Impact of Dynamic Model Learning on Classification of Human Motion

Vladimir Pavlović and James M. Rehg
Compaq Computer Corporation
Cambridge Research Lab
Cambridge, MA 02139
{vladimir,rehg}@crl.dec.com

Abstract

The human figure exhibits complex and rich dynamic behavior that is both nonlinear and time-varying. However, most work on tracking and analysis of figure motion has employed either generic or highly specific hand-tailored dynamic models superficially coupled with hidden Markov models (HMMs) of motion regimes. Recently, an alternative class of learned dynamic models known as switching linear dynamic systems (SLDSs) has been cast in the framework of dynamic Bayesian networks (DBNs) and applied to analysis and tracking of the human figure. In this paper we further study the impact of learned SLDS models on analysis and tracking of human motion and contrast them to the more common HMM models. We develop a novel approximate structured variational inference algorithm for SLDS, a globally convergent DBN inference scheme, and compare it with standard SLDS inference techniques. Experimental results on learning and analysis of figure dynamics from video data indicate the significant potential of the SLDS approach.

1. Introduction

The human figure exhibits complex and rich dynamic behavior. Dynamics are essential to the analysis of human motion (e.g. gesture recognition) as well as to the synthesis of realistic figure motion in computer graphics. In visual tracking applications, dynamics can provide a powerful cue in the presence of occlusions and measurement noise.

Although the use of kinematic models in figure tracking is now commonplace, dynamic models have received relatively little attention. The kinematics of the figure specify its degrees of freedom (e.g. joint angles and torso pose) and define a state space. A dynamic model imposes additional structure on the state space by specifying which state trajectories are possible (or probable) and by specifying the speed at which a trajectory evolves.

One promising approach is to learn dynamic models from a training corpus of observed state space trajectories. In cases where sufficient training data is available, the learning approach promises flexibility and generality. Many different modeling frameworks are possible. Previous work by a number of authors have applied Hidden Markov Models (HMMs) to motion classification. In more recent work, switching linear dynamic system (SLDS) models have been applied to human motion modeling [6, 13, 15]. In SLDS models, the Markov process controls an underlying linear dynamic system, rather than a fixed Gaussian measurement model.

By mapping discrete hidden states to piecewise linear measurement models, the SLDS framework has potentially greater descriptive power than an HMM. Offsetting this advantage is the fact that inference in SLDS is considerably more complex than inference in HMM's, which in turn complicates SLDS learning.

In this paper we describe the results of an empirical comparison between SLDS and HMM models on two common tasks: classification and one-step ahead prediction of motion sequences. We derive three different approximate inference schemes for SLDS: Viterbi [16], variational, and GPB2 [2]. We compare the performance of these schemes to that of conventional HMM models.

We demonstrate that even on fairly simple motion sequences, the SLDS model class consistently outperforms standard HMMs on classification and continuous state estimation tasks. These preliminary results suggest that SLDS models are a promising tool for figure motion analysis. In addition to our experimental results, the derivations we provide for the three SLDS inference schemes should be useful to other researchers who are interested in these models. Moreover, our variational inference algorithm is novel.

2. Switching Linear Dynamic System Model

A switching linear dynamic system (SLDS) model describes the dynamics of a complex, nonlinear physical pro-

cess by switching among a set of linear dynamic models over time. The system can be described using the following set of state-space equations:

$$\begin{aligned} x_{t+1} &= A(s_{t+1})x_t + v_{t+1}(s_{t+1}), \\ y_t &= Cx_t + w_t, \text{ and} \\ x_0 &= v_0(s_0) \end{aligned}$$

for the physical system, and

$$\begin{aligned} Pr(s_{t+1} = i|s_t = j) &= \Pi(i,j), \text{ and} \\ Pr(s_0 = i) &= \pi_0(i) \end{aligned}$$

for the switching model. The meaning of the variables is as follows: $x_t \in \Re^N$ denotes the hidden state of the LDS, and v_t is the state noise process. Similarly, $y_t \in \Re^M$ is the observed measurement and w_t is the measurement noise. Parameters A and C are the typical LDS parameters: the state transition matrix and the observation matrix, respectively. We assumed that the LDS models a Gauss-Markov process. Hence, the noise processes are independently distributed Gaussian:

$$\begin{aligned} v_t(s_t) &\sim \mathcal{N}(0, Q(s_t)),\ t > 0 \\ v_0(s_0) &\sim \mathcal{N}(x_0(s_t), Q_0(s_t)) \\ w_t &\sim \mathcal{N}(0, R). \end{aligned}$$

The switching model is assumed to be a discrete first order Markov process. State variables of this model are written as s_t. They belong to the set of S discrete symbols $\{0, \ldots, S-1\}$. The switching model is defined with the state transition matrix Π whose elements are $\Pi(i,j) = Pr(s_{t+1} = i|s_t = j)$, and an initial state distribution vector π_0.

Coupling between the LDS and the switching process stems from the dependency of the LDS parameters A and Q on the switching process state s_t. Namely,

$$\begin{aligned} A(s_t = i) &= A_i \\ Q(s_t = i) &= Q_i \end{aligned}$$

In other words, switching state s_t determines which of S possible plant models is used at time t.

The complex state space representation is equivalently depicted by the DBN dependency graph in Figure 1. The dependency graph implies that the *joint distribution* P over the variables of the SLDS can be written as

$$\begin{aligned} P(\mathcal{Y}_T, \mathcal{X}_T, \mathcal{S}_T) = Pr(s_0) \prod_{t=1}^{T-1} Pr(s_t|s_{t-1}) \\ Pr(x_0|s_0) \prod_{t=1}^{T-1} Pr(x_t|x_{t-1}, s_t) \\ \prod_{t=0}^{T-1} Pr(y_t|x_t), \end{aligned} \quad (1)$$

where $\mathcal{Y}_T, \mathcal{X}_T$, and $\mathcal{S}_T$ denote the sequences (of length T) of observations and hidden state variables. For instance, $\mathcal{Y}_T = \{y_0, \ldots, y_{T-1}\}$. From the Gauss-Markov assumption on the LDS and the Markov switching model assumption, we can expand Equation 1 into the parameterized joint pdf of the SLDS of duration T.

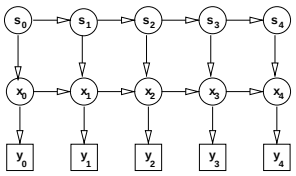


Figure 1. Bayesian network representation (dependency graph) of the SLDS. s denote instances of the discrete valued action states switching the physical system models with continuous valued states x and observations y.

Learning in complex DBNs can be formulated as the problem of ML learning in general Bayesian networks. Hence, a generalized EM algorithm [14] can be used to find optimal values of DBN parameters $\{A, C, Q, R, \Pi, \pi_0\}$. The expectation (E) step of EM is the task of inference. Inference, which is addressed in the next section, is the most difficult step in SLDS learning. Given the sufficient statistics from the inference phase, the *parameter update equations* in the maximization (M) step are obtained by maximizing the expected log of Equation 1 with respect to the LDS and MC parameters. Derivations can be found in [16].

3. Inference in SLDS

The goal of inference in complex DBNs is to estimate the posterior probability of the hidden states of the system (s_t and x_t) given some known sequence of observations $\mathcal{Y}_T$ and the known model parameters. Specifically, we need to find the *sufficient statistics* of the posterior $P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T)$. Given the form of P it is easy to show that these are the first and the second order statistics: mean and covariance among hidden states $x_t, x_{t-1}, s_t, s_{t-1}$.

If there were no switching dynamics, the inference would be straightforward – we could infer $\mathcal{X}_T$ from $\mathcal{Y}_T$ using LDS inference (RTS smoothing [1]). However, the presence of switching dynamics embedded in matrix Π makes exact inference more complicated. To see that, assume that the initial distribution of x_0 at $t = 0$ is Gaussian, at $t = 1$ the pdf of the physical system state x_1 becomes a mixture of S Gaussian pdfs since we need to marginalize over S possible but unknown plant models. At time t we will have a mixture of S^t Gaussians, which is clearly intractable for even moderate sequence lengths. It is therefore necessary to explore approximate inference techniques that will result in a tractable learning method.

An approximate Viterbi inference algorithm was presented in [16] and evaluated experimentally. We briefly review it in Section 3.1. We then describe two additional approximation techniques: variational inference (Section 3.2) and generalized Pseudo Bayesian inference (Section 3.3).

3.1. Approximate Viterbi Inference

The task of Viterbi approximation approach is to find the most likely sequence of switching states s_t for a given observation sequence $\mathcal{Y}_T$. If the best sequence of switching states is denoted $\mathcal{S}_T^*$ we can then approximate the desired posterior $P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T)$ as[1]

$$P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T) \approx Pr(\mathcal{X}_T|\mathcal{S}_T, \mathcal{Y}_T)\ \delta(\mathcal{S}_T - \mathcal{S}_T^*), \quad (2)$$

i.e. the switching sequence posterior $Pr(\mathcal{S}_T|\mathcal{Y}_T)$ was approximated by its mode. It is well known how to apply Viterbi inference to discrete state hidden Markov models [17] and continuous state Gauss-Markov models [1]. Here we review an algorithm for approximate Viterbi inference in SLDSs presented in [16].

We would like to compute the switching sequence $\mathcal{S}_T^*$ such that $\mathcal{S}_T^* = \arg\max_{\mathcal{S}_T} Pr(\mathcal{S}_T|\mathcal{Y}_T)$. Define first the following probability up to time t of the switching state sequence being in state i at time t given the measurement sequence $\mathcal{Y}_t$:

$$J_{t,i} = \max_{\mathcal{S}_{t-1}} Pr\left(\mathcal{S}_{t-1}, s_t = i, \mathcal{Y}_t\right) \quad (3)$$

If this quantity is known at time T the probability of the most likely switching sequence $\mathcal{S}_T^*$ is simply $Pr(\mathcal{S}_T^*|\mathcal{Y}_T) \propto \max_i J_{T-1,i}$. In fact, a recursive procedure can be used to obtain the desired quantity:

$$\begin{aligned} J_{t,i} \quad \approx \quad & \max_j \left\{ Pr\left(y_t|s_t = i, s_{t-1} = j, S_{t-2}^*(j), \mathcal{Y}_{t-1}\right) \right. \\ & \left. Pr\left(s_t = i|s_{t-1} = j\right) J_{t-1,j} \right\}. \end{aligned} \quad (4)$$

We call the two terms next to $J_{t-1,j}$ the "transition probability" from state j at time $t-1$ to state i at time t, and denote it by $J_{t|t-1,i,j}$. Also, $S_{t-2}^*(i)$ is the "best" switching sequence up to time $t-1$ when SLDS is in state i at time $t-1$: $S_{t-2}^*(i) = \arg\max_{S_{t-2}} J_{t-1,i}$.

Hence, the switching sequence posterior at time t can be recursevly computed from the same at time $t-1$. The two scaling components in $J_{t|t-1,i,j}$ are the likelihood associated with the transition $i \to j$ from t to $t-1$, and the probability of discrete SLDS switching from j to i.

To find the likelihood term note that concurrently with the recursion of Equation 4, for each pair of consecutive switching state i, j at times $t, t-1$ one can obtain the following statistics using the Kalman filter [1]: $\hat{x}_{t|t,i}$, the "best" filtered LDS state estimate at t when the switch is in state i at time t and a sequence of t measurements, $\mathcal{Y}_t$, has been processed; $\hat{x}_{t|t-1,i,j}$ and $\hat{x}_{t|t,i,j}$, the one-step predicted LDS state and the "best" filtered state estimates at time t, respectively, given that the switch is in state i at time t and in state j at time $t-1$ and only $t-1$ measurements are known. Similar definitions are used for filtered and predicted state variance estimates, $\Sigma_{t|t,i}$ and $\Sigma_{t|t-1,i,j}$ respectively. See [16] for details. The likelihood term can then be easily computed as the probability of innovation $y_t - C\hat{x}_{t|t-1,i,j}$ of $j \to i$ transition, $Pr(y_t|_{s_t=i,s_{t-1}=j,S_{t-2}^*(j)}) = \mathcal{N}\left(y_t; C\hat{x}_{t|t-1,i,j}, C\Sigma_{t|t-1,i,j}C' + R\right)$.

Obviously, for every current switching state i there are S possible previous switching states where the system could have originated from. To maximize the overall probability at every time step t and for every switching state i one "best" previous state j is selected: $\psi_{t-1,i} = \arg\max_j \left\{J_{t|t-1,i,j}J_{t-1,j}\right\}$. The index of this state is kept in the state transition record $\psi_{t-1,i}$. Consequently, we now obtain a set of S best filtered LDS states and variances at time t: $\hat{x}_{t|t,i} = \hat{x}_{t|t,i,\psi_{t-1,i}}$ and $\Sigma_{t|t,i} = \Sigma_{t|t,i,\psi_{t-1,i}}$.

Once all T observations $\mathcal{Y}_{T-1}$ have been fused to decode the "best" switching state sequence one uses the index of the best final state, $i_{T-1}^* = \arg\max_i J_{T-1,i}$, and then traces back through the state transition record $\psi_{t-1,i}$, setting $i_t^* = \psi_{t,i_{t+1}^*}$. The switching model's sufficient statistics are now simply $Pr(s_t = i) = 1$ if $i = i_t^*$ and $Pr(s_t = i, s_{t-1} = j) = 1$ if $i = i_t^*$ and $j = i_{t-1}^*$. Given the "best" switching state sequence the sufficient LDS statistics can be easily obtained using Rauch-Tung-Streiber (RTS) smoothing [1]. The Viterbi inference algorithm for complex DBNs can now be summarized as

```
Initialize LDS state estimates x̂_{0|-1,i} and Σ_{0|-1,i};
Initialize J_{0,i};
for t = 1 : T − 1
    for i = 1 : S
        for j = 1 : S
            Predict and filter LDS state estimates
                x̂_{t|t,i,j} and Σ_{t|t,i,j};
            Find j → i "transition probability" J_{t|t-1,i,j};
        end
        Find best transition ψ_{t-1,i} into state i;
        Update sequence probabilities J_{t,i}
          and LDS state estimates x̂_{t|t,i} and Σ_{t|t,i};
    end
end
Find "best" final switching state i*_{T-1};
Backtrack to find "best" switching state sequence i*_t;
Find DBN's sufficient statistics;
```

3.2. Approximate Variational Inference

A general structured variational inference technique for Bayesian networks is described in [10]. The basic idea is to construct a parameterized distribution Q which is in some sense close to the desired conditional distribution P, but is easier to compute. One can then employ Q as an approximation of P,

$$P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T) \approx Q(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T).$$

[1] $\delta(x) = 1$ for $x = \emptyset$ and zero otherwise.

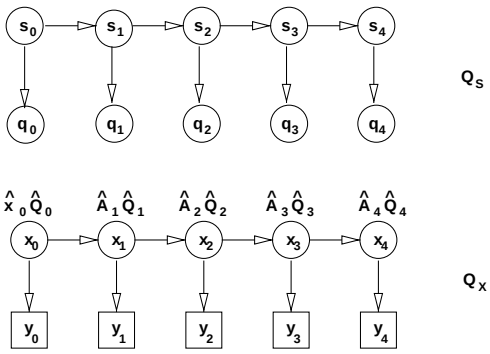


Figure 2. Factorization of the original SLDS. Factorization reduces the fully coupled model into a seemingly decoupled pair of a HMM (Q_s) and a LDS (Q_x).

Namely, for a given set of observations $\mathcal{Y}_T$, a distribution $Q(\mathcal{X}_T, \mathcal{S}_T|\eta, \mathcal{Y}_T)$ with an additional set of *variational parameters* η is defined such that Kullback–Leibler divergence between $Q(\mathcal{X}_T, \mathcal{S}_T|\eta, \mathcal{Y}_T)$ and $P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T)$ is minimized with respect to η:

$$\eta^* = \arg\min_{\eta} \sum_{\mathcal{S}_T} \int_{\mathcal{X}_T} Q(\mathcal{X}_T, \mathcal{S}_T|\eta, \mathcal{Y}_T) \times \log \frac{P(\mathcal{X}_T, \mathcal{S}_T|\mathcal{Y}_T)}{Q(\mathcal{X}_T, \mathcal{S}_T|\eta, \mathcal{Y}_T)}. \quad (5)$$

The dependency structure of Q is chosen such that it closely resembles the dependency structure of the original distribution P. However, unlike P the dependency structure of Q is designed to allow computationally efficient inference. In our case we define Q by decoupling the switching and LDS portions of SLDS as shown in Figure 2. The two subgraphs of the original network are a Hidden Markov Model (HMM) Q_S with variational parameters $\{q_0, \ldots, q_{T-1}\}$ and a time-varying LDS Q_X with variational parameters $\{\hat{x}_0, \hat{A}_0, \ldots, \hat{A}_{T-1}, \hat{Q}_0, \ldots, \hat{Q}_{T-1}\}$. factorized, allowing for independent inference inference: $Q(\mathcal{X}_T, \mathcal{S}_T|\eta, \mathcal{Y}_T) = Q_X(\mathcal{X}_T|\eta, \mathcal{Y}_T)Q_S(\mathcal{S}_T|\eta)$. This is also reflected in the sufficient statistics of the posterior defined by the approximating network, e.g. $\langle x_t x_t' s_t = i\rangle = \langle x_t x_t'\rangle Pr(s_t = i)$.

The optimal values of the variational parameters η are obtained by setting the derivative of the KL-divergence w.r.t. η to zero. For example, we can then arrive at the following optimal variational parameters:

$$\hat{Q}_t^{-1} = \sum_{i=0}^{S-1} Q_i^{-1} Pr(s_t = i) + \sum_{i=0}^{S-1} A_i' Q_i^{-1} A_i Pr(s_{t+1} = i) - \hat{A}_{t+1}' \hat{Q}_{t+1}^{-1} \hat{A}_{t+1}$$

$$\hat{A}_t = \hat{Q}_t \sum_{i=0}^{S-1} Q_i^{-1} A_i Pr(s_t = i) \quad (6)$$

$$\log q_t(i) = -\frac{1}{2}\left\langle (x_t - A_i x_{t-1})' \hat{Q}_i^{-1} (x_t - A_i x_{t-1}) \right\rangle - \frac{1}{2}\log|\hat{Q}_{t,i}| \quad (7)$$

To obtain the terms $Pr(s_t) = Pr(s_t|q_0, \ldots, q_{T-1})$ we use the inference in the HMM with output "probabilities" q_t, as described in [17]. Similarly, to obtain $\langle x_t\rangle = E[x_t|\mathcal{Y}_T]$ we perform LDS inference in the decoupled time-varying LDS via RTS smoothing. Since $\hat{A}_t, \hat{Q}_t$ in the decoupled LDS Q_X depends on $Pr(s_t)$ from the decoupled HMM Q_S and q_t depends on $\langle x_t\rangle, \langle x_t x_t'\rangle, \langle x_t x_{t-1}'\rangle$ from the decoupled LDS, Equations 6 and 7 together with the inference solutions in the decoupled models form a set of fixed-point equations. Solution of this fixed-point set yields a tractable approximation to the intractable inference of the original fully coupled SLDS.

The variational inference algorithm for fully coupled SLDSs can now be summarized as:

error = ∞;
Initialize $Pr(s_t)$;
while (error > maxError) {
  Find $\hat{Q}_t, \hat{A}_t, \hat{x}_0$ from $Pr(s_t)$ using Equations 6;
  Estimate $\langle x_t\rangle, \langle x_t x_t'\rangle$ and $\langle x_t x_{t-1}'\rangle$ from y_t using time-varying LDS inference;
  Find q_t from $\langle x_t\rangle, \langle x_t x_t'\rangle$ and $\langle x_t x_{t-1}'\rangle$ using Equations 7;
  Estimate $Pr(s_t)$ from q_t using HMM inference.
  Update approximation error (KL divergence);
}

Interpretation of recursions for variational parameters in Equations 6 and 7 is not immediately clear. LDS parameters $\hat{A}_t$ and $\hat{Q}_t^{-1}$ are, roughly, averages of the corresponding switching system parameters weighted by the estimates of the switching states $P(s_t)$. HMM variational paremeters $\log q_t$ measure the agreement of each individual LDS with the data. As an example, we considered variational inference in a simple three state SLDS. The SLDS parameters were chosen to be:

$$A_0 = \begin{bmatrix} 1 & .6 \\ 0 & 1 \end{bmatrix} \quad Q_0 = Q_1 = Q_2 = \begin{bmatrix} .08 & 1.25 \\ 1.25 & 25 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 1 & -.6 \\ 0 & .4 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix} \; R = 5000$$

$$A_2 = \begin{bmatrix} 1 & .1 \\ -.7 & -.4 \end{bmatrix} \quad \Pi = \begin{bmatrix} 0.98 & 0 & 0 \\ .02 & .99 & 0 \\ 0 & .01 & 1 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

The SLDS was simulated over 140 time steps to produce a sequence of switching states, continuous states and measurements. Variational inference was then used to infer distributions of switching and continuous states from the simulated measurements. Figure 3.2 depicts state estimates and

variational parameters for the first and third iteration of variational inference.

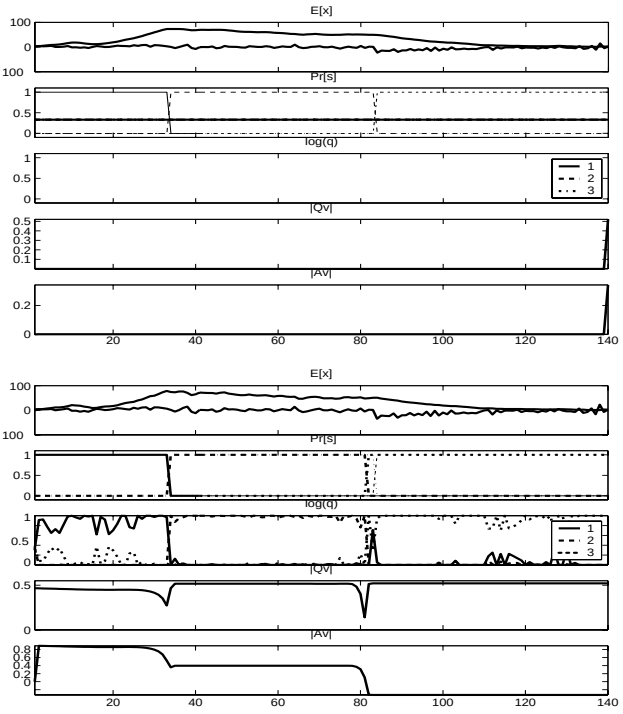


Figure 3. Iterations 1 and 3 of variational inference. The graphs depict (top-down): continuous state estimates $E[x]$, switching state estimates $Pr(s)$ and true state (thin lines), HMM variational parameter $\log q$, and determinants of LDS variational parameters $|Q_v|$ and $|A_v|$.

Initial uncertain switching state distribution $Pr(s_t)$ leads to low variational state noise variance $\hat{Q}$ (whose determinant is indicated by $|Q_v|$ in Figure 3.2) and low variational state transition matrix (whose determinant is indicated by $|A_v|$ in Figure 3.2). Through further iterations the variational inference algorithm converges to the true switching state sequence.

3.3. Approximate Generalized Pseudo Bayesian Inference

The Generalized Psuedo Bayesian [2, 11] (GPB) approximation scheme is based on the general idea of "collapsing", i.e. representing a mixture of M^t Gaussians with a mixture of M^r Gaussians, where $r < t$ (see [13] for a detailed review). While there are several variations on this idea, our focus is the GPB2 algorithm, which maintains a mixture of M^2 Gaussians over time and can be reformulated to include smoothing as well as filtering[2].

[2]Other similar pseudo Bayesian algorithms of [2], GBP1 and IMM, do not have an obvious smoothing reformulation.

GPB2 is closely related to the Viterbi approximation of Section 3.1. It differs in that instead of picking the most likely previous switching state j at every time step t and switching state i, we collapse the M Gaussians (one for each possible value of j) down into a single Gaussian.

Consider the filtered and predicted means $\hat{x}_{t|t,i,j}$ and $\hat{x}_{t|t-1,i,j}$, and their associated covariances, which were defined in Section 3.1. Assume, in addition, that for each switching state i and pairs of states (i,j) the following distributions are defined at each time step:

$$Pr(s_t = i|\mathcal{Y}_t)$$
$$Pr(s_t = i, s_{t-1} = j|\mathcal{Y}_t).$$

It is easy to show (see [13]) that a regular Kalman filtering update can be used to fuse the new measurement y_t and obtain S^2 new SLDS states at t for each S states at time $t-1$, in a similar fashion to the one in Section 3.1.

Unlike the Viterbi approximation which picks one best switching transition for each state i at time t, GPB2 "averages" over S possible transitions from $t-1$. Namely, it is easy to see that

$$Pr(s_t = i, s_{t-1} = j|\mathcal{Y}_t) \sim Pr(y_t|\hat{x}_{t,i,j})\Pi(i,j)Pr(s_{t-1} = j|\mathcal{Y}_{t-1}).$$

From there it follows immediately that the current distribution over the switching states is $Pr(s_t = i|\mathcal{Y}_t) = \sum_j Pr(s_t = i, s_{t-1} = j|\mathcal{Y}_t)$ and that each previous state j now has the following posterior

$$Pr(s_{t-1} = j|s_t = i, \mathcal{Y}_t) = \frac{Pr(s_t = i, s_{t-1} = j|\mathcal{Y}_t)}{Pr(s_t = i|\mathcal{Y}_t)}.$$

This posterior is important because it allows one to "collapse" or "average" the S transitions into each state i into one average state, e.g.

$$\hat{x}_{t|t,i} = \sum_j \hat{x}_{t|t,i,j} Pr(s_{t-1} = j|s_t = i, \mathcal{Y}_t).$$

Analogous expressions can be obtained for the variances $\Sigma_{t|t,i}$ and $\Sigma_{t,t-1|t,i}$.

Smoothing in GPB2 is unfortunately a more involved process that includes several additional approximations. Details of this can be found in [13]. We note at this point that, effectively, an assumption is made that decouples the MC model from the LDS when smoothing the MC states. Smoothed MC states are obtained directly from $Pr(s_t|\mathcal{Y}_t)$ estimates, namely $Pr(s_t = i|s_{t+1} = k, \mathcal{Y}_T) \approx Pr(s_t = i|s_{t+1} = k, \mathcal{Y}_t)$. Additionally, it is assumed that $\hat{x}_{t+1|T,i,k} \approx \hat{x}_{t+1|T,k}$. Armed with the two assumptions a set of smoothing equations for each transition (i,k) from $t+1$ to t can be obtained that obey an RTS smoother, followed by collapsing similar to the filtering step.

The GPB2 algorithm can now be summarized as the following pseudo code

Initialize LDS state estimates $\hat{x}_{0|-1,i}$ and $\Sigma_{0|-1,i}$;
Initialize $Pr(s_0 = i| - 1) = \pi(i)$;
for $t = 1 : T - 1$
  for $i = 1 : S$
    for $j = 1 : S$
      Predict and filter LDS state estimates $\hat{x}_{t|t,i,j}$, $\Sigma_{t|t,i,j}$;
      Find switching state distributions $Pr(s_t = i|\mathcal{Y}_t)$, $Pr(s_{t-1} = j|s_t = i, \mathcal{Y}_t)$;
      Collapse $\hat{x}_{t|t,i,j}$, $\Sigma_{t|t,i,j}$ to $\hat{x}_{t|t,i}$, $\Sigma_{t|t,i}$;
    end
    Collapse $\hat{x}_{t|t,i}$ and $\Sigma_{t|t,i}$ to $\hat{x}_{t|t}$ and $\Sigma_{t|t}$;
  end
end
Do GPB2 smoothing; Find sufficient statistics;

The inference process of GPB2 is clearly more involved than those of the Viterbi or the variational approximation. Unlike Viterbi, GPB2 provides soft estimates of switching states at each time t. Like Viterbi GPB2 is a local approximation scheme and as such does not guarantee global optimality inherent in the variational approximation. However, some recent work (see [4]) on this type of local approximation in general DBNs has emerged that provides conditions for it to be globally optimal.

4. Previous Work

SLDS models and their equivalents have been studied in statistics, time-series modeling, and target tracking since early 1970's. See [16, 13] for a review. Ghahramani [7] introduced a DBN-framework for learning and approximate inference in one class of SLDS models. His underlying model differs from ours in assuming the presence of S independent, white noise-driven LDSs whose measurements are selected by the Markov switching process. An alternative input-switching LDS model was proposed by Pavlovic et al. [15] and utilized for mouse motion classification. A switching model framework for particle filters is described in [9] and applied to dynamics learning in [3]. Manifold learning [8] is another approach to constraining the set of allowable trajectories within a high dimensional state space. An HMM-based approach is described in [5].

5. Experimental Results

There are two important empirical questions that should be addressed for the class of SLDS models:

- Which approximation inference scheme in SLDS results in the best learning performance?
- How does the performance of learned SLDS models compare to that of HMM models on tasks such as classification, tracking, and synthesis?

In this section we report some early progress in addressing these questions.

We applied an HMM and three SLDS frameworks which differed in the approximate inference technique (Viterbi, GPB2, and variational) to the analysis of two categories of fronto-parallel motion: walking and jogging. Fronto-parallel motions exhibit interesting dynamics and are free from the difficulties of 3-D reconstruction. Experiments can be conducted easily using a single video source, while self-occlusions and cluttered backgrounds make the tracking problem non-trivial.

We learned HMM and SLDS models from our data set, and evaluated their classification performance. Classification is an important task in its own right, and it is particularly useful in comparing SLDS and HMM models. The LDS component of the SLDS model provides more flexibility in fitting the underlying measurements, in comparison to HMMs. Classification accuracy is one way to measure the value of this additional modeling power.

We adopted the 2-D Scaled Prismatic Model proposed by Morris and Rehg [12] to describe the kinematics of the figure and define the state space for learning. The kinematic model lies in the image plane, with each link having one degree of freedom (DOF) in rotation and another DOF in length. A chain of SPM transforms can model the image displacement and foreshortening effects produced by 3-D rigid links. The appearance of each link in the image is described by a template of pixels which is manually initialized and deformed by the link's DOF's.

In our experiments we have analyzed the motion of the legs, torso, and head, and ignoring the arms. Our kinematic model had eight DOF's, corresponding to rotations at the knees, hip, and neck. Our dataset consists of 18 sequences of six individuals jogging (two examples of three people) and walking at a moderate pace (two examples of six people.) Each sequence was approximately 50 frames duration. We created the SPM measurements in each frame by hand, so as to guarantee fidelity to the observed motion.

5.1. Learning

The first task we addressed was learning an SLDS model for walking and running. Each of the two motion types were each modeled as multi–state[3] HMM and SLDS models and then combined into a single complex jog-walk model. In addition, each SLDS motion model was assumed to be of either the first or the second order[4]. Hence, a total of three models (HMM, first order SLDS, and second order SLDS) were considered for each switching state order.

HMM models were initially assumed to be fully connected. Their parameters were then learned using the standard EM learning, initialized by k-means clustering (see [17] for details.) HMM models were in turn used

[3] We explored models with one, two, and four states.

[4] Second order SLDS models imply $x_t = A_1(s_t)x_{t-1} + A_2(s_t)x_{t-2}$.

to initial switching state segmentations for more complex SLDS models. For SLDS models, the measurement matrix in all cases was assumed to be identity, $C = I$. The SLDS parameters of the model $(A, Q, R, x_0, \Pi, \pi_0)$ were then reestimated using the EM-learning framework. The E-step (inference) in SLDS learning was accomplished using the three approximated methods outlined in Section 3: Viterbi, GPB2, and variational inference.

Results of SLDS learning using either of the three approximate inference methods did not produce significantly different models. This can be explained by the fact that initial segmentations using the HMM and the initial SLDS parameters were all very close to a locally optimal solution and all three inference schemes indeed converged to the same or similar posteriors. Therefore, only the models learned using the Viterbi inference scheme were employed in the analyses of the next two sections.

5.2. Classification

We considered classification of unknown motion sequences as the first step in testing the impact of different dynamic models. Unknown motion sequences were considered to be the ones of *complex* motion, i.e., motion consisting of alternations of "jog" and "walk."[5] Identification of different motion "regimes" was conducted using the HMM inference under the learned HMM model and the approximate Viterbi, GPB2, and variational inference under the SLDS model. Estimates of "best" switching states $Pr(s_t)$ indicated which of the two models can be considered to be driving the corresponding motion segment.

Figure Figure 4 shows classification results for a complex motion sequence of jog and walk motion using different order HMM and SLDS models and different SLDS inference schemes. For instance, Figure 4(a) depicts true sequence of jog-walk motions in the top graph, followed by Viterbi, GPB2, variational, and HMM classifications. In this case each motion type (jog and walk) is modeled using one switching state SLDS (HMM). Furthermore, the LDS part of the SLDS model is of the second order. Figure 4(d) shows jog-walk classification using jog and walk models who contain four switching states each, and where SLDS models contain second order LDS.

The accuracy of classification increases as the order of the switching states and the SLDS model order increase. More interesting, however, is that the HMM model consistently yields lower segmentation accuracy then the SLDS model using any inference scheme. This is of course to be expected because the HMM model does not impose continuity across time in the plant state space (x), which does indeed exist in a natural figure motion (joint angles evolve continuously in time.) Analysis of different SLDS inference schemes indicates that Viterbi and variational schemes do seem to yield appealing classifications. However, GPB2 does not considerably lack behind the mentioned schemes and sometimes even outperforms the first two. Moreover, GPB2 clearly provides "soft" state estimates, while the Viterbi scheme does not. Variational inference tends to produce somewhat soft decisions, but is more often similar to Viterbi. In terms of computational complexity, Viterbi does seem to be the clear winner among the SLDS schemes.

[5] Test sequences were constructed by concatenating in random order randomly selected and noise corrupted training sequences. Transitions between sequences were smoothed using B-spline smoothing.

6. Conclusions

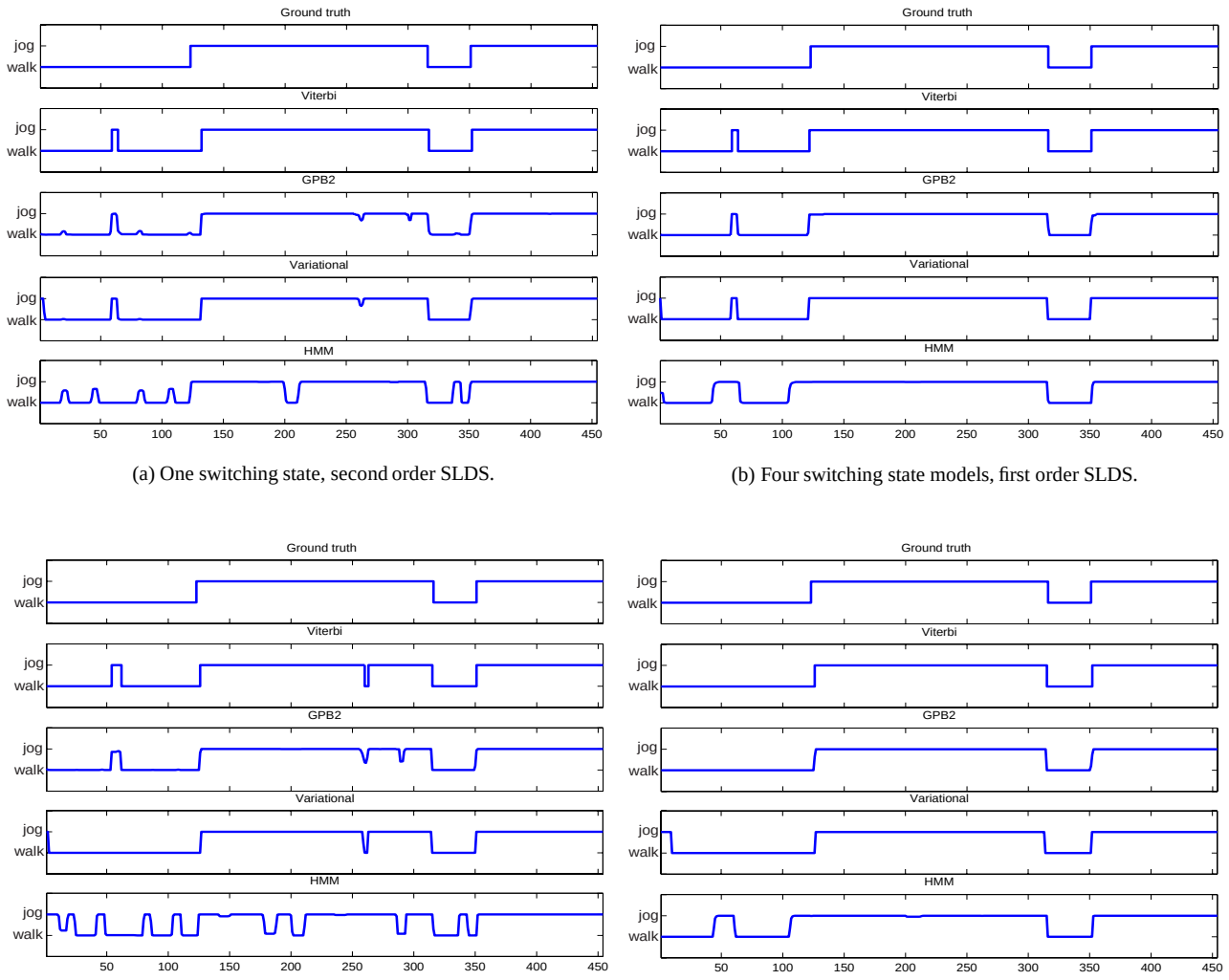
We have explored the impact of learned SLDS models on analysis of figure motion. We have compared the SLDS models using three different inference schemes (Viterbi, GPB2, and variational) to the more common HMM models. One of the considered inference scheme, approximate variational approximation, is novel in the SLDS domain.

Our comprehensive classification experiments have demonstrated promising results in the use of SLDS models for modeling of the human figure motion. We demonstrated accurate discrimination between walking and jogging motions. We showed that SLDS models provide more robust classification performance than the more commonly used HMM models. The fact that these models can be learned from data may be an important advantage in figure tracking, where accurate physics-based dynamical models may be prohibitively complex.

We are currently conducting additional experiments that will shed more light on the predictive qualities of SLDS models. This evaluation is crucial step in studying the use of learned models in applications such as figure tracking. We also plan to build SLDS models for wide variety of motions and performers and evaluate their performance.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal filtering*. 1979.
[2] Y. Bar-Shalom and X.-R. Li. *Estimation and tracking: principles, techniques, and software*. 1998.
[3] A. Blake, B. North, and M. Isard. Learning multi-class dynamics. In *NIPS '98*, 1998.
[4] X. Boyen, N. Firedman, and D. Koller. Discovering the hidden structure of complex dynamic systems. In *Proc. Uncertainty in Artificial Intelligence*, pages 91–100, 1999.
[5] M. Brand. Pattern discovery via entropy minimization. Technical Report TR98-21, MERL, 1998.
[6] C. Bregler. Learning and recognizing human dynamics in video sequences. In *CVPR*, pages 568–574, 1997.
[7] Z. Ghahramani and G. E. Hinton. Switching state-space models. submitted for publication, 1998.
[8] N. Howe, M. Leventon, and W. Freeman. Bayesian reconstruction of 3d human motion from single-camera video. In *NIPS'99*, 1999.
[9] M. Isard and A. Blake. A mixed-state CONDENSATION tracker with automatic model-switching. In *ICCV*, pages 107–112, 1998.



(a) One switching state, second order SLDS.

(b) Four switching state models, first order SLDS.

(c) Two switching state models, second order SLDS.

(d) Four switching state models, second order SLDS.

Figure 4. Comparison of classification results on complex jog-walk sequence using SLDS (Viterbi, GPB2, and variational inference) and HMM models (exact inference). Figures (a) through (d) depict impact of different model orders on this classification. See text for more details.

[10] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In *Learning in graphical models*. 1998.

[11] C.-J. Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60:1–22, 1994.

[12] D. D. Morris and J. M. Rehg. Singularity analysis for articulated object tracking. In *CVPR*, pages 289–296, 1998.

[13] K. P. Murphy. Learning switching kalman-filter models. Technical Report 98-10, Compaq Cambridge Research Lab., 1998.

[14] R. M. Neal and G. E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. In *Learning in graphical models*, pages 355–368. 1998.

[15] V. Pavlović, B. Frey, and T. S. Huang. Time-series classification using mixed-state dynamic Bayesian networks. In *CVPR*, pages 609–615, 1999.

[16] V. Pavlović, J. M. Rehg, T.-J. Cham, and K. P. Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *ICCV*, pages 94–101, 1999.

[17] L. R. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.