Real-Time Distributed Multi-Object Tracking Using Multiple Interactive Trackers and a Magnetic-Inertia Potential Model

Wei Qu, *Member, IEEE*, Dan Schonfeld, *Senior Member, IEEE*, and Magdi Mohamed, *Associate Member, IEEE*

***Abstract*—This paper presents a method which avoids the common practice of using a complex joint state-space representation and performing tedious joint data association for multiple object tracking applications. Instead, we propose a distributed Bayesian formulation using multiple interactive trackers that requires much lower complexity for real-time tracking applications. When the objects' observations do not interact with each other, our approach performs as multiple independent trackers. However, when the objects' observations exhibit interaction, defined as close proximity or partial and complete occlusion, we extend the conventional Bayesian tracking framework by modeling such interaction in terms of potential functions. The proposed *"magnetic-inertia"* model represents the cumulative effect of virtual physical forces that objects undergo while interacting with each other. It implicitly handles the *"error merge"* and *"object labeling"* problems and thus solves the difficult object occlusion and data association problems in an innovative way. Our preliminary simulations have demonstrated that the proposed approach is far superior to other methods in both robustness and speed.**

***Index Terms*—Bayesian tracking, data association, multiple object tracking, object occlusion, particle filter.**

I. Introduction

MULTIPLE object tracking (MOT) in video sequences has been an intensive area of research due to its numerous applications including target identification, surveillance, video coding and communications. Most of the early efforts for MOT relied on the use of multiple trackers to track multiple objects—one tracker per object. Sophisticated Bayesian tracking methods have been adopted for MOT applications using techniques such as Condensation [1] and sequential Monte Carlo [2]. In the particle filtering implementation of this approach, for example, each tracker has its own particles which evolve within a single object state space. MOT for distinctive objects is much easier since the objects can be tracked independently by using multiple trackers. However, MOT of identical or similar objects in appearance (such as pedestrians) fails when the objects are in close proximity or present occlusions. In such circumstances, modeling the *interaction* among objects and solving the *data association* problem [3] (i.e., establishing a correspondence between objects and observations) are the most critical problems for MOT. Without an effective scheme to model the interaction among observations and solve the data association problem, multiple independent single object trackers suffer from the well-known *"error merge"* problem, where the tracker loses its associated object and falsely coalesces with other objects, as well as the classical *"object labeling"* problem, where trackers falsely switch labels between objects after occlusion.

A widely accepted approach to MOT that addresses many of the difficulties inherent in this complex task is based on a centralized solution. In particular, it introduces a joint state space representation which concatenates all of the objects' states together to form a large meta state. It infers the joint data association by characterization of all possible associations between objects and observations. Various methods have been developed for centralized MOT [3]–[6]. The joint probabilistic data association filter (JPDAF) [3] combines all measurements associated with the objects to form a joint observation model. In [4], the likelihood of each particle is calculated by enumerating all possible association hypotheses. MacCormick and Blake [5] use a binary variable to identify foreground objects in order to solve the association problem. BraMBLe [6] combines a multi-blob likelihood function with the Condensation filter. Several MOT particle filters have been proposed in a context where measurements are available from sensors such as radars [7]. Recently, Dellaert *et al.* [8] proposed an MCMC-based particle filter which uses a Markov random field to model motion interaction. Although this framework is very powerful, its current implementation cannot handle severe occlusions. An important characteristic shared by all of the above centralized trackers is that although they can handle, in principle, the *"error merge"* problem, they require a tremendous computational cost due to the complexity introduced by the high dimensionality of the joint state representation. The complexity of most implementations based on a joint state-space representation grows exponentially in terms of the number of objects tracked. A centralized tracker based on a joint state space representation is therefore unsuitable for real-time tracking applications. Indeed, in order to minimize the computational costs of a centralized tracker, most of these schemes invoke the joint state representation only when the objects are in close proximity to each other. This notion is carried further in the work of Yu *et al.* [9] which allows for collaboration among different trackers by modeling the joint prior of objects using a Markov Random Network to solve the *"error merge"* problem. This approach provides the

Manuscript received June 25, 2005; revised May 25, 2006. The associate editor coordinating the review of this paper and approving it for publication was Prof. Yi-Ping Phoebe Chen.

W. Qu is with Network Research COE, Motorola Labs, Schaumburg, IL 60196 USA (e-mail: wei.qu@motorola.com).

D. Schonfeld is with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: ds@ece.uic.edu).

M. Mohamed is with Physical Realization Research COE, Motorola Labs, Schaumburg, IL 60196 USA (e-mail: Magdi.Mohamed@motorola.com).

Color versions of Figs. 3 and 5–8 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2006.886266

most promising direction to resolve the problems with the high complexity of the joint state space representation. However, it does not deal with the data association between objects and observations or the labeling problem.

In this paper, we present a distributed multiple object tracking (DMOT) framework using multiple interactive trackers. This formulation discards the joint state representation which introduces high complexity and requires expensive computational costs. Instead, a novel conditional density propagation framework has been derived for each object by modeling the interaction among the objects' observations in a distributed scheme. Moreover, by estimating the *interactive likelihood* and the state transition density using a *"magnetic-inertia potential"* model in the particle filtering implementation, we implicitly handle the *"error merge"* and *"object labeling"* problems and thus solve the difficult object occlusion and data association problems in an innovative way. The paper is organized as follows: In Section II, we present the formulation of our DMOT framework. In Section III, we provide a particle filtering implementation based on the *"magnetic-inertia potential"* model. Computer simulations and experimental results are presented in Section IV.

II. Distributed Multi-Object Tracking Using Multiple Interactive Trackers

We used a four dimension parametric ellipse to model objects. The state of an individual object is denoted by $x_t^i = \left(cx_t^i, cy_t^i, a_t^i, \rho_t^i\right)$, where $i = 1, \ldots, M$ is the index of objects, t is the time index, (cx, cy) is the center of the ellipse, a is the major axis, and ρ is the orientation in radians. The ratio of the major and minor axis of the ellipse is kept constant equal to its value computed in the initialization. We also denote the image observation of x_t^i by z_t^i, the set of all states up to time t by $x_{0:t}^i$ where x_0^i is the initialization prior, the set of all observations up to time t by $z_{1:t}^i$. We further denote the interactive observations of z_t^i at time t by $z_t^{J_t}$ where $J_t = \{j_{l_1}, j_{l_2}, \ldots\}$. The elements $j_{l_1}, j_{l_2}, \ldots \in \{1, \ldots, M\}, j_{l_1}, j_{l_2}, \ldots \neq i$ are the indexes of objects whose observations interact with z_t^i. Empty set $J_t = \emptyset$ means there is no interaction of z_t^i with other observations at time t. Similarly, $z_{1:t}^{J_{1:t}}$ represents the collection of the interactive observation sets up to time t. Since the interactive relationship among observations is changing, J may also be different over time.

A. Analysis of the Difficulties of Multi-Object Tracking

When multiple objects move close or present occlusions, it is generally difficult for the trackers to segment and distinguish these spatially adjacent objects from image observations, namely, the interactive observations are not independent $p\left(z_t^1, \ldots, z_t^M\right) \neq \prod_{i=1}^M p\left(z_t^i\right)$. As a result, we cannot simply factorize the posteriors of different objects, $p\left(x_t^1, \ldots, x_t^M | z_t^1, \ldots, z_t^M\right) \neq \prod_{i=1}^M p(x_t^i | z_t^i)$. This conditional dependency of objects is the main reason why multiple independent trackers cannotcope with the *"error merge"* problem, let alone the *"object labeling"* problem.

In the implementation of multiple independent trackers, we notice the occurrence of the *"error merge"* problem in two cases: 1) when two objects move closer or begin to present occlusion, the object with the stronger observation (in the sense of a larger visual image) *"pulls"* the tracker of the object with weaker observation and 2) after occlusion, when two objects move apart, their trackers often cannot detach and remain bonded while tracking the object with stronger observation. In these situations, it seems that there is an invisible force among interactive trackers attracting them to merge together when objects move closer and preventing them from disjointing when objects move apart. This phenomenon has inspired us to seek a physical explanation to explain this behavior. We believe that classical *gravitation theory* can provide a good analogy to understand the underlying principle of the *"error merge"* problem. *Newton's Universal Law of Gravitation* [10] states that *"Any two objects attract each other by a gravitational force that is proportional to the product of their masses and inversely proportional to the square of the distance between them"*. In our scenario, the object's observation can be imagined as the associated tracker's *"mass"*. When objects are far apart, the *"gravitational force"* between their trackers is very weak and can be ignored. However, when objects are adjacent or occluded, this attractive force becomes very strong. This is called an *"interaction"*. Although the relative forces between two trackers are equal according to *Newton's Third Law* [10], when their *"masses"* are different, they have different accelerations by *Newton's Second Law* [10]. As a result, after several frames, the tracker with small *"mass"* (big acceleration) will be attracted to the merge with the object with bigger observation (small acceleration) and thus *"error merge"* happens. To resist the excessive attraction incurred by the *"gravitational force"* and thus handle the *"error merge"* problem, the above analogy has further inspired us to introduce a repulsive force between interactive trackers. When objects move closer, this force can prevent the trackers from falsely merging. When objects move away, it can help trackers detach from each other. *Magnetic field theory* [11] provides a suitable analogy that can be used to model such a force. We present our magnetic repulsion model later.

B. Graphical Modeling of Multi-Object Tracking

In Fig. 1, we illustrate our dynamic graphical model with two consecutive frames for multiple objects with interactive observations. It has two layers. The hidden layer has circle nodes that represent the states of objects x^i. The observable layer represents the observations z^i associated with the hidden states. The directed link between consecutive states associated with the same object represents the state transition density which is a Markov chain. Here, we release the usual *first-order Markov chain* assumption in regular Bayesian tracking approaches and allow higher order Markov chain for generality. The directed link from object x^i to its observation z^i represents the *"generative"* relationship and can be characterized by the local likelihood $p(z^i|x^i)$. The undirected link between observation nodes represents the interaction. The structure of the observation layer at each time depends on the spatial relations among objects' observations. If two objects' observations are sufficiently close or even occluded, an undirected link between them is constructed to represent the dependency.

C. Distributed Graphical Decomposition for Each Object and Conditional Independence Analysis

The graphical model like Fig. 1 is very complicated for analysis. Since we use multiple interactive trackers, one tracker per object, for MOT simultaneously, we further decompose the

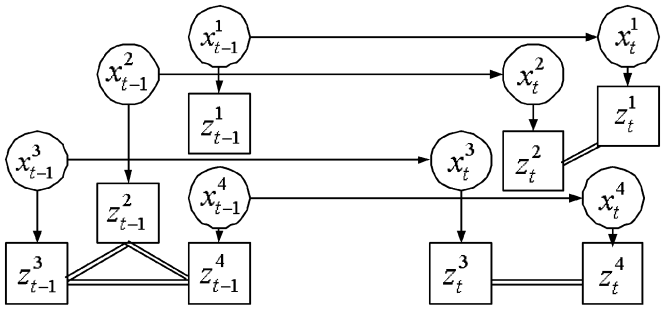


Fig. 1. Dynamic graphical model for multiple objects with interactive observations.

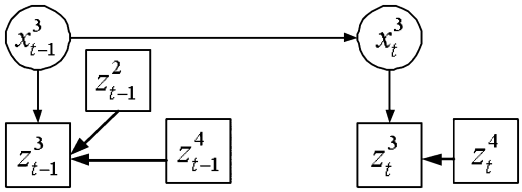


Fig. 2. Decomposition results for object 3 in Fig. 1.

graphical model of M objects into M submodels by three rules: 1) each submodel focuses on one object; 2) only the interactive observations which have direct links to the analyzed object's observation are kept; the noninteractive observation and all other objects' state nodes are removed; and 3) each undirected link between two interactive observations is decomposed into two directed links. The direction is from the other object's observation to the analyzed object's observation. Fig. 2 illustrates part of the decomposition results for Fig. 1. Although we neglect the temporal state correlation of certain interactive observations z^j when we consider object i, this information has been taken into account when we consider object j. Therefore, if we run all the trackers simultaneously, the decomposed submodels together can keep all the information (nodes, links) from the original model. The effectiveness of this simplification has been demonstrated by our experiments.

The decomposed graphs are all *Directed Acyclic Independence Graphs* [12, p. 56–82]. By exploiting the *Separation Theorem* [12, p. 67] to the associated *Moral Graphs* [12, p. 75], we can get the *Markov Properties* [12, p. 68–71], namely, the conditional independence of the decomposed graphs. It's easy to verify the following conditional independence properties: $p\left(z_t^i|x_{0:t}^i, z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right) = p\left(z_t^i|x_t^i, z_t^{J_t}\right)$, $p\left(z_t^{J_t}|x_t^i, x_{0:t-1}^i\right) = p\left(z_t^{J_t}|x_t^i\right)$, and $p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) = p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i\right)$.

D. Bayesian Formulation for Each Interactive Tracker

Now the problem is how to model the density propagation for each object. We estimate the posterior based on all involved observations $p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$. When z^i has no interaction with other objects' observations all the time, namely, $z_{1:t}^{J_{1:t}} = \emptyset$, $p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) = p\left(x_{0:t}^i|z_{1:t}^i\right)$. In this case, our formulation is consistent with the regular Bayesian tracker [1], [13].

Parallel with the derivation of *Sequential Importance Sampling Algorithm* [13], we formulate the density propagation for each interactive tracker as follows:

$$\begin{aligned} &p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) \\ &= \frac{p\left(z_t^i|x_{0:t}^i, z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right) p\left(x_{0:t}^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right)}{p\left(z_t^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right)} \\ &= \frac{p\left(z_t^i|x_t^i, z_t^{J_t}\right) p\left(x_{0:t}^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right)}{p\left(z_t^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right)}. \end{aligned} \tag{1}$$

In (1), we use the conditional independence property

$$p\left(z_t^i|x_{0:t}^i, z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right) = p\left(z_t^i|x_t^i, z_t^{J_t}\right).$$

Here, $p\left(z_t^i|x_t^i, z_t^{J_t}\right)$ is the *interactive* likelihood; $p\left(x_{0:t}^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right)$ is the *interactive* prior density. We further develop these two densities respectively as follows.

For the interactive likelihood, we have

$$p\left(z_t^i|x_t^i, z_t^{J_t}\right) = p\left(z_t^i|x_t^i\right) \frac{p\left(z_t^{J_t}|x_t^i, z_t^i\right)}{p\left(z_t^{J_t}|x_t^i\right)}. \tag{2}$$

Local likelihood $p\left(z_t^i|x_t^i\right)$ characterizes the *"gravitation"* between interactive observations. The mutual attraction in the gravitational model between interactive observations of different objects is the main cause for both the *"error merge"* and *"object labeling"* problems as discussed in Section II-A.

For the interactive prior density of $x_{0:t}^i$, we have

$$\begin{aligned} p\left(x_{0:t}^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right) &= \frac{p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i, z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)}{p\left(z_t^{J_t}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)} \\ &\quad \times p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \\ &= \frac{p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i\right)}{p\left(z_t^{J_t}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)} \\ &\quad \times p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \qquad (3) \\ &= \frac{p\left(z_t^{J_t}|x_t^i, x_{0:t-1}^i\right)}{p\left(z_t^{J_t}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)} p\left(x_t^i|x_{0:t-1}^i\right) \\ &\quad \times p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \\ &= \frac{p\left(z_t^{J_t}|x_t^i\right)}{p\left(z_t^{J_t}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)} p\left(x_t^i|x_{0:t-1}^i\right) \\ &\quad \times p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right). \qquad (4) \end{aligned}$$

In (3), we use the conditional independence property $p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) = p\left(x_t^i, z_t^{J_t}|x_{0:t-1}^i\right)$. In (4), we use the property that $p\left(z_t^{J_t}|x_t^i, x_{0:t-1}^i\right) = p\left(z_t^{J_t}|x_t^i\right)$.

By substituting the (4) and (2) back into (1) and rearranging the order, we have

$$\begin{aligned} & p\left(x_{0:t}^{i}|z_{1:t}^{i}, z_{1:t}^{J_{1:t}}\right) \\ &= p\left(z_t^i|x_t^i\right) p\left(x_t^i|x_{0:t-1}^i\right) p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \\ &\quad \cdot p\left(z_t^{J_t}|x_t^i, z_t^i\right) \\ &\quad \cdot \frac{1}{p\left(z_t^i|z_{1:t-1}^i, z_{1:t}^{J_{1:t}}\right) p\left(z_t^{J_t}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)} \qquad (5) \\ &= k_t p\left(z_t^i|x_t^i\right) p\left(x_t^i|x_{0:t-1}^i\right) p\left(x_{0:t-1}^i|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \\ &\quad \cdot p\left(z_t^{J_t}|x_t^i, z_t^i\right). \qquad (6) \end{aligned}$$

Densities in the denominator of (5) are unrelated with x^i and thus the fraction in the second line of (5) becomes a normalization constant k_t. In (6), $p\left(z_t^i|x_t^i\right)$ is the local likelihood, $p\left(x_t^i|x_{0:t-1}^i\right)$ is the state transition density, which are similar as the regular Bayesian tracking methods. The main novelty of our framework is that we introduce a new density $p\left(z_t^J|x_t^i, z_t^i\right)$ called an *"interactive likelihood"* to characterize the interaction among objects' observations. When not activating the interaction among objects' observations, our formulation will degrade to multiple independent particle filters [13]. This can easily be achieved by switching $p\left(z_t^J|x_t^i, z_t^i\right)$ to a uniform distribution.

III. Density Estimation Using Particle Filtering

Different density estimation methods [14] can be used to estimate the derived posterior. In this paper, we exploit a *sequential importance sampling method* [13] as the paradigm.

We denote $\left\{x_{0:t}^{i,n}, w_t^{i,n}\right\}_{n=1}^{N_s}$ as a random measure that characterizes the posterior density $p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$, where $\{x_{0:t}^{i,n}, n = 1, \ldots, N_s\}$ is a set of support particles with associated weights $\left\{w_t^{i,n}, n = 1, \ldots, N_s\right\}$. The weights are normalized so that $\sum_n w_t^{i,n} = 1$. Therefore, the posterior density at t can be approximated as

$$p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) \approx \sum_{n=1}^{N_s} w_t^{i,n} \delta\left(x_{0:t}^i - x_{0:t}^{i,n}\right) \qquad (7)$$

where $\delta(\cdot)$ is the Dirac delta measure. We thus have a discrete weighted approximation to the true posterior density $p\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$. The weights are chosen according to the *importance sampling* theory [15]. If the particles $x_{0:t}^{i,n}$ were drawn from an importance density $q\left(x_{0:t}^i|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$, then the corresponding weights in (7) are

$$w_t^{i,n} \propto \frac{p\left(x_{0:t}^{i,n}|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)}{q\left(x_{0:t}^{i,n}|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)}. \qquad (8)$$

Back to the sequential case, if the importance density is chosen to factorize such that

$$\begin{aligned} & q\left(x_{0:t}^{i,n}|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) \\ &= q\left(x_t^{i,n}|x_{0:t-1}^{i,n}, z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) q\left(x_{0:t-1}^{i,n}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right) \qquad (9) \end{aligned}$$

then we can obtain particles $x_{0:t}^{i,n} \sim q\left(x_{0:t}^{i,n}|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$ by augmenting each of the existing particles $x_{0:t-1}^{i,n} \sim q\left(x_{0:t-1}^{i,n}|z_{1:t-1}^i, z_{1:t-1}^{J_{1:t-1}}\right)$ with the new state $x_t^{i,n} \sim q\left(x_t^{i,n}|x_{0:t-1}^{i,n}, z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$

By substituting (6) and (9) into (8), the weight update equation can be shown to be

$$w_t^{i,n} \propto w_{t-1}^{i,n} \frac{p(z_t^i|x_t^{i,n}) p\left(x_t^{i,n}|x_{0:t-1}^{i,n}\right) p\left(z_t^{J_t}|x_t^{i,n}, z_t^i\right)}{q\left(x_t^{i,n}|x_{0:t-1}^{i,n}, z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)}. \qquad (10)$$

If $q\left(x_t^i|x_{0:t-1}^i, z_{1:t}^i, z_{1:t}^{J_{1:t}}\right) = q\left(x_t^i|x_{t-1}^i, z_t^i, z_t^{J_t}\right)$, the importance density is only dependent on $x_{t-1}^{i,n}$ and z_t^i, $z_t^{J_t}$. This is particularly useful in the common case when only a filtered estimate of $p\left(x_t^{i,n}|z_{1:t}^i, z_{1:t}^{J_{1:t}}\right)$ is required at each time. Moreover, if $p\left(x_t^{i,n}|x_{0:t-1}^{i,n}\right) = p\left(x_t^{i,n}|x_{t-1}^{i,n}, x_{t-2}^{i,n}\right)$, namely, a *"second-order"* state transition density, the current state is only dependent on previous two states. From this point on, we will accept the above assumptions. In such scenarios, only x_t^n, x_{t-1}^n, and x_{t-2}^n need to be stored. We can disregard the path $x_{0:t-3}^n$ and history of observations $z_{1:t-1}$. Then the modified weight is

$$w_t^{i,n} \propto w_{t-1}^{i,n} \frac{p\left(z_t^i|x_t^{i,n}\right) p\left(x_t^{i,n}|x_{t-1}^{i,n}, x_{t-2}^{i,n}\right) p\left(z_t^{J_t}|x_t^{i,n}, z_t^i\right)}{q\left(x_t^{i,n}|x_{t-1}^{i,n}, z_t^i, z_t^{J_t}\right)}. \qquad (11)$$

Modeling the densities in (11) is definitely very critical. We further present our *"magnetic potential"* model to estimate the *interactive likelihood* $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$ and the *"inertia potential"* model to estimate the state transition density $p\left(x_t^i|x_{t-1}^i, x_{t-2}^i\right)$.

A. Magnetic Potential Model

As discussed, we seek to model the interactive likelihood by introducing a *"repulsion force"* to resist excessive attraction among the interactive observations. We therefore use a model motivated by *Magnetic field theory* [11, p. 224–322] to provide a suitable framework.

Let's first consider a simple case where $z_t^{J_t} = \{z_t^j\}$. We assume that the two objects i and j are two magnetic monopoles with the same polarity. The analogy to the magnetic model is motivated by the repulsive magnetic field that can be used to represent a repulsive force between objects' observations. These assumptions are consistent with the former assumptions imposed in the graphical model, i.e., different objects' states (magnets in this case) at a certain time are independent while they interact with each other only through their observations (magnetic fields in this case). In this analogy, the interactive likelihood $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$ characterizes the mutual interaction between two magnetic fields. Pursuing this analogy further, just like in the concept of *potential difference* in magnetic theory [11, p. 224–322], the interactive likelihood $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$ is also related to the distance between the analyzed object's observation z_t^i and its interactive observation z_t^j. We must keep in mind, however, that the proposed analogy between the magnetic potential and interactive likelihood model represents

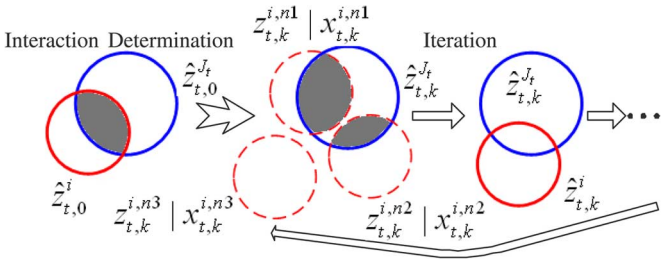


Fig. 3. Process of calculating *"magnetic weights"* in half of one iteration.

different entities. We observe that the magnetic potential difference is inversely proportional to the square of the distance between two points as captured by the *Biot-Savart Law* [11, p. 225–232]. Whereas, in the case of the interactive likelihood model $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$, we want the probability that two interactive objects remain in close proximity to be proportional to the squared distance between their interactive observations. Thus, we do not rely on the magnetic repulsion force directly, but simply use this as inspiration for the interactive likelihood model to counteract the attractive effect introduced by the local observation model. In particular, in the particle filter implementation, the density $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$ determines the *"magnetic weight"* for each particle. By assigning lower weights to particles that are closer to the reference observation z_t^j and higher weights to particles that are farther, the cumulative effect of this *"punish-close and reward-far"* scheme is equivalent to introducing a magnetic repulsion force between interactive observations. We define the *"magnetic weight"* for particle $x_t^{i,n}$ as

$$\varphi_t^{i,n}\left(z_t^{J_t}, z_t^i|x_t^{i,n}\right) = 1 - \frac{1}{\alpha_1}\exp\left\{-\frac{d_{i,n,t}^2}{\sigma_1^2}\right\} \tag{12}$$

where α_1 is a normalization constant, σ_1 is a prior constant which characterizes the allowable maximal interaction distance, and $d_{i,n,t}$ is the distance between the current particle's observation and the interactive observation z_t^j (for example, it can be the Euclidean distance $d_{i,n,t} = \|z_t^j - z_t^i|x_t^{i,n}\|$). For simplicity, we use a dissimilarity measure for representation of the distance based on the reciprocal of the area of the overlapping region (e.g., the shadow region illustrated in Fig. 3) in our experiments. Clearly, the larger the overlapping region, the smaller the dissimilarity measure between the interactive observations. The choice of the parameters α_1 and σ_1 needed to get a good function curve is currently obtained by trial-and-error. The larger the value of σ_1, the smaller the repulsion effect. The interactive likelihood can therefore be estimated as

$$p\left(z_t^{J_t}|x_t^i, z_t^i\right) = \varphi_t^i(\cdot) \approx \sum_{n=1}^{N_s} \frac{\varphi_t^{i,n}}{\sum_{n'=1}^{N_s}\varphi_t^{i,n'}}\delta\left(x_t^i - x_t^{i,n}\right). \tag{13}$$

Since usually it is very hard to distinguish the observation z_t^i from its interactive observations $z_t^{J_t}$ directly when they are in occlusion, which has an analog that the equilibrium between gravitation attraction and magnetic repulsion cannotachieve at a draught but after a process of oscillation, therefore, we need to recursively locate the interactive observations and iterate the repulsion process until it *"calms down"*. In Fig. 3, we illustrate half of one repulsion iteration cycle. The subscript $k = 1, \ldots, K$ represents the iteration time. The dash ellipses represent the particles while the solid ellipses represent the temporary estimates of objects' observations. At the beginning of iteration at time t, we first roughly estimate the observations' regions $\hat{z}_{t,0}^i$ and $\hat{z}_{t,0}^{J_t}$ using two independent trackers. If they have an overlapping area (shadow region), we determine that they are interacted and then trigger the recursive estimation. After that, each particle's observation of object i, $z_{t,k}^i|x_{t,k}^{i,n}$ is *"repelled"* by the temporary estimate $\hat{z}_{t,k}^j$ by calculating the *"magnetic weight"*. The weighted mean of all particles specifies the new temporary estimate of object i's observation $\hat{z}_{t,k}^i$. After that, we can similarly calculate the *"magnetic weight"* for object j's particles and thus estimate $\hat{z}_{t,k}^{j,n}$ to complete one iteration cycle. In our simulation, $K = 4 \sim 6$ is enough to guarantee a convergence.

When z_t^i has two interactive observations, $z_t^{J_t} = \{z_t^{j_1}, z_t^{j_2}\}$, it should be *"repelled"* by the other two simultaneously. Therefore, we have to revise (12) to be

$$\varphi_t^{i,n}(\cdot) = \left(1 - \frac{1}{\alpha_{11}}\exp\left\{-\frac{d_{i,j_1,n,t}^2}{\sigma_{11}^2}\right\}\right) \times \left(1 - \frac{1}{\alpha_{12}}\exp\left\{-\frac{d_{i,j_2,n,t}^2}{\sigma_{12}^2}\right\}\right) \tag{14}$$

where α_{11} and α_{12} are normalization constants; σ_{11} and σ_{12} are also prior constants; and $d_{i,j_1,n,t}$ and $d_{i,j_2,n,t}$ are distances between current particle's observation $z_t^i|x_t^{i,n}$ and other interactive observations $z_{t,k}^{j_1}$ and $z_{t,k}^{j_2}$ respectively. Similarly, we can derive the formulas for the cases where z_t^i has more interactive observations.

By exploiting the *"magnetic potential"* model, the interactive likelihood $p\left(z_t^{J_t}|x_t^i, z_t^i\right)$ reduces the probability that objects' estimates occupy the same position in the feature space. Therefore, our *"gravitation attraction versus magnetic repulsion"* scheme can be regarded as a *"competitive exclusion principle"*.

B. Inertia Potential Model

By using the *"magnetic potential"* model to estimate the *interactive likelihood*, our tracker can successively separate the observation in occlusion and thus solve the *"error merge"* problem. However, the blind *"mutual repulsion"* between interactive observations may lead false *"object labeling"* especially after severe occlusion. This further inspired us to use more information to solve the *"object labeling"* problem.

Different cues such as motion consistence, the three-dimensional (3-D) model, and depth ordering have been exploited to solve the *"object labeling"* problem in the literature [5]. Some of them can be integrated with our DMOT framework straightforward. For example, speed and depth ordering can be incorporated into our tracker by extending the state's dimensionality. However, considering the higher complexity introduced by the extra dimensionality, in this paper, we propose an *ad hoc "first-order inertia Markov chain"* to estimate the second-order

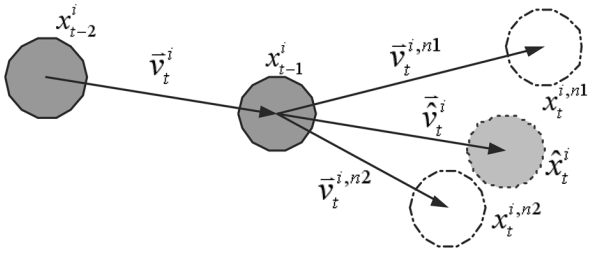


Fig. 4. Process of calculating inertia weights.

state transition density $p\left(x_t^i|x_{t-1}^i, x_{t-2}^i\right)$ and solve the *"object labeling"* problem with a much lower computation cost as follows:

$$\begin{aligned} p\left(x_t^i|x_{t-1}^i, x_{t-2}^i\right) &= p\left(x_t^i|x_{t-1}^i\right) \frac{p\left(x_{t-2}^i|x_t^i, x_{t-1}^i\right)}{p\left(x_{t-2}^i|x_{t-1}^i\right)} \\ &= p\left(x_t^i|x_{t-1}^i\right) \phi_t^i\left(x_t^i, x_{t-1}^i, x_{t-2}^i\right). \end{aligned} \quad (15)$$

The state transition density $p\left(x_t^i|x_{t-1}^i\right)$ can be modeled by a *first-order Markov chain* as usual in the regular Bayesian tracking methods [1]. It can be estimated by either a constant acceleration model or a Gaussian random walk model. We call $\phi_t^i(\cdot)$ the *"inertia function"* and further estimate it as follows. $\phi_t^i(\cdot)$ is related with two posteriors. Since x_{t-2}^i, x_{t-1}^i, and x_t^i are propagated forward, what information can we infer backwards without any prior knowledge? The answer is *Inertia*. The *Inertia Law* [10] in physics states that *"An object at rest tends to stay at rest and an object in motion tends to stay in motion with the same speed and in the same direction unless acted upon by an unbalanced force"*. Fig. 4 illustrates the analysis of object i 's motion in three consecutive frames where shadow ellipses represent the states and dash line ellipses represent the particles. $\vec{\mathbf{v}}_t^i = x_{t-1}^i - x_{t-2}^i$ is the reference motion vector from x_{t-2}^i to x_{t-1}^i. By shifting $\vec{\mathbf{v}}_t^i$ along its direction, we get the *"inertia state"* $\hat{x}_t^i$ and its *"inertia motion vector"* $\vec{\hat{\mathbf{v}}}_t^{\,i}$ for the current frame. Without any external force, the new state x_t^i should be same as $\hat{x}_t^i$ according to the *Inertia Law*. Even if there are external forces, as long as the frame rate is high enough, we can always assume that x_t^i is not far away $\hat{x}_t^i$. $x_t^{i,n1}$, $x_t^{i,n2}$ are particles of state x_t^i. $\vec{\mathbf{v}}_t^{i,n1}$ and $\vec{\mathbf{v}}_t^{i,n2}$ are the associated motion vectors. We define the *inertia weights* as

$$\begin{aligned} \phi_t^{i,n}\left(x_t^{i,n}, x_{t-1}^i, x_{t-2}^i\right) \propto \frac{1}{\alpha_2} \exp\left\{-\frac{\left(\theta_{t,n}^i\right)^2}{\sigma_{21}^2}\right\} \\ \cdot \exp\left\{-\frac{\left(\|\vec{\mathbf{v}}_t^{i,n}\| - \|\vec{\hat{\mathbf{v}}}_t^{\,i}\|\right)^2}{\sigma_{22}^2}\right\} \end{aligned} \quad (16)$$

where α_2 is a normalization term, and σ_{21} and σ_{22} are also prior constants which characterize the allowable variances of the motion vector's direction and speed, respectively. In (16), $\vec{\mathbf{v}}_t^{i,n} = x_t^{i,n} - x_{t-1}^i$, $\vec{\hat{\mathbf{v}}}_t^{\,i} = x_{t-1}^i - x_{t-2}^i$; $\theta_{t,n} = \angle(\vec{\mathbf{v}}_t^{i,n}, \vec{\hat{\mathbf{v}}}_t^{\,i})$ is the angle between $\vec{\mathbf{v}}_t^{i,n}$ and $\vec{\hat{\mathbf{v}}}_t^{\,i}$. $\|\vec{\mathbf{v}}_t^{i,n}\|$ and $\|\vec{\hat{\mathbf{v}}}_t^{\,i}\|$ are the Euclidean norms. Therefore, the *inertia function* can be approximated as

$$\phi_t^i\left(x_t^i, x_{t-1}^i, x_{t-2}^i\right) \approx \sum_{n=1}^{N_s} \frac{\phi_t^{i,n}}{\sum_{n'=1}^{N_s} \phi_t^{i,n'}} . \delta\left(x_t^i - x_t^{i,n}\right). \quad (17)$$

C. Local Likelihood Model and Importance Density

1) Local Likelihood Model: Different image cues such as gradient, color, and motion have been exploited to estimate the local observation likelihood $p(z_t|x_t)$ in the literature [1], [16], [17]. In this paper, we combine the color histogram model [16] and the PCA-based model [17] to efficiently estimate the local likelihood

$$p\left(z_t^i|x_t^{i,n}\right) = p_c \cdot p_p \quad (18)$$

where p_c and p_p are the likelihood densities estimated by the color histogram and PCA models, respectively.

2) Importance Density: It has been widely accepted that better importance density according to different criteria could give more efficient particles and thus improves the tracker's performance [13]. In this paper, we accept the most common choice, which is intuitive and simple for implementation and allows us to perform a better comparison with other MOT methods using the same choice

$$q\left(x_t^i|x_{t-1}^{i,n}, z_t^i, z_t^{J_t}\right) = p\left(x_t^i|x_{t-1}^{i,n}\right). \quad (19)$$

D. Implementation Issues

A pseudocode of our DMOT using particle filtering implementation for one object at one time step is shown in Table I. We only illustrate the implementation for pairwise repulsion. The repulsion among more interactive observations can be added similarly. After tracking at time t, all trackers resample their associated particles again to alleviate the *degeneracy* problem [13].

IV. Experimental Results

The performance of the proposed DMOT algorithm has been demonstrated on several video databases including different synthetic, indoor, and outdoor video sequences. Due to the limited space available, we will only present a few examples of our results in this paper.

A. Tracking Results

We have compared the performance of the proposed DMOT with multiple independent conventional particle filters (MIPF) [1] and mean field Monte Carlo (MFMC) [9]. For all methods, different number of particles have been tested to achieve a balance between tracking stability and speed as discussed later in this section. The results reported are obtained by using 50 particles per object for each tracker using each method. Different

TABLE I
PSEUDOCODE OF THE PROPOSED DMOT ALGORITHM

For i=1:M
* Set k=0; Draw particles $x_t^{i,n} \sim q(x_t^i | x_{t-1}^{i,n}, z_t^i, z_t^{J_t})$;
* Calculate the initial $w_t^{i,n}$ without $\varphi_t^{i,n}(\cdot), \phi_t^{i,n}(\cdot)$;
* Normalize the weights $w_t^{i,n}$;
* Temporary estimate $\hat{z}_{t,0}^i \sim \hat{x}_{t,0}^i = \sum_{n=1}^{N_s} w_t^{i,n} x_t^{i,n}$;
* FOR j=1:i-1 // Search pairwise links.
  - IF ($\hat{z}_{t,0}^i$ and $\hat{z}_{t,0}^j$ have interaction)
    - Save the interaction link(i,j)=1;
    - FOR k=1:K // Magnetic-Inertia Iteration.
      - Compute the magnetic weight $\varphi_t^{i,n}(\cdot)$ and the inertia weight $\phi_t^{i,n}(\cdot)$;
      - Update weights $w_t^{i,n}$ by $w_t^{i,n} \varphi_t^{i,n}(\cdot) \phi_t^{i,n}(\cdot)$;
      - Normalize the weights $w_t^{i,n}$;
      - Estimate $\hat{z}_{t,k}^i \sim \hat{x}_{t,k}^i = \sum_{n=1}^{N_s} w_t^{i,n} x_t^{i,n}$;
      // Repel Back
      - Compute the magnetic weight $\varphi_t^{j,n}(\cdot)$ and the inertia weight $\phi_t^{j,n}(\cdot)$;
      - Update weights $w_t^{j,n}$ by $w_t^{j,n} \varphi_t^{j,n}(\cdot) \phi_t^{j,n}(\cdot)$;
      - Normalize the weights $w_t^{j,n}$;
      - Estimate $\hat{z}_{t,k}^j \sim \hat{x}_{t,k}^j = \sum_{n=1}^{N_s} w_t^{j,n} x_t^{j,n}$;
      - Resample $\{x_t^{i,n}, w_t^{i,n}\}$, $\{x_t^{j,n}, w_t^{j,n}\}$;
    - END FOR k
  - END IF
* END FOR j
END FOR i

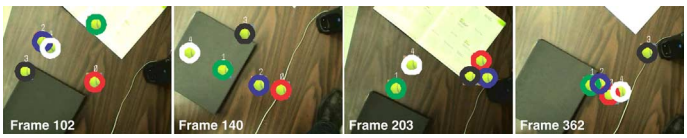


Fig. 5. Tracking results of the proposed DMOT for the synthetic sequence TennisBalls.

color and indexes corresponding to ellipses are used to label the tracked objects.

1) Conceptual Demonstration: The synthetic sequence `Tennisballs` was developed by Yu *et al.* [9]. It contains five identical tennis balls moving independently at constant velocity and bouncing at the image boundary. Due to the changing background in the video sequence, tracking algorithms that use background subtraction are unsuitable. We manually initialize all trackers. In Fig. 5, we present the tracking results of the proposed DMOT. As we can see, by modeling the interaction among observations and introducing the *"inertia function"*, the performance is relatively robust.

2) Indoor Scenario: The video `Lab` was captured with a resolution of 320 by 240 pixels and a frame rate of 25 frame per second (fps). It contains four persons moving across each other in our lab. It challenges many existing algorithms because of

Fig. 6. Comparisons of MIPF (first row), MFMC (second row), and the proposed DMOT (third row) for the indoor scenario sequence Lab.

Fig. 7. Tracking results of the proposed DMOT for sequence SecurityCAM. Background subtraction is used to initialize the trackers. Frame 7110 illustrates a failure mode in the dash line rectangular region.

the frequent presence of occlusions. All of the objects are initialized manually for all methods. In Fig. 6, we compare the results. MIPF (first row) suffers from severe *"error merge"* problems due to the lack of an object interaction model. By modeling the motion interaction, the performance of MFMC (second row) is improved, but the results are not accurate after occlusion and the objects' labels are incorrect. DMOT (third row), on the other hand, performs well in solving both the *"error merge"* and *"object labeling"* problems even in the severe occlusion.

3) Outdoor Scenario: The test video `SecurityCAM` is from MERL Video Database. It is a long sequence (7433 frames) with a resolution of 400 by 300 pixels and a frame rate of 25 fps. Many people walk or run into and out of the camera view, presenting a lot of different occlusions. Background subtraction [18] is used to initialize the objects dynamically. When a new object is detected, our DMOT can easily assign a new tracker to it without reinitialization of the other objects, thus exploiting its distributed formulation. When an object goes out of the scene, the associated tracker is deleted. This is another important advantage of our approach compared with centralized methods which need reinitialization to monitor the presence of new objects as well as existing objects that are no longer in the scene. In Fig. 7, we illustrate the tracking results of our DMOT. In most situations, the proposed DMOT presents an impressive performance by resolving multiple object occlusions very well. A failure case is illustrated in frame 7110 (the dash line rectangular) where background subtraction does not initialize the trackers correctly due to the continuous occlusion of the two people in the scene when they walk into the camera view. Automatic object initialization is a classical problem for video tracking but is beyond the scope of this paper. By exploiting more sophisticated initialization schemes, we believe such failure case could be substantially improved.

Fig. 8. Tracking results of the proposed DMOT for sequence TrafficCAM. Background subtraction is used to initialize the trackers.

TABLE II
AVERAGE COMPUTATION TIME OF DIFFERENT OPERATIONS PER FRAME

Local Likelihood	Repulsion Weight $\varphi(\cdot)$	Inertia Weighting $\phi(\cdot)$
34.5 millisecond	3.6 millisecond	2.1 millisecond

TABLE III
SPEED ANALYSIS OF DIFFERENT RESOLUTIONS AND NUMBER OF PARTICLES PER OBJECT

Video Resolution	Particles Per Object	Speed
128×96	50	$31 \sim 33$ fps
320×240	50	$21 \sim 24$ fps
	200	$9 \sim 11$ fps
	1000	$0.7 \sim 0.8$ fps

Another long sequence TrafficCAM (6510 frames) is used to to represent the wide possibility for real-world applications. It was captured by a traffic surveillance camera on a highway with a resolution of 320 by 240 pixels and a frame rate of 30 fps. In Fig. 8, we show the tracking results of the proposed DMOT.

B. Computational Cost and Speed Analysis

The experiments were performed using C++ on a 3.2-GHz Pentium IV PC. Without code optimization, our algorithm can achieve rates of $13 \sim 25$ fps for robust tracking of $4 \sim 20$ objects simultaneously. The computational cost of the proposed DMOT algorithm using the particle filtering implementation is impacted by four main factors: 1) the computational complexity of the local likelihood $p(z_t|x_t)$; 2) the computation of the magnetic and inertia weights; 3) the video resolution; and 4) the number of particles used for tracking each object. In Table II, we compare the computational demands of the local likelihood and the proposed *"magnetic-inertia"* weights by providing the average computation time required in processing one frame. The data is obtained by running DMOT on the TennisBalls sequence with five magnetic-repulsion iterations. It can be observed that most of the computation time is devoted to local likelihood weighting using image features whereas the cost of magnetic and inertia weighting are around 1/10 and 1/16, respectively, of the cost of local likelihood weighting. This observation clearly demonstrates the efficiency of our DMOT algorithm since this approach only introduces the additional cost of calculating magnetic and inertia weights beyond the computations required for multiple independent trackers. In Table III, we analyze the effect of different video resolutions and particle number used for tracking each object on the computation speed of DMOT on the Labvideo sequence. As we can see, lower resolutions can produce higher computation speed. However, compared with video resolution, the number of particles used for each object is a more critical factor for tracking speed.

TABLE IV
SPEED AND STABILITY COMPARISON OF DIFFERENT PARTICLE FILTER BASED APPROACHES ON THE "LAB" SEQUENCE

Method	Optimal Particle Number		Speed	Position	Label
	Each Tracker	Total	(fps)	Failures	Failures
MIPF	50	200	$22 \sim 25$	493	511
MIPF-I	50	200	$22 \sim 24$	446	453
MFMC	50	200	$17 \sim 19$	9	536
DMOT	50	200	$21 \sim 24$	3	0
MCPF	1300†	1300	$0.9 \sim 1.2$	51	64

† MCPF uses only one tracker to track all objects.

The number of particles used plays an important role in tracking stability as well. Although using fewer particles can achieve faster speed, lacking a sufficient number of particles will result in unstable and inaccurate tracking. On the other hand, in principle, increasing the number of particles can improve the tracking robustness at the expense of lower speed. Therefore, an "optimal" number of particles has to be determined in order to achieve a balance between tracking stability and speed. However, since performance evaluation for multiple object tracking is still an open problem [19], a method to select the "optimal" number of particles is a challenging problem that remains unresolved. In our experiments, we select the number of particles from a predefined set, e.g., $\{30, 50, 75, \ldots\}$. By visually comparing the tracking performance, the smallest number of particles that does not significantly degrade the tracking quality is selected. In Table IV, we compare the speed and tracking performance for MIPF [1], multiple independent particle filter with inertia model (MIPF-I) which is a simplified version of our approach without the *"magnetic potential"* model, MFMC [9], our DMOT and MCPF [8], an MCMC-based particle filter using joint state representation. The data is obtained by invoking the different tracking algorithms on the Lab video sequence. We define a position failure as the absence of a tracker (ellipse) associated with one of the tracking objects. If a tracker is associated with an incorrect object, we regard it as a label failure. With the inertia model alone, MIPF-I cannot improve the performance much compared with MIPF. It still suffers from the *"error merge"* problem since it does not handle the interaction among the objects' observations. Compared with MIPF, MIPF-I, and MFMC, the proposed DMOT has greatly improved the tracking performance with much fewer position and label failures, while still achieving almost near-real-time speed due to the efficiency of the *"magnetic-inertia"* model. The methods which rely on joint state space representation, such as MCPF, require an exponential growth in particles as the number of objects increases. In contrast, because DMOT uses multiple interactively distributed trackers, one tracker per object for MOT, the number of particles needed by DMOT increases linearly with the number of objects as shown in Table IV, i.e., $N_t = N_s \cdot M$, where N_t is the total number of particles used for tracking M objects, and N_s is the number of particles for each tracker (here, $N_s = 50$ and $M = 4$).

ACKNOWLEDGMENT

The authors would like to thank Y. Wu and T. Yu from Northwestern University and F. Porikli from Mitsubishi Electric Research Laboratories for sharing the test sequences.

REFERENCES

[1] M. Isard and A. Blake, "Condensation—Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.

[2] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for bayesian filtering," *Statist. Comput.*, vol. 10, pp. 197–208, 2000.

[3] Y. Bar-Shalom and A. G. Jaffer, *Tracking and data association*. San Diego, CA: Academic, 1998.

[4] N. Gordon, "A hybrid bootstrap filter for target tracking in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 1, pp. 353–358, Jan. 1997.

[5] J. Maccormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *Int. J. Comput. Vis.*, vol. 39, no. 1, pp. 57–71, 2000.

[6] M. Isard and J. MacCormick, "Bramble: A bayesian multiple-blob tracker," in *Proc. Int. Conf. Computer Vision (ICCV'01)*, Vancouver, BC, Canada, Jul. 2001, vol. 2, pp. 34–41.

[7] C. Hue, J. P. L. Cadre, and P. Pérez, "Sequential monte carlo methods for multiple target tracking and data fusion," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 309–325, Feb. 2002.

[8] Z. Khan, T. Balch, and T. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets," in *Proc. 8th European Conf. Computer Vision (ECCV'04)*, Prague, Czech Republic, May 2004.

[9] T. Yu and Y. Wu, "Collaborative tracking of multiple targets," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04)*, Washington, DC, 2004.

[10] I. S. Grant and W. R. Phillips, *The Elements of Physics*. Oxford, U.K.: Oxford Univ. Press, 2001.

[11] W. H. Hayt and J. A. Buck, *Engineering Electromagnetics*. New York: McGraw-Hill, 2001.

[12] J. Whittaker, *Graphical Models in Applied Mathematical Multivariate Statistics*. New York: Wiley, 1990.

[13] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian baysian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[14] B. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman and Hall, 1986.

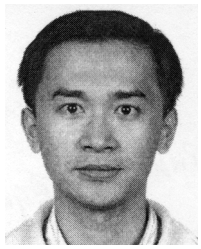
[15] A. Doucet, On sequential monte carlo methods for bayesian filtering Cambridge Univ. Eng. Dept., Cambridge, U.K., 1998, Tech. Rep.

[16] K. Nummiaro, E. B. Koller-Meier, and L. V. Gool, "Object tracking with an adaptive color-based particle filter," in *Symp. Pattern Recognition of the DAGM*, 2002, pp. 355–360.

[17] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.

[18] O. Tuzel, F. Porikli, and P. Meer, "A Bayesian approach to background modeling," in *Proc. IEEE Int. Workshop on Machine Vision for Intelligent Vehicles*, San Diego, CA, 2005.

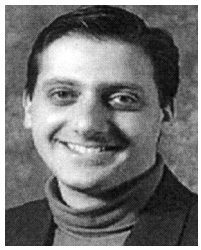
[19] J. Black, T. J. Ellis, and P. Rosin, "A novel method for video tracking performance evaluation," in *Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Nice, France, 2003.



Wei Qu (S'04–M'06) received the B.S. degree in electrical and computer engineering from the Beijing Institute of Technology, China, in 2000, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Chicago in 2005 and 2006, respectively.

He was a Research Assistant at Institute of Automation, Chinese Academy of Science, from 2000 to 2002. During the summers of 2005 and 2006, he was a research intern with Mitsubishi Electric Research Laboratories, Cambridge, MA, and Siemens Medical Solutions, Inc., Malvern, PA, respectively. He is currently a Senior Researcher at Motorola Labs, Schaumburg, IL. He has authored over 20 technical papers in various journals and conferences and has two U.S. patents pending.

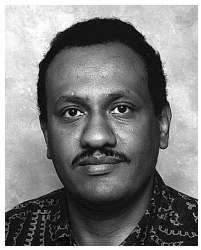
Dr. Qu received the Best Student Paper Award at the IEEE International Conference on Image Processing, 2006. He has also served regularly as a reviewer for different journals and conferences, such as IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Neural Network*, etc.



Dan Schonfeld (S'86–M'90–SM'05) received the B.S. degree in electrical engineering and computer science from the University of California at Berkeley, and the M.S. and Ph.D. degrees in electrical and computer engineering from The Johns Hopkins University, in 1986, 1988, and 1990, respectively.

In 1990, he joined the University of Illinois at Chicago, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. He has authored over 100 technical papers in various journals and conferences. His current research interests are in signal, image, and video processing; video communications, retrieval, and networks; image analysis and computer vision; and genomic signal processing.

Dr. Schonfeld was co-author of papers that won the Best Student Paper Awards in Visual Communication and Image Processing 2006 and IEEE International Conference on Image Processing 2006. He is currently serving as associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He has served as an associate editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING as well as an associate editor for nonlinear filtering of the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Magdi A. Mohamed (S'93–M'93–A'95) received the B.Sc. degree in electrical engineering from the University of Khartoum, Sudan, and the M.S. (computer science) and Ph.D. (electrical and computer engineering) degrees from the University of Missouri-Columbia, in 1983, 1990 and 1995, respectively.

He worked as a Visiting Professor in the Computer Engineering and Computer Science Department at the University of Missouri-Columbia from 1995 to 1996. He has been with Motorola, Inc. since 1996, and is currently a Principal Staff Engineer at Motorola Labs, Physical Realization Research Center of Excellence, Schaumburg, IL. His research interests include digital signal and image processing, computer vision, fuzzy sets and systems, neural networks, parallel and distributed computing, nonlinear and adaptive modeling techniques.