ENERGETIC PARTICLE FILTER FOR ONLINE MULTIPLE TARGET TRACKING

Abir El Abed [1], *Séverine Dubuisson* [1], *Dominique Béréziat* [2]

[1] Laboratoire d'Informatique de Paris 6, Université Pierre et Marie Curie, 104 avenue du Président Kennedy, 75016 Paris, France

[2] LIP6/UPMC, Clime project/INRIA Rocquencourt, B.P. 105 78153 Le Chesnay Cedex, France

abir.elabed@lip6.fr

ABSTRACT

Online target tracking requires to solve two problems: data association and online dynamic estimation. Usually, association effectiveness is based on prior information and observation category. However, problems can occur for tracking quite similar targets under the constraints of missing data and complex motions. The lack in prior information limits the association performance. To remedy, we propose a novel method for data association inspired from the evolution of target's dynamic model and given by a global minimization of an energy. The concept amounts to measure the absolute geometric accuracy between features. The main advantage of our approach is that it is parameterless. We also integrate our method into the classical particle filter, that leads to what we call the Energetic Particle Filter (EPF).

Index Terms— Energy minimization, particle filter, missing data, online tracking, adaptive parametrization.

1. INTRODUCTION

Usually, association effectiveness is based on prior information and observation category. If we have a lack in prior information, the association task becomes difficult. Such case can occur when the observed system is deformed over time, moreover, if no information about motion is available when we track multiple and quite similar (even non distinguishable) objects. Difficulties increase if there is a considerable interval of time between observations and if the observer has no prior information about the motion. Likewise, if we only observe target positions, it is possible for a measurement to be equidistant from several targets: all target association probabilities are relatively the same and it is difficult to associate the good measurement with the good target. As far as, no association method can handle all the cases illustrated previously. The literature contains some classical approaches for data association: the deterministic approaches and the probabilistic ones. The simplest deterministic method is the Nearest-Neighbor Standard Filter (NNSF) [1] that selects the closest validate measurement to a predicted target. In some tracking applications, the color is also exploited and the histogram intersection technique is used to compute the color histogram difference between measurements and targets. Unfortunately, the color metric is not sufficient in many cases: for deformable objects, which color distribution may differ from one frame to another, or in case of several quite identical objects. Probabilistic approaches are based on posterior probability and make an association decision using the probability error. We can cite the most general one, called Multiple Hypothesis Tracking [2], for which multiple hypothesis are formed and propagated, implying the calculation of every possible hypothesis. Another strategy for multiple target tracking is the Joint Probability Data Association (JPDA) [3] which uses a weighted sum of all measurements near the predicted state, each weight corresponding to the posterior probability for a measurement to come from a target. JPDAF provides an optimal data solution in the Bayesian framework. However, the number of possible hypothesis increases rapidly with the number of targets.

In this paper, we propose a novel method for data association where a measurement is associated to a target if the amplitude of its total energy is minimized. The main advantages of this energy are it is parameterless and does not require prior knowledges. Only one information about a target is used: its position. We integrate this association method in the classical particle filter to achieve the task of multiple object tracking. Furthermore, the parameters of the dynamic model are estimated in an adaptive and automated way by using a cubic B-Spline interpolation [4]. This combination of the particle filter with the energy minimization approach, whose dynamic parameters are updated online each time an observation is available, leads to the Energetic Particle Filter (EPF). The outline of this paper is as follows. In section 2, we expose the energy minimization approach, derive its geometrical representation and its mathematical model. In section 3, we develop the EPF algorithm for multiple target tracking. The proposed method is then tested on several sequences in section 4. Finally, concluding remarks and perspectives are given in section 5.

2. ENERGY MINIMIZATION APPROACH

We observe a video sequence describing a dynamical scene using a sensor which can deliver exactly one observation at instant t. It contains at least one measurement which can be associated with a specific object or can be a false alarm. Let's set $y = (y_1, ..., y_M)$, where y is the vector containing the M measurements, also called observation. Our goal is to associate one measurement per target.

In this paper, we propose an algorithm for data association restricted to one category of measurement: the position. Furthermore, we affirm the total lack of prior information concerning targets: exclusively the two anterior predicted positions are used as input for our algorithm. We will first give the concept of our approach before starting its mathematical modeling. If we only consider the linear translation in one direction, the data association problem is limited to the computation of the Mahalanobis distance energy. Thus, in case of complex dynamics such as non linear displacements, oscillatory motions or non-constant velocities, we incorporate a second energy

which measures the absolute accuracy between the dynamic features and indicates how much their parameters are close. Moreover, we distinguish some dynamic cases, that will be clarified by geometric descriptions afterward, when we need to improve the data association problem by adding a third energy, also called the proximity energy. We define the energy between the k^{th} target and the j^{th} measurement y_j by: $E(k,y_j) = \frac{1}{\sqrt{3}}\sqrt{\sum_{l=1}^{3}\alpha_l^2\left(E^l(k,y_j)\right)^2}$, where $\alpha_l = \frac{1}{\sum_{k=1}^{K} E^l(k,y_j)}$ is a weighted factor introduced to sensibly emphasize the relative importance attached to the energy quantities E^l. Before interpreting each energy, we will explain some notations. If we consider a target A and a measurement y_j available at instant t, we call (see Figure 1 for illustrations):

• $\hat{A}(t-2)$, $\hat{A}(t-1)$, $\hat{A}(t)$ and $\hat{A}_1(t+1)$: predictions of A at $t-2$, $t-1$, t and $t+1$ using the initial dynamic model;

• $\hat{A}_2(t+1)$: prediction of A at $t+1$ using the updated dynamic model. In this case, y_j is associated with A at instant t and the parameters of the dynamic model are updated according to y_j.

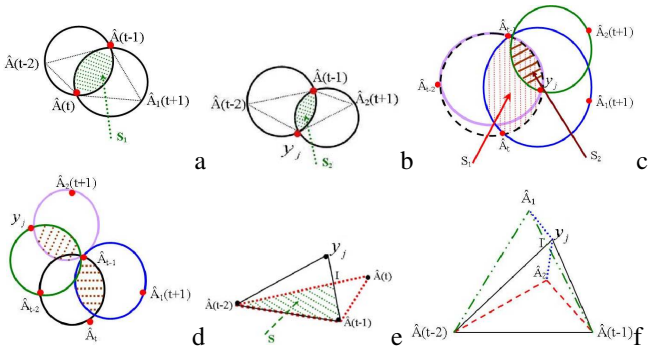

Fig. 1. (a-b-e) Intersection surfaces $\{S_1, S_2, S\}$; (c-d) Difference between the surfaces S_1 and S_2 extracted from two dynamical models; (f) Intersection surfaces when two predictions at instant t, $\hat{A}_1$ and $\hat{A}_2$, are equidistant from y_j.

The components of $(E(k,y_j)$ are:

1. The Mahalanobis distance energy, $E^1(k,y_j)$, measures the distance between a measurement y_j available at t and the prediction of A at $(t-1)$. This energy is sufficient if the motion is limited to translations in one direction (case of linear displacements).

$$E^1(k,y_j) = \sqrt{(y_j - \hat{A}(t-1))^T \hat{\Sigma}_k^{-1} (y_j - \hat{A}(t-1))}$$

$\hat{\Sigma}_k$: covariance matrix of target k (designed by A in the equation).

2. To consider the case of complex dynamics, such as oscillatory motions or non-constant velocities, we have added the absolute accuracy evolution energy $E^2(k,y_j)$. It introduces the notion of the geometric accuracy between two sets of features whose dynamic evolutions are different:

• The not updated dynamic model predicts a new state at $(t+1)$ without considering the presence of any measurement, *i.e.* without updating the parameters of the dynamic model.

• The updated dynamic model considers the association between measurements and target at t to update its parameters and then predict new states at $(t+1)$.

$E^2(k,y_j)$ extends a numerical estimation of the closeness between two dynamic models. Our idea aims to evaluate the parameters of the dynamic model in two steps if the measurement y_j arises from a target or no. We first predict the states $\hat{A}_1(t+1)$ and $\hat{A}_2(t+1)$ of the target. We then determine S_1, the intersection surface between the two circumscribed circles of triangles $(\hat{A}(t-2), \hat{A}(t-1), \hat{A}(t))$ and $(\hat{A}(t-1), \hat{A}(t), \hat{A}_1(t+1))$, and S_2, the intersection surface between the two circumscribed circles of triangles $(\hat{A}(t-2), \hat{A}(t-1), y_j)$ and $(\hat{A}(t-1), y_j, \hat{A}_2(t+1))$ (see Figures 1.(a-b)).

$E^2(k,y_j) = |S_1 - S_2|$ is minimized when the similarity between both dynamic models is maximized.

A question might be asked: is the component E^2 able to handle all type of motions?

Indeed, E^2 evaluates a numerical measure of similarity between dynamic models. This measurement depends on the difference between two surfaces. It is considered as reliable if both positions $\hat{A}(t)$ and y_j are on the same side comparing to axis $(\hat{A}_{t-2}\hat{A}_{t-1})$, see Figure 1.c. In Figure 1.d, we show the case where both surfaces S_1 and S_2 are quite similar, which implies that E^2 is null. This case can occur when the position of $\hat{A}(t)$ and y_j are diametrically opposite or in different side comparing to axis $(\hat{A}_{t-2}\hat{A}_{t-1})$. In such cases, the energy is not a sufficient information source to achieve the task of association. To compensate this energy, we incorporate a third energy E^3.

3. The proximity evolution energy $E^3(k,y_j) = \frac{1}{S}$ evaluates the absolute accuracy between the prediction $\hat{A}(t)$ and the measurement y_j at instant t, and corresponds to the inverse of the surface S defined by the intersection of the triangles $(\hat{A}(t-2), \hat{A}(t-1), y_j)$ and $(\hat{A}(t-2), \hat{A}(t-1), \hat{A}(t))$ (see the dotted area of Figure 1.e). Increasing S means that the prediction and the measurement at instant t are close.

Another question could be asked: why using the intersection surface instead of only calculating the distance between the measurement y_j and the prediction of target's position at instant t? In Figure 1.f, we have two predictions at instant t, $\hat{A}_1$ and $\hat{A}_2$ that are both equidistant from the measurement y_j. If we compute the proximity energy by only measuring distances, we will get that both models have the same degree of similarity with the initial model defined by the dynamic model of points $(\hat{A}(t-2), \hat{A}(t-1), y_j)$. This result leads to a contradiction with the reality. This problem can be explained by the fact that if they have both the same degree of similarity with the third dynamic model, we can conclude that their corresponding targets have the same dynamic. For this reason, we have chosen to evaluate the similarity by computing the intersection surface between triangles. We can remark in Figure 1.f that these intersection surfaces are very different, which leads to a different measure in the degree of similarity. Finally, y_j is associated to target k if its energy magnitude is minimized:

$$\mathcal{D}_{y_j \to k} = \left\{ \min_{k=1,\ldots,K} \left(\frac{1}{\sqrt{3}} \sqrt{\sum_{l=1}^{3} \alpha_l^2 E^l(k,y_j)^2} \right) \right\} \qquad (1)$$

with $0 \leq \alpha_l \leq 1$ and $0 \leq E(k,y_j) \leq 1$.

We have described a novel approach for data association based on the minimization of an energy magnitude whose components are extracted from geometrical representations constructed with measurements, previous states and predictions. The purpose of choosing a geometrical definition for these energies refers to:

• show the geometrical continuity of the system between predictions and previous states using two different dynamic models;

• measure the similarity between predictions, at a particular time for the same object, using two different dynamic models, that logically must be quite similar because they represent the same system.

3. ENERGETIC PARTICLE FILTER (EPF)

Given a video sequence depicting K moving targets, the tracking consists in estimating their states $X_{k,t}$ in frame t. Particle filter is an inference process which can be considered as a generalization of the Kalman filter [5]. It aims at estimating the unknown states $X_{k,t}$ from a set of noisy observations that occur sequentially, $Y_{1:t} = (y_1, ..., y_t)$. The two important components of this approach are the state transition and observation models whose most general forms can respectively be given by $X_{k,t} = F_{k,t}(X_{k,t-1}, U_{k,t})$ and $Y_{k,t} = G_{k,t}(X_{k,t}, V_{k,t})$. $U_{k,t}$ is the system noise, $F_{k,t}$ the kinematics of the k^{th} target, $V_{k,t}$ the observation noise, and $G_{k,t}$ the observation model. In this paper, we combine the particle filter with the energy minimization approach which leads to the Energetic Particle Filter (EPF). This filter is a global framework for multiple target tracking under the restrictions of lack in prior information about motion and measurements only corresponding to positions. The algorithm of EPF proceeds as follows:

1. Compute $E(k, y_j)$ for target candidates so that y_j falls in their validation regions. Associate y_j with k if $E(k, y_j)$ is minimized. If y_j is not included in any validation region, consider it as clutter. The validation region is an ellipsoid that contains a given probability mass under the Gaussian assumption. Its center is the mean of the target. The minor and major axes are respectively given by the largest and smallest eigenvalues of the covariance matrix and their directions by the corresponding eigenvectors.
2. If y_j is generated by target k, estimate the new parameters of the dynamic model by using the cubic B-spline which control points are extracted from the available observations [4].
3. For $k = 1...K, n = 1...N$, initialize the samples for target states $S_{k,0} = \left\{ s_{k,0}^{(n)}, w_{k,0}^{(n)} \right\}$ where $s_{k,0}^{(n)} \sim P(X_{k,0})$ and $w_{k,0}^{(n)} = \frac{1}{N}$.
4. Generate new samples $\tilde{s}_{k,t}^n$ from $P(X_{k,t}|X_{k,t-1} = s_{k,t-1}^n, y_j)$.
5. Compute and normalize the particle weights:
$\tilde{w}_{k,t}^n = w_{k,t-1}^n \frac{P(\tilde{s}_{k,t}^n | s_{k,t-1}^n) P(y_t | \tilde{s}_{k,t}^n)}{f(\tilde{s}_{k,t}^n | s_{k,t-1}^n, y_t)}$ and $w_{k,t}^n = \frac{\tilde{w}_{k,t}^n}{\sum_{n=1}^N \tilde{w}_{k,t}^n}$.
6. Approximate the density function by $\sum_{n=1}^N w_{k,t}^n \delta_{s_{k,t}^n}$.

It has been shown that the variance of the particle weights always increases over time, which causes a weight degeneracy. To reduce this effect, we use the multinomial resampling approach [6] to resample the particle weights in an adaptive way when their effective number estimated by $N_{\text{eff}} = \frac{1}{\sum_{n=1}^N (w_{k,t}^n)^2}$ is under a given threshold.

4. EXPERIMENTAL RESULTS

Van-Plane test: Two measurements $\{M_1, M_2\}$ are available at t, each one is a position in the target space. The first row in Figure 2 contains the frames at $\{t-2, t-1, t\}$ in which two targets $\{T_1, T_2\}$, the van and the plane, are present. If we look at the position of these measurements on the frame at t (right image in Figure 2), we observe that M_1 is closer to T_1 and M_2 to T_2. In the second and third rows, we show the predictions of both targets by evaluating the two different dynamic models. We point out that the horizontal and the vertical axis of the frame are represented by the y-axis and x-axis. To associate these measurements, we first compute the Mahalanobis distance energy and get $\{\alpha_1 E^1(T_1, M_2) = 0.32\} < \{\alpha_1 E^1(T_2, M_2) = 0.68\}$ which means that the distance from M_2 to T_1 is smallest than the one from M_2 to T_2. Hence, the NNSF association method would associate M_2 to T_1 which is a contradiction with the reality. To remedy, we compute the energies E^2 and E^3 and obtain $E(T_1, M_1) < E(T_2, M_1)$ and $E(T_2, M_2) < E(T_1, M_2)$, implying that the energies magnitude are minimized if M_1 and M_2 are respectively associated to T_1 and T_2 (see equations below).

$$
\begin{aligned}
E(T_1, M_1) &= \frac{1}{\sqrt{3}} \sqrt{(0.31)^2 + (0.02)^2 + (0.34)^2} = \underline{\underline{0.27}} \\
E(T_1, M_2) &= \frac{1}{\sqrt{3}} \sqrt{\underline{(0.32)^2} + (0.72)^2 + (0.86)^2} = 0.68 \\
E(T_2, M_1) &= \frac{1}{\sqrt{3}} \sqrt{\underline{(0.69)^2} + (0.98)^2 + (0.66)^2} = 0.79 \\
E(T_2, M_2) &= \frac{1}{\sqrt{3}} \sqrt{\underline{(0.68)^2} + (0.28)^2 + (0.14)^2} = \underline{\underline{0.43}}
\end{aligned}
$$

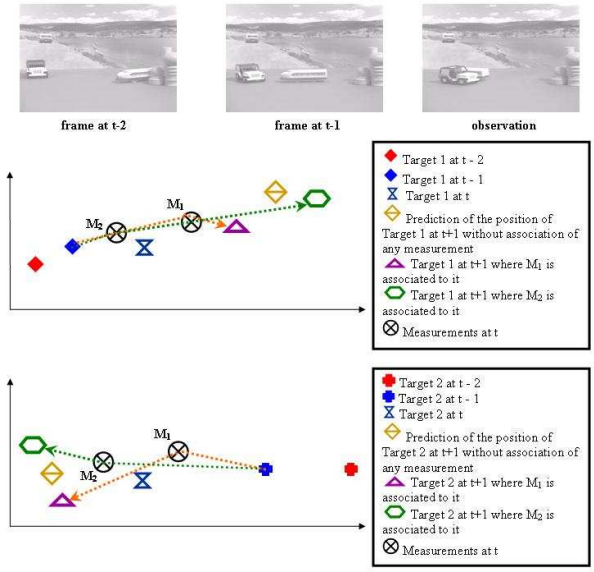

Fig. 2. Van Plane test. First row: frames at $\{t-2, t-1, t\}$. The 2nd and 3rd rows show the positions of both targets at different instants.

Fig. 3. Walking men test. Frames at $\{t-2, t-1, t, t+1\}$.

Walking men test: Figure 3 shows from left to right the frames at $\{t-2, t-1, t, t+1\}$. At $t-2$, three men are walking close. At $t-1$, two men (T_2 and T_3) continue walking in the same direction, while the third one (T_1) takes the opposite direction. The available observation at t contains three measurements corresponding to positions in the target space. At t, we observe a partial occlusion between two walking men. We notice that the observed positions of the cross men are very close. At $t+1$, these men change their directions.

	M_1	M_2	M_3	$E(T_i, M_j)$		
T_1	0.144 0.03 0	0.70 0.16 0.5	0.63 0.0037 0	<u>0.085</u>	0.51	0.363
T_2	0.41 0.82 0	<u>0.15</u> 0.22 0.5	0.21 0.76 0	0.53	<u>0.32</u>	0.455
T_3	0.444 0.16 0	<u>0.15</u> 0.62 0	0.16 0.235 0	0.47	0.37	0.16

Table 1. Each column represents the computation of energy components when a measurement M_j is associated to a target T_i. The second part of this table shows their energy magnitude.

We remark from Table 1 that the measurement M_2 is equidistant from targets T_2 and T_3 ($\alpha_1 E^1(T_2, M_2) = \alpha_1 E^1(T_3, M_2) = 0.15$). We also remark that most of time the third energy is null, this effect is due to the presence of a linear displacement (motion limited to a displacement in the directions x and y). Once the energies are computed, the energy magnitude $E(T_i, M_j)$ is minimized when measurement M_j is associated to target T_i, see the column of $E(T_i, M_j)$ in Table 1. Despite the change in illumination, the measurements were correctly associated to targets by using the approach of energy minimization.

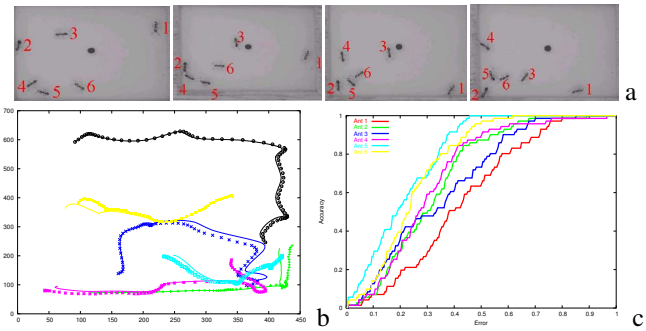

Fig. 4. Ant sequence. (a) Frames at $\{t-2, t-1, t, t+1\}$; (b) Tracking result obtained with EPF; (c) REC curves.

Ant sequence test: In this sequence, the ants are quite similar even non-distinguishable and characterized by the same gray level distribution. The sensor, at t, provides an observation containing six measurements corresponding to positions in the (x, y) space. In such scene, only the motion information is used. We remark from Figure 4 that their displacement is erratic with non-constant velocity. They change their direction, accelerate, decelerate, stop moving, do rotation around their axis, *etc*. Figure 4.a shows the acquisitions at $\{t-2, t-1, t, t+1\}$ corresponding to frames $\{10, 25, 35, 45\}$. Notice that the frame at t is the available observation. To track these ants under the assumption of missing data, we only use five frames from the sequence as observations without having any prior knowledge about the motion or the trajectory in the interval of time that separates two successive observations.

	M_1	M_2	M_3	M_4	M_5	M_6	$E(T_i, M_j)$					
T_1	6.5	47.2	25.8	43.9	48.7	46.6	3.9	38.1	14.9	26.7	40.1	36.7
	1.5	1.8	1.4	1.1	11.3	1.1						
	0.03	45.8	0.8	14.1	48.1	43.2						
T_2	22.5	5.8	21.4	6.5	2.5	12.1	13.7	3.4	18.9	6.7	32.6	10.5
	3.1	1.2	3.1	9.6	54.5	1.6						
	6.8	0.01	24.6	0.2	14.6	13.6						
T_3	15.1	24.9	4.1	17.7	23.7	19.7	17.5	15.2	2.6	10.8	14	18.1
	14.2	0.2	0.3	0.2	1.3	0.1						
	22.2	8.5	1.8	5.6	4.7	24.5						
T_4	21.4	1.7	21.6	9.4	5.4	10.2	25.9	8.7	20.4	5.4	17.6	9.5
	6.6	2.2	2.7	0.3	24.1	4.4						
	38.8	14.8	27.8	0.7	17.8	12.1						
T_5	18.3	6.1	17.9	12.2	9.3	5.9	44.4	50.1	54.7	56.1	7.1	53.6
	74.4	85.8	92.2	96.2	6.9	92.4						
	8.5	10.5	12.6	3.2	4.05	6.2						
T_6	16.1	13.5	9.3	10.9	10.3	5.5	16.6	13.9	19.4	44.5	8.7	3.2
	0.3	0.25	0.2	1.1	1.9	0.4						
	23.7	20.1	32.4	76.3	10.8	0.35						

Table 2. Computation of the energy components when a measurement M_j is associated to a target T_i and their energy magnitude.

Table 2 contains the numerical values of the energy components when a measurement M_j is associated to a target T_i. The NNSF method associates measurements $\{M_2, M_4, M_5\}$ respectively to observations $\{T_4, T_2, T_2\}$ which leads to a contradiction with the reality (see $\alpha_1 E^1(T_i, M_j)$ in Table 2). We remark from Table 2 that $\alpha_2 E^2(T_2, M_2) < \alpha_2 E^2(T_4, M_2)$ and $\alpha_3 E^3(T_2, M_2) < \alpha_3 E^3(T_4, M_2)$ which compensates the error given by $\alpha_1 E^1(T_2, M_2)$. Finally, $E(T_i, M_2)$ is minimized when M_2 is associated with target T_2. Lets take another example to show the necessity of using the energy E^3 in our formulation. If we only use the energies $\alpha_1 E^1(T_i, M_j)$ and $\alpha_2 E^2(k, M_i)$ to associate data, we will get $E(T_6, M_5) < E(T_5, M_5)$ and the measurement M_5 will be associated with target T_6 which is wrong. We can remark from Table 2 that $\alpha_3 E^3(T_5, M_5) < \alpha_3 E^3(T_6, M_5)$ which compensates the other energy error. Finally, we observe that each measurement is well associated with its corresponding target. We notice that our approach is not a time-consumer. In Matlab, the total time of computation of all these energies is 0.25 seconds. We apply the EPF to estimate their mean, $\hat{X}_t^i$, which represents their translation vector. The dotted lines in Figure 4.b show the estimated mean for each ant. Figure 4.c shows the Regression Error Characteristic (REC) curve for each ant: the error rate is on the x-axis and the accuracy is on the y-axis. Accuracy is defined as the percentage of points that are fit within the tolerance. The error here is defined as the difference between the actual value and its prediction. As the error increases, the accuracy increases. The accuracy goes to 1 when the error becomes large enough.

5. CONCLUSION

In this paper, we have proposed a new method for data association based on an energy minimization which can handle complex motions and highly non-linear systems, and deals with the lack in prior knowledge. The main advantages of our method is that it requires only one information, the position, and it is not a time consumer. The geometric illustration of energy components allows to measure the accuracy between two dynamic models and to define their degree of similarity. The integration of this energy minimization approach for data association in the particle filter, leads to our Energetic Particle Filter and contributes to a new framework for multiple object tracking. As a perspective, we suggest to add some energy components, as necessary, to handle the case of multimodal observations and to integrate them within the particle filter to build a tracking framework for multimodal observations.

6. REFERENCES

[1] L.X. Rong and Y. Bar-Shalom, "Tracking in clutter with nearest neighbor filter: analysis and performance," *IEEE transactions on aerospace and electronic systems*, 1996.

[2] J. Vermaak, S.J. Godsill, and P. Pérez, "Monte carlo filtering for multi-target tracking and data association," *IEEE Transactions on Aerospace and Electronic Systems*, 2005.

[3] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE Journ. Oceanic Engineering*, 1983.

[4] A. El Abed, S. Dubuisson, and D. Béréziat, "Comparison of statistical and shape-based approaches for non-rigid motion tracking with missing data using a particle filter," *Lecture Notes in Computer Science, Advanced Concepts for Intelligent Vision Systems*, vol. 4179/2006, pp. 185–196, September 2006.

[5] A. Doucet, N. Gordon, and J.D. Freitas, "An introduction to sequential monte carlo methods," *Springer-Verlag*, 2001.

[6] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputation," *J.Amer. Statist. Assoc*, 1995.