Variational Bayesian object tracking and smoothing

BMVC 2011 Submission # 193

Abstract

Probabilistic models for visual tracking of objects often involve likelihoods that lead to intractable integrals in Bayesian inference. Vermaak *et al*. (2003) introduced an approximation algorithm that combines variational Bayesian approximation with importance sampling in an EM-like algorithm. Their diffusion model is a linear dynamical system and the algorithm is an efficient approximation to the Kalman filter. In this paper we show that their model does not lead to an effective Kalman smoother and propose a new velocity model using a similar Kalman filter and a new Kalman smoother algorithm. This is shown to produce a smoother track for an object in a real video example.

1 Introduction

The aim of video tracking is to identify one or more objects in a video sequence and then to trace their movement through successive frames to identify the path taken and perhaps to identify behaviour or interactions. Our interest lies in tracking crickets (*Gryllus campestris*) in their natural habitat (see figure 2) [1]. In these videos the camera is static, though the background is subject to small variations due to changing light conditions and the effect of natural forces such as wind.

Probabilistic state space models have become popular for video tracking and Bayesian inference may be used to determine posterior distributions for the state of the tracked object $\mathbf{x}_t$ in frame t given the observations $\mathbf{y}_t$ (the video frames) up to t, as the *filtered state*, $\mathrm{p}(\mathbf{x}\,|\,\mathbf{y}_{1:t})$. The integrals involved in the Bayesian formulation are often analytically intractable, requiring an approximation scheme to be used. One popular scheme is particle filtering [5, 6, 8]. It is simple to implement, but quickly becomes computationally expensive as the number of dimensions increases and it is not robust where the likelihood is very sharply peaked, as it is in this video tracking problem. An attractive alternative, introduced by Vermaak *et al*. [2003], is a variational Bayesian approximation, in which the posterior densities are approximated by a product of simpler parametrized densities, as briefly reviewed in section 1.1. Vermaak *et al*. model the uncertainty around the central state by a Gaussian density whose mean diffuses from timestep to timestep. The linear-Gaussian formulation means that the expectations required for the variational approximation associated with the state evolution are equivalent to the efficient Kalman filter recursions. These are combined with a novel importance sampling method to evaluate expectations associated with the highly nonlinear likelihood term.

In a retrospective analysis information from observations after $\mathbf{y}_t$ can be incorporated to estimate the *smoothed state* by $\mathrm{p}(\mathbf{x}_t\,|\,\mathbf{y}_{1:T})$. However, we find straightforward extension of the

Vermaak *et al*. scheme is ineffective because the diffusion variance necessary to capture the changes in the tracked object's state is so large that the future observations make negligible contribution to the estimate of the current state. In this paper we (a) show how the diffusive model may be replaced with a 'constant velocity' model for the state evolution accounting separately for the uncertainty in the state and its dynamics and (b) describe a variational approximation scheme that combines past and future observations in an even-handed manner to permit effective smoothing.

We start by giving a very brief summary of the variational Bayes method. In section 2, we summarise the diffusion model described by Vermaak *et al*. [2003] and show why the straightforward Kalman smoothing recursions are not effective here. In section 3 the new velocity model is introduced, including the new approximate Kalman smoother. Section 4 illustrates the effect of the new algorithm on the tracking of crickets in real videos and conclusions are drawn in section 5.

1.1 Variational Bayes

Very briefly, variational Bayes (for tutorials see [7, 10] and [4, chapter 10]) seeks approximate posterior distributions $\mathrm{q}(\Omega_i) \approx \mathrm{p}(\Omega_i \,|\, \mathbf{y}_{1:T})$ (where $\Omega_i \in \Omega$ represents one of the model's parameter variables) that minimise the Kullback-Leibler (KL) divergence [9] between q and p, where

$$\mathrm{KL}(\mathrm{q} \,\|\, \mathrm{p}) = \int \mathrm{q}(\Omega) \log\left(\frac{\mathrm{q}(\Omega)}{\mathrm{p}(\Omega \,|\, \mathbf{y}_{1:T})}\right) d\Omega \tag{1}$$

The KL divergence is non-negative, and only zero when q and p are equal. An elegant method provided by Waterhouse *et al*. [13] (see also [1, 2, 3]) exploits the assumed factorisation of the approximate posterior and leads to

$$\log(\mathrm{q}(\Omega_i)) = \mathbb{E}_{j \neq i}[\log(\mathrm{p}(\mathbf{y}_{1:T}, \Omega))] \tag{2}$$

where the expectation on the right-hand side is with respect to all variables other than Ω_i. When conjugate priors are chosen for the variables Ω_i the resulting posterior distributions are of the same family and their parameters are expressed in terms of the expectations of the other variables in the problem. Suitable posterior parameter values are found by iteratively calculating each of the $\mathrm{q}(\Omega_i)$ in terms of the others until convergence.

2 Vermaak *et al*.'s diffusion model

With $\mathbf{y}_t$ representing the observation and $\mathbf{x}_t$ the state at time t, the prior for $\mathbf{x}_t$ is defined as

$$\mathrm{p}(\mathbf{x}_t \,|\, \mu_t, \lambda_t) = \mathcal{N}(\mathbf{x}_t \,|\, \mu_t, \lambda_t^{-1}) \tag{3}$$

where the precision variable, λ_t, represents the uncertainty of the estimate of the state and is assigned a Wishart prior (the conjugate prior for a Gaussian precision):

$$\mathrm{p}(\lambda_t) = \mathcal{W}(\lambda_t \,|\, \bar{\mathbf{Y}}, \bar{n}) \tag{4}$$

The evolution of the expectation of the state is modelled as a diffusive process:

$$\mathrm{p}(\mu_t \,|\, \mu_{t-1}, \bar{\lambda}) = \mathcal{N}(\mu_t \,|\, \mu_{t-1}, \bar{\lambda}^{-1}) \tag{5}$$

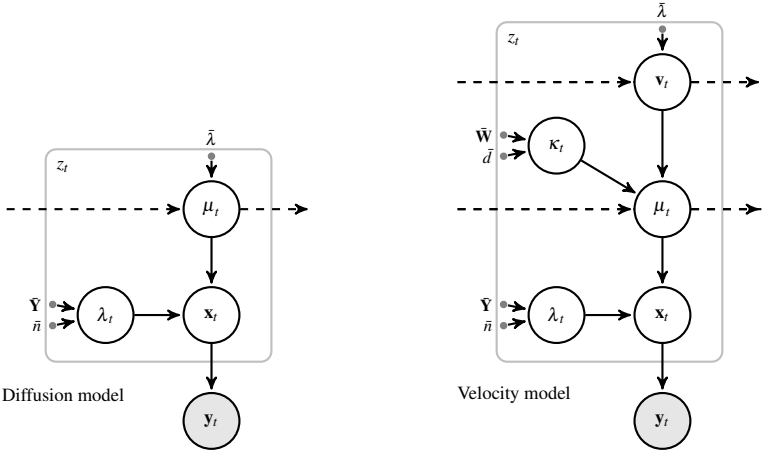

Figure 1 Graphical representation of the priors for (left) Vermaak *et al.*'s [12] diffusion model and (right) the velocity model.

where $\bar{\lambda}$ is a (fixed) parameter that defines the magnitude of the diffusion, i.e. the range of change in the state that is probable between frames. Note that the precisions in λ_t and $\bar{\lambda}$ are full matrices. A graphical representation of the model is shown in figure 1. For notational convenience the full state at time t is collected into $z_t = \{\mathbf{x}_t, \mu_t, \lambda_t\}$.

Given a possible state and different probability distributions for the colours of background and foreground pixels, the likelihood, $\mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t)$, may be calculated.

2.1 Filtering

Posterior probabilities for the state variables may be calculated using the forward recursions as follows (we define a further prior, $\mathrm{p}(\mu_0)$, to express beliefs about the initial state):

$$\mathrm{p}(z_t \,|\, \mathbf{y}_{1:t}) \propto \int \mathrm{p}(\mathbf{y}_t \,|\, z_t)\, \mathrm{p}(z_t \,|\, z_{t-1})\, \mathrm{p}(z_{t-1} \,|\, \mathbf{y}_{1:t-1})\, dz_{t-1} \tag{6}$$

This is not tractable, mainly because $\mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t)$ cannot be expressed in terms of $\mathbf{x}_t$, but also because of the interactions between the variables in z_t. So Vermaak *et al.* use the factorised variational Bayesian method [1, 3, 13] to calculate approximate posterior distributions (denoted by $\mathrm{q}(\cdot)$) and make the following assumption regarding factorisation of the posterior:

$$\mathrm{p}(z_t \,|\, \mathbf{y}_{1:T}) = \mathrm{p}(\mathbf{x}_t, \mu_t, \lambda_t \,|\, \mathbf{y}_{1:T}) \approx \mathrm{q}(\mathbf{x}_t)\, \mathrm{q}(\mu_t)\, \mathrm{q}(\lambda_t) \tag{7}$$

Following through the factorised variational Bayesian method for each variable results in the following approximate posterior distributions. In the forward sweep, with $t = 1, \ldots, T$, the variables associated with μ are updated according to the standard Kalman filter update: $\mathrm{q}^{\alpha}(\mu_t) \approx \mathrm{p}(\mu_t \,|\, \mathbf{y}_{1:t})$ is the Gaussian distribution $\mathcal{N}(\mu_t \,|\, \mathbf{m}_t^{\alpha}, \mathbf{S}_t^{\alpha})$, where

$$\mathbf{S}_t^{\alpha} = \left(\langle \lambda_t \rangle + \bar{\lambda} \right)^{-1} \qquad \mathbf{m}_t^{\alpha} = \mathbf{S}_t^{\alpha} \left(\langle \lambda_t \rangle \langle \mathbf{x}_t \rangle + \bar{\lambda} \langle \mu_{t-1} \rangle \right) \tag{8}$$

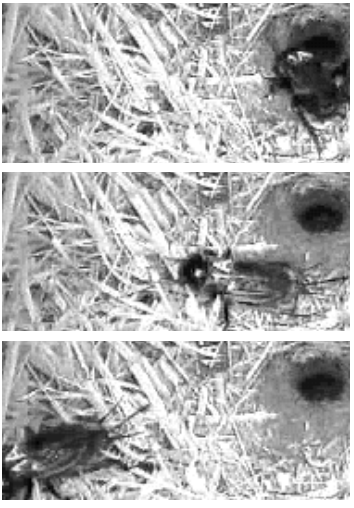

Figure 2 Three example frames from a cricket video sequence.

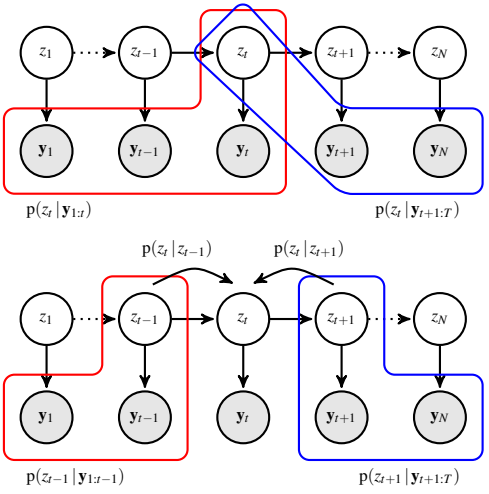

Figure 3 Graphical representations of (top) the traditional and (bottom) the new forward and backward sweeps for filtering and smoothing.

Here $\langle f(a) \rangle$ denotes the expectation of $f(a)$ with respect to the posterior distribution of a. For the posterior of the state precision, $\mathrm{q}(\lambda_t)$, the Wishart $\mathscr{W}(\lambda_t \,|\, \mathbf{Y}_t, n_t)$ is obtained, with

$$n_t = \bar{n} + 1 \qquad \mathbf{Y}_t = \left(\bar{\mathbf{Y}}^{-1} + \langle \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \rangle + \langle \mu_t \mu_t^{\mathrm{T}} \rangle - \langle \mathbf{x}_t \rangle \langle \mu_t \rangle^{\mathrm{T}} - \langle \mu_t \rangle \langle \mathbf{x}_t \rangle^{\mathrm{T}} \right)^{-1} \tag{9}$$

Standard derivations for Gaussian and Wishart distributions give the expectations

$$\langle \mu_t \rangle = \mathbf{m}_t^{\alpha} \qquad \langle \mu_t \mu_t^{\mathrm{T}} \rangle = \mathbf{S}_t^{\alpha} + \mathbf{m}_t^{\alpha} \mathbf{m}_t^{\alpha \mathrm{T}} \qquad \langle \lambda_t \rangle = n_t \mathbf{Y}_t \tag{10}$$

Due to the intractable form of the likelihood, this method cannot be used to calculate $\mathrm{q}(\mathbf{x}_t)$. Instead it is simplified to

$$\mathrm{q}(\mathbf{x}_t) \propto \mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t) \mathscr{N}(\mathbf{x}_t \,|\, \langle \mu_t \rangle, \langle \lambda_t \rangle^{-1}) \tag{11}$$

and importance sampling for $\mathbf{x}_t$ is used to give a Monte Carlo approximation for $\mathrm{q}(\mathbf{x}_t)$. With

$$\mathbf{x}_t^{(i)} \sim \mathscr{N}(\mathbf{x}_t \,|\, \langle \mu_t \rangle, \langle \lambda_t \rangle^{-1}) \qquad w_t^{(i)} = \frac{\mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t^{(i)})}{\sum_{j=1}^{N} \mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t^{(j)})} \tag{12}$$

the result is the following approximate expectations with respect to $\mathrm{q}(\mathbf{x}_t)$:

$$\langle \mathbf{x}_t \rangle \approx \sum_{i=1}^{N} w_t^{(i)} \mathbf{x}_t^{(i)} \qquad \langle \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \rangle \approx \sum_{i=1}^{N} w_t^{(i)} \mathbf{x}_t^{(i)} \mathbf{x}_t^{(i)\mathrm{T}} \tag{13}$$

where N is the number of samples.

2.2 Smoothing

In the traditional implementation of the smoothing recursions (see figure 3, top diagram), the forward recursion $\mathrm{p}(z_t \,|\, \mathbf{y}_{1:t})$ is combined with a backward recursion $\mathrm{p}(z_t \,|\, \mathbf{y}_{t+1:T})$, to give the

posterior probability distribution $\mathrm{p}(z_t \,|\, \mathbf{y}_{1:T})$. The backward recursive sweep is defined, for t from $T-1$ ($\mathrm{q}^{\alpha}(z_T)$ is conditioned on all the observations and hence is equal to $\mathrm{q}(z_T)$) down to 1, as

$$\mathrm{p}(z_t \,|\, \mathbf{y}_{t+1:T}) \propto \int \mathrm{p}(\mathbf{y}_{t+1} \,|\, z_{t+1})\, \mathrm{p}(z_{t+1} \,|\, z_t)\, \mathrm{p}(z_{t+1} \,|\, \mathbf{y}_{t+2:T})\, dz_{t+1} \tag{14}$$

The only term involving the current time-step is $\mathrm{p}(\mu_{t+1} \,|\, \mu_t)$, hence variationally $\mathrm{q}^{\beta}(\mu_t) \approx \mathrm{p}(\mu_t \,|\, \mathbf{y}_{t+1:T})$ is the Gaussian distribution $\mathcal{N}(\mu_t \,|\, \langle \mu_{t+1} \rangle, \bar{\lambda}^{-1})$. With $\mathrm{q}(\cdot)$ denoting the full approximate posterior distribution to $\mathrm{p}(z_t \,|\, \mathbf{y}_{1:T})$ (i.e. the result after the forward and backward sweeps), we have $\mathrm{q}(\mu_t) = \mathrm{q}^{\alpha}(\mu_t)\, \mathrm{q}^{\beta}(\mu_t)$, which is the Gaussian $\mathcal{N}(\mu_t \,|\, \mathbf{m}_t, \mathbf{S}_t)$ where

$$\mathbf{S}_t = \left((\mathbf{S}_t^{\alpha})^{-1} + \bar{\lambda} \right)^{-1} \qquad \mathbf{m}_t = \mathbf{S}_t^{-1} \left((\mathbf{S}_t^{\alpha})^{-1} \mathbf{m}_t^{\alpha} + \bar{\lambda} \mathbf{m}_{t+1} \right) \tag{15}$$

The approximate posterior expectation for μ_t is a average of contributions from the forward sweep ($\mathbf{m}_t^{\alpha}$) and backward sweep ($\mathbf{m}_{t+1}$), weighted according to the expected precision of each contribution. Unfortunately the backward sweep contribution is weighted by the diffusion precision $\bar{\lambda}$, which is necessarily small as a large covariance is required to model the magnitude of the state movement between video frames, while the forward sweep is weighted by an uncertainty precision, which turns out to be the relatively large precision in $(\mathbf{S}_t^{\alpha})^{-1}$. The result is that the "smoothed" posterior distribution is highly weighted in favour of the forward (filter) contribution and almost no smoothing occurs. To avoid this problem, the diffusion precision in this variational model needs to be separated from the state expectation in μ_t. This, and the necessity to track multiple objects that come into close proximity, leads us to the constant velocity model that is described in the next section.

3 Velocity model

In this new model, shown in graphical form on the right in figure 1, a new variable, $\mathbf{v}_t$, is introduced to represent the change in state between observations. The fixed diffusion precision in $\bar{\lambda}$ is now associated with this velocity variable, while the state expectation in μ_t is assigned the uncertainty precision κ_t, for which a posterior distribution is calculated as part of the variational process. All the state variables are collected together into z_t which is now defined as $\{\mathbf{x}_t, \mu_t, \lambda_t, \mathbf{v}_t, \kappa_t\}$.

We define variables associated with the forward and backward evolution as follows:

forward

$$\mathrm{p}(\mu_t \,|\, \mu_{t-1}^{\alpha}, \mathbf{v}_t^{\alpha}, \kappa_t^{\alpha}) = \mathcal{N}(\mu_t \,|\, \mu_{t-1}^{\alpha} + \mathbf{v}_t^{\alpha}, (\kappa_t^{\alpha})^{-1}) \tag{16}$$

$$\mathrm{p}(\mathbf{v}_t^{\alpha} \,|\, \mathbf{v}_{t-1}^{\alpha})) = \mathcal{N}(\mathbf{v}_t^{\alpha} \,|\, \mathbf{v}_{t-1}^{\alpha}, \bar{\lambda}^{-1}) \tag{17}$$

$$\mathrm{p}(\kappa_t^{\alpha}) = \mathcal{W}(\kappa_t^{\alpha} \,|\, \bar{\mathbf{W}}, \bar{d}) \tag{18}$$

backward

$$\mathrm{p}(\mu_t \,|\, \mu_{t+1}^{\beta}, \mathbf{v}_t^{\beta}, \kappa_t^{\beta}) = \mathcal{N}(\mu_t \,|\, \mu_{t+1}^{\beta} - \mathbf{v}_t^{\beta}, (\kappa_t^{\beta})^{-1}) \tag{19}$$

$$\mathrm{p}(\mathbf{v}_t^{\beta} \,|\, \mathbf{v}_{t+1}^{\beta}) = \mathcal{N}(\mathbf{v}_t^{\beta} \,|\, \mathbf{v}_{t+1}^{\beta}, \bar{\lambda}^{-1}) \tag{20}$$

$$\mathrm{p}(\kappa_t^{\beta}) = \mathcal{W}(\kappa_t^{\beta} \,|\, \bar{\mathbf{W}}, \bar{d}) \tag{21}$$

Note that while κ_t^α is the precision of μ_t based on μ_{t-1} and κ_t^β is the precision of μ_t based on μ_{t+1}, the $\mathbf{v}_t^\alpha$ and $\mathbf{v}_t^\beta$ represent different transitions ($\mathbf{v}_t^\alpha$ is equivalent to $\mathbf{v}_{t-1}^\beta$).

We then calculate variational approximations to $\mathrm{p}(z_{t-1}\,|\,\mathbf{y}_{1:t-1})$ and $\mathrm{p}(z_{t+1}\,|\,\mathbf{y}_{t+1:T})$, which are combined below to yield an approximation for $\mathrm{p}(z_t\,|\,\mathbf{y}_{1:T})$ (cf. figure 3). The forward expressions for the approximate posteriors for μ_t, $\mathbf{v}_t$ and κ_t are dependent on the forward expectations of each variable:

$$\mathrm{q}^\alpha(\mu_t) = \mathcal{N}(\mu_t\,|\,\mathbf{m}_t^\alpha, \mathbf{S}_t^\alpha) \tag{22}$$
$$\mathbf{S}_t^\alpha = \left(\langle\lambda_t\rangle + \langle\kappa_t^\alpha\rangle\right)^{-1} \tag{23}$$
$$\mathbf{m}_t^\alpha = \mathbf{S}_t^\alpha\left(\langle\lambda_t\rangle\langle\mathbf{x}_t\rangle + \langle\kappa_t^\alpha\rangle(\langle\mu_{t-1}^\alpha\rangle + \langle\mathbf{v}_t^\alpha\rangle)\right) \tag{24}$$
$$\mathrm{q}^\alpha(\mathbf{v}_t) = \mathcal{N}(\mathbf{v}_t\,|\,\mathbf{u}_t^\alpha, \mathbf{C}_t^\alpha) \tag{25}$$
$$\mathbf{C}_t^\alpha = \left(\langle\lambda_t\rangle + \bar{\lambda}\right)^{-1} \tag{26}$$
$$\mathbf{u}_t^\alpha = \mathbf{C}_t^\alpha\left(\langle\lambda_t\rangle(\langle\mu_t^\alpha\rangle - \langle\mu_{t-1}^\alpha\rangle) + \bar{\lambda}\langle\mathbf{v}_{t-1}^\alpha\rangle\right) \tag{27}$$
$$\mathrm{q}^\alpha(\kappa_t) = \mathcal{W}(\kappa_t\,|\,\mathbf{W}_t^\alpha, d_t^\alpha) \tag{28}$$
$$d_t^\alpha = \bar{d} + 1 \tag{29}$$
$$\mathbf{W}_t^\alpha = \Big(\bar{\mathbf{W}}^{-1} + \langle\mu_t^\alpha\mu_t^{\alpha\mathrm{T}}\rangle + \langle\mu_{t-1}^\alpha\mu_{t-1}^{\alpha\mathrm{T}}\rangle + \langle\mathbf{v}_t^\alpha\mathbf{v}_t^{\alpha\mathrm{T}}\rangle - \langle\mu_t^\alpha\rangle(\langle\mu_{t-1}^\alpha\rangle + \langle\mathbf{v}_t^\alpha\rangle)^\mathrm{T} - (\langle\mu_{t-1}^\alpha\rangle + \langle\mathbf{v}_t^\alpha\rangle)\langle\mu_t^\alpha\rangle^\mathrm{T} + \langle\mu_{t-1}^\alpha\rangle\langle\mathbf{v}_t^\alpha\rangle^\mathrm{T} + \langle\mathbf{v}_t^\alpha\rangle\langle\mu_{t-1}^\alpha\rangle^\mathrm{T}\Big)^{-1} \tag{30}$$

while backward expressions are dependent on the backward expectations of each variable:

$$\mathrm{q}^\beta(\mu_t) = \mathcal{N}(\mu_t\,|\,\mathbf{m}_t^\beta, \mathbf{S}_t^\beta) \tag{31}$$
$$\mathbf{S}_t^\beta = \left(\langle\lambda_t\rangle + \langle\kappa_t^\beta\rangle\right)^{-1} \tag{32}$$
$$\mathbf{m}_t^\beta = \mathbf{S}_t^\beta\left(\langle\lambda_t\rangle\langle\mathbf{x}_t\rangle + \langle\kappa_t^\beta\rangle(\langle\mu_{t+1}^\beta\rangle - \langle\mathbf{v}_t^\beta\rangle)\right) \tag{33}$$
$$\mathrm{q}^\beta(\mathbf{v}_t) = \mathcal{N}(\mathbf{v}_t\,|\,\mathbf{u}_t^\beta, \mathbf{C}_t^\beta) \tag{34}$$
$$\mathbf{C}_t^\beta = \left(\langle\kappa_t^\beta\rangle + \bar{\lambda}\right)^{-1} \tag{35}$$
$$\mathbf{u}_t^\beta = \mathbf{C}_t^\beta\left(\langle\kappa_t^\beta\rangle(\langle\mu_{t+1}^\beta\rangle - \langle\mu_t^\beta\rangle) + \bar{\lambda}\langle\mathbf{v}_{t+1}^\beta\rangle\right) \tag{36}$$
$$\mathrm{q}^\beta(\kappa_t) = \mathcal{W}(\kappa_t\,|\,\mathbf{W}_t^\beta, d_t^\beta) \tag{37}$$
$$d_t^\beta = \bar{d} + 1 \tag{38}$$
$$\mathbf{W}_t^\beta = \Big(\bar{\mathbf{W}}^{-1} + \langle\mu_t^\beta\mu_t^{\beta\mathrm{T}}\rangle + \langle\mu_{t+1}^\beta\mu_{t+1}^{\beta\mathrm{T}}\rangle + \langle\mathbf{v}_t^\beta\mathbf{v}_t^{\beta\mathrm{T}}\rangle - \langle\mu_t^\beta\rangle(\langle\mu_{t+1}^\beta\rangle - \langle\mathbf{v}_t^\beta\rangle)^\mathrm{T} - (\langle\mu_{t+1}^\beta\rangle - \langle\mathbf{v}_t\rangle^\beta)\langle\mu_t^\beta\rangle^\mathrm{T} - \langle\mu_{t+1}^\beta\rangle\langle\mathbf{v}_t^\beta\rangle^\mathrm{T} - \langle\mathbf{v}_t^\beta\rangle\langle\mu_{t+1}^\beta\rangle^\mathrm{T}\Big)^{-1} \tag{39}$$

These may be combined to give a smoothed approximate posterior distribution for the state expectation, $\mathrm{q}(\mu_t) \approx \mathrm{p}(\mu_t\,|\,\mathbf{y}_{1:T})$, using the transition probabilities to z_t from the previous

and following steps (cf. figure 3).

$$\begin{aligned} \mathrm{q}(\mu_t) = \int & \mathrm{p}(\mathbf{y}_t \,|\, \mathbf{x}_t)\ \mathrm{p}(\mathbf{x}_t \,|\, \mu_t, \lambda_t)\ \mathrm{q}^{\alpha}(\mu_{t-1})\ \mathrm{p}(\mu_t \,|\, \mu^{\alpha}_{t-1}, \mathbf{v}^{\alpha}_t, \kappa^{\alpha}_t)\ \mathrm{q}^{\alpha}(\mathbf{v}_t)\ \mathrm{q}^{\alpha}(\kappa_t) \\ & \mathrm{q}^{\beta}(\mu_{t+1})\ \mathrm{p}(\mu_t \,|\, \mu^{\beta}_{t+1}, \mathbf{v}^{\beta}_t, \kappa^{\beta}_t)\ \mathrm{q}^{\beta}(\mathbf{v}_t)\ \mathrm{q}^{\beta}(\kappa_t)\, \mathrm{q}(\lambda_t) \\ & d\mathbf{x}_t\ d\mu^{\alpha}_{t-1}\ d\mathbf{v}^{\alpha}_t\ d\kappa^{\alpha}_t\ d\mu^{\beta}_{t+1}\ d\mathbf{v}^{\beta}_t\ d\kappa^{\beta}_t\ d\lambda_t \end{aligned} \tag{40}$$

This integral is, again, analytically intractable, so we use the variational approximation, resulting in a Gaussian distribution for the posterior of μ_t as follows:

$$\mathrm{q}(\mu_t) = \mathcal{N}(\mu_t \,|\, \mathbf{m}_t, \mathbf{S}_t) \tag{41}$$

$$\mathbf{S}_t = \left(\langle \lambda_t \rangle + \langle \kappa^{\alpha}_t \rangle + \langle \kappa^{\beta}_t \rangle \right)^{-1} \tag{42}$$

$$\mathbf{m}_t = \mathbf{S}_t \left(\langle \lambda_t \rangle \langle \mathbf{x}_t \rangle + \langle \kappa^{\alpha}_t \rangle \left(\langle \mu^{\alpha}_{t-1} \rangle + \langle \mathbf{v}^{\alpha}_t \rangle \right) + \langle \kappa^{\beta}_t \rangle \left(\langle \mu^{\beta}_{t+1} \rangle - \langle \mathbf{v}^{\beta}_t \rangle \right) \right) \tag{43}$$

The approximate posterior expectation of μ_t is now a weighted average of contributions from the past ($\mu^{\alpha}_{t-1} + \mathbf{v}^{\alpha}_t$), the present ($\mathbf{x}_t$) and the future ($\mu^{\beta}_{t+1} - \mathbf{v}^{\beta}_t$), with each contribution weighted according to its expected precision. Note that the expected value of μ_t no longer includes terms in the diffusion precision $\bar{\lambda}$, only in precisions that measure uncertainty, which are hopefully large. A similar procedure results in posterior distributions for $\mathbf{v}_t$ and κ_t, as follows:

$$\mathrm{q}(\mathbf{v}_t) = \mathcal{N}(\mathbf{v}_t \,|\, \mathbf{u}_t, \mathbf{C}_t) \tag{44}$$

$$\mathbf{C}_t = \left(2\bar{\lambda} + \langle \kappa_t \rangle \right)^{-1} \tag{45}$$

$$\mathbf{u}_t = \mathbf{C}_t \left(\bar{\lambda} \langle \mathbf{v}^{\alpha}_{t-1} \rangle - \bar{\lambda} \langle \mathbf{v}^{\beta}_{t+1} \rangle + \langle \kappa_t \rangle \left(\langle \mu_t \rangle - \langle \mu_{t-1} \rangle \right) \right) \tag{46}$$

$$\mathrm{q}(\kappa_t) = \mathcal{W}(\kappa_t \,|\, \mathbf{W}_t, d_t) \tag{47}$$

$$d_t = \bar{d} + 1 \tag{48}$$

$$\begin{aligned} \mathbf{W}_t = \Big(& \bar{\mathbf{W}}^{-1} + \langle \mu_t \mu_t^{\mathrm{T}} \rangle + \langle \mu_{t-1} \mu_{t-1}^{\mathrm{T}} \rangle + \langle \mathbf{v}_t \mathbf{v}_t^{\mathrm{T}} \rangle - \langle \mu_t \rangle \left(\langle \mu_{t-1} \rangle + \langle \mathbf{v}_t \rangle \right)^{\mathrm{T}} \\ & - \left(\langle \mu_{t-1} \rangle + \langle \mathbf{v}_t \rangle \right) \langle \mu_t \rangle^{\mathrm{T}} + \langle \mu_{t-1} \rangle \langle \mathbf{v}_t \rangle^{\mathrm{T}} + \langle \mathbf{v}_t \rangle \langle \mu_{t-1} \rangle^{\mathrm{T}} \Big)^{-1} \end{aligned} \tag{49}$$

As shown in Algorithm 1, the approximate posterior for λ_t may be calculated in the forward and backward sweeps and then, again, in the combining sweep, exactly as for the diffusion model (9). The combination and backward sweeps may be performed simultaneously.

4 Results

For the cricket tracking problem, the cricket was modelled as an ellipse. The 5-dimensional state defined the ellipse's location through the x and y coordinates of its centre, and its size and orientation through the lengths of its axes and a rotation angle. Each video frame, $\mathbf{Y}_t$, was segmented into a background and a foreground by projecting it onto the first principal component calculated across a sliding window of 20 frames and subtracting the projection (i.e. the background) from the original colour image, resulting in the foreground or movement image, $\mathbf{M}_t$. This procedure helps the algorithm to distinguish between a dark (moving)

Algorithm 1 Smoothing for the velocity model

Locate the object in the first frame (using the diffusion model) to determine $\langle \mu_0 \rangle$.

forward sweep

for $t = 1$ to T **do**
  Initialise $\langle \mu_t^\alpha \rangle$ and $\langle \lambda_t \rangle$ from $\mathrm{p}(z_t \,|\, \mathbf{y}_{1:t-1})$.
  for $iter = 1$ to convergence **do**
    Use importance sampling to estimate $\langle \mathbf{x}_t \,|\, \mathbf{y}_{1:t} \rangle$ and $\langle \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \,|\, \mathbf{y}_{1:t} \rangle$.
    Calculate $\mathrm{q}^\alpha(\mu_t)$, $\mathrm{q}^\alpha(\mathbf{v}_t)$, $\mathrm{q}^\alpha(\kappa_t)$ and $\mathrm{q}(\lambda_t)$ dependent on forward expectations.
  end for
end for

backward sweep

for $t = T-1$ down to 1 **do**
  Initialise $\langle \mu_t^\beta \rangle$ and $\langle \lambda_t \rangle$ from $\mathrm{p}(z_t \,|\, \mathbf{y}_{t+1:T})$.
  for $iter = 1$ to convergence **do**
    Use importance sampling to estimate $\langle \mathbf{x}_t \,|\, \mathbf{y}_{t:T} \rangle$ and $\langle \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \,|\, \mathbf{y}_{t:T} \rangle$.
    Calculate $\mathrm{q}^\beta(\mu_t)$, $\mathrm{q}^\beta(\mathbf{v}_t)$, $\mathrm{q}^\beta(\kappa_t)$ and $\mathrm{q}(\lambda_t)$ dependent on backward expectations.
  end for
end for

combination sweep

$\mathrm{q}(\mu_1) = \mathrm{q}^\beta(\mu_1)$.
$\mathrm{q}(\mu_T) = \mathrm{q}^\alpha(\mu_T)$.
for $t = 2$ to $T-1$ **do**
  Initialise $\langle \mu_t \rangle$ and $\langle \lambda_t \rangle$ from $\mathrm{p}(z_t \,|\, \mathbf{y}_{1:t-1})$.
  for $iter = 1$ to convergence **do**
    Use importance sampling to estimate $\langle \mathbf{x}_t \,|\, \mathbf{y}_t \rangle$ and $\langle \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \,|\, \mathbf{y}_t \rangle$.
    Calculate $\mathrm{q}(\mu_t)$, $\mathrm{q}(\mathbf{v}_t)$, $\mathrm{q}(\kappa_t)$ and $\mathrm{q}(\lambda_t)$ dependent on combined expectations.
  end for
end for

elliptical cricket and its dark, but stationary, elliptical burrow (cf. top right figure 4). For a given ellipse, the likelihood of a single pixel, i, is given by

$$\mathrm{p}(i) = \begin{cases} \mathcal{N}(Y_t^{(i)} \,|\, \mu_{bg}, \lambda_{bg}^{-1})\, \mathcal{N}(M_t^{(i)} \,|\, 0, 1/30) & i \notin \text{ellipse} \\ \mathcal{G}(Y_t^{(i)} \,|\, \mu_{fg}, \lambda_{fg}^{-1})\, \mathcal{N}(M_t^{(i)} \,|\, -100, 1/75) & i \in \text{ellipse} \end{cases} \tag{50}$$

where μ_{bg} and λ_{bg} are calculated by fitting a Gaussian distribution to all the background pixels in the video sequence, and μ_{fg} and λ_{fg} were determined empirically. The video was recorded at 2 frames per second; we remark that in 500 ms the crickets can move well over their own body lengths.

Figure 4 shows the forward, backward and combined tracks produced by each of the models over 30 frames, overlaid with ellipses representing the state estimated for the frame shown in the background (which has been lightened for clarity). Notice how for the diffusion model the combined track is indistinguishable from the forward track, while for the velocity model it is quite separate.

Figure 5 shows the combined tracks for each algorithm, offset vertically to aid comparison, demonstrating the smoother track produced by the velocity model.

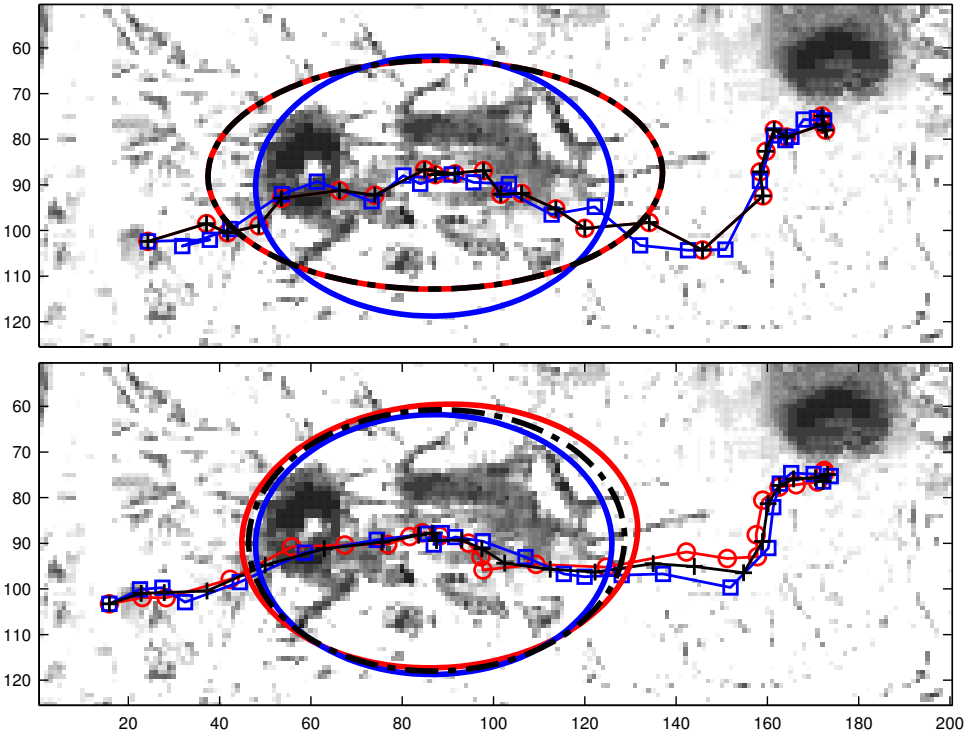

Figure 4 The forward (red circles), backward (blue squares) and combined (black crosses) tracks produced by (top) the diffusion model and (bottom) the velocity model over 30 frames. The ellipses show the state estimated for the frame shown in the background, which has been lightened for clarity.

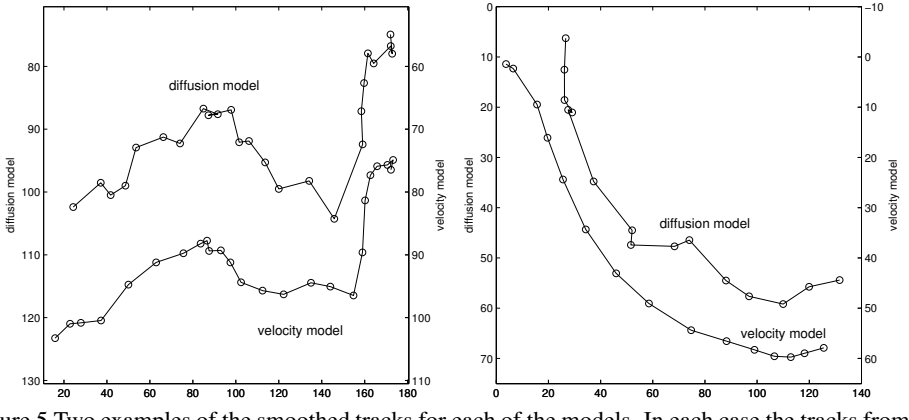

Figure 5 Two examples of the smoothed tracks for each of the models. In each case the tracks from the diffusion and velocity models have been offset vertically to aid comparison.

5 Conclusions

Traditional smoothing recursions are ineffective when the transition between states is modelled as a diffusion which represents both uncertainty in the state and the genuine change in the state. We have presented a straightforward variational approach to Bayesian smoothing which provides smoother state estimates, by including an effective smoother step, and explicitly modelling the genuine state change. As a result the new model is more robust to minor occlusions of the tracked cricket by, for example, blown blades of grass.

References

[1] H.T. Attias. A variational Bayesian framework for graphical models. *Advances in Neural Information Processing Systems*, 12:209–215, 2000.

[2] M. Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London, 2003.

[3] M. Beal and Z. Ghahramani. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. In *Bayesian Statistics*, volume 7. Oxford University Press, 2002.

[4] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.

[5] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. In D. Crisan and B. Rozovskii, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 656–704. Oxford University Press, 2011.

[6] N. Gordon, D. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140:107–113, 1993.

[7] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183, 1999.

[8] G. Kitigawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25, 1996.

[9] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[10] H. Lappalainen and J.W. Miskin. *Advances in Independent Component Analysis*, chapter Ensemble Learning, pages 75–92. Springer-Verlag, Berlin, 2000.

[11] R. Rodríguez-Muñoz, A. Bretman, J. Slate, C.A. Walling, and T. Tregenza. Natural and sexual selection in a wild insect population. *Science*, 328:1269–1272, 2010.

[12] J. Vermaak, N.D. Lawrence, and P. Pérez. Variational inference for visual tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 773–780, 2003.

[13] S.R. Waterhouse, D.J.C. MacKay, and A.J. Robinson. *Advances in Neural Information Processing Systems 7*, chapter Bayesian methods for mixtures of experts, pages 351–357. MIT Press, 1995.