Fusion of Nonlinear Motion Dynamics Using Fokker-Planck Equation and Projection Filter

Gang Qian, Khurram Shafique*, Ping Wang
ObjectVideo, Inc.
11600 Sunrise Valley Drive
Reston, VA 20191 USA
{gqian, kshafique, pwang}@objectvideo.com

Abstract—This paper presents a novel approach to the fusion of nonlinear motion dynamics for 2D target tracking applications in video analytics by using Fokker-Planck equation (FPE) and the projection filter. The motion dynamics of the target is first conveniently represented using the corresponding FPE for the state posterior. The motion dynamics is then fused into target tracking process to reliably predict the target position and velocity in 2D target tracking by solving the corresponding FPE of the state posterior using a projection filter. Once the next system measurement is available, the state posterior is then updated using the Bayes' rule in the projection filter framework as well. Experiments using synthetic and real aerial surveillance video data show that the proposed FPE-based target tracker is able to reliably track targets in the present of nonlinear motion dynamics, and that the proposed FPE-based tracker outperforms traditional nonlinear filters in target tracking such as the Kalman filter (including extended Kalman filter and unscented Kalman filter) and the particle filter.

Keywords—motion dynamics; Fokker-Planck equation; projection filter; nonlinear filtering; target tracking

I. INTRODUCTION

Target tracking from video is an integral component of video analysis and it finds important applications in security, surveillance, and sports video analysis. Target tracking from video is essentially a state (e.g., position and velocity) estimation problem of a dynamic system from observations (e.g., detected possible target locations). In the presence of nonlinearities in the system dynamics (e.g., target motion) and measurements, it becomes challenging to obtain reliable target tracking. Fortunately in many situations, the dynamics of the target motion is available because of the underlying constraints imposed to the target motion by specific applications. For example, in vehicle tracking from aerial videos, the motion trajectories of the vehicles usually overlap with the road network because in most cases vehicles travel on road. A rough target motion dynamic model in the form of a system state transition equation can be constructed either by integrating the geometric shape of the road network and the speed limits information or learned from training data. The availability of such target motion dynamics is valuable in robust target tracking from video in the sense that it can be used to predict the system state from the current time instant to the next time instant. On the other hand, such motion dynamic models are often highly nonlinear. The key challenge is to find an effective fusion mechanism to fuse the system dynamics in the tracking framework.

In the nonlinear state estimation and filtering literature, there are three major approaches trying to tackle this challenge, including the family of Kalman filters (KF) [1, 2], the particle filters [3-5], and projection filters [6, 7]. The KF family, including the original KF, the extended KF, and the unscented KF, attempts to represent the state probability density function using the mean and covariance. Different approximation techniques such as Taylor expansion and unscented transformation have been introduced to propagate the mean and covariance the system state posterior through the nonlinear system transition equation. In fact, the KF family is a special case of a more general class of nonlinear filters, often known as the closure-filters [8-10], which try to characterize the state posterior density function using a set of its low-order moments. KF is a closure-filter utilizing only the first order (mean) and the second order (covariance) moments of the state posterior density function. Although higher order moments can be included, the closure-filters often suffer from similar poor tracking performance and divergence problems as the KF family does in the presence of high nonlinearity in the system dynamics and measurements when the state posterior density cannot be simply characterized by a limited set of moments.

Over the past two decades stochastic sampling-based approaches such as particle filters have been proposed to further tackle this challenge and have been widely used for density estimation and filtering for nonlinear dynamic systems. In such stochastic sampling-based approaches the state posterior density function is approximated by a large set of weighted samples. These samples are properly weighted so that their weighted mean asymptotically approaches the actual mean of the state posterior (i.e., the minimum mean square error estimate of the unknown state) when the sample size goes to infinity. The sequential importance sample (SIS) procedure is used to propagate samples of the system state posterior between measurements. In SIS, the system dynamics in the form of the conditional density of current state given the
previous state is used to generate new samples for the current time instant. Then the current observation is used to update the weights of the current state samples. An additional resampling step is often necessary after weight update using the most recent observation to combat the sample degeneration problem and keep samples in the more probable areas in the state space. To obtain a decent approximation of the state posterior density, usually a large amount of samples are required, depending on the dimensionality of the state space.

Projection filter is another powerful nonlinear filtering technique. In projection filter, the state density function is projected on a basis function space and approximated by a linear combination of such basis functions. Projection filters are often realized using the Galerkin method \([6, 7, 11]\), which is a classical method for converting a continuous problem (e.g., a differential equation) to a discrete problem, represented in a function subspace spanned by a finite set of basis functions. In state prediction, the combination coefficients of the predicted density can be found by solving the Fokker-Planck equation (FPE) that describes the state dynamics of the system. In state update, the related coefficients of the state posterior density can be solved using the Bayes’ rule. Many numerical methods for solving the FPE are available. Among them, the most commonly used methods are grid-based methods that use a grid of points in state space and time to approximate the target density. Nevertheless, such grid-based methods face a fundamental dilemma, i.e., the grid must be large enough to cover different possibilities of the state and yet dense enough to yield good approximation. If a fixed grid is used, it must be defined purely according to prior information. This fundamental dilemma inherently leads to a curse of dimensionality for high-dimensional problems. To tackle this challenge, in this paper we propose to use a grid adaptation method that that moves the grid around based on the estimated distribution. In our research, we implemented a projection filter using an adaptive-grid finite element method. In our experiments of target tracking using the projection filter, it has been observed that the FPE-based projection filter produces better tracking results than the KF family and the particle filters.

II. APPROACH

A. System Model and Fokker-Planck Equation

In our implementation, we have focused on nonlinear density filtering for target tracking from video in mixed-time (i.e., with continuous-time dynamics and discrete-time observations). The mixed-time setup is not only efficient but is also closer to the real systems where the dynamics are more accurately described in continuous time and the use of digital systems requires discrete-time observations.

In target tracking from video, the state vector at time \(t\) is \(x_t = (u_t, v_t)\), where \(u_t\) and \(v_t\) are the 2D position and velocity of the target in the 2D image pixel coordinate system. The nonlinear system dynamic of the target motion can be described by a stochastic differential equation (SDE) as the following

\[
dx_t = f(t, x_t)dt + G(t, x_t)dw_t, \quad t \geq t_0
\]

where \(\{w_t, t \geq t_0\}\) is the Wiener process with \(E[\text{d}w(t)\text{d}w(t)'] = Q(t)dt\). Let \(z_k\) be the observation at discrete-time \(t_k\)

\[
z_k = h(x_{t_k}, t_k) + n_k
\]

where \(\{n_k, k \geq 1\}\) is a white Gaussian dynamic noise sequence independent of \(dw_t\) with covariance \(R_k\). For example, in target tracking from video, the observation \(z_k\) can be the observed possible target location from a target detector. It can also be a small image patch centered at the candidate target location. Different types of observations require different measurement equations. Define the observation sequence up to \(t\) as \(Z_t = \{z_k, t_k \leq t\}\). The goal of nonlinear filter is to find the state posterior \(p(x_t|Z_t) \equiv p(t, x|Z_t)\) based on the system dynamic and measure equations from the observation sequence.

Between two adjacent observation time \(t_k\) and \(t_{k+1}\) (i.e., after the arrival of \(z_k\) but before the arrival of \(z_{k+1}\)), the evolution of the system posterior density \(p(t, x|Z_t)\) is completely governed by the following Fokker-Planck equation (FPE) for the state posterior density, also known as Kolmogorov forward equation. FPE specifies that

\[
\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial (pf_i)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 [p(GG^T)_{ij}]}{\partial x_i \partial x_j}
\]

where \(p \equiv p(x_t|Z_t)\). The FPE (3) can be derived from the system SDE (2) using the Feynman-Kac formula. By solving the FPE (3), the predicted density \(p(t_{k+1}, x|Z_{t_k})\) can be approximated by using the previous posterior \(p(t_k, x|Z_{t_k})\) as the initial condition. Once the observation \(z_{k+1}\) at \(t_{k+1}\) becomes available, it is then used to update the predicted density to obtain the posterior \(p(x_{k+1}|Z_{k+1})\) using the Bayes’ rule.

\[
p(t_{k+1}, x|Z_{t_{k+1}}) = \frac{p(z_{k+1}|x)p(t_{k+1}, x|Z_{t_k})}{\int p(z_{k+1}|x)p(t_{k+1}, x|Z_{t_k}) dx}
\]

B. State Posterior Prediction by Sovling FPE Using Galerkin Method

In nonlinear projection filter, the above prediction step is often achieved by solving the system FPE using the Galerkin method. In our research, we have mainly followed \([6]\) and \([7]\) to derive the projection filter solution for the FPE using the Galerkin method. The Galerkin method assumes that the underlying density function can be well approximated as a linear combination of a set of basis functions and then solves for the corresponding combination coefficients by minimizing the \(L_2\) norm of the FPE residual projected onto the basis function set. In this way, the original FPE as a nonlinear partial differential equation (PDE) in both time and space can be simplified into a linear ordinary differential equation.
linear ODEs and can be further written in matrix notation.

\[
p_N(t,x) = \sum_{i=0}^{N-1} c_i(t) \phi_i(x) \tag{5}\]

where \(p_N(t,x)\) is the density approximation using the set of basis functions \(\{\phi_i(x), l = 0, \ldots, N - 1\}\) \cite{7}. In practice, different types of basis functions have been used, including the Fourier basis, cosine basis, and the linear nodal basis functions. In our implementation, the following linear nodal basis function shown in equation (6) has been adopted.

\[
\phi_i(x_m) = \begin{cases} 
1 & \text{if } m = l \\
0 & \text{otherwise}
\end{cases} \tag{6}
\]

where \(x_m\)'s are the grid points. Such linear nodal basis function is one of the commonly used linear basis functions in finite element analysis. The optimal coefficients \(c_i(t)\) can be found by minimizing the energy of the projection of the FPE residual onto the basis function space. Hence the optimization problem is written in matrix notation as

\[
\int_{\Omega} \left( \frac{\partial p_N}{\partial t} + \sum_{l=1}^{n} \frac{\partial (p_N f_l)}{\partial x_l} \right) \phi_q dx = 0 \tag{7}
\]

for \(q = 0, \ldots, N - 1\), where \(\Omega\) is the support of the basis functions. The support is given by a grid defined in the state space, represented by a set of nodes and mesh elements. After interchanging summation and differentiation, (6) becomes

\[
\sum_{i=0}^{N-1} c_i \int_{\Omega} \phi_i \phi_q dx = 0
\]

\[
= \sum_{i=0}^{N-1} c_i \left\{ - \sum_{j=1}^{n} \int_{\Omega} \frac{\partial [\phi_i f_j]}{\partial x_j} \phi_q dx + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{\Omega} \frac{\partial^2 [p_N (GQG^T)_{ij}]}{\partial x_i \partial x_j} \phi_q dx \right\} \tag{8}
\]

for \(q = 0, \ldots, N - 1\). Equation (7) is basically a system of \(N\) linear ODEs and can be further written in matrix notation. Define the following vector and matrices involved in solving the FPE.

\[
c = [c_0, \ldots, c_{N-1}]^T,
\]

\[
[M]_{i,j} = \int_{\Omega} \phi_i \phi_j dx,
\]

\[
[A_1(t)]_{ij} = -\sum_{k=1}^{n} \int_{\Omega} \frac{\partial [\phi_j f_k]}{\partial x_k} \phi_i dx,
\]

\[
[A_2(t)]_{ij} = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} \int_{\Omega} \frac{\partial^2 [\phi_j (GQG^T)_{kl}]}{\partial x_k \partial x_l} \phi_i dx
\]

Then the linear ODE system (7) can be written as

\[
[M - \frac{\Delta t}{2} A(t)] c(t_k + \Delta t) = \left( M + \frac{\Delta t}{2} A(t) \right) c(t_k)
\]

This linear system can be easily derived by first formulating the Taylor series expansion of \(c(t_k + \Delta t)\) over \(t\), and then omitting the second and higher terms of \(\Delta t\). In practice, when the time interval between \(t_k\) and \(t_{k+1}\) is large, a number of iterative predictions need to be carried out using small time step size \(\Delta t\) to reduce the approximation error introduced by removing the second and higher order terms in the Taylor series expansion. In our experiments, we used five iterative steps for the synthetic data and 15 iterative steps for the real data since the real data has a lower frame rate than the synthetic data.

C. State Posterior Update Using Bayes’ Rule

When a new observation becomes available at \(t_{k+1}\), the combination coefficients of the approximation of the posterior \(p(t_{k+1}, x|Z_{t_{k+1}})\) can be found using the Bayes’ rule by applying the Galerkin method and replace \(p(t_{k+1}, x|Z_{t_{k+1}})\) by \(P_N(t_{k+1}, x|Z_{t_{k+1}})\). Define the following updated approximate posterior

\[
p(t_{k+1}, x|Z_{t_{k+1}}) = \frac{p(z_{k+1} | x)p(x, t_{k+1} | Z_{t_k})}{\int_{\Omega} p(z_{k+1} | x)p(x, t_{k+1} | Z_{t_k}) dx}
\]

and project both sides of the equation onto the basis function set \(\{\phi_i(x), l = 0, \ldots, N - 1\}\). We have the following linear system with linear equations with \(c_i(t_{k+1}), i = 1, \ldots, N - 1\) as the unknowns.

\[
\sum_{i=0}^{N-1} c_i(t_{k+1}) \int_{\Omega} \phi_i \phi_q dx = \frac{\sum_{i=0}^{N-1} c_i(t_{k+1}) \int_{\Omega} p(z_{k+1} | x) \phi_i \phi_q dx}{\sum_{i=0}^{N-1} c_i(t_{k+1}) \int_{\Omega} p(z_{k+1} | x) \phi_i dx} \_tag{13}
\]

for \(q = 0, \ldots, N - 1\). Likewise, this linear system can be written in matrix notation as

\[
c(t_{k+1}) = \frac{M^{-1}\int_{\Omega} p(z_{k+1} | x) \phi_i \phi_q dx}{\sum_{i=0}^{N-1} \int_{\Omega} p(z_{k+1} | x) \phi_i \phi_q dx} \tag{14}
\]

where
\[
[Y(z_{k+1})]_{q,t} = \int_Y p(z_{k+1}|x) \phi_q(x) \, dx \\
[u(z_{k+1})]_t = \int_I p(x_{k+1}|x) \phi_t(x) \, dx
\] (15)

By using the above prediction-update recursion, the posterior density \( p(t, x|Z_t) \) can be propagated and updated over time.

D. Projection Filtering using Adaptive Mesh

In our research, we employed an adaptive-grid finite element method to solve for the combination coefficients. The goal of utilizing adaptive grid is to refine the grid used for density approximation. Our grid adaptation algorithm consists of two steps: mesh trimming and mesh expansion. In the mesh trimming step, the mesh grid (i.e., \( \Omega \) in the Galerkin method) are adaptively trimmed according to the current approximated density. In the mesh expansion step, the mesh grid is expanded

![Fig. 1. Tracking results of a target following a nonlinear 2D sinusoidal motion model. Subplot (a) shows the ground-truth and measured target trajectories as well as tracking results obtained using KF, EKF, UF, particle filter and the FPE tracker. Subplot (b) shows the average tracking errors from various trackers. Subplots (c) through (g) show the predicted target location density at five time steps between frames 33 and 34. The updated posterior of the target location at frame 34 is shown in subplot (h). The marker arrows in these density plots indicate the ground truth target locations.](image-url)
according to the system dynamics to create a proper support for the next predicted density.

The following further explains the two steps in detail. During the tracking process using FPE, the approximated density evolves over time. After one iteration of prediction, the approximated density shifts from the previous estimation. As a result, not the entire mesh grid is necessary to present the approximated density. In other words, the current mesh grid can be trimmed. Given the current approximated density represented by the nodes, mesh elements, and the coefficients of the corresponding nodal basis functions, the mean and covariance matrix of the approximated density can be estimated. In fact, it can be shown that the coefficient computed for the nodal basis function in equation (6) at any node \( x_m \) is identical to the approximated density value at the node. According to the covariance matrix of the approximated density, the necessary support of the underlying density can be estimated. In our approach, the minimum bounding box \( \Omega_\kappa \) covering the \( \kappa \)-sigma ellipse of the estimated density is taken to approximate the necessary support. In our experiment, \( \kappa = 4 \). The part of the current mesh grid that is outside \( \Omega_\kappa \) is then trimmed. In the mesh expansion step, the current mesh grid is expanded according to a predicted displacement vector. Given the system dynamics, e.g., equation (1), the displacement vector from the current time instant to the next time instant can be approximated using the motion model at the mean location of the current density. The current mesh grid is then expanded in the direction and magnitude of the displacement vector. This adaptive-grid algorithm effectively maintains the necessary

Fig. 2. Tracking results of a target following a nonlinear zigzag motion model. Subplot (a) shows the ground-truth and measured target trajectory as well as tracking results obtained using KF, EKF, UF, particle filter and the FPE tracker. Subplot (b) shows the average tracking errors from various trackers. Subplots (c) to (g) show the predicted densities of target location at five time steps between frames 37 and 38. The updated posterior of the target location at frame 38 is shown in subplot (h). The marker arrows in these density plots indicate the ground truth target locations.
mesh grid support for the approximated density during the FPE tracking process.

III. EXPERIMENTAL RESULTS

We have tested the FPE-based nonlinear projection filter using both synthetic and real video data.

A. Experiments Using Synthetic Data

In Experiment 1, a nonlinear 2D sinusoidal motion model was used to generate synthetic ground-truth and noisy measurement of a target moving trajectory. The measurement data rate is 20 Hz. We obtained the tracking results using KF, EKF, UF, PF, and the FPE-based projection filter. All filters were initialized using the truth data. The PF used is a basic bootstrapping filter with 10,000 samples. For the FPE-filter, five iterative prediction steps were applied to propagate the posterior between observations. Figure 1 (a) shows the ground-truth data, noisy observation data, and tracking results from different trackers. The average tracking errors over time at different frames from all the trackers are shown in Figure 1 (b). It can be seen that the FPE tracker outperforms the KF family and the particle filter. Figures 1(c) to 1(g) present the predicted density of the target location at five time steps between frames 33 and 34. The units of the x and y axes in these plots are in pixels. Finally Figure 1(h) shows the updated posterior of the target location at frame 34 once the new observation is available. The marker arrows in these density plots indicate the ground truth target locations.

In Experiment 2, a nonlinear zigzag motion model was used to generate the synthetic target trajectory. Similar to Experiment 1, five iterative prediction steps using the FPE-filter were applied to propagate the posterior between observations. The tracking results, average tracking errors, and sample predicted and updated target location densities are given in Figure 2. It can be seen from Figure 2(b) that the tracking result using the FPE tracker is superior to those from the other trackers. On the other hand, in our research FPE was observed to be more computational expensive than the other filters.

B. Experiments Using Real Data

We also used real wide area aerial surveillance video data to test the implemented FPE tracker. The data contained both measurement nonlinearities (from the motion of the sensor) and the dynamic nonlinearities. In our experiment, our test video contains 16 frames of aerial images over 8 seconds. Each aerial image is in the resolution of 744x1520. The scene captured by this aerial surveillance video has a road network. In our experiment, the geometric shape of the road network in an urban environment is represented by various linear and nonlinear models (e.g., using Bezier curves). For each road, a specified motion model is adopted to describe the motion dynamics of the vehicles travelling on the road. Based on the current location of the target vehicle in the road network, the corresponding dynamic model can be adopted and used for target tracking using the proposed FPE tracker. In our experiment, we successfully applied the FPE tracker to track a vehicle as shown in Figure 3. In this experiment, 15 iterative prediction steps using the FPE-filter were applied to propagate the posterior between two frames of observations. Figure 3(a) presents a sample image frame superimposed with the bounding boxes of the tracked vehicle provided as the observation. This observation is noisy. Figure 3(b) shows the zoomed-in version of the image patch containing the target vehicle. Figure 3(c) shows the FPE tracking result (red crosses), which demonstrates that the proposed FPE tracker is able to reliably track moving vehicles from real aerial surveillance videos using the road models. The tracking result from Kalman filter is also shown in Figure 3(c) as green circles. It can be seen that in this case the tracking results from FPE and KF are very similar to each other due to the simple liner road model. In our experiments, we have also used additional real aerial video data to test the proposed FPE tracker. Using these additional videos, we have observed that
the proposed FPE tracker was able to successfully track vehicles following a more complicated travel pattern in the road network, e.g., making turns at road intersections.

IV. CONCLUSIONS AND FUTURE WORK

Our research shows that the proposed approach using Fokker-Planck equation and projection filter is effective to model motion dynamics and conduct posterior propagation and update for 2D target tracking in video analytics. Experimental results obtained using synthetic and real aerial surveillance video data show that the proposed FPE-based target tracker can reliably track targets in the present of nonlinear motion dynamics. It has been observed form the experiments that the FPE-based projection filter produces better tracking results than the KF family and the particle filters. In the presence of highly nonlinear system dynamics, the FPE-based tracker is expected to work better than the KF filters and particle filters, as the expense of higher computational cost. One of the limitations of the current tracking framework is that it tracks single targets and assumes that data association has been done separately by a data association module. In our future work, we would like to extend the proposed framework to handle multiple target tracking and integrate data association into the tracking framework, for example, by adopting a more general observation model that also takes into account of appearance similarity of the targets.

ACKNOWLEDGMENT

This research was supported by the Air Force Research Laboratory (AFRL), under contract #FA8750-12-C-0058. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

REFERENCES