Variational Mixture Smoothing for Non-Linear Dynamical Systems

Cristian Sminchisescu and Allan Jepson, Department of Computer Science, University of Toronto

Contributions
⇒ Mixture smoother for non-linear (non-Gaussian) dynamical systems
⇒ No existing algorithm appears to compute this type of model
⇒ General formulation, unrestricted dynamics, noise and observation functions
⇒ Accurate, compact, interpretable multimodal trajectory approximation
⇒ Efficiently applies to weakly identifiable models with strong multimodal distributions common in visual tracking
  • Monocular 3D human motion estimation, scene reconstruction, etc.
  • Multimodality persists after both filtering and smoothing

Existing Methodology
⇒ Iterated Kalman smoothers are Gaussian, unimodal (thus not applicable)
⇒ MCMC or particle smoothers mix slowly and do not provide accurate compact multimodal approximations (mean states are often uninformative)

Approach
⇒ Variationally refine a mixture approximation to the trajectory distribution
⇒ High-dimensional problem, initialization is critical. The key construction is an embedded network that allows the efficient location of multiple trajectories
⇒ Four layers of computation: filtering, dynamic programming (DP) in the embedded network, sparse non-linear optimization and variational refinement.

Non-Linear Dynamical System

\[ \begin{align*}
&x_t = f(x_{t-1}, \eta_t) + \eta_t, \\
&P_x(x_t | x_{t-1}) = \int p(x_t | x_{t-1}, \eta_t) p(\eta_t) \, d\eta_t
\end{align*} \]

Formulation
⇒ The trajectory distribution of a non-linear dynamical system, over \( T \) timesteps is:
\[ P(X, R) = p(x_{t-1}) p(r_t | x_{t-1}) \prod_{t=2}^{T} p(x_t | x_{t-1}, p(r_t | x_{t-1})) \]
where \( p(x_t) \) is the state space prior, \( X = (x_1, \ldots, x_T) \) is the joint state vector, \( R = (r_1, \ldots, r_T) \) is the joint observation vector, \( p(x_t | x_{t-1}) \) is the dynamic transition probability and \( p(r_t | x_{t-1}) \) is the observation model
⇒ Compute a tractable trajectory approximation \( q^* \) (here a Gaussian mixture), that minimizes relative entropy to \( P(X, R) \)
⇒ Equivalent to minimizing the variational free energy
\[ F(q^*, P) = D(q^* \| P) - \log P(R) = \frac{1}{2} \int q^*(X) \log \frac{q^*(X)}{P(X|R)} \, dX + \frac{1}{2} \int q^*(X) \log \frac{1}{P(X)} \, dX \]

Experiments
⇒ 2.5 s monocular video sequence
⇒ 32 d.o.f. articulated human model
⇒ Tracking (filtering) using Kinematic Jump Sampling
⇒ Distribution represented with up to 8 modes

Step 1. Multiple Trajectories over an Embedded Network
⇒ Let the temporal observation likelihood be approximated by mixtures
\[ p(r_t | x_{t-1}) = \sum_{i=1}^{m} \pi_i m_i(x_t, r_t | x_{t-1}) \]
⇒ Construct Embedded Network: Lift the dynamics and observation likelihood from point-wise (as defined in a non-linear dynamical system) to mode-wise

Time \( t-1 \) \( t \) \( t+1 \)
\[ \begin{align*}
p(r_t | x_{t-1}) &= \sum_{i=1}^{m} \pi_i m_i(x_t, r_t | x_{t-1}) \\
p(r_t | x_{t-1}) &= \int m_i(x_t, r_t | x_{t-1}) p(x_t | x_{t-1}) \, dx_t
\end{align*} \]
⇒ Use DP to compute most probable trajectories between nodes at time 1 and \( T \)
⇒ Trajectories are optimal with respect to the fixed embedded network

Step 2. Mixture of continuous MAP estimates (means + inv. Hessians) of \( P(X,R) \)
⇒ Initialize using the DP trajectories in \( \text{Step 1} \)
⇒ Sparse robust second-order optimizer

Step 3. Bayesian variational refinement of \( F(q^*, P) \) with \( q^* \) initialized from \( \text{Step 2} \)
⇒ 2° per parameter change but … Qualitative state change
⇒ Reduced uncertainty for longer sequences

Constraints preserved
⇒ Good image match

Plausible Trajectory #1
Plausible Trajectory #2