Y_{G,\theta_j}) | \hat{y}_c, P_{y_c})$$

where \hat{y}_c and P_{y_c} are the predicted position of the centroid and the corresponding error covariance given by the Kalman estimator and $\xi(Y_{G,\theta_j})$ is the centroid of the measurements Y_{G,θ_j} .

The mass on the number of measurements is assumed being of gaussian form as well:

$$m_N(\theta_j) = k_N \cdot e^{-\frac{1}{2} \cdot \left(\frac{|Y_{G,\theta_j}| - \hat{N}}{\lambda_N} \right)^2}$$

where \hat{N} is the estimated number of measurements, $|Y_{G,\theta_j}|$ is the cardinality of Y_{G,θ_j} and λ_N is chosen based on the expected variability of the number of measurements in a group.

Finally, in order to have state estimates which actually represent groups, the updating process should consider only configurations of measurements satisfying the group constraint which we take into account through the following mass:

$$m_g(\theta_j) = k_g \cdot p_g(Y_{G,\theta_j})$$

where p_g is given by:

$$p_g(Y_{G,\theta_j}) = \begin{cases} 1 & \text{if } |\mathcal{G}_y(Y_{G,\theta_j})| = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

The overall mass for a given association θ_j is chosen proportional to the product of the single factors introduced above:

$$m_{tot}(\theta_j) \propto m_d(\theta_j) \cdot m_N(\theta_j) \cdot m_g(\theta_j) \quad (4.2)$$

which takes into account dynamics, cardinality and the group constraint.

Note that configurations of measurements that do not satisfy the group constraint are given a null mass. In particular, if all associations for a group are given a null mass for a certain number of consecutive frames the group is eliminated and the measurements are re-grouped and initialized as described in section 3.3.

4.2.1. Computational efficiency. Verifying the group constraint for each set Y_{G,θ_j} involves the computation of $\mathcal{G}_0(Y_{G,\theta_j})$ which clusters many measurements into connected components. $\mathcal{G}_0(Y_{G,\theta_j})$ can be efficiently computed as follows. The set Y_{G,θ_j} can be written in the following way:

$$Y_{G,\theta_j} = Y_{G,\theta_{j-1}} \cup \{y_j\}$$

where y_j is the j -th sorted measurement. Assume $\mathcal{G}_0(Y_{G,\theta_{j-1}})$ has been computed:

$$\mathcal{G}_0(Y_{G,\theta_{j-1}}) = \{Y_{G_1,\theta_{j-1}}, \dots, Y_{G_M,\theta_{j-1}}\}$$

Let's indicate with I_R the indices of the connected components $Y_{G_i,\theta_{j-1}}$ which contain at least one element related to y_j :

$$I_R \triangleq \{i \mid \exists y \in Y_{G_i,\theta_{j-1}} \text{ y } \mathcal{R}_0 y_j\}$$

and with $\overline{I_R}$ the remaining indices $\overline{I_R} \triangleq \{1, \dots, M\} - I_R$. The set Y_{G,θ_j} can be partitioned as follows:

$$\mathcal{G}_0(Y_{G,\theta_j}) = \{\{Y_{G_i,\theta_{j-1}}\}_{i \in \overline{I_R}}, \{y_j, \bigcup_{i \in I_R} Y_{G_i,\theta_{j-1}}\}\}$$

The computation of the centroid of Y_{G,θ_j} is efficiently performed as well starting from the centroid of $Y_{G,\theta_{j-1}}$.

Given the predictions at time k , \forall group G :

- Y_G are the measurements in the search region (2.2) of G . $Y_{G,s} \subset Y_G$ and $Y_{G,ns} \subset Y_G$ are the shared and unshared measurements respectively
- The measurements $Y_{G,s}$ are sorted in ascending order according to their Mahalanobis distance from the centroid of G yielding the set $\{y_1, \dots, y_{N_s}\}$
- $Y_{G,\theta_0} \triangleq Y_{G,ns}$
- $\forall j = 0, \dots, |Y_{G,s}|$
 - $Y_{G,\theta_j} = Y_{G,\theta_{j-1}} \cup \{y_j\}$ is associated to G
 - The conditional state estimate \hat{x}_{k,θ_j} is computed (sec. 4.3)
 - $m_{tot}(\theta_j)$ is computed through (4.2) and in particular $p_g(Y_{G,\theta_j})$ is computed through (4.1)
- If $m_{tot}(\theta_j) = 0 \forall j$ the group G **splits** into new groups obtained through the initialization process (sec. 3.3) applied to the set Y_G
- else the overall state estimate for group G is computed:

$$\hat{x}_k = \sum_j \hat{x}_{k,\theta_j} \cdot m_{tot}(\theta_j)$$

We have now a collection of groups C_G along with their state estimates. The **merging** process is applied:

- $\mathcal{G}_x(C_G) = \{\{G_i\}_{i \in I_1}, \dots, \{G_i\}_{i \in I_M}\}$
- $\forall j = 1, \dots, M \quad \bigoplus_{i \in I_j} G_i = G_{m,j}$ (eq. 3.1)
- For each new merged group $G_{m,j}$ the predictions for time $k + 1$ are computed

Figure 1. Summary of the algorithm.

4.3. State estimation

Given an association θ , the conditional state estimate of a group G is obtained updating the Kalman filter using the measured position of the centroid $\xi(Y_{G,\theta})$. The cardinality $|Y_{G,\theta}|$ and the sample covariance $cov(Y_{G,\theta})$ are used in (2.4).

The overall estimate is obtained computing the weighted sum of the single conditional estimates with weights being the masses of the corresponding associations as is done in the PDAF.

The overall data association and estimation process is summarized in figure 1.

5. Experiments

We tested our algorithm in video sequences of groups of people merging and crossing in an outdoor scene. The pixel coordinates of these groups are obtained at each frame through background subtraction techniques. Experiments demonstrate that taking into ac-

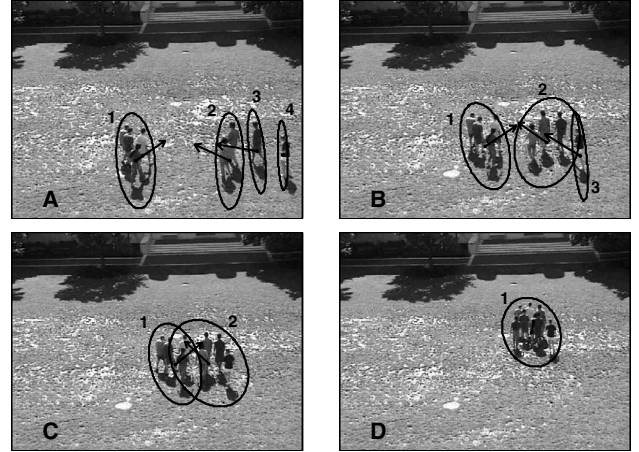


Figure 2. Groups of people merging into a single group.

count velocity improves in general the tracking performances in situations in which groups cross each other. In section 5.2.1 it is shown how the data association methodology proposed in section 4 contributes in obtaining good tracking results.

5.1. Groups merging into a single group

In the first video sequence groups of people are approaching and merging together. Four frames of the sequence are reported in figure 2. Ellipsoids and arrows represent search regions and velocities of groups respectively. In frame A, three groups are coming from the lower right corner and approaching group 1 coming from the opposite side. Groups 2 and 3 in frame A which are closely spaced and moving with similar velocities are merged together into group 2 of frame B. In the same way, groups 2 and 3 of frame B merge into group 2 of frame C. People of groups 2 and 3 in frame C are changing direction in order to move together as a single group. The estimated velocities take few seconds to change accordingly. At that point the groups are merged together as shown in frame D.

5.2. Groups crossing each other

In the second video sequence groups of people are moving along opposite directions and crossing each other. We report four frames of the sequence in figure 3. In frame A group 1 is moving toward the upper right corner while groups 2 and 3 are coming from the opposite direction. In frame B groups 2 and 3 are crossing group 1. The algorithm fails in tracking group 3, which

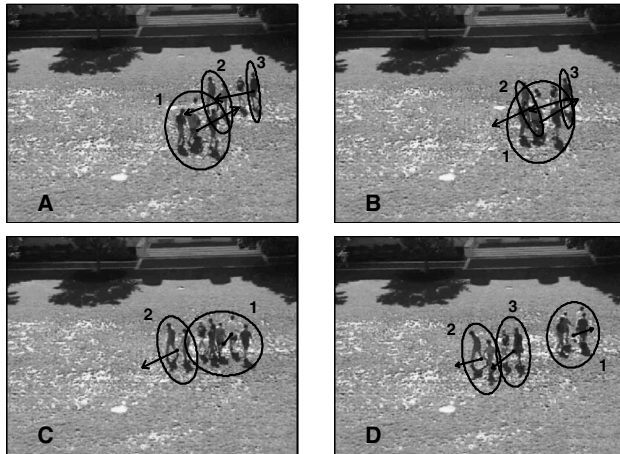


Figure 3. Groups of people moving along opposite directions and crossing each other.

is eliminated while it is passing through group 1 (see frame C). On the other hand the algorithm succeeds in tracking group 2. Taking into account the velocity is in general advantageous in dealing with such challenging situations. In fact, although group 1 and group 2 are partially overlapped, the difference in the corresponding velocities prevents them from being merged together and when the crossing phase is over, groups 1 and 2 are still tracked reliably (see frame C). In frame D, group 3 is tracked again when splits apart from group 1.

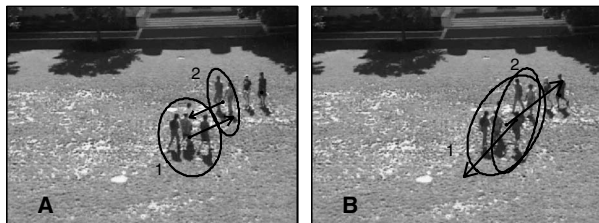


Figure 4. Groups of people moving along opposite directions and crossing each other. No proper data association is performed resulting in inferior tracking performances.

5.2.1. Data association. Data association techniques are crucial in obtaining good tracking performances. We processed the video sequence of groups crossing each other eliminating the data association step, i.e. assigning all the validated measurements

Y_G to the group G . A couple of key frames are reported in figure 4. In frame A groups 1 and 2 start crossing. During this interaction, all the measurements generated by group 2 and being in the search region of group 1 are assigned to group 1 and viceversa. This results in group 1 engulfing all objects of group 2. The same happens to group 2 (see frame B).

6. Conclusions and future work

In the present work we introduced a formal definition of group and a general merging process defined in the phase space. The data association process dynamically segments the measurements at pixel level on the basis of the predictions and it is properly modified in order to embed the additional group constraint. The algorithm has been tested on real data yielding good tracking results in challenging situations such those of groups crossing each other. The approach is promising and it can be further improved. Integration of additional features such as pixels color and online identification procedures will be considered.

References

- [1] Y. Bar-Shalom and X.R. Li, "Multitarget-Multisensor Tracking: Principles and Techniques", YBS Publishing, 1995.
- [2] S. S. Blackman, "Multiple Target Tracking with Radar Applications", Artech House, Norwood, MA, 1986.
- [3] Stephen J. McKenna, "Tracking groups of people", *Computer vision and Image Understanding*, Vol. 80, 42-56, 2000.
- [4] Proc. of 1st Int. Workshops on Performance Evaluation of Tracking Systems, PETS 2000.
- [5] Proc. of 2nd Int. Workshops on Performance Evaluation of Tracking Systems, PETS 2001.
- [6] 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.
- [7] T. KirubaraJan, Y. Bar Shalom, K.R. Pattipati, "Multi-assignment for Tracking a Large Number of Overlapping Objects", *IEEE Trans. on Aerospace and Electronic Systems*, Vol.37, No.1, Jan. 2001.