Generative Model for Maneuvering Target Tracking

XIN FAN, Member, IEEE
GUOLIANG FAN, Senior Member, IEEE
Oklahoma State University

JOSEPH P. HAVLICEK, Senior Member, IEEE
University of Oklahoma

We consider the challenging problem of tracking highly maneuverable targets with unknown dynamics and introduce a new generative maneuvering target model (GMTM) that, for a rigid body target, explicitly estimates not only the kinematics, here considered as effect variables, but also the underlying causative dynamic variables including forces and torques acting on the rigid body target in a Newtonian mechanics framework. We formulate relationships between the dynamic and kinematic state variables in a novel graphical model that naturally facilitates the feedback of physical constraints from the target kinematics to the maneuvering dynamics model in a probabilistic form, thereby achieving improved tracking accuracy and efficiency compared to competing techniques. We develop a sequential Monte Carlo (SMC) inference algorithm that is embedded with Markov chain Monte Carlo (MCMC) steps to generate probabilistic samples amenable to the feedback constraints. The proposed algorithm can estimate both maneuvering dynamics and target kinematics simultaneously. The robustness and efficacy of this approach are illustrated by experimental results obtained from noisy video sequences of both simulated and real maneuvering ground vehicles.

Manuscript received July 20, 2007; revised June 27, 2008; released for publication December 4, 2008.

IEEE Log No. T-AES/46/2/936803.

Refereeing of this contribution was handled by D. Salmond.

This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grants W911NF-04-1-0221 and W911NF-08-1-0293.

This paper was presented in part at the 4th IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS07), Minneapolis, MN, June 22, 2007.

Authors' current addresses: X. Fan, School of Software, Dalian University of Technology, Road 8# Kai Fa Qu, Dalian, Liaoning, 116620; G. Fan, School of Electrical and Computer Engineering, Oklahoma State University, 202 Engineering South, Stillwater, OK 74078, E-mail: (glfan@okstate.edu); J. P. Havlicek, School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK.

I. INTRODUCTION

The effective and robust tracking of maneuvering objects of interest, or targets, in video sequences acquired from imaging sensors is a challenging problem that is central to a variety of important applications ranging from intelligent surveillance to military guidance, threat warning, situational awareness, and fire control. This is a difficult problem that has been studied extensively for decades, with notable progress occurring recently in several areas including dynamic motion modeling [1–3], target/background representation [4, 5], and high-dimensional state estimation [6–8]. In the most difficult cases where the targets of interest are highly maneuverable and no accurate target/background representation is available, it is absolutely critical to devise a sophisticated, robust motion model. A major challenge in this regard lies in effectively modeling the relationship between the maneuvering dynamics and the target kinematics [9–11]. Here, we specifically distinguish between kinematics, which refers to motion over time, and dynamics, which refers to the forces and torques that give rise to temporal changes in the kinematic quantities. For the case of a maneuvering vehicle, the dynamics in particular often result at least in part from the intentions of a human operator and are consequently unpredictable in a general sense. In a recent comprehensive survey of motion modeling, it was stated that "a good model is worth a thousand pieces of data" [3].

In this paper, we introduce a new generative maneuvering target model (GMTM) based on explicit online estimation of the maneuvering dynamics that directly explain the target kinematics through the laws of Newtonian mechanics. In addition to the improvements gained by modeling the dynamics directly, this approach enables us to leverage additional physical constraints between the dynamics and kinematics to further enhance tracking performance. For targets such as ground vehicles and aircraft, the maneuvering actions are largely due to the forces and torques present in the engine system and a mechanical drivetrain or flight control system [12]. Indeed, it is the dynamic quantities that cause the kinematics to vary over time. The kinematic variables include velocities and positions that are estimated statistically by traditional target tracking systems. Many existing tracking algorithms estimate and predict the kinematic variables based on a motion model applied to a point target model or to a centroid obtained from the detection process [3]. This may not be appropriate in cases where the target orientation and aspect are significant. The GMTM approach we propose in this paper is based on a more complete model of the full dynamics of a rigid body target and the physical laws that govern its motion, where both linear and angular motions are integrated into one formulation [12]. This results in a realistic dynamic/motion model that is capable of accommodating substantial maneuvering behavior. In addition, we account for the limited power that can be delivered by any practical engine by introducing a probabilistic constraint between the velocity and the driving force and use this to generate probable hypotheses of driving forces during inference.

Graphical models provide a general framework to represent the underlying probabilistic structure of a complex system in the presence of uncertainty [13]. For GMTM, we propose a generative graphical model to represent two types of latent variables describing the target state: the kinematics, which we interpret as effect variables, and the dynamic forces and torques, which we interpret as causative variables that alter the kinematics. We consider a sequence of video frames acquired from an imaging sensor as observations associated with the latent states. The conditional densities of the effect variables are dependent on the causative variables through the principles of Newtonian dynamics. Hence the dynamics underlying the maneuvering actions are incorporated into the motion model in a probabilistic way, so that target tracking and dynamics analysis can be formulated as a Bayesian estimation problem. Moreover, additional physical constraints that relate the kinematics and dynamics, such as the velocity-force constraint mentioned in the preceding paragraph, can be accommodated via conditional probabilities between the latent variables. We use a sequential Monte Carlo (SMC) method (e.g., a particle filter) to infer the posterior densities of the latent variables given the observations. This is in view of the fact that the effectiveness of the SMC methods for nonlinear/non-Gaussian estimation problems has been well established recently [2, 9, 14–16]. In addition, we use a Markov chain Monte Carlo (MCMC) step [17–19] to rejuvenate the particles associated with the causative variables given those of the effect variables. The kinematic states as well as those of forces reflecting the maneuvering actions of the target can be simultaneously obtained by this inference algorithm.

A key idea behind the new GMTM proposed in this paper is that of formulating the dependencies between the latent dynamic and kinematic variables of a rigid body target into a graphical model. This idea is sufficiently general that it could be used as the basis for a new graphical approach to the general object tracking problem at a fundamental level. Doing so in complete generality is beyond our present scope, however. The reason for this is that the particular conditional dependencies between the observations and the latent dynamic and kinematic variables must be considered explicitly in developing a specific practical inference engine for performing state estimation on the graphical model, and clearly these relationships are dependent upon the particular sensor technologies used to acquire the observations.

Here, we choose rather to focus on the particular case of an imaging electro-optical (EO) sensor that delivers a temporal sequence of two-dimensional (2D) video frames integrated at, e.g., infrared or visible wavelengths. This restriction of scope relative to the completely general problem provides us with a concrete foundation on which to investigate physical principles such as Newton rigid body dynamics [6, 7] in conjunction with practical engineering constraints and to formulate them together in a concrete probabilistic framework that facilitates the development of demonstrable SMC estimation algorithms on specific graphical structures and probability distributions. However, it should be kept in mind that the graphical model approach is general and not limited to observations acquired from any specific sensor types. Thus, addressing the more general problem in order to support other types of sensors such as radar and sonar as well as sensor fusion between, e.g., radar and imaging EO sensors [20] is an important problem that merits further investigation but is beyond the scope of this paper.

The remainder of the paper is organized as follows. Important related work is briefly reviewed in Section II. In Section III, we formulate the GMTM model for ground vehicle motion and discuss it in relation to several existing, well-known motion models. Our generative graphical model for representing the maneuvering dynamics is proposed in Section IV, followed by a development of the inference algorithm for state estimation in Section V. Experimental results including both simulated and real data sequences are presented in Section VI, where we compare the proposed algorithm with two recent techniques. Finally, conclusions and a brief discussion of future research directions are given in Section VII.

II. RELATED WORK

Practical target tracking systems are generally based on models of the available observations, which typically consist of measurements delivered by one or more sensors and of the target motion. The relationships between the observation model and the motion model depend on the specific types of sensors that are used. Throughout this paper we assume an imaging EO sensor, where the goal of the observation model is to provide an informative and compact representation of the target appearance. A considerable variety of appearance models have been studied for visual target tracking, including kernel based appearance models [21, 22], eigen appearance [23, 24], and silhouettes [25, 26]. A comprehensive recent review of techniques for vision-based tracking was given in [27]. Our main focus in this paper is on development of a new GMTM that explicitly considers the dependencies between kinematics and dynamics. In order to illustrate the complementary roles of observation and motion models while demonstrating how tracking performance and robustness can be enhanced by a sophisticated motion model, we make use of a relatively simple appearance model.

The number of dynamic motion models that have been proposed for target tracking is also large [3]. One of the most popular among these, particularly for practical deployment tracking systems, has been the constant velocity model where the motion of a point target (or of the centroid obtained from the detection processing for an extended target) is modeled as a white Gaussian noise acceleration (WGNA) process [1, 2]. In [6], [7], Miller, et al., utilized the Newtonian equations of three-dimensional (3D) rigid body dynamics as the motion model and proposed a joint tracking and recognition algorithm. They provided an elegant and unified framework that explicitly models the target motion and rotation under arbitrary dynamics and proposed a jump-diffusion process for statistical inferencing in the resulting high-dimensional kinematic state space. Unfortunately, the state space grows exponentially with respect to increases in the observation sequence. Thus, the associated jump-diffusion random sampling algorithm results in infeasible computational complexity for many practical tracking problems. A similar approach using the Metropolis algorithm instead of jump Markov MCMC was proposed by Runnalls in [28].

The new GMTM we introduce in this paper leverages recent advances in graphical models and particle filtering to dramatically improve efficiency and reduce the computational burden relative to [6], [7] for tracking ground vehicles. In particular, we simplify the Newtonian dynamics for ground vehicles and, more importantly, obtain enhanced state estimation efficiency by introducing the velocity-force constraint. We also invoke a graphical model approach to characterize the probabilistic dependencies between latent variables in an explicit way, where the Markovian structure is assumed in order to support efficient sequential state estimation.

Multiple-model techniques have been developed for tracking maneuvering targets where a discrete process is introduced to generate maneuvering switches among a finite set of continuous linear dynamic models [29]. This general approach has also been well studied in the context of a jump Markov linear system (JMLS) where the discrete switching process is assumed to be a Markov chain [29, 30]. Graphical models can also be used to capture probabilistic dependencies similar to those of a JMLS. When expressed in terms of graphical models, such jump motion models are collectively referred to as switching linear dynamic systems (SLDS) [31]. The parameters of an SLDS are often adjusted (or learned) a priori using precollected training data [32, 33]. Neither JMLS nor SLDS techniques explicitly represent the maneuvering dynamics. Rather, they represent maneuvering actions as discrete switching variables with instantaneous values that indicate different motion models. For the GMTM approach proposed in this paper, we investigate the underlying physical laws of maneuvering actions to build a graphical motion model that is able to encode the dynamics in terms of clear physical meanings.

In a recent study, Godsill, et al., also developed a dynamic model in terms of the driving force [9]. The force is expressed as a piecewise constant deterministic function and the maneuvering time when the force changes act on the target is explicitly modeled, which demands deterministic interpolation of the maneuvering state process to match the measurement time. This deterministic interpolation is used to drive a stochastic particle filtering method to achieve the state estimation. In our approach, continuous force values are stochastically generated from the probability distribution that is conditionally dependent on current velocities, which are derived from physical observations. This probabilistic expression of force together with the dependencies between the kinematic variables are formulated into a unified probabilistic graphical model framework where the stochastic particle filter is a natural choice for performing statistical inference. Compared to the method of [9], our approach provides a flexible and general formulation where physical laws as well as practical engineering constraints can be seamlessly incorporated together for target tracking.

Blom's interacting multiple-model (IMM) algorithm is one of the most widely used estimation algorithms for multiple-model techniques where the constituent continuous models are linear and Gaussian [34, 35]. However, the standard IMM algorithm fails to accurately estimate non-Gaussian probabilities. Moreover, exact inference techniques for graphical models, such as belief propagation (BP), are not applicable to hybrid state estimation with discrete and continuous variables; they also are not easily adapted to the case of complex non-tree structures [36]. Thus, variants of IMM [30, 37, 38] and BP [39–41] have been proposed to achieve state estimation in more general cases.

These algorithms are all related in the sense that they employ sampling-based approximation in the form of particle filters (SMC). The SMC method can easily propagate particles in a sequential manner according to the motion model for recursive Bayesian state estimation. Researchers have explored a great number of applications of the SMC method to navigation [8], fault and change detection [14], data fusion [42], and visual tracking [25, 43] since it was first introduction by Gordon, et al. [44]. Meanwhile, many variants of the SMC method have been proposed to improve its performance

[15, 16]. The common philosophy underlying these various SMC-based algorithms is to devise an effective proposal density that generates particles, approximating the true probability distribution as closely as possible while at the same time maintaining the diversity of the particles. An MCMC method, which can generate samples of any target probability distribution by constructing a Markov chain, is embedded into each time step of the SMC process to overcome the impoverishment problem of the particle set (e.g., all particles have nearly identical values) [18, 19]. However, performance can only be improved with the introduction of MCMC moves that have the correct stationary distribution. Therefore, for the GMTM proposed in this paper, the MCMC scheme is specifically designed for consistency with the physical constraints that exist between the target kinematics and the maneuvering dynamics.

III. DYNAMICS OF RIGID TARGET MOTION

We consider the dynamics of a maneuvering ground vehicle in 3D rigid motion which, as discussed below, may be approximated as a special constrained case of general 3D motion. The Newtonian equations governing the general 3D motion of a rigid body are given by [12]

$$\begin{aligned}\dot{\mathbf{p}} &= \mathbf{f}\\ \dot{\mathbf{h}} &= \boldsymbol{\tau}\end{aligned} \tag{1}$$

where the vectors $\mathbf{p}$ and $\mathbf{h}$ are the linear and angular momentum of the body, respectively, and the vectors $\mathbf{f}$ and $\boldsymbol{\tau}$ are the force and torque acting on the body. It should be noted that the Newtonian equations in (1) are valid when the motion is resolved in a fixed inertial reference frame, in which the linear and angular momentum are dependent on the linear and angular velocities, respectively. The dynamics given by (1) indicate that forces and torques cause the motion of a rigid body having kinematics represented by position $\mathbf{r}$, linear velocity $\mathbf{v}$, and orientation and angular velocity $\boldsymbol{\omega}$. Our generative graphical model for GMTM is constructed based on an explicit model of how forces and torques generate kinematical state changes.

To characterize the motion of a ground vehicle, here considered as an object of interest that is to be tracked, we define three principal axes along the body frame as shown in Fig. 1 and express the dynamics (1) according to [7]

$$\dot{\mathbf{v}} + \tilde{\boldsymbol{\omega}}\mathbf{v} = \frac{\mathbf{f}}{M} \tag{2}$$

$$\mathbf{J}\dot{\boldsymbol{\omega}} + \tilde{\boldsymbol{\omega}}\mathbf{J}\boldsymbol{\omega} = \boldsymbol{\tau} \tag{3}$$

where $\mathbf{f} = [f_x\ \ f_y\ \ f_z]^T$ and $\mathbf{v} = [v_x\ \ v_y\ \ v_z]^T$ are composed of the forces and linear velocities along the body axes, and where $\boldsymbol{\tau} = [\tau_x\ \ \tau_y\ \ \tau_z]^T$ and $\boldsymbol{\omega} = [\omega_x\ \ \omega_y\ \ \omega_z]^T$ are the torques and angular velocities expressed with respect to the body axes. The mass M in (2) and inertial momentum matrix $\mathbf{J}$ in (3) are fixed for a given object. The skew-symmetric matrix $\tilde{\boldsymbol{\omega}}$ appearing in (2) and (3) is defined by

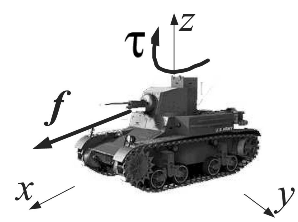

Fig. 1. Coordinate system defined relative to body frame of ground vehicle. The x axis is given by direction of forward motion, and y axis is its perpendicular within sustaining plane. The z axis is defined so that it forms a right-handed coordinate system with sustaining plane axes.

$$\tilde{\boldsymbol{\omega}} = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix}.$$

Motion of the vehicle with respect to the ground is caused primarily by f_x and τ_z. The following list provides a discussion of each of the six forces and torques appearing in (2) and (3).

1) The driving force, f_x: The total force parallel to the x axis is the summation of engine boosting and friction, i.e., driving and resistant forces. This force causes changes in the linear velocity.

2) The sliding force, f_y: When a vehicle such as the one shown in Fig. 1 travels along a steep transverse slope, it may occasionally slide in the y direction due to the component of gravity along the y axis. Even less frequently, translation in the y direction may result from a loss of contact with the ground, collisions with other vehicles, or ordnance impacts. However, these cases are not typical. Therefore, in the interest of reasonably constraining the complexity of the model without unduly compromising sophistication, we henceforth assume that the vehicle remains in contact with the ground and that the gravitational as well as any other applied forces along the y axis are in equilibrium with the frictional forces acting on the vehicle so that $f_y = 0$.

3) The supporting force, f_z: Similar to the discussion above, it is possible for a vehicle such as the one shown in Fig. 1 to become momentarily airborne when traversing rough terrain at high speeds or to be momentarily lifted from the ground by a severe collision or ordnance impact. However, these are relatively rare occurrences. Hence, we assume here that $f_z = 0$.

4) The rolling torque, τ_x, and pitching torque, τ_y: Consistent with our discussion of f_y and f_z above, we assume that the vehicle remains in contact with

the ground in an upright posture most of the time. The drivetrain of a vehicle such as the one shown in Fig. 1 is not designed to exert rolling and pitching torques as part of normal maneuvers. Uneven terrain may give rise to momentary unbalanced torques τ_x and τ_y. However, if these torques were substantial and persistent they would result in the vehicle being overturned, which is not typical. Therefore, in the interest of simplicity we assume here that $\tau_x = \tau_y = 0$.

5) The turning torque, τ_z: Torque about the z axis results from the engine and drivetrain of the vehicle, giving rise to turning maneuvers. Like f_x, τ_z is a major contributor to kinematical state changes.

Based on the discussion above, we assume that, in relation to the coordinate system in Fig. 1, the object motion is locally planar in the x-y sustaining plane over time scales that are both realistic and appropriate for the tracking problem. Thus, (2) and (3) may be simplified according to

$$\begin{bmatrix} \dot{v}_x \\ \dot{v}_y \end{bmatrix} + \begin{bmatrix} 0 & -\omega_z \\ \omega_z & 0 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} f_x \tag{4}$$

and

$$J_z \dot{\omega}_z = \tau_z \tag{5}$$

which are expressed relative to the object-centric coordinate frame of Fig. 1. The position $\mathbf{r}$ of the object is related to the velocities $\mathbf{v}$ and ω_z by

$$\dot{\mathbf{r}} = \mathbf{A}(\phi)\mathbf{v} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} \tag{6}$$

where ϕ is the azimuthal orientation of the object. As we assume that the object exhibits nontrivial rotation about the z axis, the orientation ϕ is related to the angular velocity ω_z by

$$\dot{\phi} = \omega_z. \tag{7}$$

Although the force and velocity are resolved in the coordinate system relative to the body frame as shown in Fig. 1, it should be noted that the position $\mathbf{r}$ and orientation ϕ are expressed relative to some predefined, fixed reference point and direction.

Together, the differential equations (4)–(7) characterize how f_x and τ_z drive the kinematical state to change over time. It is worth restating that many motion models, including the widely used WGNA models [2] for multi-aspect target tracking, are obtained from the Newtonian equations of the point motion, where the object is considered as a point with mass. However, the WGNA models neglect the constraints between linear and angular velocities that can be captured in the dynamics of the rigid body motion model adopted here. Equation (4) shows that a typical ground vehicle would not tend to undergo a large increase in linear velocity while simultaneously exhibiting non-zero angular velocity. These constraints (causes) can be observed in the motion of ground vehicles so widely that we regard such objects as rigid bodies instead of simple points.

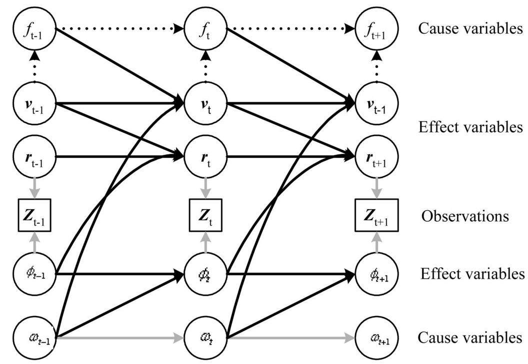

Fig. 2. Proposed generative graphical motion model for maneuvering target tracking in GMTM. Circles and boxes denote latent variables and observations, respectively. Forces and torques are causative variables, while changes in kinematics are effect variables. Bold arrows show how forces and torques cause kinematic variations, whereas dotted arrows depict effects of velocity on realizable forces.

IV. GENERATIVE GRAPHICAL MODELS

In this section, we develop a generative graphical model, depicted in Fig. 2, to describe the uncertainties associated with the maneuvering actions of an object of interest. The dynamics (4)–(7) reveal how the maneuvering forces and torques generate kinematics variations in future observations, which imply the structure of the graphical model. The associated conditional probabilities are derived from these equations, thereby encoding the physical constraints underlying the object motion.

A. Probabilistic Models of Maneuvering Variables

Similar to the main ideas of jump Markov linear (JML) models [30] and SLD models [33], we define multiple models indicated by $\mathcal{M}_t \in \{0,1,2\}$ for the driving force $f_{x,t}$ at time t with the goal of describing distinct maneuvers characterized by approximately constant velocity ($\mathcal{M}_t = 0$), by significant acceleration ($\mathcal{M}_t = 1$), or by significant deceleration ($\mathcal{M}_t = 2$). Each model is initialized to a normal density with mean $\mu_{k,0}$ and variance $\sigma^2_{k,0}$, $0 \le k \le 2$. Rather than learning the Gaussian parameters a priori as was done in [33], we generate the hypotheses of the acceleration and deceleration models ($\mathcal{M}_t = 1,2$) in future time steps by a conditional Rayleigh distribution on the current velocities. Use of the conditional Rayleigh distribution is justified in Section IVC below. For the constant velocity model ($\mathcal{M}_t = 0$), we assume throughout that $f_{x,t} \sim N(0, \sigma^2_{0,0})$. Although estimation algorithms exist for the case of an unknown transition probability matrix (TPM) [45], we specify a fixed prior TPM $T_{i,j} \triangleq \Pr(\mathcal{M}_t = i \mid \mathcal{M}_{t-1} = j)$, e.g., $T_{i,j} = 0.5$ when $i = j$ and $T_{i,j} = 0.25$ when $i \neq j$, since we find that our estimation algorithm is reasonably insensitive to the exact choice of TPM. Our model for linear motions is analogous to the JML and SLD models in the sense of having a discrete switching variable, but the continuous motion models accommodating acceleration and deceleration allow for adaptively updating the state based on the current kinematics estimation.

We assume that the torque τ_z randomly generates a certain angular velocity ω_z at a single time step and that the angular velocity remains constant over the sampling interval. Thus, we do not model the behavior of the torque directly; rather, we characterize the distribution of the angular velocity ω_z by the following ternary-uniform mixture [3]: the object maintains its current direction ($\omega_z = 0$) with probability P_0, rotates counter clockwise at a rate uniformly distributed in $[0, \omega_{\max}]$ with probability $(1 - P_0)/2$, or rotates clockwise at a rate uniformly distributed in $[-\omega_{\max}, 0]$ with probability $(1 - P_0)/2$. The value of P_0 may reflect some prior knowledge about the possible motion pattern. The ternary-uniform mixture was originally proposed for modeling the magnitude and duration of target maneuvers when no prior knowledge on the maneuvering time is available in [46]. We apply this distribution to model angular maneuvers. The distribution is easy to sample and, as illustrated in Section VIB, produces satisfactory estimation results in our experiments with real image sequences in which the target executes abrupt turns. A limitation of this approach as it is implemented here is that the parameters of the ternary-uniform mixture must be specified manually. However, with some appropriate additional constraints these parameters could be adaptively updated online.

B. Temporal Constraints

The temporal constraints relating the kinematical variables in two consecutive time steps are determined by (4)–(6). We denote the set of hidden variables $\mathbf{v}$, $\mathbf{r}$, ϕ in Fig. 2 together in a kinematical vector $\mathbf{K} = [\mathbf{v}^T, \mathbf{r}^T, \phi]^T$. We assume that the sampling interval T is sufficiently small that we may solve the differential equations (4)–(6) using Euler integration according to

$$\mathbf{r}_t = \mathbf{r}_{t-1} + T\mathbf{A}(\phi_{t-1})\mathbf{v}_{t-1} + \mathbf{n}^r_{t-1} \tag{8}$$

$$\mathbf{v}_t = \mathbf{v}_{t-1} - T\tilde{\boldsymbol{\omega}}_{z,t-1}\mathbf{v}_{t-1} + \mathbf{f}_{t-1} + \mathbf{n}^v_{t-1} \tag{9}$$

where the velocity vector $\mathbf{v}$ comprises the velocities along the x and y axes, the skew-symmetric matrix $\tilde{\boldsymbol{\omega}}_{z,t-1}$ is formed by the angular velocity with respect to z as in (4), and where

$$\phi_t = \phi_{t-1} + T\omega_{z,t-1} + n^{\phi}_{t-1}. \tag{10}$$

The noise variables $\mathbf{n}^r$, $\mathbf{n}^v$, and $\mathbf{n}^{\phi}$ in (8)–(10), which act as "driving" noise (or process noise) to accommodate numerical errors and uncertainties in the motion processes, are assumed to be independent identically distributed (IID) and Gaussian with zero means and known variances (in practice, we have found that tracking performance is not terribly sensitive to the exact tuning of these parameters; in the simulation experiments of Section VIA, we take the standard deviations of all three process noises equal to the parameter $f_{\max}$ given below in (11)). The one-step conditional probabilities of the current kinematical variables $p(\mathbf{K}_t \mid \mathbf{K}_{t-1})$, graphically represented by the bold arcs in Fig. 2, can be obtained by (8)–(10) and the given noise statistics. If the directed arcs from $\boldsymbol{\omega}$ to $\mathbf{v}$ are omitted in Fig. 2, then the proposed graphical model degenerates to the WGNA or constant acceleration dynamic models that are currently widely used in the context of target tracking [2, 3]. In addition to temporally evolving the linear and angular motion independently, our model couples the nonlinear angular effects into the linear motion.

C. Velocity-Force Constraints

In the JML [30] and SLD [33] models, an independent discrete Markov chain is introduced as the driving process of the continuous kinematic states in tracking applications. The specific trajectory of the evolving kinematical variables in any given realization provides cues by which to infer the underlying driving forces, but does not directly influence the evolution of the driving process itself. We observe, however, that the engine of any practical vehicle has limited output power given by the product of force with velocity [47], rendering certain state transitions impossible or highly unlikely. In the GMTM approach proposed here, we leverage this constraint by directing a dotted arc from $\mathbf{v}_t$ to $\mathbf{f}_t$ in Fig. 2, thereby enabling the kinematic state to drive the trajectory of the causative (dynamical) force variables. It is worth noting in this regard that forces and torques of practical aircraft are strongly dependent on the velocities [12], implying that this model could be generalized to one that is applicable to conventional aircraft, missiles, and space vehicles.

In defining the conditional probability $p(f_{x,t} \mid \mathbf{v}_t, f_{x,t-1})$, we explicitly utilize the observation that, typically, only small increases in engine power are available when the vehicle is operating near maximum speed. For an object with increasing acceleration, large acceleration increases are more likely at low speeds than at high speeds, whereas small acceleration increases are correspondingly more likely at high speeds than at low speeds. Similarly, for a decelerating object, the reverse is typically true. Large decelerations are more likely at high speeds than at low speeds due to the larger resistant forces present at high speeds, whereas small decelerations are correspondingly more likely

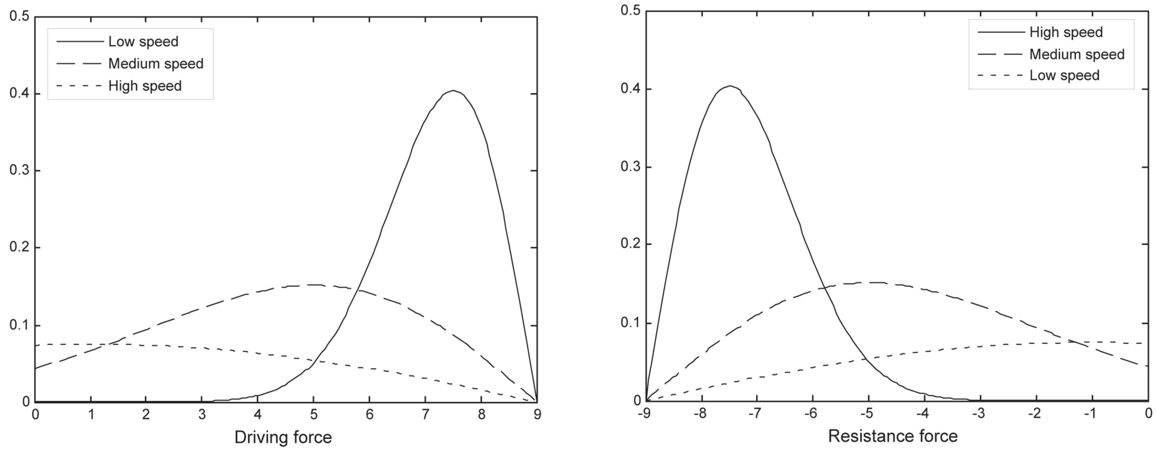

Fig. 3. Rayleigh conditional distributions of driving force (left) and resistant force (right) for several speeds. Left figure shows that large accelerating forces are more likely when target moves at lower speed, while right figure shows that large decelerating forces are more likely at high speeds.

at low speeds where the resistant forces tend to be smaller. These relationships between the driving and resistant forces are dependent on the kinematics and require an asymmetric distribution for a probabilistic representation. The Rayleigh distribution, which is asymmetric, was successfully used in [48] to estimate the probability parameters of acceleration from the velocity. In addition, the Rayleigh distribution can be efficiently sampled, facilitating practical implementation of the inference algorithm. Therefore, we model the velocity-force constraints using conditional Rayleigh distributions defined by

$$p(f_{x,t} \mid \mathbf{v}_t, f_{x,t-1}) = \begin{cases} \dfrac{f_{\max} - f_{x,t}}{c_t^2} \exp\left(-\dfrac{(f_{\max} - f_{x,t})^2}{2c_t^2}\right) & \\ \quad \times \mathbb{U}(f_{\max} - f_{x,t}), & f_{x,t-1} > 0 \\ \dfrac{f_{x,t} - f_{-\max}}{c_t^2} \exp\left(-\dfrac{(f_{x,t} - f_{-\max})^2}{2c_t^2}\right) & \\ \quad \times \mathbb{U}(f_{x,t} - f_{-\max}), & \text{otherwise} \end{cases} \tag{11}$$

as shown in Fig. 3, where $\mathbb{U}(\cdot)$ is the unit step function; $f_{\max}$ and $f_{-\max}$ are the forces that generate maximum acceleration and deceleration, respectively; and c_k is a parameter that depends on the current velocity $\mathbf{v}_t$ according to

$$c_t = \begin{cases} \dfrac{f_{\max}\sqrt{2/\pi}}{1 + \exp(|\mathbf{v}_t| - |\mathbf{v}_{\max}|)}, & f_{x,t-1} > 0 \\ \dfrac{f_{-\max}\sqrt{2/\pi}}{1 + \exp(|\mathbf{v}_{\max}| - |\mathbf{v}_t|)}, & \text{otherwise} \end{cases}.$$

The maximum velocity $\mathbf{v}_{\max}$, as well as $f_{\max}$ and $f_{-\max}$, are determined by the configuration of a vehicle and are manually specified for the experiments given in Section VI. The first case in (11) corresponding to the condition $f_{x,t-1} > 0$ generates the acceleration hypotheses $\mathcal{M}_t = 1$, while the second case corresponding to $f_{x,t-1} \leq 0$ generates deceleration hypotheses $\mathcal{M}_t = 2$.

Zhou and Kumar used a Rayleigh distributed mean-adaptive acceleration model similar to (11) for deriving the variance of the predicted acceleration in a Kalman estimator [48]. In that case, however, the proposed Rayleigh density depended only on the previous acceleration and not on the current velocity. The model in (11) also shares some common characteristics with the idea of data-driven sampling approaches [32, 49], where the feedback constraints of the effect (e.g., the velocity) on the cause (e.g., the force) are considered. But our model is distinct from the data-driven paradigm in that the data-driven paradigm utilizes cues from the observed data to define a better proposal that has a larger overlap with the target distribution. As shown in Fig. 2, this consideration is explicitly reflected by the conditional distribution between the latent variables in our generative model.

V. INFERENCE ALGORITHM

Due to the complicated topology of the proposed graphical model shown in Fig. 2, it is nontrivial to perform exact inference by the standard BP algorithm [36] in this case. For the GMTM approach, we resort to a SMC-based algorithm [2, 8, 9, 14–16, 30, 42, 44, 50–53] to estimate the kinematics and the underlying dynamics. It is well understood that the performance of SMC can be improved by embedding MCMC moves having the correct stationary distribution [17–19, 30, 54–65]. Hence, we apply MCMC to generate the samples of the conditional Rayleigh distribution (11) which enforces practical constraints between the velocity and the driving force as described in Section IVC. Observing that the evolution of the kinematic variables $\mathbf{K}$ may be represented by a Markov chain and that nonlinearity

TABLE I
The SMC-Based Inference Algorithm

Initialize
N_t = no. time steps; N_s = no. particles;
initialize particle set $\{\mathcal{M}_0^i, f_{x,0}^i, \omega_{z,0}^i, \mathbf{K}_0^i\}_{i=1}^{N_s}$;
For $t = 1, \ldots, N_t$
Prediction
sample $\{\mathbf{K}_t^i\}_{i=1}^{N_s}$ using (8)–(10);
Weighting
evaluate weights $\{\tilde{w}_t^i\}_{i=1}^{N_s}$ using (12);
Normalization
normalize the weights: $\left\{w_t^i = \tilde{w}_t^i / \sum_{k=1}^{N_s} \tilde{w}_t^k\right\}_{i=1}^{N_s}$;
Selection
resample $\{\mathcal{M}_{t-1}^i, f_{x,t-1}^i, \omega_{z,t-1}^i, \mathbf{K}_t^i\}_{i=1}^{N_s}$ according to importance weights $\{w_t^i\}_{i=1}^{N_s}$, and then set $\{w_t^i = 1/N_s\}_{i=1}^{N_s}$
MCMC step
For $i = 1, \ldots, N_s$
sample $\omega_{z,t}^i$ using ternary-uniform mixture;
sample $\mathcal{M}_t^i$ using TPM conditioned on $\mathcal{M}_{t-1}^i$;
If $\mathcal{M}_t^i = 0$
sample $f_{x,t}^i$ from $N(0, \sigma_{0,0}^2)$;
Else
sample $f_{x,t}^i$ from (11);
End
End

is involved in the dynamic equations due to (8), we apply the SMC algorithm to sequentially approximate the posterior density $p(\mathbf{K}_t \mid \mathbf{Z}_t)$ at the current time step given the previous density $p(\mathbf{K}_{t-1} \mid \mathbf{Z}_{t-1})$ by a particle set $\{\mathbf{K}_t^i, w_t^i\}_{i=1}^{N_s}$. The procedure for updating the particle set is given in Table I, where the prediction $p(\mathbf{K}_t^i \mid \mathbf{K}_{t-1}^i)$ is formulated using (8)–(10).

The weights w_t^i in Table I are evaluated using a likelihood function $p(\mathbf{Z}_t \mid \mathbf{K}_t^i)$ that depends on an object template $\mathbf{L}$. This template contains a standardized instance of the spatial target signature that is expected to occur in a video frame acquired from the imaging sensor currently in use. It could be generated on the fly from an appearance model of arbitrary complexity including dependencies on the particular sensor technology, estimated object type, and estimated pose and magnification parameters. Alternatively, the template $\mathbf{L}$ could be retrieved from a stored library of expected signatures. Depending on the estimated magnification, the template will generally have a spatial support that is smaller than an entire video frame. Thus, the probability that an object (e.g., target) with kinematical variables $\mathbf{K}_t^i$ is present in the observed frame $\mathbf{Z}_t$ may be quantified by the agreement between the template $\mathbf{L}$ and a region of interest having the same size and shape as $\mathbf{L}$ that is extracted from the location $\mathbf{r}_t^i$ in the observed frame $\mathbf{Z}_t$. Since our objective in this paper is to demonstrate the performance of the proposed GMTM, we restrict our attention to extremely simple appearance models for generating the template $\mathbf{L}$ in the experiments of Section VI. However, it should be noted that this is not a general limitation; in any particular application, further performance gains can be obtained through the use of a more sophisticated appearance model.

Given the template $\mathbf{L}$, the likelihood function for the SMC-based inference algorithm of Table I is given by

$$p(\mathbf{Z}_t \mid \mathbf{K}_t^i) = C \exp\left(-\frac{\|T_{\mathbf{K}}(\mathbf{Z}_t) - \mathbf{L}\|^2}{2\sigma_o^2}\right) \qquad (12)$$

where C is a normalization constant, σ_o^2 is the variance of the observation noise, $\|\cdot\|^2$ is the squared Euclidean norm, and $T_{\mathbf{K}}$ is an operator which extracts a region of interest from the observed frame $\mathbf{Z}_t$ based on $\mathbf{r}_t^i$ and performs a 2D transformation depending on ϕ_t^i to account for rotation of the observed object signature relative to the template $\mathbf{L}$. Thus, $T_{\mathbf{K}}(\mathbf{Z}_t)$ is a subframe having the same size and shape as $\mathbf{L}$. More generally, the transformation $T_{\mathbf{K}}$ might also account for magnification differences between the observed object and the template $\mathbf{L}$; however, since significant magnification changes are not considered in the experiments of Section VI we neglect the magnification parameter here in the interest of simplicity.

The SMC algorithm maintains a running discrete simulation of the posterior density $p(\mathbf{K}_t \mid \mathbf{Z}_{1:t})$, from which we generate samples $f_{x,t}^i$ and $\mathbf{v}_t^i$ in each sampling interval. These samples are obtained by applying the Metropolis algorithm [17] to draw from a distribution $p(f_{x,t} \mid \mathbf{K}_t^i)$ that is given by either the Rayleigh distribution (11) if $\mathcal{M}_t^i > 0$ or, as we stated in Section IVA, from $N(0, \sigma_{0,0}^2)$ if $\mathcal{M}_t^i = 0$. One sample is generated in each iteration as follows. We first sample a proposal density $q(\cdot)$ to obtain $f^* \sim q(\cdot)$. We then accept f^* as $p(f_{x,t}^i)$ with probability

$$\min\left\{1, \frac{q(f_t^*)}{p(f_t^{i-1} \mid \cdot)}\right\}$$

where the proposal $q(\cdot)$ is Gaussian. This procedure is illustrated for the case $\mathcal{M}_t = 1$ in Fig. 4, where histograms of the generated samples are shown overlayed on the true conditional Rayleigh distributions. The details of the inference algorithm are specified by pseudo-code in Table I.

VI. EXPERIMENTAL RESULTS

To illustrate the performance of the proposed GMTM approach, we conducted tracking experiments using both simulated and real video sequences. In the simulated data cases, the simulated frame rate was 24 frames per second (fps). The real video sequences were acquired with a commercial visible wavelength camera having a frame rate of 30 fps. To avoid any confusion arising from the two different frame rates, we refer to digital video frames by the sequential frame number throughout this section; in the case of the synthetic examples the kth frame occurred at a

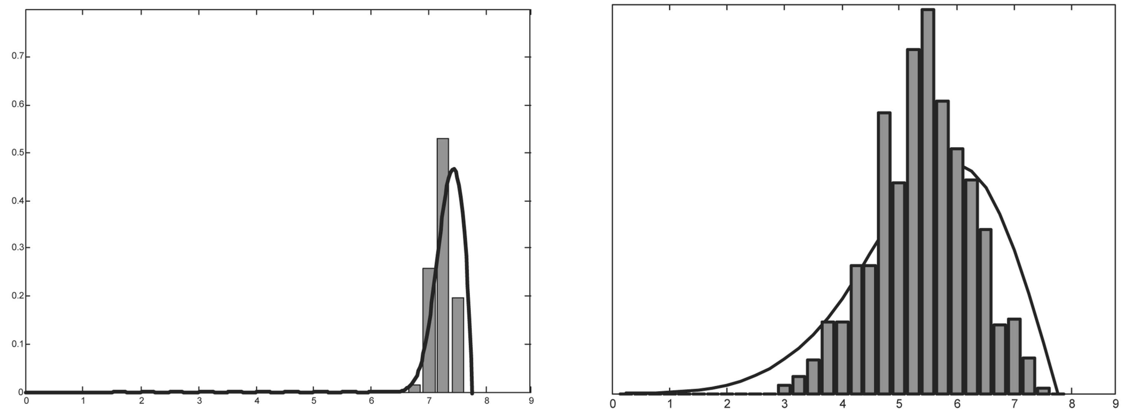

Fig. 4. MCMC approximations for conditional Rayleigh distribution of driving force conditioned on velocity (as shown in Fig. 3), corresponding to a low-speed target (left) and a high-speed target (right). Solid curves denote desired ideal Rayleigh density functions, while histograms give approximations for 500 samples generated by MCMC algorithm.

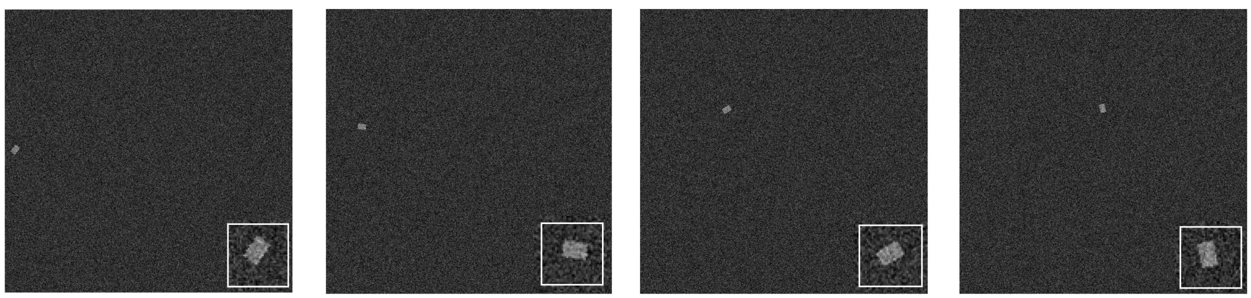

Fig. 5. Example simulated video sequence generated by proposed GMTM. Every fifteenth frame is shown. SNR = −0.5 dB. Zoomed view of target is shown inset at lower right corner of each frame.

time $k/24$ s past the beginning of the sequence, and in the case of the real video examples, the kth frame occurred at a time $k/30$ s past the beginning of the sequence. In Figs. 6–10, the time axis is expressed not in seconds, but rather in units of frame number.

For every example given in this section, the number of particles used in the SMC-based inference algorithm was $N_s = 500$. The variance of the estimates delivered by SMC is a decreasing function of N_s, so it is desirable to have a large number of particles to ensure stability and reduce the estimation errors. However, the computational bandwidth required to implement the algorithm is an increasing function of N_s, making large particle sets infeasible from a practical standpoint. Although we did not study the design tradeoff between large and small N_s with rigor, we did run examples with $N_s = 100$ which generally resulted in unstable estimates with large run-to-run variance and with $N_s = 1{,}000$, which resulted in unacceptably long run times using Matlab on a typical current generation commercial PC. In every case, the kinematics were manually initialized in the first frame.

For the simulation studies, the object of interest was a uniform rectangular block corrupted by additive white Gaussian noise maneuvering against an IID Gaussian background process. The object positions and orientations were generated sequentially according to the Newtonian rigid body dynamics (8)–(10), and the physical constraints, given in Section IVC, where the variance of the Gaussian noise processes $\mathbf{n}^r$, $\mathbf{n}^v$, and $\mathbf{n}^\phi$ in (8)–(10) were set equal to the parameter $f_{\max}$ in (11). The background noise power was not set directly. Rather, this parameter was calculated based on the desired signal-to-noise ratio (SNR) as follows. The background-noise-free target signature was a subimage having the same size as the lower-right insets shown in Fig. 5 and containing an image of the rectangular white target corrupted by additive white Gaussian noise against a black background, but without any injected background noise. The target power was defined as the variance of this background-noise-free target signature, which was calculated at run time in the simulations. The variance for the injected background noise process was then set such that the ratio of the target power to the background noise power was equal to the desired SNR. The SNR was specified as low as −0.5 dB for the simulated examples given in this section. For all of the simulated examples, the background-noise-free target signature described above was used as the object template $\mathbf{L}$ in (12). An example simulated video sequence is given in Fig. 5, where every fifteenth frame is shown. In the simulated data experiments described below in Section VIA, we compare the GMTM tracking results with those obtained by an estimation algorithm developed for JML systems [30, 66].

The real data experiment was performed on a video sequence depicting a top-down view of a highly maneuverable radio-controlled tank moving on an indoor planar surface. This sequence consists of 300 frames acquired at 30 fps. The tank was controlled to perform significant linear accelerations and turns. Additive white Gaussian background noise was injected into the acquired video frames to achieve an SNR of 0 dB, where SNR was defined as described in the paragraph above, in order to evaluate performance in a high-noise scenario where the tracker is required to rely more heavily on the dynamic model as opposed to the appearance model. The object template $\mathbf{L}$ in (12) was generated by manually extracting the background-noise-free target signature from the initial frame prior to injecting background noise. Several example frames are shown in Fig. 11. Comparison of the GMTM results against those of the relatively simple yet widely applied constant velocity constant turn (CVCT) model [3], [35] demonstrate the power of a sophisticated motion model for tracking highly maneuvering targets in low quality video where no strong appearance models are available.

A. Simulated Data Experiments

In this section, we present quantitative tracking results obtained with GMTM and a jump-Markov particle filter (JMPF) algorithm for simulated video sequences where the object exhibits coupled linear and angular motion and velocity-force constraints. The specific JMPF algorithm, due to Doucet, et al. [30], is a particle filtering approach to the optimal estimation of JML systems in which non-Gaussian probabilities are involved. JML systems [30, 66] are multiple models where the linear system parameters evolve according to a finite discrete Markov chain; they share several common features with SLD systems [31]. In our JMPF implementation, the switching term accommodating linear motions takes one of three discrete values $(0, 2, -2)$ corresponding to constant velocity, maximum acceleration (driving force), and maximum deceleration (resistant force). For the angular motion model, the transition probabilities of this three-state Markov chain are set identical to those given for GMTM in Section IVA. For the linear kinematic state transitions, the parameters are set to those given in [2], [30], [66], where coupling between the linear and angular velocities is excluded. For the sake of experimental validation, we restate that GMTM differs from JMPF in the following main respects:

1) GMTM formulates explicit probabilistic models for the causative variables, as opposed to several fixed switching values as in JMPF.

2) GMTM models rigid body dynamics as opposed to point mass dynamics, including an explicit model of the coupling between angular and linear velocities.

3) GMTM explicitly incorporates physically meaningful velocity-force constraints.

In the following two sections, we present the forces and kinematics estimated by the proposed GMTM technique in comparison with those obtained by the JMPF method for several different motion scenarios. Each scenario was run ten times, and the results obtained from a randomly selected exemplar run are given in Figs. 6–9.

1) *Tracking with Coupled Linear and Angular Motion*: In order to compare how the generative model and JMPF cope with coupled linear and angular motion, we generated two video sequences in which the targets have linear velocity changes both with and without angular motions. In the first sequence, the target accelerates in frame 40 without angular motion. Angular motion occurs in subsequent frames, but without linear acceleration. The dynamics and kinematics estimated by GMTM and by JMPF are shown in Fig. 6. As can be seen in Fig. 6(a), both GMTM and JMPF track the force changes that occur in frame 40 after a delay. Both approaches also provide satisfactory kinematics estimation, as shown in Fig. 6(b)–(d) and (f)–(h).

In the second sequence, the target simultaneously turns and decelerates beginning in frame 33. It should be noted that such coupled turning and deceleration is widely observed in the motion of practical ground vehicles. In this case, there is a tangible performance difference between GMTM and JMPF. The dynamic and kinematic estimates for both algorithms are given in Fig. 7. The velocity and orientation estimates obtained by JMPF deviate significantly from ground truth from frame 34 onwards, as shown in Fig. 7(b) and (c). As shown in Fig. 7(a), however, JMPF does successfully capture the drop in force that occurs in frame 33 after a delay, as does GMTM. The inaccuracies in the kinematic estimates delivered by JMPF in Fig. 7 result directly from the exclusion of an explicit model for the dependency between linear and angular motion, which is significant in the second sequence. By contrast, since the rigid body motion model of GMTM explicitly couples angular effects into the linear motion through (4), GMTM appropriately allocates particles in a way that leads to accurate kinematic estimation as shown in Fig. 7.

2) *Tracking with Velocity-Force Constraints*: The third simulated video sequence was designed to illustrate the performance of GMTM and JMPF in a scenario where velocity-force constraints play a significant role. Beginning in frame 22, the target rapidly accelerates. This cannot be maintained indefinitely, however, and from frame 30 on the

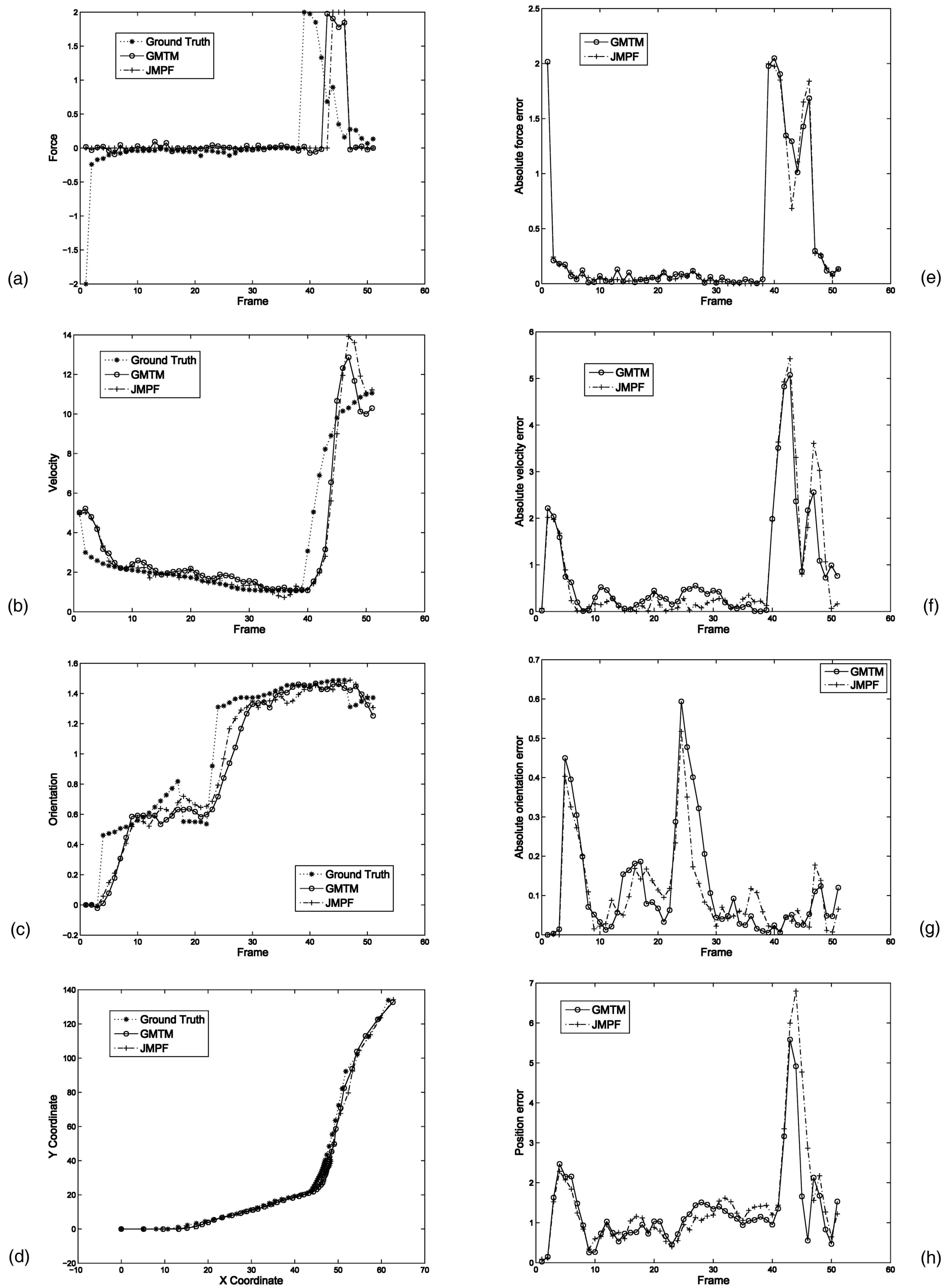

Fig. 6. Dynamic and kinematic estimation results for first simulated video sequence. (a) Force. (b) Velocity. (c) Orientation. (d) Position. (e)–(g) Absolutely error plots. Linear acceleration occurs in frame 40 without angular motion, where force has an abrupt change shown in (a). Both JMPF and GMTM perform satisfactory estimation on kinematics, shown in (b)–(d) and (f)–(h).

speed levels off at near maximum. In addition, there is mild angular motion throughout the sequence and an abrupt turn in frames 41 through 45. The dynamic and kinematic estimation results are shown in Fig. 8. Both algorithms track well for the first 30 frames. However, when the velocity starts to level out in frame 30, JMPF begins to fail. As shown in Fig. 8(a), JMPF does not follow the drop in force

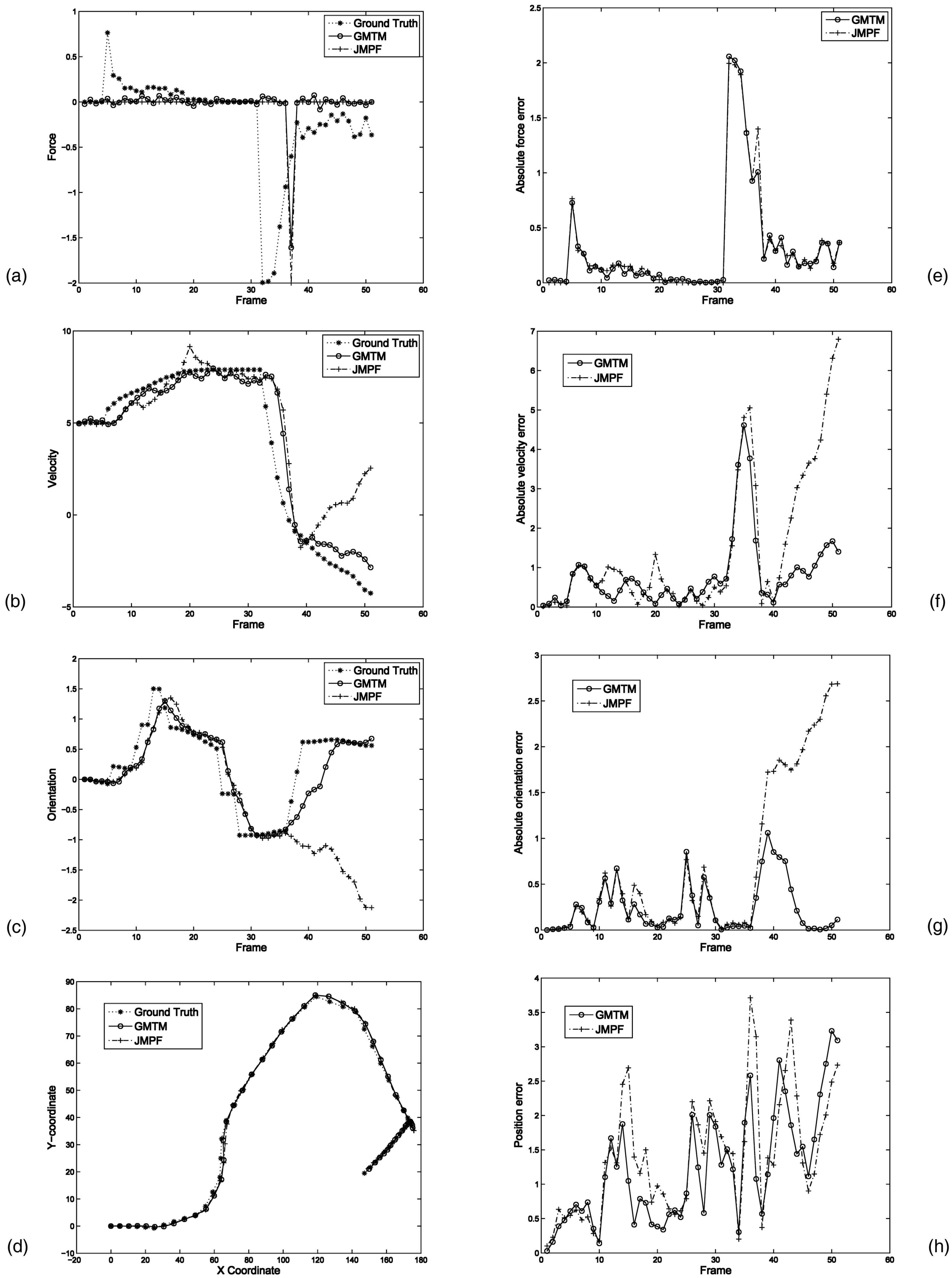

Fig. 7. Dynamic and kinematic estimation results for second simulated video sequence. (a) Force. (b) Velocity. (c) Orientation. (d) Position. (e)–(h) Absolute error plots. Simultaneous turning shown in (c) and deceleration shown in (a) occur beginning in frame 33. JMPF, denoted as pluses in plots, captures force drop but fails for orientation changes, resulting in inaccurate kinematic estimation on velocity ((b) and (f)) in this case.

that occurs in frame 24, whereas GMTM captures the drop (after a delay) due to the feedback constraints between velocity and force in the generative model. As a result, the GMTM velocity estimates shown in Fig. 8(b) remain accurate, while the JMPF velocity estimates first overshoot, then undershoot the ground truth. Consequently, when the abrupt turn occurs in frames 41 through 45, JMPF totally loses both the

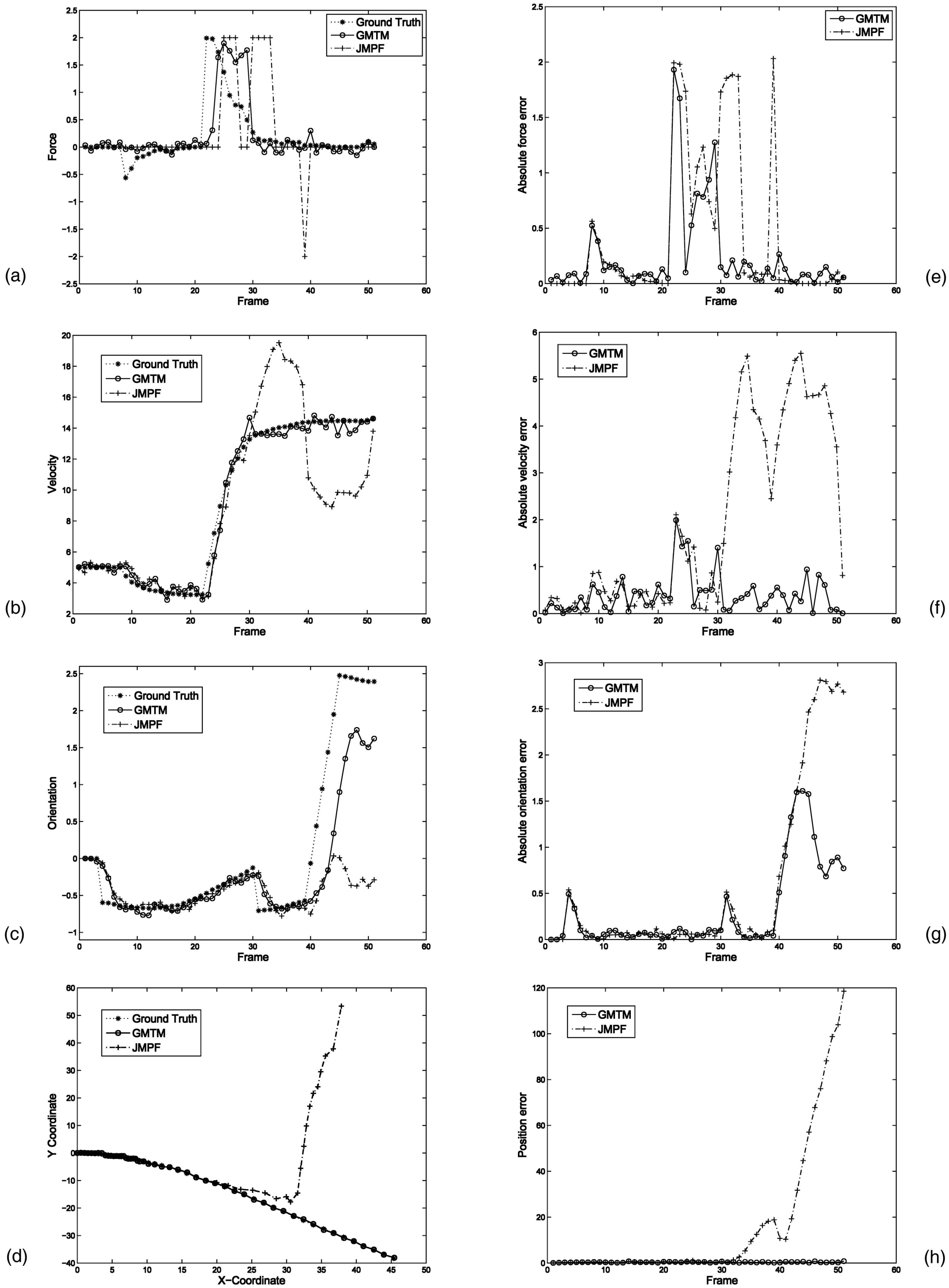

Fig. 8. Dynamic and kinematic estimation results for third simulated video sequence. (a) Force. (b) Velocity. (c) Orientation. (d) Position. (e)–(h) Absolute error plots. Target exhibits strong acceleration in frame 22, with speed leveling out near maximum by frame 30, see (a). There is an abrupt turn from frame 41 through 45 shown in (c). JMPF fails starting in frame 30 when velocity starts to level out (see (b) and (f)), and totally loses tracks when abrupt turn occurs in frames 41 through 45, shown in (d) and (h), whereas GMTM remains accurate.

orientation and position tracks, as shown in Fig. 8(c) and (d), while GMTM delivers accurate simultaneous estimates of both dynamics and kinematics throughout the sequence.

3) *Discussion*: The results of the simulated data tracking experiments given in Figs. 6–8 show that the SMC-based GMTM inference algorithm produces delayed estimates of force and velocity.

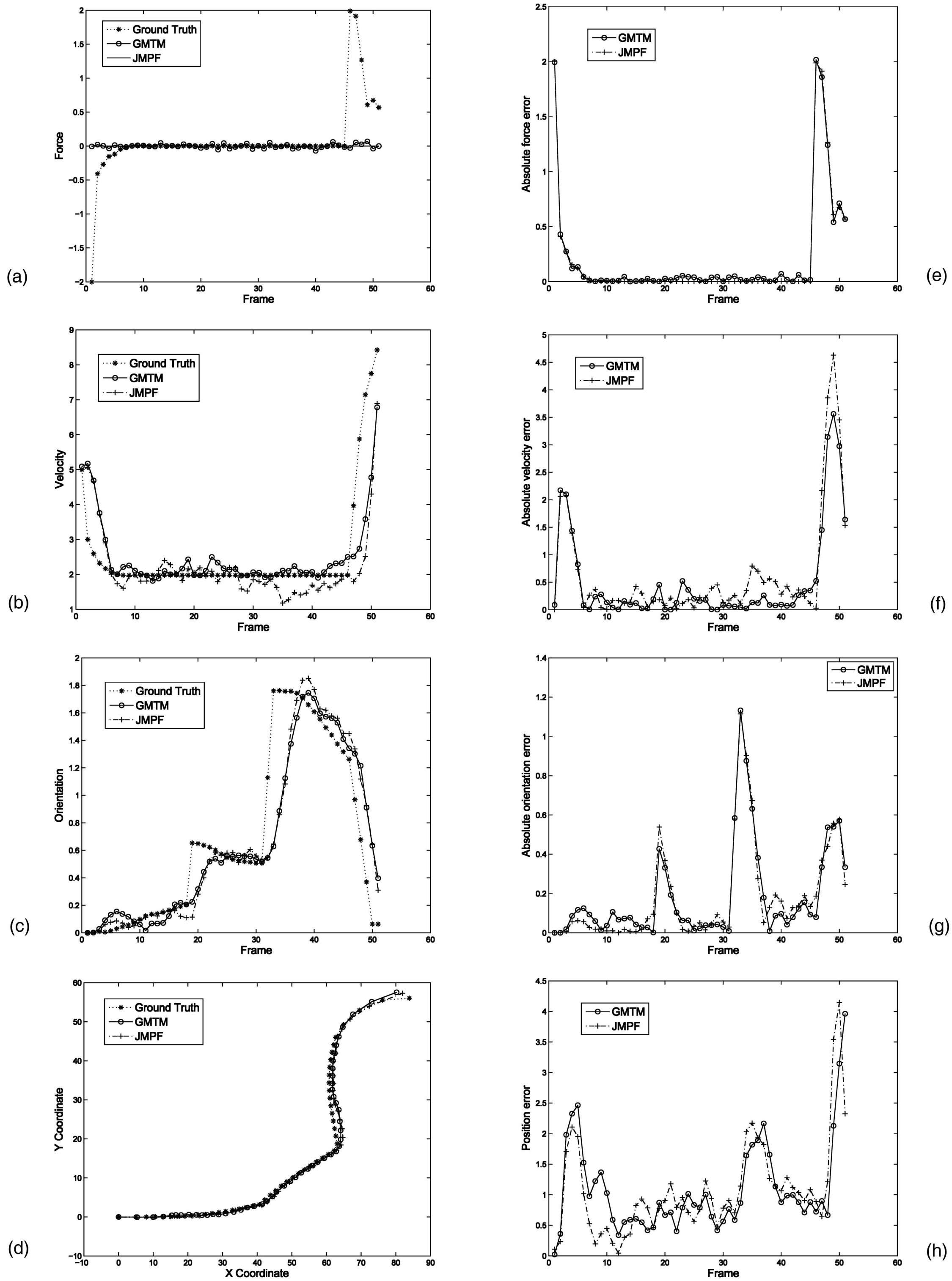

Fig. 9. Dynamic and kinematic estimation results for fourth simulated video sequence. (a) Force. (b) Velocity. (c) Orientation. (d) Position. (e)–(h) Absolute error plots. Significant changes in force occur only in final frames of sequence, shown in (a). Due to inherent delay in state estimation, GMTM provides no advantage over JMPF. Both techniques fail to track abrupt force changes starting from frame 45.

This may be understood in terms of the structure of the generative model (depicted in Fig. 2), in which the causative variables from the previous time step induce the current effect variables and observations. This implies that there is some delay in the state estimation. In most cases, this delay is not expected to compromise the overall tracking performance. However, the GMTM algorithm fails to accurately capture accelerations that occur in the last frames of a sequence, as illustrated in Fig. 9(a). In such cases there are no subsequent observations by which to correct the state estimation; consequently, GMTM provides no tangible performance advantage compared to JMPF for the final frames of a sequence.

B. Real Data Example

A typical tracking system is composed of two complementary components, viz., a representation of the expected target signature and a motion model [21]. In many practical systems, the main approach for maintaining performance in the presence of heavy clutter, noise, and unknown motion patterns is to devise a sophisticated target representation [5, 67]. By contrast, our emphasis in this paper has been on the development of a sophisticated motion model. In this section, we apply the proposed GMTM tracking algorithm to a real-world video sequence depicting an overhead view of a highly maneuverable radio controlled tank. The tank exhibits abrupt turns with substantial changes in orientation as well as rapid decelerations. Although the sequence was acquired with a commercial visible wavelength camera, we inject substantial synthetic additive white Gaussian noise to degrade the target SNR to 0 dB. We also track the sequence with the CVCT model that assumes the target moves with a constant speed and a constant rate of turn up to uncorrelated drift noises [3]. Both the GMTM and CVCT methods use the likelihood function given by (12). The results suggest that the combination of an elaborate motion model with a simple target representation can provide accurate tracking performance even in highly challenging real-world scenarios where the simple motion model performs poorly by comparison.

Fig. 10(a) shows the tracked position trajectories obtained by GMTM and by the CVCT model against ground truth, where the estimation errors are given in Fig. 10(b). The GMTM and CVCT orientation estimates are given in Fig. 10(c) with ground truth, and the respective absolute errors are shown in Fig. 10(d). We manually labeled four vertices of the target region in each frame and derived the ground truth of the target position (centroid) and orientation. Finally, the GMTM and CVCT velocity estimates are given in Fig. 10(e), where no ground truth is readily available in this case. Representative video frames are shown in Fig. 11, where a closeup view of the noise-free target is inset in the lower left of each frame. The estimated track gates delivered by the CVCT method are shown overlayed on the closeup target views in Fig. 11(a). Likewise, the GMTM estimated track gates are shown in Fig. 11(b). The proposed GMTM method provides robust tracking throughout the sequence, producing accurate position and orientation estimates even as the target executes abrupt turns and accelerations in severe noise. By contrast, the CVCT orientation estimates begin to fail almost immediately. In frame 180, which is marked by a heavy dot in Fig. 10(a), there is a sudden deceleration accompanied by an abrupt turn. As shown in Figs. 10 and 11, the GMTM algorithm maintains accurate tracking through this maneuver while the position and orientation estimates of the CVCT method rapidly degrade. It is interesting to note that the GMTM algorithm occasionally misclassifies a portion of the target shadow as being part of the target itself, as shown, e.g., in frame 150 of Fig. 11(b). This problem, which generally perturbs the position and orientation estimates, could be ameliorated by incorporating a more sophisticated observation model than the one given by (12).

VII. CONCLUSION

In this paper, we considered the problem of tracking a highly maneuverable object with unknown dynamics. Based on the principles of Newtonian mechanics, we introduced a new GMTM where the target is considered to be a rigid body as opposed to a point mass. This approach enabled us to develop a new generative graphical model incorporating useful constraints in the form of explicit coupling between the linear and angular velocities and explicit feedback of the kinematic state variables to the evolving dynamics. This model also incorporates realistic velocity-force constraints based on the fact that the engine output power of any practical ground vehicle is limited; thus, e.g., large increases in velocity are unlikely when the vehicle is operating near maximum speed. Based on the generative graphical model, we formulated probabilistic relationships between the kinematic state variables, including position, velocity, and orientation, here considered to be effect variables, and the causative dynamic state variables, including forces and torques, in a Bayesian estimation framework. We developed an SMC inference algorithm embedded with an MCMC step to estimate both the target kinematics and the maneuvering dynamics simultaneously.

Performance of the proposed GMTM algorithm was evaluated on several simulated video sequences designed to highlight the unique features of the approach and demonstrate the specific differences between GMTM and the popular JMPF method. In particular, we demonstrated that the explicit incorporation of velocity-force constraints and coupling between the linear and angular velocities can lead to tangible improvements in tracking performance for highly maneuverable targets. Robustness of the GMTM approach was also demonstrated on a real video sequence depicting an overhead view of a highly maneuverable radio controlled tank.

Our focus in this paper was on dynamic modeling for target tracking, specifically on the dependencies between latent kinematic and dynamic

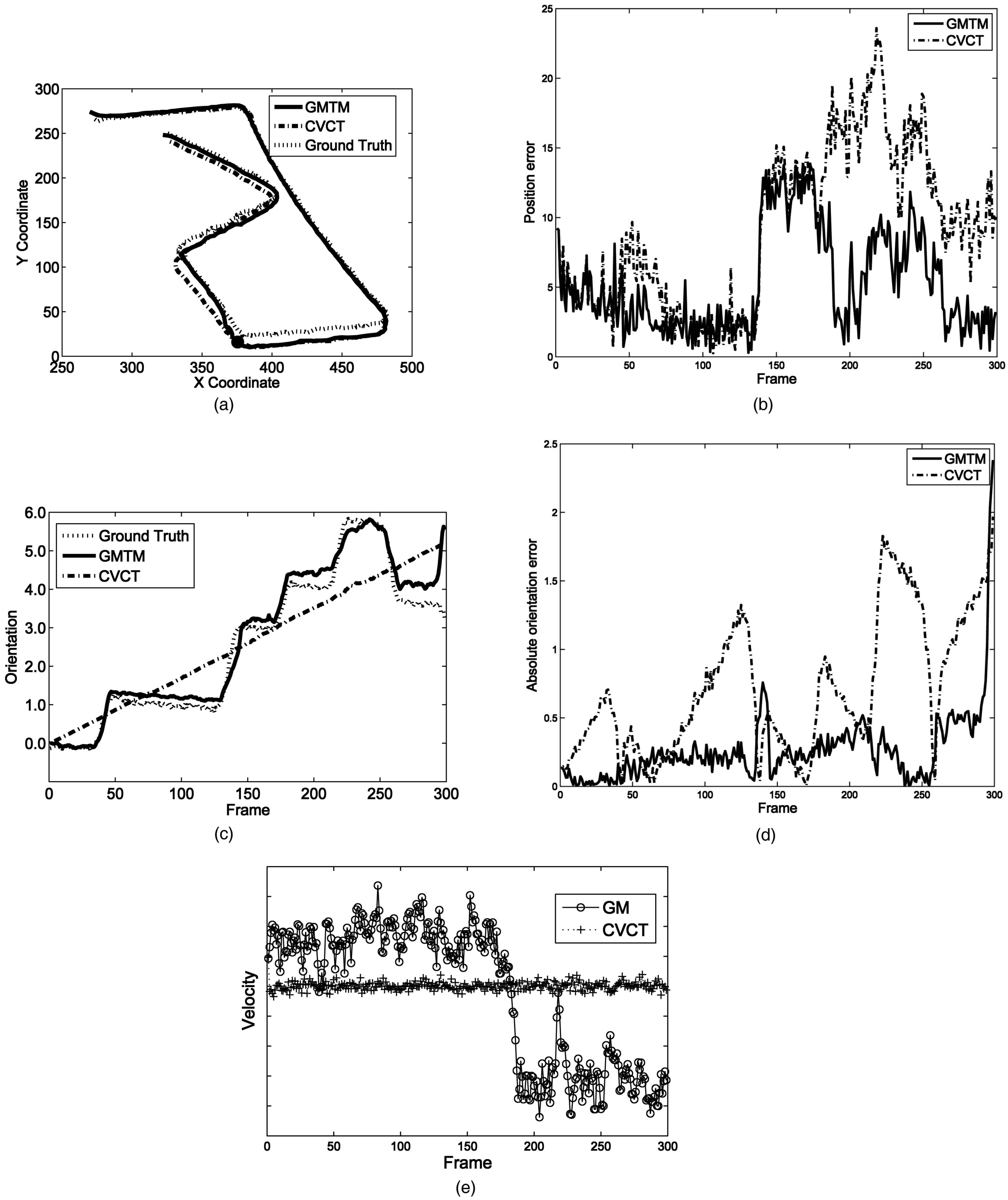

Fig. 10. Kinematic tracking results for real video sequence. (a) Position. Estimates delivered by GMTM and CVCT are shown along with ground truth in pixel coordinates. (b) Absolute errors in position estimates delivered by GMTM and CVCT expressed in units of pixels. (c) Orientation estimates delivered by GMTM and CVCT with ground truth. (d) Absolute errors in orientation estimates. (e) Velocity estimates delivered by GMTM and CVCT. There is no ground truth in this case.

variables. Other measurement models that describe the relationships between the observations and the latent variables, such as Poisson models [4] for example, could also be incorporated into the proposed graphical model in order to track targets in other modes or in multiples modes [42]. In a multi-mode tracking system, for example, radar sensors could provide direct observations of velocity or speed, which could improve the accuracy of the velocity estimation by explicitly

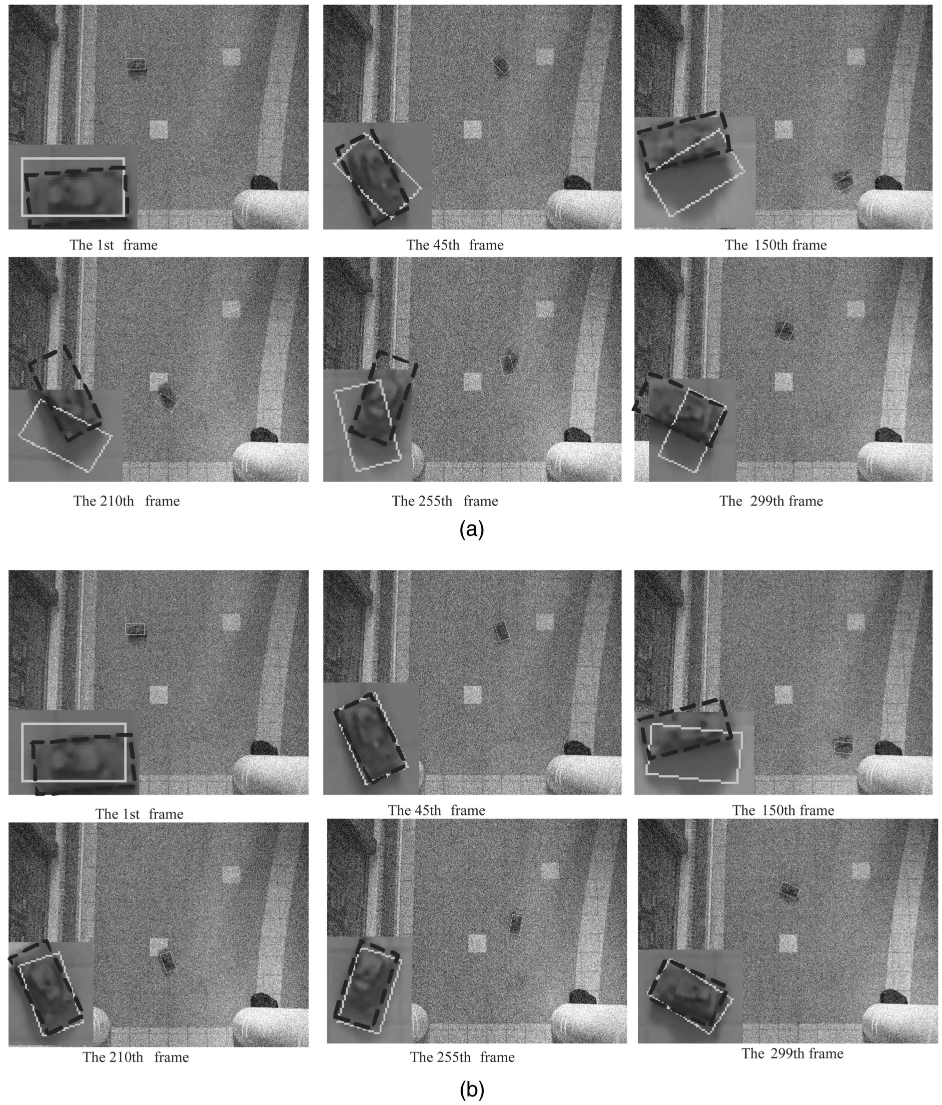

Fig. 11. Tracking results on 8 representative frames obtained by (a) CVCT model and (b) GMTM. Inset in lower left corner of each frame shows estimated track gate (white box) overlayed on closeup of target (black dash) without injected noise. Position and orientation estimates delivered by GMTM are clearly superior, as expected.

treating the likelihood function of the velocity variable. With such additional observations, one could generate more probable force hypotheses based on the velocity-force constraint from the model proposed here, thereby improving both the velocity and position estimates. In addition, since the GMTM is developed based on rigid body motion dynamics in the context of a 3D body frame, the technique could be generalized to track ground vehicles from any perspective by incorporating a 3D camera model and a 3D target appearance model [68, 69]. The GMTM technique could also be extended for multiple target tracking by incorporating data association techniques for appearance modeling, such as the joint probability data association (JPDA) method that has been used successfully with other multiple-model approaches [70, 71].

REFERENCES

[1] Bar-Shalom, Y., Li, X., and Kirubarajan, T.
Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software.
Hoboken, NJ: Wiley, 2001.

[2] Bruno, M. G. S.
Bayesian methods for multiaspect target tracking in image sequences.
IEEE Transactions on Signal Processing, **52**, 7 (2004), 1848–1861.

[3] Li, X., and Jilkov, V. P.
Survey of maneuvering target tracking. Part I: Dynamic models.
IEEE Transactions on Aerospace and Electronic Systems, **39**, 4 (2003), 1333–1363.

[4] Gilholm, K., Godsill, S., Maskell, S., and Salmond, D.
Possion models for extended target and group tracking.
In *Proceedings of SPIE Signal and Data Processing of Small Targets*, 2005.

[5] Jepson, A., Fleet, D., and El-Maraghi, T.
Robust online appearance models for visual tracking.
In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, 415–422.

[6] Miller, M. I., Srivastava, A., and Grenander, U.
Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition.
IEEE Transactions on Signal Processing, **43**, 11 (1995), 2678–2690.

[7] Miller, M. I., Grenander, U., Osullivan, J. A., and Snyder, D. L.
Automatic target recognition organized via jump-diffusion algorithms.
IEEE Transactions on Image Processing, **6**, 1 (1997), 157–174.

[8] Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P-J.
Particle filters for positioning, navigation, and tracking.
IEEE Transactions on Signal Processing, **50**, 2 (2002), 425–437.

[9] Godsill, S., Vermaak, J., Ng, W., and Li, J.
Models and algorithms for tracking of maneuvering objects using variable rate particle filters.
In *Proceedings of IEEE*, **95**, 5 (May 2007), 925–952.

[10] Maskell, S.
Joint tracking of maneuvering targets and classification of their maneuverability.
EURASIP Journal of Applications on Signal Processing, **1** (2004), 2339–2350.

[11] Vanpeperstraete, J-M., and Le Cadre, J-P.
Robust detection of target maneuvers.
In *Proceedings of 9th International Conference on Information Fusion*, Florence, Italy, July 10–13, 2006, 1–8.

[12] Friedland, B.
Control System Design: An Introduction to State-Space Methods.
Colombus, OH: McGraw-Hill, 1986.

[13] Jordan, M.
Graphical models.
Statistical Science, **19**, 1 (2004), 140–155.

[14] Andrieu, C., Doucet, A., Singh, S. S., and Tadic, V. B.
Particle methods for change detection, system identification, and control.
In *Proceedings of IEEE*, **92**, 3 (2004), 423–438.

[15] Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T.
A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking.
IEEE Transactions on Signal Processing, **50**, 2 (2002), 174–188.

[16] Cappé, O., Godsill, S., and Moulines, E.
An overview of existing methods and recent advances in sequential Monte Carlo.
In *Proceedings of IEEE*, **95**, 5 (2007), 899–924.

[17] Andrieu, C., Freitas, N. D., Doucet, A., and Jordan, M. I.
An introduction to MCMC for machine learning.
Machine Learning, **50** (2003), 5–43.

[18] Godsill, S., and Clapp, T.
Improvement strategies for Monte Carlo particle filters.
In A. Doucet, N. de Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, New York: Springer, 2001, ch. 23, 479–497.

[19] Andrieu, C., De Freitas, N., and Doucet, A.
Sequential MCMC for Bayesian model selection.
In *Proceedings of IEEE Signal Processing Workshop on Higher-Order Statistics*, 1999, 130–134.

[20] Maskell, S. R., Everitt, R. G., Wright, R., and Briers, M.
Multi-target out-of-sequence data association: Tracking using graphical models.
Information Fusion, **7**, 4 (2006), 434–447.

[21] Comaniciu, D., Ramesh, V., and Meer, P.
Kernel-based object tracking.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **25**, 5 (2003), 564–577.

[22] Fan, Z., Yang, M., and Wu, Y.
Multiple collaborative kernel tracking.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **29**, 7 (2007), 1268–1273.

[23] Black, M., and Jepson, A.
Eigentracking: Robust matching and tracking of articulated objects using a view-based representation.
In *Proceedings of the European Conference on Computer Vision*, 1996, 329–342.

[24] Khan, Z., Balch, T., and Dellaert, F.
A Rao-Blackwellized particle filter for Eigen Tracking.
In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2004, 930–986.

[25] Isard, M., and Blake, A.
Contour tracking by stochastic propagation of conditional density.
In *Proceedings of the European Conference on Computer Vision*, 1996, 343–356.

[26] Cheung, G. K-M., Baker, S., and Kanade, T.
Shape-from-silhouette across time. Part I: Theory and algorithms.
International Journal on Computer Vision, **62**, 3 (2005), 221–247.

[27] Yilmaz, A., Javed, O., and Shah, M.
Object tracking: A survey.
ACM Computing Surveys, **38**, 4 (2006), 13.

[28] Runnalls, A.
Monte Carlo Markov chain methods for tracking.
In *Proceedings of IEE Colloquium on Algorithms for Target Tracking*, London, May 16, 1995, 5–8.

[29] Li, X., and Jilkov, V.
Survey of maneuvering target tracking. Part V: Multiple-model methods.
IEEE Transactions on Aerospace and Electronic Systems, **41**, 4 (2005), 1255–1321.

[30] Doucet, A., Gordon, N., and Krishnamurthy, V.
Particle filters for state estimation of jump Markov linear systems.
IEEE Transactions on Signal Processing, **49**, 3 (2001), 613–624.

[31] Ghahramani, Z., and Hinton, G.
Variational learning for switching state-space models.
Neural Computation, **12**, 4 (1998), 963–996.

[32] Oh, S. M., Rehg, J. M., Balch, T., and Dellaert, F.
Data-driven MCMC for learning and inference in switching linear dynamic systems.
In *Proceedings of 20th National Conference on Artificial Intelligence*, 2005.

[33] Oh, S. M., Rehg, J. M., Balch, T., and Dellaert, F.
Parameterized duration modeling for switching linear dynamic systems.
In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, 1694–1700.

[34] Blom, H. A. P., and Bar-Shalom, Y.
The interacting multiple model algorithm for systems with Markovian switching coefficients.
IEEE Transactions on Automatic Control, **33** (1988), 780–783.

[35] Yuan, X., Han, C., Duan, Z., and Lei, M.
Adaptive turn rate estimation using range rate measurements.
IEEE Transactions on Aerospace and Electronic Systems, **42**, 4 (2006), 1532–1541.

[36] Yedidia, J., Freeman, W., and Weiss, Y.
Understanding belief propagation and its generalizations.
MERL, Technical Report TR2001-022, 2001.

[37] Blom, H. A. P., and Bloem, E. A.
Exact Bayesian and particle filtering of stochastic hybrid systems.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 1 (2007), 55–70

[38] McGinnity, S., and Irwin, G.
Maneuvering target tracking using a multiple-model bootstrap filter.
In A. Doucet, N. de Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, New York: Springer, 2001, ch. 23, 479–497.

[39] Fan, X., Qi, C., Liang, D., and Huang, H.
Probabilistic contour extraction using hierarchical shape representation.
In *Proceedings of International Conference on Computer Vision*, vol. 1, 2005, 302–308.

[40] Isard, M.
PAMPAS: Real-valued graphical models for computer vision.
In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003, 613–620.

[41] Sigal, L., Bhatia, S., Roth, S., Black, M. J., and Isard, M.
Tracking loose-limbed people.
In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, 421–428.

[42] Pérez, P., Vermaak, J., and Blake, A.
Data fusion for visual tracking with particles.
In *Proceedings of IEEE*, **92**, 3 (2004), 495–513.

[43] Deutscher, J., and Reid, I.
Articulated body motion capture by stochastic search.
International Journal on Computer Vision, **61**, 2 (2005), 185–205.

[44] Gordon, N., Salmond, D., and Smith, A.
Novel approach to nonlinear/non-Gaussian Bayesian state estimation.
IEE Proceedings of Radar Signal Processing, **140** (1993), 107–113.

[45] Doucet, A., and Ristic, B.
Recursive state estimation for multiple switching models with unknown transition probabilities.
IEEE Transactions on Aerospace and Electronic Systems, **38**, 3 (2002), 1098–1104.

[46] Singer, R.
Estimating optimal tracking filter performance for manned maneuvering targets.
IEEE Transactions on Aerospace and Electronic Systems, **6**, 4 (1970), 473–483.

[47] Blackwood, O. H.
General Physics (2nd ed.).
New York: Wiley, 1955.

[48] Zhou, H., and Kumar, K.
A ‘current’ statistical model and adaptive algorithm for estimating maneuvering targets.
AAAI Journal on Guidance, **7**, 5 (1984), 596–602.

[49] Tu, Z., and Zhu, S.
Image segmentation by data-driven Markov chain Monte Carlo.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**, 5 (2002), 657–673.

[50] Ristic, B., Arulampalam, M., and Gordon, A.
Beyond Kalman Filters: Particle Filters for Target Tracking.
Norwood, MA: Artech House, 2004.

[51] Doucet, A., de Freitas, N., and Gordon, N. (Eds.)
Sequential Monte Carlo Methods in Practice.
New York: Springer, 2001.

[52] Liu, J. S.
Monte Carlo Strategies in Scientific Computing.
New York: Springer, 2001.

[53] Liu, J., and Chen, R.
Sequential Monte-Carlo methods for dynamic systems.
Journal of the Royal Statistical Society, Series B, **93** (1998), 1032–1044.

[54] Cappé, O., Moulines, E., and Rydén, T.
Inference in Hidden Markov Models.
New York: Springer, 2005.

[55] Robert, C. P., and Casella, G.
Monte Carlo Statistical Methods (2nd ed.).
New York: Springer, 2004.

[56] Gilks, W. R., and Berzuini, C.
Following a moving target—Monte Carlo inference for dynamical Bayesian models.
Journal of Royal Statistical Society, Series B, **63**, 1 (2001), 127–146.

[57] Berzuini, C., and Gilks, W. R.
Resample-move filtering with cross-model jumps.
In A. Doucet, N. de Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, New York: Springer, 2001.

[58] MacEachern, S. N., Clyde, M., and Liu, J.
Sequential importance sampling for nonparametric Bayes models: The next generation.
Canadian Journal of Statistics, **27** (1999), 251–267.

[59] Fearnhead, P.
Markov chain Monte Carlo, sufficient statistics and particle filter.
Journal of Computational and Graphical Statistics, **11**, 4 (2002), 848–862.

[60] Khan, Z., Balch, T., and Dellaert, F.
MCMC-based particle filtering for tracking a variable number of interacting targets.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 11 (Nov. 2005), 1805–1918.

[61] Tierney, L.
Markov chains for exploring posterior distributions (with discussion).
Annals of Statistics, **22**, 4 (1994), 1701–1762.

[62] Doucet, A., Logothetis, A., and Krishnamurthy, V.
Stochastic sampling algorithms for state estimation of jump Markov linear systems.
IEEE Transactions on Automatic Control, **45**, 2 (Feb. 2000), 188–202.

[63] Everett, N., Tham, S-S., and Salmond, D.
A novel MCMC tracker for stressing scenarios.
In *Proceedings of the SPIE Conference on Signal and Data Processing of Small Targets*, 2004.

[64] Oh, S., Russell, S., and Sastry, S.
Markov chain Monte Carlo data association for general multiple-target tracking problems.
In *Proceedings of the 43rd IEEE Conference on Decision and Control*, vol. 1, 2004, 735–742.

[65] Oh, S., Kim, J., and Sastry, S.
A sampling-based approach to nonparametric dynamic system identification and estimation.
In *Proceedings of the 2004 American Control Conference*, vol. 1, 2004, 873–879.

[66] Bar-Shalom, Y., and Li, X.
Multitarget-Multisensor Tracking: Principles and Techniques (1st ed.).
Storrs, CT: YBS Publishing, 1995.

[67] Chan, L. A., and Nasrabadi, N. M.
Clutter rejection using Eigenspace transformation.
Army Research Laboratory, Adelphi, MD, Technical Report, 1999.

[68] Rosenhahn, B., and Sommer, G.
Pose estimation in conformal geometric algebra, Part I: The stratification of mathematical spaces.
Journal of Mathematical Imaging and Vision, **22**, 1 (2005), 27–48.

[69] Rosenhahn, B., and Sommer, G.
Pose estimation in conformal geometric algebra, Part II: Real-time pose estimation using extended feature concepts.
Journal of Mathematical Imaging and Vision, **22**, 1 (2005), 49–70.

[70] Hue, C., Le Cadre, J-P., and Pérez, P.
Posterior Cramer-Rao bounds for multi-target tracking.
IEEE Transactions on Aerospace and Electronic Systems, **42**, 1 (2006), 37–49.

[71] Puranik, S., and Tugnait, J. K.
Tracking of multiple maneuvering targets using multiscan JPDA and IMM filtering.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 1 (2007), 23–35.

Xin Fan (M'06) was born in 1977. He received the B.E. and Ph.D. degrees in information and communication engineering from Xi'an Jiaotong University, China, in 1998 and 2004 respectively.

From May 2006 to December 2007, he worked in Oklahoma State University as a postdoctoral research fellow. In January 2008, he moved to the University of Texas Southwestern Medical Center, Dallas, for his second postdoctoral training. Since 2005, he has also been an assistant professor in Dalian Maritime University, China. His research interests include Bayesian estimation and machine learning, and their applications to target tracking, image super-resolution, and DTI-MR image analysis.

Guoliang Fan (S'97—M'01—SM'05) received his B.S. degree in automation engineering from Xi'an University of Technology, Xi'an, China, M.S. degree in computer engineering from Xidian University, Xi'an, China, and Ph.D. degree in electrical engineering from the University of Delaware, Newark, in 1993, 1996, and 2001, respectively.

From 1996 to 1998, he was a graduate assistant in the Department of Electronic Engineering, Chinese University of Hong Kong. Since 2001, he has been with the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater, where he is currently an associate professor. His research interests are image processing, computer vision, biomedical imaging, and remote sensing applications.

Dr. Fan was awarded the First Prize in 1997 IEEE Hong Kong Section Postgraduate Student Paper Contest and the First Prize in 1997 IEEE Region 10 (Asia-Pacific) Postgraduate Paper Contest. He is a recipient of the National Science Foundation (NSF) CAREER award (2004). He received the Halliburton Excellent Young Teacher Award (2004) and Halliburton Outstanding Young Faculty award (2006) from OSU.

Joseph P. Havlicek (S'84—M'88—SM'98) received the B.S. and M.S. degrees from Virginia Polytechnic Institute and State University, Blacksburg, and the Ph.D. degree from the University of Texas at Austin, all in electrical engineering.

He was a software developer with Management Systems Laboratories, Blacksburg, VA, from 1984 to 1987 and with IBM, Austin, TX, throughout 1993 where he was affiliated with Ralph Kirkley Associates. From 1987 to 1997 he was with the Naval Research Laboratory, Washington, D.C., where he was associated with SFA, Inc., from 1987 to 1989. He was a recipient of the Department of the Navy Award of Merit for Group Achievement in 1990 for this work. He joined the School of Electrical and Computer Engineering at the University of Oklahoma in 1997 where he currently holds the rank of professor. His research interests include signal, image, and video processing and intelligent transportation systems.

Dr. Havlicek has been a member of the organizing committee of the IEEE Southwest Symposium on Image Analysis and Interpretation since 1998, serving as technical program cochair in 2004, 2006, and 2008, as well as general cochair for 2010. He served as publications chair on the organizing committee of ICIP 2007. He is a member of Tau Beta Pi, Phi Kappa Phi, and Eta Kappa Nu.