Two-Hand Gesture Recognition using Coupled Switching Linear Model

Abstract

We present a method coupling multiple switching linear models. The coupled switching linear model is an interactive process of two switching linear models. Coupling is given through causal influence between their hidden discrete states. The parameters of this model are learned via EM algorithm. Tracking is performed through the coupled-forward algorithm based on Kalman filtering and a collapsing method. A model with maximum likelihood is selected out of a few learned models during tracking. We demonstrate the application of the proposed model to tracking and recognizing two-hand gestures.

1. Introduction

Gesture recognition plays an important role in a host of man-machine interaction applications. Although some gestures are expressed by one hand, many of them are done by two hands. To model these two-hand gestures, we have to consider interactions between the two hands. We assume that a two-hand gesture is an interacting process of the two hands whose shapes and motions are described by the switching linear dynamics [2,3,6], and propose a coupled switching linear dynamic model to capture interactions between the two hands.

CHMM (coupled hidden Markov model) [4] has been proposed to deal with interacting processes. However, since CHMM inherits from HMM, it has a limitation in treating time series having dependencies like shape-changing hand gestures.

Reynard [5] has introduced a coupling concept to track complex motions, however that means just a coupling of two kinds of continuous state variables in a single process, and is essentially different from interaction considered here.

We demonstrate an application of the coupled switching linear model to tracking and recognizing two hands whose shapes change during their motion. The presented coupling scheme enables tracking both hands even when one of them is not observed well in images by occlusions or complex backgrounds. And it also gives probabilistic explanation to recognition of gestures by combination of two hands.

2. Coupled switching linear model

2.1 Model specification

To represent a variety of shapes of a hand, it may be an efficient way that outlines of the hand are parameterized by active contour model using B-spline, which was well established in [1]. A curve is parameterized into a control vector composed of B-spline control points. A control vector is transformed to a low-dimensional shape vector on a specific shape space formed with some key control vectors. Then the shape vector, s_t, is considered as a state vector in switching linear dynamics:

$$x_t = \begin{pmatrix} s_t \\ \dot{s}_t \end{pmatrix}$$

Switching linear model can be seen as a hybrid model of the linear state-space model and HMM. It is described using the following set of state-space equations:

$$\begin{aligned} & x_t = F_{m_t} x_{t-1} + D_{m_t} + u_t,\ u_t \sim N(0, Q_{m_t}) \\ & \Phi_{m_t, m_{t+1}} = p(m_{t+1} \mid m_t) \\ & \pi_{m_1} = p(m_1) \end{aligned} \quad (1)$$

In the above equations, x_t is a hidden continuous state vector. u_t is independently distributed on the Gaussian distribution with zero-mean and covariance Q_{m_t}, π_{m_1}, F_{m_t} and D_{m_t}, which are typical parameters of linear dynamic model, denote the prior probability of a discrete state, the continuous state transition matrix, and the offset, respectively. The parameters with the subscript m_t are dependent on the discrete state variable m_t indexing a linear dynamic model. And the switching process between discrete states obeys the first Markov process and is defined with the discrete state transition matrix Φ.

Coupled switching linear model is an interactive process of two switching linear models. Coupling is given through causal influence between their hidden discrete states. The complex state space representation is equivalently depicted by dependency graph in figure 1.

To accommodate another interacting process, it seems good enough to consider a single lumped system with dimension-increased state variables. However, there exist a few problems. Due to increased number of discrete states, the computational cost is prohibitive, and sufficient data can rarely be obtained for estimation of parameters, usually resulting in under-sampling and numerical underflow errors [4]. Consequently, the suggested coupling scheme, as shown in figure 1, offers conceptual advantages of parsimony and clarity with computational benefits in efficiency and accuracy. This is revealed in the following sections.

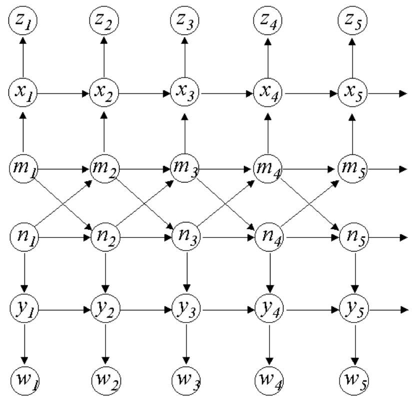


Figure 1. Coupled switching linear model. m_t and n_t denote discrete state variables. x_t and y_t denote continuous state variables. z_t and w_t denote observation vectors.

In the coupled switching linear model, since transition between discrete states is Markov process, it follows that

$$p(m_t, n_t \mid m_1, \cdots, m_{t-1,} n_1, \cdots, n_{t-1}) = p(m_t, n_t \mid m_{t-1}, n_{t-1})$$

Assuming

$$p(m_t, n_t \mid m_{t-1}, n_{t-1}) \propto p(m_t \mid m_{t-1}) \cdot p(m_t \mid n_{t-1}) p(n_t \mid n_{t-1}) p(n_t \mid m_{t-1}) \tag{2}$$

referred to in [4], transition probability of joint discrete states can be parameterized as

$$p(m_t, n_t \mid m_{t-1}, n_{t-1}) = k_c \Phi_{m_{t-1} m_t} \Gamma_{n_{t-1} m_t} \hat{\Phi}_{n_{t-1} n_t} \hat{\Gamma}_{m_{t-1} n_t} \tag{3}$$

where k_c is a normalizing constant, Γ is the state transition matrix representing causal influences between two switching linear system, and superscript ^ denotes the lower switching linear system in figure 1.

2.2 Coupled-forward algorithm

Following [3], given the known parameters of switching linear dynamics, the predicted joint-continuous state variable and the corresponding covariance are defined dependently on $m_{t-1} = i$ and $m_t = j$:

$$\begin{aligned} x_{t|t-1}^{(i,j)} &= F_j x_{t-1|t-1}^{(i)} + D_j \\ P_{t|t-1}^{(i,j)} &= F_j P_{t-1|t-1}^{(i)} F_j' + Q_j \end{aligned} \tag{4}$$

where $x_{t-1|t-1}^{(i)}$ and $P_{t-1|t-1}^{(i)}$ are the filtered continuous states and its covariance at time $t-1$ based on information up to time $t-1$. Now the filtered joint-continuous state $x_{t|t}^{(i,j)}$ and its covariance $P_{t|t}^{(i,j)}$ are estimated by the conventional Kalman updating algorithm. In particular, we follow Kalman filtering application of [1] to active contour model.

From the above fact, as noted by [3], switching linear dynamic model requires computing a Gaussian mixture with M^t components at time t for M switching states. If coupled with a N-switching linear system as shown in figure 1, typically $M^t + N^t$ computations are required, which is clearly intractable for moderate sequence length. It is necessary to introduce some approximations to solve the intractable computation problem.

We collapse $M^2 + N^2$ jointed continuous state variables into $M + N$ state variables at each time, and can avoid prohibitive increase of computational cost. The following collapsing method is given: Expediently only in terms of the upper system in figure 1.

$$x_{t|t}^{(j)} = \frac{\sum_{i=1}^{M} \left(\sum_{ii=1, jj=1}^{N} p \begin{pmatrix} m_{t-1} = i, n_{t-1} = ii, \\ m_t = j, n_t = jj \mid O_t \end{pmatrix} \cdot x_{t|t}^{(i,j)} \right)}{p(m_t = j \mid O_t)} \tag{5}$$

$$P_{t|t}^{(j)} = \frac{\sum_{i=1}^{M} \left(\begin{array}{l} \sum_{ii=1, jj=1}^{N} p(m_{t-1} = i, n_{t-1} = ii, m_t = j, n_t = jj \mid O_t) \\ \cdot \left(P_{t|t}^{(i,j)} + (x_{t|t}^{(j)} - x_{t|t}^{(i,j)})(x_{t|t}^{(j)} - x_{t|t}^{(i,j)})' \right) \end{array} \right)}{p(m_t = j \mid O_t)}$$

where O_t is a sequence $(o_1, o_2, \cdots, o_t)$ and o_t is an observation vector (z_t, w_t).

The filtered coupled-joint distribution of discrete states is defined by

$$\begin{aligned} & p(m_{t-1}, n_{t-1}, m_t, n_t \mid O_t) \\ &= k_t p(z_t \mid x_{t|t-1}^{(m_{t-1}, m_t)}) p(w_t \mid y_{t|t-1}^{(n_{t-1}, n_t)}) p(m_{t-1}, n_{t-1}, m_t, n_t \mid O_{t-1}) \end{aligned} \tag{6}$$

where k_t is a normalizing constant. From (2) and (3) the prediction step given sequence up to time t gives

$$\begin{aligned} p(m_t, n_t, m_{t+1}, n_{t+1} \mid O_t) &= k_p \Phi_{m_t m_{t+1}} \Gamma_{n_t m_{t+1}} \\ & \cdot \hat{\Phi}_{n_t n_{t+1}} \hat{\Gamma}_{m_t n_{t+1}} \sum_{m_{t-1}, n_{t-1}} p(m_{t-1}, n_{t-1}, m_t, n_t \mid O_t) \end{aligned} \tag{7}$$

$$p(m_{t+1}, n_{t+1} \mid O_t) = \sum_{m_t, n_t} p(m_t, n_t, m_{t+1}, n_{t+1} \mid O_t) \tag{8}$$

$$p(m_t \mid O_t) = \sum_{m_{t-1}, n_t, n_{t-1}} p(m_t, n_t, m_{t-1}, n_{t-1} \mid O_t)$$

$$x_{t|t} = \sum_{m_t=1}^{M} p(m_t \mid O_t) x_{t|t}^{(m_t)}$$

where k_p is a normalizing constant. Now (6) and (7) are iterated during filtering process.

2.3 Coupled -backward algorithm

While the coupled-forward algorithm is a filtering process given sequence up to current time, the coupled-backward algorithm is a smoothing process given sequence of full length. Like the conventional Kalman smoothing method, joint-continuous state variable and its covariance are smoothed [3]. And the collapsing is similarly performed using the following probability of the smoothed coupled-joint discrete states:

$$p(m_t,n_t,m_{t+1},n_{t+1} \mid O_T) = p(m_t,n_t,m_{t+1},n_{t+1} \mid O_t) \cdot \frac{p(m_{t+1},n_{t+1} \mid O_T)}{p(m_{t+1},n_{t+1} \mid O_t)} . \tag{9}$$

From (9) the followings are obtained as

$$p(m_t,n_t \mid O_T) = \sum_{m_{t+1},n_{t+1}} p(m_t,n_t,m_{t+1},n_{t+1} \mid O_T) \tag{10}$$

$$p(m_t \mid O_T) = \sum_{n_t} p(m_t,n_t \mid O_T) \tag{11}$$

3. EM learning

EM algorithm is a general iterative technique for finding maximum likelihood parameter estimates in problems where some variables are unobserved [7]. It is natural to use EM algorithm for our problem, in which unobserved variables are continuous state variables x_t, y_t and discrete state variables m_t, n_t.

Assuming that the probability density for an observation sequence is parameterized using λ, which consists of $\{F,Q,\pi,\Phi,\Gamma\}$ and $\{\hat{F},\hat{Q},\hat{\pi},\hat{\Phi},\hat{\Gamma}\}$, its auxiliary log-likelihood is given by

$$\begin{aligned} L &= \sum_{M_T,N_T} \int_{X_T,Y_T} \overline{p} \log p(M_T,N_T,X_T,Y_T,O_T \mid \lambda) dX_T dY_T \\ &= E_{\overline{p}}[\log p(M_T,N_T,X_T,Y_T,O_T \mid \lambda)] \end{aligned} \tag{12}$$

where (M_T,N_T) and (X_T,Y_T), are sequences(of length T) of discrete states and continuous states, respectively, $\overline{\lambda}$ is the parameter set estimated previously, and $\overline{p} = p(M_T,N_T,X_T,Y_T \mid O_T,\overline{\lambda})$. EM algorithm starts with some initial guess and proceeds by applying the following two steps repeatedly until the likelihood converges:

E-step On the condition given the observation sequence of full length O_T and the previous parameter set $\overline{\lambda}$, we estimate the hidden continuous states and discrete states through the backward process following the forward process described in sections 2.2 and 2.3.

M-step If L is expressed by λ and the estimations from E-step, then we estimate λ maximizing L.

4. Recognition

Recognition of hand gestures can be considered as the problem to determine which model tracks a hand gesture well. Therefore, a given sequence of hand gestures can be recognized by means of the likelihood values of candidate models.

In order to track and recognize hand gestures simultaneously, we have to compute the likelihood of each model while tracking is being performed with the coupled-forward algorithm.

If the coupled switching linear model is represented by the parameter set λ, log-likelihood $\widetilde{L}_\tau$ of λ at time τ is obtained by

$$\widetilde{L}_\tau = \sum_{t=1}^{\tau} \log(\frac{1}{k_t}) \tag{13}$$

where k_t has been computed in (6).

5. Experimental Result

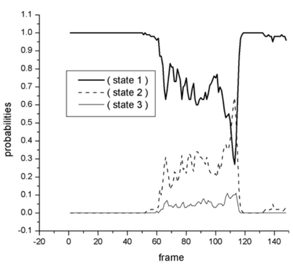
An observation, o_t, is shown as edges detected by line searching along the normal direction at sample points on a hand contour [1]. For robustness and accuracy of the one dimensional edge detection, we use a Mahalanobis distance from a mean color of a hand with its covariance rather than common gray intensities. However, the edge detection using color models fails well when a hand moves in front of a face, which is frequent in hand gestures considered. Figure 2-(a) shows two saparate trackers of hands using conventional switching linear model [6]. The left hand tracker was not able to catch a finger's moves due to failure in the detection of edges. However, in figure 2-(b) where the presented coupling method was applied, since the right hand tracker forces the left one to operate its own switching dynamics, the left tracker could track successfully regardless of failures in edge detection. We can confirm it in figure 3 which shows transition between discrete states in the left hand tracker. The left hand model has been trained to have three states: State 1 corresponds to moving one's fist, state 2 describes shape changes from stone to scissors, and state 3 corresponds to moving the scissors to the origin. If coupling is applied, switching discrete states is well performed as shown in figure 3-(b).

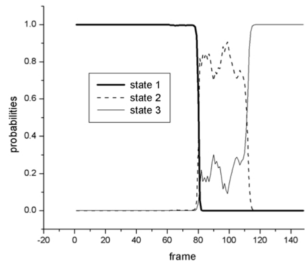
We have prepared four models for four two-hand gestures. For the purpose of recognition during tracking, Tracking is performed through the coupled-forward algorithm with respect to all models. At the same time, likelihood values for all models are computed by (13). Accordingly, an observed sequence is recognized as the model with the minimum value. Figure 4 shows results in recognition during tracking.

(a) without coupling

(b) with coupling

Figure 2. Tracking two hands



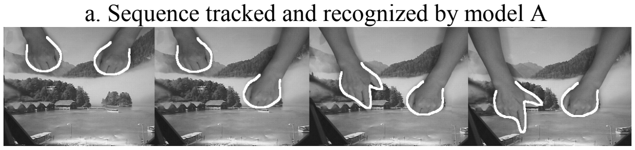


(a) Without coupling (b) With coupling

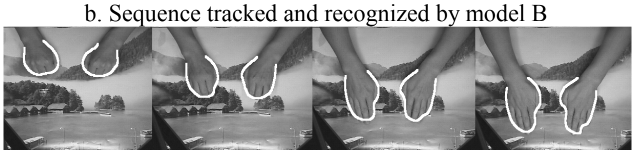
Figure 3. Transition between discrete states.

In (a) and (b) of figure 4, although each corresponding hand has similar motion, both two-hand gestures can be discriminated as shown in figure 5. This confirms that the proposed coupled switching linear model well explains the interaction between two hands.

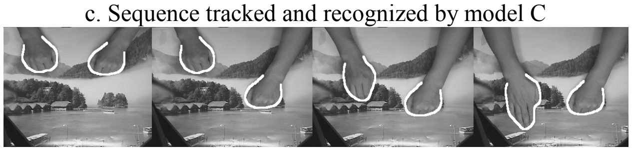
a. Sequence tracked and recognized by model A



b. Sequence tracked and recognized by model B



c. Sequence tracked and recognized by model C



d. Sequence tracked and recognized by model D

Figure 4. Recognition of two-hand gestures during tracking.

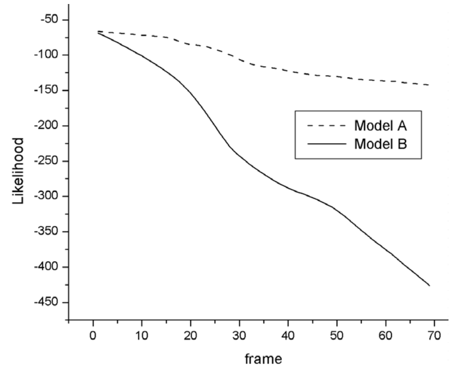


Figure 5. Likelihood vs. frame. Given the sequence in figure 5-a, likelihood values are plotted with respect to model A and B.

7. Conclusion

We have proposed a coupled switching linear model, which is an interacting process between two switching linear models and presented its EM learning method using a collapsing method. We have applied the proposed scheme to recognizing two-hand gestures. The presented method showed the effectiveness in tracking shape-changing hand under failures in feature detection. And it also showed satisfactory results that two-hand gestures are recognized and tracked simultaneously.

References

[1]Blake,A. and Isard,M., *Active contour*,Springer-Verlag, 1998.

[2]Ghahramani,Z. and Hinton,G. E., "Variational Learning for Switching State-Space Models", *CRG-TR-96-3 of Toronto Univ.*, 1996.

[3]Kim,C-J., "Dynamic Linear Models with Markov-Switching", *Journal of Econometrics*, Vol. 60, pp. 1-22, 1994.

[4]Brand,M., Oliver,N. and Pentland,A.,1997, "Couple Hidden Markov Models for Complex Action Recognition", *Proc.,IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'97.

[5] Reynard David, Andrew Wildenberg, Andrew Blake and John Marchant, "Learning Dynamics of Complex Motions from Image Sequences", *Proc. European Conf. on Computer Vision*, vol. 1, pp. 357-368, Cambridge UK, 1996.

[6] Jeong,M.H., Kuno,Y., Shimada,N., and Shirai,Y., "Recognition of Shape-Changing Hand Gestures Based on Switching Linear Model", ICIAP2001, 2001.

[7]Dempster, A., Laird, M. and Rubin, D., "Maximum Likelihood from Incomplete Data via the EM Algorithm", *J. of the Royal Statistical Society*, B(39), pp.1-38, 1977.