ACCV2002: The 5th Asian Conference on Computer Vision, 23--25 January 2002, Melbourne, Australia

Complex Gesture Recognition using Coupled Switching Linear Model

Mun Ho Jeong
Dept. Computer-Controlled Mech. Systems, Osaka Univ.
mhjeong@cv.mech.eng.osaka-u.ac.jp

Yoshinori Kuno
Dept. Information and Computer Sciences, Saitama Univ.
kuno@cv.ics.saitama-u.ac.jp

Nobutaka Shimada, Yoshiaki Shirai
Dept. Computer-Controlled Mech. Systems, Osaka Univ.
shimada@mech.eng.osaka-u.ac.jp
shirai@mech.eng.osaka-u.ac.jp

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. A well-known method in gesture recognition is HMM (Hidden Markov Model) [11,13,14,15], which is essentially a quantization of time series (observation sequence) into a small number of discrete states with transition probabilities between states.

In HMM-based gesture recognition schemes, there are two bottlenecks. One is a limitation in treating with time series having dependencies because HMM-based schemes are based on distributions of statistically independent observations or measurements. The other is a difficulty in dealing with multiple interacting processes. HMM is ill-suited to this problem because it has a single state variable (hidden discrete states). Many interesting man-machine interfaces are composed of multiple interacting processes. This is typically the case for systems that have structure both in time and space [16].

We adopt a dynamic process to explain dependencies between spatio-temporal configurations of the sequence. In fact, if the dynamic model is known, we might be able to infer states, for example, the positions and shapes of the hand over time. However, real cases are not so simple since shape-changing hand gestures exhibit complex and rich dynamic behaviors. To model such shape-changing hand gestures, we introduce switching linear dynamics, which has been developed in fields ranging from econometrics to engineering [3,5,6,7, 12]. It combines the discrete transition structure of HMM with the stochastically linear dynamic model of state-space model. Therefore, it may be able to overcome the first bottleneck of HMM. However, the second remains in this method. CHMM (coupled hidden Markov model) [16,17,19] has been proposed to deal with interacting processes. However, since CHMM inherits from HMM, it also has a limitation in considering trajectory information in time.

Our goal in this paper is to track and recognize two-hand gestures simultaneously. The process of two-hand gestures can be considered as interacting processes of two one-hand gestures. We propose a method coupling switching linear model to overcome the second bottleneck in HMM-based gesture recognition. Two switching linear models are coupled through causal influences between their hidden discrete states. Reynard [18] has introduced a coupling concept to track complex motions, however this means just a coupling of two kinds of continuous state variables in a single process, and is essentially different from interaction considered here.

A well-known problem in switching linear model is that the presence of Markov switching makes exact inference impossible. In this paper we use an approximate inference based on a collapsing method to avoid the problem. To estimate the parameters of switching linear model we present an EM learning process into which approximate inference using the collapsing method is well incorporated.

Hand contours are parameterized into shape vectors by the active contour model, and the shape vectors are considered as state vectors in the switching linear model. We demonstrate an application of the coupled switching linear model to tracking and recognizing two hands whose shapes change during their motion. By considering the coupling of two hand motions, the method can track both hands even when one of them is not observed well in images by occlusions or complex backgrounds.

The paper is organized as follows: In the following section we address the coupled switching linear model. The forward and the backward algorithm for approximate inference are explained. In section 3, we concern practical problems in applying active contour model to complex hand gestures. In section 4, we explain the EM learning process using a collapsing method for the coupled switching linear model. In section 5, we address the recognition process where AIC criterion [10] is put in use for online selection of the model. The experimental results are shown in section 6. Finally, we conclude with section 7.

2. Coupled switching linear model

We change the shapes of our hands as well as move our hands (arms) when we make gestures. 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 the shapes and motions of hands and interactions between the two hands. Here, 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, and proposes a coupled switching linear dynamic model to capture interactions between the two hands.

2.1 Model specification

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:

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

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. This model can be expressed graphically in the form of figure 1.

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 2.

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 undersampling and numerical underflow errors [16]. Consequently, the suggested coupling scheme, as shown in figure 2, offers conceptual advantages of parsimony and clarity with computational benefits in efficiency and accuracy. This is revealed in the following sections.

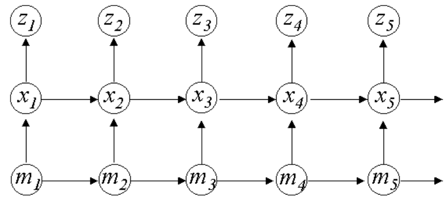


Figure 1. Switching linear model.

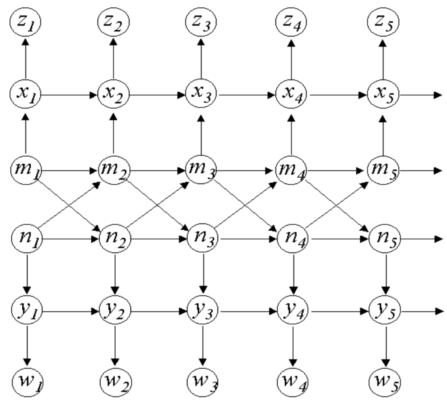


Figure 2. Coupled switching linear model. n_t, y_t and w_t denote a discrete state variable, a continuous state variable and an observation vector, respectively

In the coupled switching linear model, since coupled transitions of discrete states have 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

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

referred to in [16], coupling transition probability of discrete states can be parameterized as

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

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 2.

2.2 Coupled-forward algorithm

Given known parameters of the coupled switching linear model, $\{F, Q, D, \pi, \Phi, \Gamma\}$, $\{\hat{F}, \hat{Q}, \hat{D}, \hat{\pi}, \hat{\Phi}, \hat{\Gamma}\}$, we can perform tracking or filtering, which means estimations of continuous states and coupled-joint probabilities of hidden discrete states. In this section we describe a filtering method called as the coupled-forward algorithm.

Following [8], given the known parameters of switching linear dynamics, the predicted joint-continuous state variable and the corresponding covariance are defined dependently on $m_t = j$ and $m_{t-1} = i$:

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

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 jointed-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] and [2] to active contour model.

From the above fact, as noted by [5], 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, typically $(MN)^t$ computations are required ($M^t + N^t$ in the case of the presented coupled switching linear model) 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. Building upon ideas introduced by [7], [5] and [8], we present the following collapsing method:

Through the paper, expediently we proceed by evolving equations only in terms of the upper system in figure 2.

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

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

where O_t is a sequence $(o_1, o_2, \cdots, o_t)$ and o_t is an observation vector (z_t, w_t). In the above collapsing, the coupled-joint probability of discrete states plays a role of weighting factor of joint-continuous state variables. To complete the collapsing, we have only to calculate the weighting factor. Now we present the coupled-forward algorithm: The filtered coupled-joint distribution of discrete states is defined by

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

where k_t is a normalizing constant, Z_{t-1} and W_{t-1} are the observation sequences up to time $t-1$ in the upper and lower system, respectively in figure 2. From (2) and (3) the prediction step given sequence up to time t gives

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

where k_p is a normalizing constant. Now the followings can be obtained as

$$
\begin{aligned}
p(m_{t+1}, n_{t+1} \mid O_t) &= \sum_{m_t, n_t} p(m_t, n_t, m_{t+1}, n_{t+1} \mid O_t) \\
p(m_t \mid O_t) &= \sum_{m_{t-1}, n_t, n_{t-1}} p(m_t, n_t, m_{t-1}, n_{t-1} \mid O_t) \\
x_{t|t} &= \sum_{m_t=1}^{M} p(m_t \mid O_t) x_{t|t}^{(m_t)}
\end{aligned} \tag{8}
$$

The above algorithm can be extended up to more coupled models easily. However, since the coupled-forward algorithm has the complexity $O(TC_1^{\,2} \cdots C_n^{\,2})$ where T is the length of an observation sequence and C_n is the number of states of each switching linear model participating in the n-coupled switching linear model, the computation cost increases sharply as the number of coupling increases. In the case of $n = 2$ considered in this paper, we have no problem in terms of computational cost only if M and N have reasonable length.

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 based on full sequence can be smoothed as follows:

Given $m_t = j$ and $m_{t+1} = k$,

$$
\begin{aligned}
x_{t|T}^{(j,k)} &= x_{t|t}^{(j)} + \tilde{P}_t^{(j,k)} (x_{k+1|T}^{(k)} - x_{k+1|t}^{(j,k)}) \\
P_{t|T}^{(j,k)} &= P_{t|t}^{(j)} + \tilde{P}_t^{(j,k)} (P_{k+1|T}^{(k)} - P_{k+1|t}^{(j,k)}) \tilde{P}_t'^{(j,k)}
\end{aligned} \tag{9}
$$

where $\tilde{P}_t^{(j,k)} = P_{t|t}^{(j)} F_k' (P_{t+1|t}^{(j,k)})^{-1}$. To calculate the smoothed continuous state variable and its covariance, given that $m_t = j$, collapsing is performed similarly to (5):

$$
x_{t|T}^{(j)} = \frac{\sum_{m_{t+1}=1}^{M} \begin{pmatrix} \sum_{n_t, n_{t+1}} p(m_t = j, n_t, m_{t+1}, n_{t+1} \mid O_T) \\ \cdot x_{t|T}^{(j, m_{t+1})} \end{pmatrix}}{p(m_t = j \mid O_T)} \tag{10}
$$

$$
P_{t|T}^{(j)} = \frac{\sum_{m_{t+1}=1}^{M} \begin{pmatrix} \sum_{n_t, n_{t+1}} p(m_t = j, n_t, m_{t+1}, n_{t+1} \mid O_T) \\ \cdot \left(P_{t|T}^{(j, m_{t+1})} + (x_{t|T}^{(j)} - x_{t/T}^{(j, m_{t+1})})(x_{t|T}^{(j)} - x_{t|T}^{(j, m_{t+1})})' \right) \end{pmatrix}}{p(m_t = j \mid O_T)}
$$

To complete (10), we turn to derivation of the probability of the smoothed coupled-joint discrete states, which is given by

$$p(m_t,n_t,m_{t+1},n_{t+1} \mid O_T) = p(m_t,n_t,m_{t+1},n_{t+1} \mid O_t) \cdot \frac{p(m_{t+1},n_{t+1} \mid O_T)}{p(m_{t+1},n_{t+1} \mid O_t)} \quad (11)$$

$p(m_{t+1},n_{t+1} \mid O_t)$ already has been computed from (8). From (11) the followings are obtained as

$$p(m_t,n_t \mid O_T) = \sum_{m_{t+1},n_{t+1}} p(m_t,n_t,m_{t+1},n_{t+1} \mid O_T) \quad (12)$$

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

3. Active contour model

Active contour models such as snakes and deformable templates have practical problems in being applied to tracking hand gestures. Although the schemes are effective to retrieve features with geometric structures, they are too sensitive to noises to track an object under a complicated background and also have difficulties in progressing into boundary concavities which are frequently seen in shape-changing hand gestures as shown in figure 3.

As a solution, there have been dynamic contour methods conjugating prior dynamic models. A shape vector is treated as a continuous state vector in dynamic models. Dynamics can provide a powerful cue in the presence of occlusions and measurement noises. Known dynamics also enables a contour to progress easily into boundary concavities as shown in figure 4.

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 [2]. 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 called as templates. Then the shape vector is considered as a state vector in switching linear dynamics.

In applying active contour model to the forward algorithm, we need to compute observation probability given a predicted state vector in (5). From the known edge variance at each sample point the observation probability is computed from the sum of normal displacements between each sample point on the contour and the observed edge on its normal line [2].

4. EM learning

EM algorithm is a general iterative technique for finding maximum likelihood parameter estimates in problems where some variables are unobserved [4]. 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.

Figure 3. B-spline curve fitting

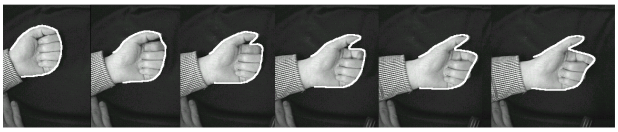


Figure 4. B-spline curve fitting with known dynamics.

Assume that the probability density for observation sequence is parameterized using λ. The log-likelihood is given by

$$\log p(O_T \mid \lambda) = \log \sum_{M_T,N_T} \int_{X_T,Y_T} p(M_T,N_T,X_T,Y_T,O_T \mid \lambda)\,dX_T\,dY_T \quad (13)$$

where (M_T,N_T) and (X_T,Y_T), are sequences(of length T) of discrete states and continuous states, respectively. Neal and Hinton [9] showed that the auxiliary log-likelihood is given by

$$L = \sum_{M_T,N_T} \int_{X_T,Y_T} \overline{p} \log p(M_T,N_T,X_T,Y_T,O_T \mid \lambda)\,dX_T\,dY_T = E_{\overline{p}}[\log p(M_T,N_T,X_T,Y_T,O_T \mid \lambda)] \quad (14)$$

where $\overline{p} = p(M_T,N_T,X_T,Y_T \mid O_T,\overline{\lambda})$ and $\overline{\lambda}$ is the parameter set estimated previously. Based on the collapsing method in the presented switching linear model, then L can be approximately represented as the followings, up to constants:

$$\begin{aligned} L \approx \widetilde{L} = &\sum_{t=2}^{T} \sum_{i,j=1}^{M} \left(\begin{array}{l} p(m_{t-1}=i, m_t=j \mid O_T)\cdot \\ \frac{1}{2}\left(\det(Q_j^{-1}) - \eta_{t|T}'^{(i,j)} Q_j^{-1} \eta_{t|T}^{(i,j)}\right) \end{array} \right) \\ &+ \sum_{t=2}^{T} \sum_{i,j=1}^{N} \left(\begin{array}{l} p(n_{t-1}=i, n_t=j \mid O_T)\cdot \\ \frac{1}{2}\left(\det(\hat{Q}_j^{-1}) - \hat{\eta}_{t|T}'^{(i,j)} \hat{Q}_j^{-1} \hat{\eta}_{t|T}^{(i,j)}\right) \end{array} \right) \quad (15) \\ &+ \sum_{t=2}^{T} \sum_{i,j=1}^{M} p(m_{t-1}=i, m_t=j \mid O_T) \log \Phi_{i,j} \\ &+ \sum_{t=2}^{T} \sum_{i,j=1}^{N} p(n_{t-1}=i, n_t=j \mid O_T) \log \hat{\Phi}_{i,j} \\ &+ \sum_{i=1}^{M} p(m_1=i \mid O_T) \log \pi_i \\ &+ \sum_{i=1}^{N} p(n_1=i \mid O_T) \log \hat{\pi}_i \end{aligned}$$

where $\eta_t^{(i,j)} = (x_t^{(j)} - F_j x_{t-1}^{(i)} - D_j)$ and d is dimension of state vectors. EM algorithm starts with some initial guess and proceeds by applying the following two steps repeatedly:

E-step On the condition given the observation sequence of full length and the previous parameter set $O_T, \overline{\lambda}$, we estimate continuous states $x_{t|T}^{(m_t)}, y_{t|T}^{(n_t)}$, the probabilities of joint-discrete states $p(m_{t-1},m_t \mid O_T)$, $p(n_{t-1},n_t \mid O_T)$ and the probabilities of discrete states $p(m_t \mid O_T)$, $p(n_t \mid O_T)$. These estimations are performed through the forward and the backward process described in sections 2.2 and 2.3.

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

The above two steps are iterated until the likelihood value converges.

5. 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.

As addressed in section 1, our goal is to track and recognize hand gestures simultaneously. So we have to compute the likelihood of each model while tracking is being performed with the coupled-forward algorithm.

The coupled switching linear model can be represented by the parameter set, λ, which consists of $\{F, Q, \pi, \Phi, \Gamma\}$ and $\{\hat{F}, \hat{Q}, \hat{\pi}, \hat{\Phi}, \hat{\Gamma}\}$. Likelihood of λ given an observation sequence can be calculated by

$$L(\lambda \mid O_\tau) = p(O_\tau \mid \lambda) = \prod_{t=1}^{\tau} p(o_t \mid O_{t-1}, \lambda). \quad (16)$$

Abbreviating λ,

$$p(o_t \mid O_{t-1}) = \sum_{\substack{m_t, m_{t-1}, \\ n_t, n_{t-1}}} \begin{pmatrix} p(m_{t-1}, n_{t-1}, m_t, n_t \mid O_{t-1}) \\ \cdot\, p(o_t \mid m_{t-1}, n_{t-1}, m_t, n_t, O_{t-1}) \end{pmatrix}. \quad (17)$$

Substituting (6) and (17) into (16), log-likelihood $\widetilde{L}$ is obtained by

$$\widetilde{L} = \sum_{t=1}^{\tau} \log\left(\frac{1}{k_t}\right) \quad (18)$$

where k_t has been computed in the coupled-forward algorithm. At time τ, the goodness of the fit of a model out of the candidate models is evaluated by AIC criterion [10]:

$$AIC = -2L_\tau + 2n \quad (19)$$

where n is the number of parameters of the model. A given sequence of hand gestures can be recognized as the model with the minimum *AIC* value.

6. Experiments

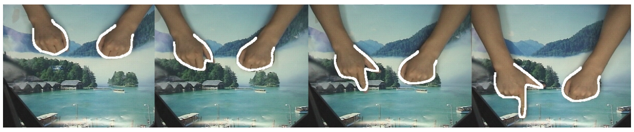
We apply the proposed coupled switching linear model to recognizing and tracking two-hand gestures simultaneously. We obtain the parameter set dominating a two-hand gesture through EM algorithm. Recognition of two-hand gestures means the problem to determine which model tracks a two-hand gesture well. Therefore, the likelihood of a model presented in section 2.4 can be a good tool to recognize two-hand gestures.

We have prepared four models for four two-hand gestures. There are two types of interaction between both hands in the gestures. One is that one hand is moving at any one time. The other is that two hands are moving at the same time. For example, as shown in (a) and (b) of figure 5, model A is actually different from model B only in terms of the interaction types: Two hands moves simultaneously in figure 5-a differently from figure 5-b.

After preparing the four models, we have performed experiments of tracking and recognition. Tracking is performed through the coupled-forward algorithm with respect to all models. At the same time, *AIC* for all models are computed by (19). Accordingly, an observed sequence is recognized as the model with the minimum value.

The motion of each hand was modeled to have three discrete states, *initial pause*, *moving* and *final pause*. For example, in the right hand in figure 5-a, classification of the sequence into the discrete states is illustrated in figure 6.

The sequences of tracked and recognized contours are shown in figure 5. In (a) and (b) of figure 5, although each corresponding hand has similar motion, both two-hand gestures can be discriminated as shown in figure 7. 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 5. Tracked and recognized two-hand gestures.

7. Conclusion

We have proposed a coupled switching linear model, which is an interacting process between two switching linear models. We have applied the proposed scheme to recognizing two-hand gestures.

The switching linear model well corresponds to temporal motion and shape of hands, and a coupling of two switching linear models can explain combination information of the two hands. Experimental results show that two-hand gestures are recognized and tracked simultaneously using the coupled switching linear model.

The proposed scheme is expected to be robust against partial failure of image feature extraction for one of the two hands caused by such as occlusion and complex backgrounds, because the shape and motion information for both hands are coupled in the model. In other words, as long as the one hand is tracked, the error for the other hand may be recovered. We are planning to perform experiments to examine this.

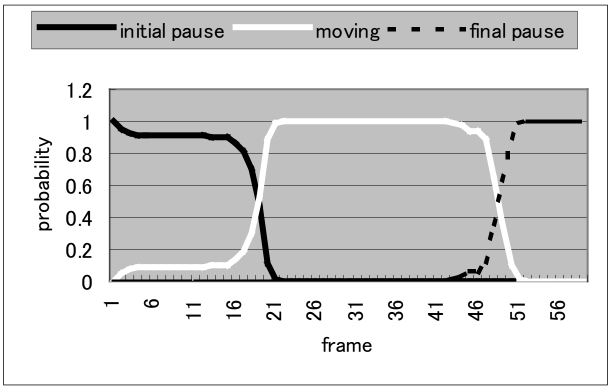


Figure 6. Classification of discrete states in the right hand in figure 5-a

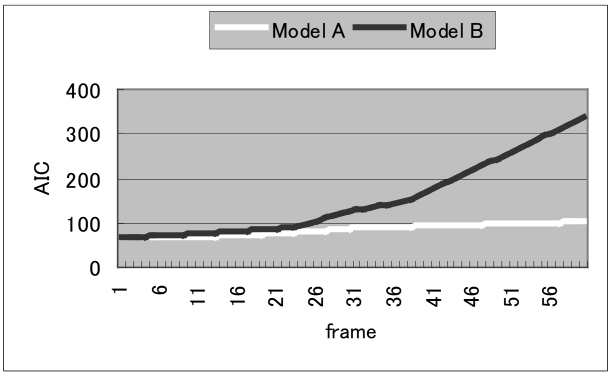


Figure 7. *AIC* vs. frame; Given the sequence in figure 5-a, *AIC* values are plotted with respect to model A and B

8. References

[1]Blake, A., Isard, M. and Rubin, D., “Learning to Track the Visual Motion of Contour"", *Some Artificial Intelligenc*, Vol. 78, pp.101-134, 1995.

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

[3]Chaer, W. S., Bishop, R. H. and Ghosh, J., “A Mixture-of-Experts Framework for Adaptive Kalman Filtering”, *IEEE Trans. on Systems, Man and Cybernetics*, 1997.

[4]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.

[5]Gordon, K. and Smith, A., “Modeling and Monitoring Discontinuous Changes in Time Series”, *Bayesian Analysis of Time Series and Dynamic Linear Model*, Dekker, M., New York, NY, pp. 359-392, 1988.

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

[7]Harrison, P. J. and Stevens, C. F., “Bayesian Forecasting”, *J. of the Royal Statistical Society*, Vol. B38, pp. 205-247, 1977.

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

[9]Neal, R. M., Hinton, G. E., “A View of the EM Algorithm that Justifies Incremental, Sparse and Other Variants”, *Learning in Graphical Model*, Dordrecht:Kluwer Academic Publishers, pp. 355-368, 1998.

[10]Sakamoto, Y., Ishiguro, M. and Kitagawa, G., *Akaike Information Criterion Statistics*, Tokyo:D.Reidel, 1986.

[11]Schlenzig, J., Hunter, E. and Jain, R., “Recursive Identification of Gesture Inputers using Hidden Markov Models”, *Proc. Second Ann. Conf. on Appl. of Comp. Vision*, pp. 187-194, 1994.

[12]Shumway, R. H. and Stoffer, D. S., “An approach to Time Series Smoothing and Forecasting using the EM Algorithm”, *J. Time Series Analysis*, Vol. 3(4), pp. 253-264, 1982.

[13]Starner, T. and Pentland, A., “Real Time American Sign Language Recognition from Video using Hidden Markov Models”, *Technical Report 375 of MIT Media Lab.*, 1995.

[14]Wilson, A. D., Bobick, A. F., “Recognition and Interpretation of Parametric Gesture”, *ICCV 98*, 1998.

[15]Yamato, J., Ohya, J., and Ishii, K., “Recognizing Human Action in Time-Sequential Images using Hidden Markov Models”, *Proc. 1992 ICCV*, IEEE Press, pp. 379-385, 1992.

[16]Brand, M.,1996, “Coupled hidden Markov models for modeling interacting processes”, TR405 of Media Lab in MIT

[17]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.

[18] 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, 199).

[19] Oliver, Nuria M., Barbara Rosario and A.P. Pentland, “A Bayesian Computer Vision System for Modeling Human Interactions”, *IEEE trans. PAMI*, vol. 22, no. 8, pp. 831-843, 2000.