Parameterized duration modeling for Switching linear dynamic systems

Sang Min Oh James M. Rehg Frank Dellaert
GVU Center, College of Computing, Georgia Institute of Technology
{sangmin, rehg, dellaert}@cc.gatech.edu

Abstract

We introduce an extension of switching linear dynamic systems (SLDS) with parameterized duration modeling capabilities. The proposed model allows arbitrary duration models and overcomes the limitation of a geometric distribution induced in standard SLDSs. By incorporating a duration model which reflects the data more closely, the resulting model provides reliable inference results which are robust against observation noise. Moreover, existing inference algorithms for SLDSs can be adopted with only modest additional effort in most cases where an SLDS model can be applied.

In addition, we observe the fact that the duration models would vary across data sequences in certain domains, which complicates learning and inference tasks. Such variability in duration is overcome by introducing parameterized duration models. The experimental results on honeybee dance decoding tasks demonstrate the robust inference capabilities of the proposed model.

1. Introduction

One of the challenging problems in computer vision is the *interpretation* of video data. Even assuming that targets can be reliably tracked, we encounter the problem of interpreting the tracks obtained. Manual interpretation, as is often done in domains such as biology, is a time-consuming and error-prone process. Thus, it is desirable to develop methods that automatically interpret the data. In this paper we are mostly interested in "*labeling*", which is to automatically segment motion according to different behavioral modes.

We take a model-based approach, in which we employ a computational model of behavior in order to interpret the data. The basic generative model we adopt is the Switching Linear Dynamic System (SLDS) model [15, 16]. In an SLDS model, there are multiple linear dynamic systems (LDS) that underly the motion. We can then model the complex behavior of the target by switching within this set of LDSs. SLDS models have become increasingly popular in the vision and graphics communities as they provide an intuitive framework for describing the continuous but non-linear dynamics of real-world motion. For example, they have been used for human motion classification [15, 16, 17] and motion synthesis [21].

Nevertheless, the modeling capabilities of a standard SLDS are *limited* by the Markov assumption which is imposed upon the switching process. This process governs the transitions between LDSs and makes it possible for an SLDS to represent nonlinear dynamics. As a consequence of the Markov assumption, however, the probability of remaining in a given switching state follows a geometric distribution with the property that a duration of one time step has the largest probability mass. Hence, if we perform inference with standard SLDSs, the result is often an over-segmentation of the labels due to the excessive importance attached to short durations.

Therefore, the use of a more flexible duration modeling technique is required for reliable video interpretation. However, naive introduction of a duration model can cause another problem. In certain domains, the duration patterns vary across data sequences due to the distinct motion characteristics being exhibited by the targets : some targets may move slowly with rare motion switchings while other targets switch their motions frequently. In such cases, the use of a fixed duration model can lead to inaccurate results when the temporal patterns of a target differ substantially from the duration patterns encoded in the model.

In this paper, we present a parameterized duration model for SLDSs and overcome the problems mentioned above. Specifically, the proposed model improves on standard SLDSs by relaxing the Markov assumption at a time-step level to a coarser segment level. The durations are first modeled explicitly and then non-stationary duration functions are derived from them, both learned from data. Consequently, the proposed model has more descriptive duration modeling power and more robust inference capabilities than standard SLDSs. Moreover, we show how we can reuse the existing inference algorithms for SLDSs in the proposed model. In addition, we adaptively find the correct duration model of a test sequence using an EM algorithm which is a modified version of the inference algorithm presented in [13]. First, the sequence is labeled in the E-step and the duration model is re-estimated in the M-step, simultaneously improving both estimates in turn.

The remainder of this paper is organized as follows. The standard SLDS model and the notations to be used are described in Section 2. In Section 3, we introduce the duration model for SLDSs, which formulates the proposed model as a semi-Markov model. Then we show the parameterization of the duration model and an EM-based inference algorithm for the model. Finally, we demonstrate improved interpretation capabilities through experimental results on the honeybee dance decoding tasks.

2. Backgrounds

2.1. Switching Linear Dynamic Systems

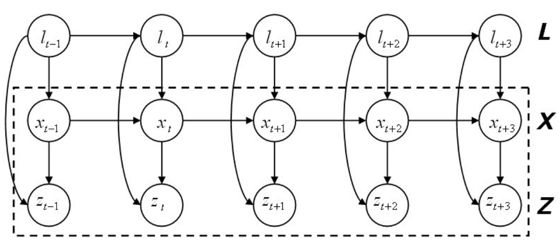


Figure 1. Switching linear dynamic systems (SLDS). The graphical model within the dashed rectangle is an LDS.

An LDS is a time-series state-space model that comprises a linear Gaussian dynamics model and a linear Gaussian observation model. The graphical representation of an LDS is shown in the dashed rectangle in Fig.1. The Markov chain in the middle represents the state evolution of the *continuous* hidden states x_t. Additionally, it assumes a prior Gaussian density p_1 on the initial state x_1. The middle chain is denoted as $X \triangleq \{x_t | 1 \le t \le T\}$, together with the observations $Z \triangleq \{z_t | 1 \le t \le T\}$ at the bottom. The linear Gaussian conditional dependencies on every edge and the Gaussian prior p_1 defines an LDS [1].

An SLDS is a natural extension of an LDS, where we assume the existence of n distinct LDS models $M \triangleq \{M_l | 1 \le l \le n\}$, where each model M_l is defined by the LDS parameters. However, we now have an additional discrete Markov chain $L \triangleq \{l_t | 1 \le t \le T\}$ at the top that determines which of the n models M_l is being used at every time-step. We call $l_t \in M$ the *label* at time t and L a *label sequence*.

In addition to a set of LDS models M, we specify two additional parameters: a probability distribution $\pi(l_1)$ over the initial label l_1 and an $n \times n$ transition matrix B that defines the switching behavior between the n distinct LDS models. In summary, a standard SLDS model is defined by the tuple $\Theta \triangleq \left\{\pi, B, M \triangleq \{M_l | 1 \le l \le n\}\right\}$.

2.2. Related work

Switching linear dynamic system (SLDS) models have been studied in a variety of research communities ranging from computer vision [3, 12, 13, 15, 16], computer graphics [17, 21], tracking [2], signal processing [4] and speech recognition [18], to econometrics [9], visualization [22], machine learning [6, 7, 11], control [20] and statistics [19].

It is formally proved that the exact inference in SLDS model is intractable [10]. Thus, there have been research efforts to derive efficient approximate inference methods. In this paper, we adopt the approximate Viterbi algorithm and variational inference method described in [15, 16].

The problem of duration modeling in HMMs has been addressed in the speech recognition communities, and several extensions to HMM models which provide enhanced duration modeling have appeared [5, 14]. The contribution of our work is that we validate the effectiveness of the idea in the SLDS framework and extend it to a parameterized form.

The idea of parameterized SLDS (P-SLDS) model appeared in [13] where an EM algorithm to perform inference with unknown global parameters was proposed. However, P-SLDS still suffers from the restriction to geometric durations. Our work improves on P-SLDS by allowing an arbitrary duration model which enhances the role of global parameters significantly.

3. SLDS with duration model

The duration models of standard SLDSs, i.e. the probability of remaining in a given switching state, are limited to a class of geometric distributions $P(d) = a^{d-1}(1-a)$ where d denotes the duration of a given switching state and a denotes the Markov transition probability of a self-transition. As a consequence, a duration of one time-step come to possess the largest probability mass (see Figure 2(b)).

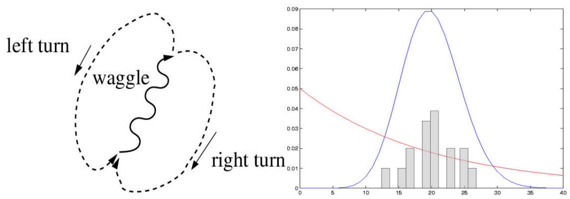


Figure 2. (a) A stylized honey bee dance. (b) A histogram of training data (gray), Gaussian (blue) and a geometric duration model (red).

In contrast, many natural temporal phenomena exhibit patterns of regularity in the duration over which a given model or regime is active. In such cases, the standard SLDS model would be unable to effectively encode the regularity of durations in data. The honey bee dance depicted in Fig.2(a) is an example: a dancer bee will attempt to stay in the waggle regime for a certain duration to correctly communicate the distance to the food sources. In such cases, it is clear that the actual duration diverges from a geometric distribution.

For example, we learned a duration model for the waggle phase using a realistic Gaussian density and a conventional geometric distribution from a real world honey bee dance data. Figure 2(b) shows the histogram of actual training data (gray bars), the learned geometric (red) and Gaussian (blue) distributions for comparison. It is observed that the learned geometric duration model does not exhibit any pattern of regularity in duration data. Hence, standard SLDS models are inappropriate for data which exhibits temporal patterns that deviate from geometric distributions.

3.1. Duration modeling in SLDS

We introduce a duration model for SLDSs which improves standard SLDSs by relaxing the Markov assumption at a time-step level to a coarser *segment level*. The durations are first modeled explicitly and then non-stationary duration functions are derived from them, both learned from data. Consequently, the resulting model has more descriptive power in modeling the duration, and provides improved inference capabilities. Furthermore, we show that one can reuse a large array of existing approximate inference algorithms for standard SLDSs in the resulting model with the advantage of duration modeling.

Conceptually, we deal with segments of finite duration, i.e. each segment $s_i \triangleq (l_i, d_i)$ is described by a tuple of label l_i and duration d_i. Within each segment, a fixed LDS model M_{l_i} is used to generate continuous state sequence for the duration d_i. Similar to SLDSs, we have an initial distribution $\pi(l_1)$ over the initial label l_1 of the first segment s_1, and an $n \times n$ semi-Markov[1] label transition matrix $\tilde{B}$ that defines the switching behavior between the segment labels. The tilde denotes that the matrix is a semi-Markov transition matrix. Additionally, however, we associate each label l with a fixed *duration model* D_l, represented as a histogram. We denote the set of n duration models as $D \triangleq \{D_l(d)|1 \leq l \leq n\}$, and refer to them as *explicit duration models* (a term we borrowed from the HMM community [5, 14]). In summary, the proposed model is defined by a tuple $\Theta \triangleq \left\{\pi, \tilde{B}, D \triangleq \{D_l|1 \leq l \leq n\}, M \triangleq \{M_l|1 \leq l \leq n\}\right\}$.

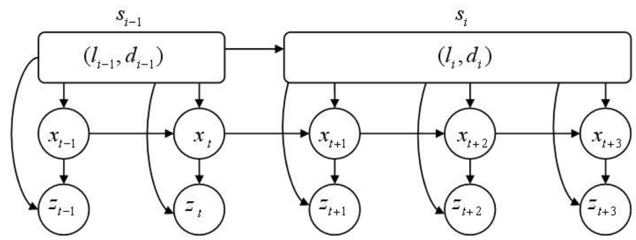


Figure 3. A schematic sketch of an S-SLDS with explicit duration models.

A schematic depiction of the model is given in Fig. 3. The top chain is a series of segments where each segment is depicted as a rounded box. In the model, the current segment $s_i \triangleq (l_i, d_i)$ generates the next segment s_{i+1} in the following manner: First, the current label l_i generates the next label l_{i+1} based on the transition matrix $\tilde{B}$; then, the next duration d_{i+1} is generated from the duration model for the label l_{i+1}, i.e. $d_{i+1} \sim D_{l_{i+1}}(d)$. The dynamics for the continuous hidden states and observations are identical to a standard SLDS : a segment s_i evolves the continuous hidden states X with a corresponding LDS model M_{l_i} for the duration d_i, then the observations Z are generated given the labels L and the continuous states X.

Now, we present a formal graphical representation of the proposed model : the conceptual generative model depicted in Fig. 3 is transformed into a concrete model that uses conventional model switching at every time-step. To maintain the same duration semantics, we introduce *counter variables* $C \triangleq \{c_t|1 \leq t \leq T\}$. The resulting graphical model is illustrated within the dashed rectangle in Fig.4, and is identical to the graphical model of an SLDS in Fig. 1, but with additional top-chain representing a series of counter variables C

[1]Semi-Markov models apply the Markov assumption at coarser segment level rather than at conventional time-step level : the label l_i of the current segment s_i only depends on the previous label l_{i-1}, however, the Markov assumption is not applied to durations, i.e. two successive durations d_i and d_{i-1} are independent given the labels.

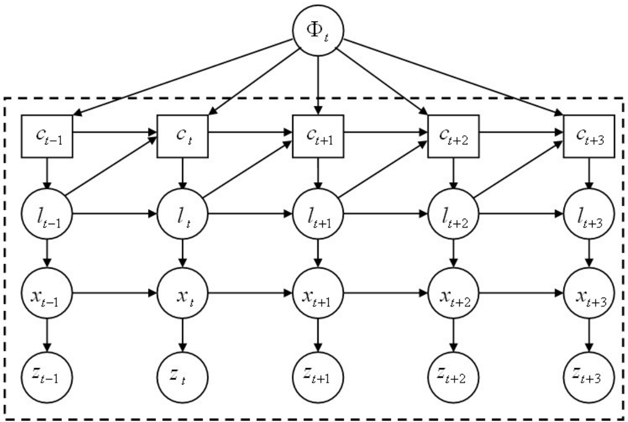


Figure 4. Graphical model within the dashed rectangle is an SLDS with duration modeling. The global temporal parameter at the top parameterizes the switching behaviors.

which is analogous to the ideas studied in the speech recognition community, e.g. see [23].

The counter chain C maintains an incremental counter which evolves based on a set of *non-stationary transition functions* (NSTFs) $U \triangleq \{U_l(c)|1 \leq l \leq n\}$. An NSTF U_l for the current label l_t defines the conditional dependency of the next counter variable c_{t+1} given the current counter variable c_t and the label l_t : $U_l(c_t) = P(c_{t+1}|c_t, l)$. The system can either increment the counter, i.e. $c_{t+1} \leftarrow c_t + 1$, or reset it to one, i.e. $c_{t+1} \leftarrow 1$. If the counter variable c_{t+1} is reset, then a label transition occurs : a new segment is initialized. A new label l_{t+1} is chosen based on the label transition matrix $\tilde{B}$. If the counter simply increments, then the new label is set to be the current label l_t, i.e. $l_{t+1} \leftarrow l_t$.

While previously introduced explicit duration models D for the model in Fig. 3 are more intuitive and readily obtained from the labeled data, it is necessary to transform the explicit duration models D into an equivalent NSTFs U to incorporate the conceptual knowledge about durations into a graphical model framework in Fig. 4 : two models are equivalent, but the latter NSTF representation provides a form in which one can systematically infer the labels at all time-steps with unknown segmentation points.

The equivalent NSTFs U are exactly evaluated from the explicit duration models D as follows :

$$U_l(c_t) = 1 - \left(D_l(c_t) / \sum_{d=c_t}^{D_l^{max}} D_l(c_t) \right) \qquad (1)$$

Above, D_l^{max} denotes the maximum duration allowed for the l th model. Intuitively, the latter composite term on the r.h.s. denotes the probability to reset the counter variable c_{t+1}. It represents the ratio of the probability of current duration c_t over the sum of probability mass in the future, i.e. durations equal or greater than c_t.

In summary, an SLDS with duration modeling is completely defined by a tuple $\Theta \triangleq \left\{\pi, \tilde{B}, U \triangleq \{U_l|1 \leq l \leq n\}, M \triangleq \{M_l|1 \leq l \leq n\}\right\}$ where the NSTFs U are obtained from the explicit duration models D.

3.2. Learning and Inference

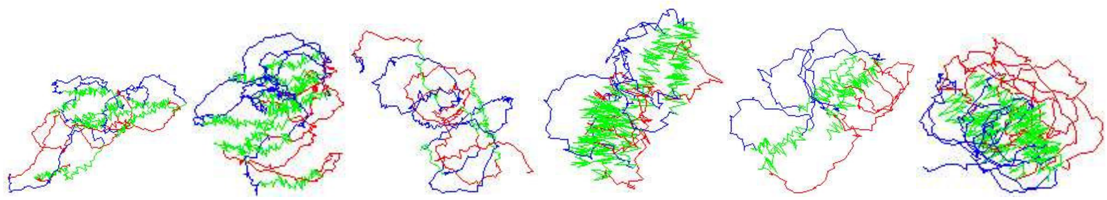
Learning in the proposed enhanced model is analogous to learning in SLDS using EM [15, 16]. The initial distribution π and LDS model parameters M are learned in exactly the same manner as in SLDS. However, it is necessary to learn the additional duration models D and the semi-Markov transition matrix $\tilde{B}$. These two additional model parameters only influence the label sequence L, and hence the ML estimates of these two parameters can be iteratively evaluated from a segmental representation of the label sequence L inferred in every E-step, i.e., $L = \cup_{j=1}^{|s|} s_j$. The specific functional forms of ML estimation depends on the choice of duration models.

Inference in the proposed model is feasible by applying any existing approximate inference algorithms for standard SLDSs. This is possible as we can convert the proposed model with duration modeling into an equivalent SLDS. The model conversion into an equivalent SLDS is possible by applying the standard technique of merging multiple discrete variables into meta variables. Specifically, all possible pairs of a label l_t and a counter value c_t are merged and form a set of "lc" variables where $\mathcal{LC} \triangleq \{(l, c_i)|1 \leq l \leq n, 1 \leq c_i \leq D_l^{max}\}$. To obtain a complete SLDS model, an equivalent $n' \times n'$ transition matrix B' where $n' \triangleq \sum_{l=1}^{n} D_l^{max}$ is constructed from the semi-Markov transition matrix $\tilde{B}$ and the NSTFs U, as follows :

$$B'_{(l_i,c_i),(l_j,c_j)} = \begin{cases} U_{l_i}(c_i) & \text{incr.} \\ \tilde{B}_{l_i,l_j}(1 - U_{l_i}(c_i)) & \text{reset} \\ 0 & \text{else} \end{cases} \qquad (2)$$

In Eq.2, the three cases differ as follows : (increment) $l_i = l_j$ and $c_j = c_i + 1$. (reset) $c_j = 1$. (otherwise) all other cases. In addition, the initial label distribution π' for the equivalent SLDS is set to be a uniform distribution as there is no guarantee that the target in the video starts its behavior in the first frame.

However, it is important to consider an efficient implementation because the naive reuse of existing algorithms may induce substantial computational overhead : it will result in an additional computational



(1) (2) (3) (4) (5) (6)

Figure 5. Bee dance sequences used in the experiments. Each dance trajectory is the output of a vision-based tracker.

cost factor of $O(D_{max}^2)$ where $D_{max} \triangleq \max\{D_l^{max}\}_{l=1}^n$, i.e. the ratio between the number of all lc states and original discrete states. This overhead applies to the inference methods that involve pairwise computations between two successive time-steps, e.g. approximate Viterbi [15] or a variational method [6, 15] with original overall performance $O(|L|^2)$. However, we observe that only $|L|+1$ transitions are allowed for every lc state : $|L|$ resets and one increment. Hence, we can achieve an overall performance of $O(D_{max}|L|^2) = O(D_{max}|L| \times (|L|+1))$ via exploiting this fact, which results in reduced overhead by a factor of $O(D_{max})$. Consequently, we can adopt the more powerful duration modeling capabilities at the cost of a modest complexity increase over the standard SLDS model.

3.3. Parameterized duration model

We propose the use of a parameterized duration model to deal with the varying duration patterns in data sequences. We refer to this model as PS-SLDS. In many applications, certain global factors Φ_t systematically vary the duration models D across the data sequences. For example, honey bees stay in waggle phase for shorter or longer times depending upon the individual sequences, i.e. the duration model of waggle phase can be represented as a Gaussian with varying means, where the means are proportional to the distances to the food sources. In such cases, it is important to identify the set of correct Gaussian means Φ_t to infer correct labels.

Such global temporal parameters Φ_t can be additionally encoded into the graphical model framework where Φ_t appears at the top of Fig. 4 and influences the incremental counter chain C. The depicted graphical model has some similarity to the parametric SLDS model (P-SLDS) in [13]. The major difference is that our work improves on P-SLDS by allowing arbitrary duration models and removing the constraint that the global temporal parameters Φ_t be restricted to the class of geometric distributions.

Inference in PS-SLDS can be performed using the EM algorithm summarized in Algorithm 1 which is an

Algorithm 1 EM for inference in PS-SLDS

- **E-step :** obtain the posterior distribution on labels L :

$$f_I^i(L) \triangleq P(L, |Z, \Theta, \Phi_t^i) \quad (3)$$

  over the hidden label sequence L and the state sequence X, using a current guess for the global temporal parameters Φ_t^i.

- **M-step :** improve the duration models by maximizing the expected log-likelihood:

$$\Phi_t^{i+1} \leftarrow \underset{\Phi_t}{\operatorname{argmax}} \ \langle \log P(L|\Theta, \Phi_t) \rangle_{f_I^i(L)} \quad (4)$$

adapted version of the inference algorithm for P-SLDS [13]. In the E-step, we infer the labels L given the current set of global temporal parameters Φ_t and the fixed SLDS parameters Θ described in Section 2.1. Then, we update the global temporal parameters Φ_t by maximizing the expected log-likelihood based on the obtained posterior distribution on labels L. Here, $\langle\cdot\rangle_p$ denotes the expectation of a function $(\cdot)$ under a distribution p. Hence, we simultaneously improve labeling results and adaptively find duration models through the iterative scheme of EM.

4. Experimental Results

The experimental results show that PS-SLDS provides improved labeling abilities over the standard SLDS. We used six dancer bee tracks shown in Fig. 5, which were obtained automatically using a vision-based tracker [8]. The 6 video sequences were of length 1058, 1125, 1054, 757, 609 and 814 frames, respectively. The observation data were a time-series sequence of vectors $z_t = [x_t, y_t, \cos(\theta_t), \sin(\theta_t)]^T$ where x_t,y_t and θ_t respectively denote the 2D coordinates and the heading angle at time t. The triangular function elements in the observations were introduced to make the system to be able to learn the location-invariant rotating motions. Note from Fig.5 that the tracks are noisy and much more irregular than the idealized stylized dance prototype shown in Fig.2(a). The red, green and blue colors represent right-turn, waggle and left-turn phases. The ground-truth labels $\bar{L}$ are marked manually for the comparison and learning purposes. The dimensionality of the continuous hidden states was set to be four.

Given the relative difficulty of obtaining this data, which has to be labeled manually to allow for a ground-truth comparison, we adopted a leave-one-out (LOO) strategy. The parameters are learned from five out of six datasets, and the learned model is applied to the left-out dataset to perform labeling. Six experiments are performed using both PS-SLDS and the standard SLDS, switching the test data sequence.

4.1. Learning from training data

The parameters of both PS-SLDS and standard SLDS are learned from the data sequences depicted in Fig. 5. The standard SLDS model parameters were learned in a standard manner without any restriction on the parameter structures based on the training labels $\bar{L}$ and observations Z, as described in [15, 16]. The covariances of duration models and the semi-Markov transition matrix $\tilde{B}$ were learned from the training data as well where we additionally provided manually found global temporal factors, i.e. means of Gaussian duration models Φ_t.

4.2. Inference on test data

In the test phase, the set of learned parameters were used to infer the labels of the left-out test sequence. An approximate Viterbi method (VI) [15] and variational approximation (VA) methods [6, 15] were used to infer the labels in standard SLDSs. The initial probability distributions for the VA method were initialized based on VI labels. Simply, VI labels were trusted by a probability of 0.8 and the other two labels at every time-step are assigned probability of 0.1 respectively.

For the inference in PS-SLDS, a VI method was used in the E-step for labeling. A VI method is adopted as it is simple and reported to be comparable to the other methods [15].

The experimental results show the superior recognition capabilities of the proposed PS-SLDS model over the SLDSs. The label inference results on all six sequences are shown in Fig.6. The four color strips in each figure represent SLDS VI, SLDS VA, PS-SLDS VI and the ground-truth (G.T.) labels from the top to the bottom. The x-axis represents time flow and the color is the label at that corresponding video frame.

The superior recognition abilities of PS-SLDS can be observed from the presented results. The PS-SLDS results on the first three sequences match closer to the

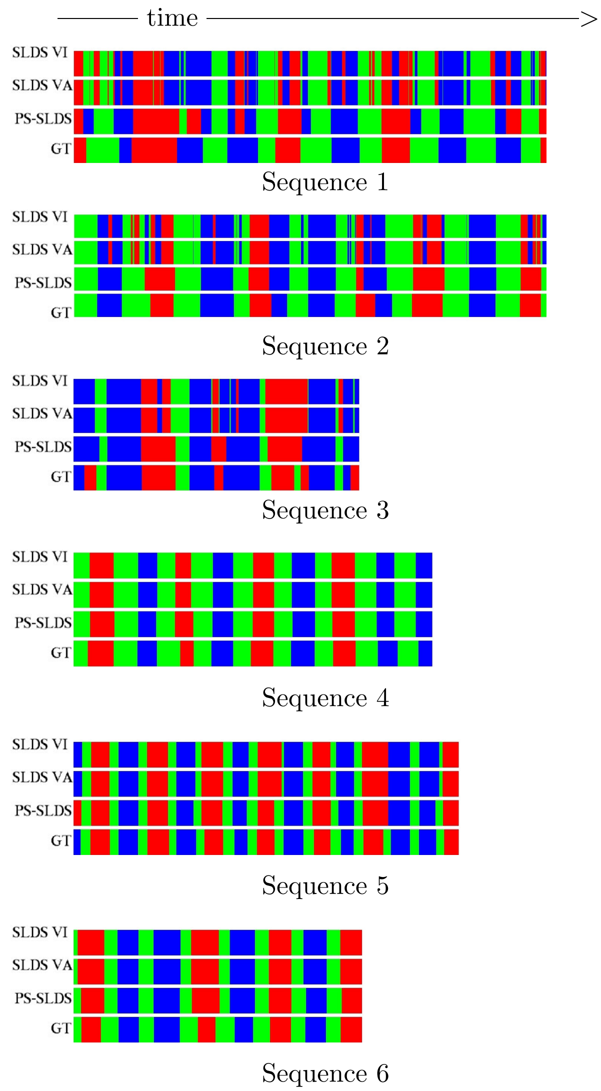


Figure 6. Label inference results. Estimates from standard SLDS and proposed models are compared to the manually-obtained ground truth (GT) labels. Key : waggle (green), right-turn (red), left-turn (blue).

ground truths than the SLDS results. This is important as the sequences 1, 2 and 3 are challenging : the observation data were more noisy and the patterns of switching in the dance modes and the durations in each dance regime are more irregular than the other sequences. The PS-SLDS results on sequences 5, 6 and 6 were mostly superior or comparable to SLDS results.

It can be observed that most of the over-segmentations that appear in the SLDS labeling results disappear in the PS-SLDS labeling results. PS-SLDS estimates still introduce some errors, especially in the sequences 1 and 3. However, given that even an expert human can introduce labeling noise, the labeling capabilities of PS-SLDS are fairly good.

Sequence	1	2	3	4	5	6
PS-SLDS	**75.9**	**92.4**	**83.1**	**93.4**	**90.4**	**91.0**
SLDS VI	71.6	82.9	78.9	92.9	89.7	89.2
SLDS VA	71.9	82.8	78.9	92.9	89.7	89.2

Table 1. Accuracy of label inference in percentage. Sequence numbers refer to Figure 5.

Finally, Table.1 shows the overall accuracy of the inferred labels in percentage, statistics from PS-SLDS and SLDS VI and SLDS VA results from top to the bottom. It can be observed that PS-SLDS provides very accurate labeling results w.r.t. the ground truth. Moreover, PS-SLDS consistently improves on standard SLDSs across all six datasets. The overall experimental results show that PS-SLDS model is promising and provides robust inference capabilities.

5. Conclusion

We presented a parameterized duration modeling technique for SLDSs. It overcomes the limitations of the simple geometric duration models induced in standard SLDSs, and actively adapts to the duration patterns in the test data effectively.

The learning and inference algorithms for the proposed model were introduced and an efficient implementation technique was discussed. The proposed model provides more powerful duration modeling capabilities than the standard SLDS at a modest cost, and its benefits have been validated experimentally.

Acknowledgment

This work is supported in part by Samsung Lee Kun Hee scholarship awarded to Sang Min Oh, NSF Award IIS-0133779 awarded to James M. Rehg, and NSF Award IIS-0219850 and NSF CAREER Award IIS-0448111 awarded to Frank Dellaert. We thank the reviewers for their helpful comments.

References

[1] Y. Bar-Shalom and X. Li. *Estimation and Tracking: principles, techniques and software*. Artech House, Boston, London, 1993.

[2] Y. Bar-Shalom and E. Tse. Tracking in a cluttered environment with probabilistic data-association. *Automatica*, 11:451–460, 1975.

[3] Christoph Bregler. Learning and recognizing human dynamics in video sequences. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1997.

[4] A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Trans. Signal Processing*, 49(3), 2001.

[5] J. Ferguson. Variable duration models for speech. In *Symposium on the Aplication of HMMs to Text and Speech*, 1980.

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

[7] A. Howard and T. Jebara. Dynamical systems trees. In *Conf. on Uncertainty in Artificial Intelligence*, pages 260–267, Banff, Canada, 2004.

[8] Z. Khan, T. Balch, and F. Dellaert. A Rao-Blackwellized particle filter for EigenTracking. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[9] C.-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1994.

[10] U. Lerner and R. Parr. Inference in hybrid networks: Theoretical limits and practical algorithms. In *Proc. 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, pages 310–318, Seattle, WA, 2001.

[11] U. Lerner, R. Parr, D. Koller, and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proc. AAAI*, Austin, TX, 2000.

[12] B. North, A. Blake, M. Isard, and J. Rottscher. Learning and classification of complex dynamics. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(9):1016–1034, 2000.

[13] S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert. Learning and Inference in Parametric Switching Linear Dynamic Systems. In *Intl. Conf. on Computer Vision (ICCV)*, 2005.

[14] M. Ostendorf, V. V. Digalakis, and O. A. Kimball. From hmm's to segment models : A unified view of stochastic modeling for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 4(5):360–378, 1996.

[15] V. Pavlović and J.M. Rehg. Impact of dynamic model learning on classification of human motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2000.

[16] V. Pavlović, J.M. Rehg, T.-J. Cham, and K. Murphy. A dynamic Bayesian network approach to figure tracking using learned dynamic models. In *Intl. Conf. on Computer Vision (ICCV)*, 1999.

[17] L. Ren, A. Patrick, A. Efros, J. Hodgins, and J. M. Rehg. A data-driven approach to quantifying natural human motion. *ACM Trans. on Graphics, Special Issue: Proc. of 2005 SIGGRAPH Conf.*, 2005.

[18] A-V.I. Rosti and M.J.F. Gales. Rao-blackwellised Gibbs sampling for switching linear dynamical systems. In *Intl. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, volume 1, pages 809–812, 2004.

[19] R.H. Shumway and D.S. Stoffer. Dynamic linear models with switching. *Journal of the American Statistical Association*, 86:763–769, 1992.

[20] R. Vidal, A. Chiuso, and S. Soatto. Observability and identifiability of jump linear systems. In *Proceedings of IEEE Conference on Decision and Control*, 2002.

[21] Y.Li, T.Wang, and H-Y. Shum. Motion texture : A two-level statistical model for character motion synthesis. In *SIGGRAPH*, 2002.

[22] O. Zoeter and T. Heskes. Hierarchical visualization of time-series data using switching linear dynamical systems. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(10):1202–1215, October 2003.

[23] G. Zweig and S. Russell. Probabilistic modeling with bayesian networks for automatic speech recognition. *Australian Journal of Intelligent Information Processing Systems*, 5(4):253–260, 1999.