

Recognition of Shape-Changing Hand Gestures Based on Switching Linear Model

Mun Ho Jeong

*Dept. Computer-Controlled
Mech. Systems, Osaka Univ.*
[mhjeong@cv.mech.eng.osaka-
u.ac.jp](mailto:mhjeong@cv.mech.eng.osaka-u.ac.jp)

Yoshinori Kuno

*Dept. Information and Com-
puter Sciences, Saitama Univ.*
kuno@cv.ics.saitama-u.ac.jp

Nobutaka Shimada,

Yoshiaki Shirai
*Dept. Computer-Controlled
Mech. Systems, Osaka Univ.*

Abstract

We present a method to track and recognize shape-changing hand gestures simultaneously. The switching linear model using active contour model well corresponds to temporal shapes and motions of hands. Inference in the switching linear model is computationally intractable, and therefore the learning process cannot be performed via the exact EM(Expectation Maximization) algorithm. However, we present an approximate EM algorithm using a collapsing method in which some Gaussians are merged into a single Gaussian. Tracking is performed through the forward algorithm based on Kalman filtering and the collapsing method. We also present the regularized smoothing, which plays a role of reducing jump changes between the training sequences of state vectors to cope with complex-variable hand shapes. The recognition process is performed by the selection of a model with the maximum likelihood from some learned models while tracking is being performed. Experiments for several shape-changing hand gestures are demonstrated.

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) [19,21,22,23], which is essentially a quantization of time series (observation sequence) into a small number of discrete states with transition probabilities between states. Most of schemes in gesture recognition are based on measurement spaces like HMM [3,5,15].

Although these showed successful results, there are two bottlenecks. First, based on the distributions of independent measurements or observations, they have a limitation in treating with time series having dependencies. Second, they also have difficulties when it is hard to measure the required information for recognition. Taking an instance of hand gestures. Taking an instance of shape

changing hand gestures, measuring the outlines of the hand is not always feasible, especially under complicated backgrounds.

We adopt a dynamic process to explain dependencies between spatio-temporal configurations of the sequence. In fact, if the motion of the hand is known in advance, that is, the dynamic model of the hand is known, we might be able to infer the positions and shapes of the hand over time. To model shape-changing hand gestures that exhibit complex and rich dynamic behaviors, we introduce switching linear dynamics [8,12,20] that consists of a few linear dynamic models with Markov switching between them, rather than a single linear dynamic model. A well-known problem in switching linear model, however, 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. For learning of the model it is necessary to collect training sequences of state vectors. When shape-changing hand gestures are considered, even though the outlines of a hand vary gradually over time, there often happen abrupt changes between the state vectors representing them due to separate parameterizations. This fact leads to poor learning or makes initial tracking impossible in the EM learning process. In this paper we propose the *regularized smoothing* method to solve this problem. It can make a training sequence of shape vectors vary gradually but the outlines of the hand remain invariant with allowable errors.

The first efforts at classification and tracking of hand gestures using active contour model and multiple dynamic model were made by Isard et al.[10]. They showed tracking of hand outlines and classification of different writing patterns. However they confined the scope of changes in hand shapes to affine transformation. Pavlovic and Rehg

applied switching linear model to tracking of human figures [17]. They used Viterbi approximation to overcome the exponential complexity of exact inference. The above approaches did not handle online recognition during tracking, but concentrated on tracking of a human gesture or the problem of where to switch to another dynamics in time domain. Pentland and Liu modeled automobile drivers' actions by Hidden Markov Dynamic Model in which Kalman filtering method is incorporated into HMM structure [16]. In their learning process, estimation of dynamic parameters was not incorporated into EM learning (Baum-Welch algorithm).

The paper is organized as follows. In the following section we address the switching linear model. In section 3, we concern practical problems in applying active contour model to complex hand gestures and present the *regularized smoothing*. In section 4, we explain the EM learning using a collapsing method for the switching linear model. In section 5, we address the recognition process where *AIC* criterion 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. Switching Linear Model

Human describes any meaning by changing shapes of a hand besides moving positions of the hand. To model these complex and rich dynamic behaviors, we introduce switching linear dynamics. 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:

$$\begin{aligned} x_t &= F_{m_t} x_{t-1} + D_{m_t} + u_t, u_t \sim N(0, Q_{m_t}) \\ \Phi_{m_t, m_{t+1}} &= p(m_{t+1} | m_t) \\ \pi_{m_t} &= p(m_t), m_t = \{1, 2, \dots, M\}. \end{aligned} \quad (1)$$

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_t} , 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 shown graphically as figure 1.

2.1 Forward Algorithm

Given known parameters of the switching linear model, $\{F, D, Q, \pi, \Phi\}$, we can perform tracking or filtering, which means estimations of continuous states and prob-

abilities of joint-discrete states here. The predicted joint-continuous state vector and its covariance are derived dependently on $m_{t-1} = i$ and $m_t = j$:

$$\begin{aligned} x_{t|t-1}^{(i,j)} &= F_j x_{t-1|t-1}^{(i)} + D_j \\ P_{t|t-1}^{(i,j)} &= F_j P_{t-1|t-1}^{(i)} F_j' + Q_j \end{aligned} \quad (3)$$

where $x_{t-1|t-1}^{(i)}$ and $P_{t-1|t-1}^{(i)}$ are estimations at time $t-1$ on the condition given observations 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 filtering. In particular, we follow Kalman filtering application to active contour model by Blake[1][2].

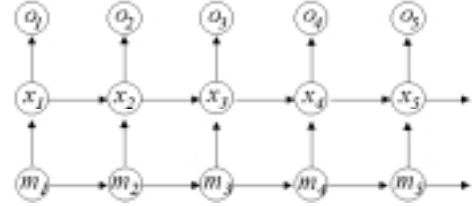


Figure 1. Switching linear model
Arrows denote probabilistic dependencies

From the above fact, as noted by Gordon and Smith[6], switching linear dynamic model requires computing a Gaussian mixture with M^t components at time t for M switching states. That leads to intractable inference for moderate sequence length. It is necessary to introduce some approximations to solve the intractable computation problem.

We collapse M^2 joint-continuous state vectors into M state variables at each time, and can avoid prohibitive increase of computational cost. Building upon ideas introduced by Harrison[9], Gordon[6] and Kim[12], the collapsing is given by

$$\begin{aligned} x_{t|t}^{(j)} &= \frac{\sum_{i=1}^M p(m_{t-1} = i, m_t = j | O_t) \cdot x_{t|t}^{(i,j)}}{p(m_t = j | O_t)} \\ P_{t|t}^{(j)} &= \frac{\sum_{i=1}^M \left(p(m_{t-1} = i, m_t = j | O_t) \cdot \left(P_{t|t}^{(i,j)} + (x_{t|t}^{(j)} - x_{t|t}^{(i,j)})(x_{t|t}^{(j)} - x_{t|t}^{(i,j)})' \right) \right)}{p(m_t = j | O_t)} \end{aligned} \quad (4)$$

where O_t is a sequence (o_1, o_2, \dots, o_t) and o_t is an observation vector. In the above collapsing, the probabilities of joint-discrete states play a role of weighting factors of joint-continuous state vectors. To complete the collapsing, we have only to calculate the weighting factors.

The probabilities of the filtered joint-discrete states are obtained as

$$p(m_{t-1}, m_t | O_t) \approx k_t p(o_t | x_{t-1|t-1}^{(m_{t-1}, m_t)}) p(m_{t-1}, m_t | O_{t-1}) \quad (5)$$

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

$$p(m_t, m_{t+1} | O_t) = k_p \prod_{m_t, m_{t+1}} \sum_{m_{t-1}=1}^M p(m_{t-1}, m_t | O_t). \quad (6)$$

$$\left. \begin{aligned} p(m_t | O_t) &= \sum_{m_{t+1}=1}^M p(m_t, m_{t+1} | O_t) \\ p(m_{t+1} | O_t) &= \sum_{m_t=1}^M p(m_t, m_{t+1} | O_t) \end{aligned} \right\} \quad (7)$$

$$x_{i|t} = \sum_{j=1}^M p(m_t = j | O_t) x_{i|t}^{(j)}. \quad (8)$$

2.2. Backward Algorithm

While the forward algorithm is a filtering process given sequence up to current time, the backward algorithm is a smoothing process given sequence of full length. Like the conventional Kalman smoothing method, the joint-continuous state vector and its covariance based on full sequence can be smoothed as follows:

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

$$x_{i|T}^{(j,k)} = x_{i|t}^{(j)} + \tilde{P}_t^{(j,k)} (x_{k+1|T}^{(k)} - x_{k+1|t}^{(j,k)}) \quad (9)$$

$$P_{i|T}^{(j,k)} = P_{i|t}^{(j)} + \tilde{P}_t^{(j,k)} (P_{k+1|T}^{(k)} - P_{k+1|t}^{(j,k)}) \tilde{P}_t'^{(j,k)}$$

where $\tilde{P}_t^{(j,k)} = P_{i|t}^{(j)} F_k'(P_{t+1|t}^{(j,k)})^{-1}$. To calculate the smoothed continuous state vector and its covariance, given that $m_t = j$, collapsing is performed similarly to (4):

$$x_{i|T}^{(j)} = \frac{\sum_{k=1}^M (p(m_t = j, m_{t+1} = k | O_T) x_{i|T}^{(j,k)})}{p(m_t = j | O_T)} \quad (10)$$

$$P_{i|T}^{(j)} = \frac{\sum_{k=1}^M \left(p(m_t = j, m_{t+1} = k | O_T) \left(P_{i|T}^{(j,k)} + (x_{i|T}^{(j)} - x_{i|T}^{(j,k)}) (x_{i|T}^{(j)} - x_{i|T}^{(j,k)})' \right) \right)}{p(m_t = j | O_T)}.$$

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

$$p(m_t, m_{t+1} | O_T) = p(m_t, m_{t+1} | O_t) \frac{p(m_{t+1} | O_T)}{p(m_{t+1} | O_t)} \quad (11)$$

From (11) the probabilities of the smoothed discrete states is obtained as

$$p(m_t | O_T) = \sum_{m_{t+1}} p(m_t, m_{t+1} | O_T). \quad (12)$$

$p(m_{t+1} | O_T)_{t=T-1}$ and $p(m_{t+1} | O_t)$ have already been computed from (7) in the forward algorithm.

3. Active Contour Model

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 con-

sidered as a state vector in switching linear dynamics.

3.1 Practical Problems

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 2.

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.



Figure 2. Sequential B-spline curve fitting in boundary Concavities

When learning dynamic models, it is necessary to prepare training sequences of state vectors. That is essential to not only general learning process using the maximum likelihood method but also the initial estimation of dynamic parameters in EM learning, which is an iterative maximum likelihood method. Even though such a sequence can be obtained in various ways, there often occurs a problem in the case of shape-changing hand gestures considered here. Generally outlines of a hand have gradual changes over time, however shape vectors representing its outlines often vary abruptly on the shape space due to separate parameterizations of outlines of the hand. For example, although outlines of the hand show gradual changes between them in figure 3, a jump change is found on the shape space from the white-colored curve in figure 4. This often leads to poor learning. Especially in the EM learning, initial tracking have to be feasible to some extent so that the dynamic model can be improved by iterative adjustment of dynamic parameters, however existence of abrupt changes between shape vectors may cause that to be impossible.



Figure 3. An initially given training sequence of contours (1st, 11th, 15th, 17th, 19th and 20th from left to right); these contours are originated from two differently parameterized contours.

3.2 Regularized Smoothing

In this section we present the *regularized smoothing* method to make a training sequence of shape vectors have gradual changes on the shape space but outlines of the hand remain invariant with allowable errors.

Suppose the state in which the index finger and the big finger are touching. This state changes to a completely different state with the separating fingers only by a small motion of the fingers. That is shown in figure 4 illustrating the first two dimensions of the eigen space representing the hand shape. The black dots clearly shows this jump.

Given a sequence of shape vectors, $\{x_1^o, x_2^o, \dots, x_T^o\}$, the new fitted shape vector at t , \hat{x}_t , is obtained by the following *regularized smoothing*:

Step 1 Initialize $x_t = x_t^o, t = 1, 2, \dots, T$

Step 2 Estimate, $t = 2, 3, \dots, T-1$

$$\hat{x}_t = \arg \min_{\hat{x}_t} \alpha \left\| \hat{x}_t - \frac{x_{t-1} + x_{t+1}}{2} \right\|^2 + \left\| \hat{r}_t - r_t^o \right\|^2$$

$$x_t = \hat{x}_t$$

Step 3 Repeat step 2

where α is the regularization constant, \hat{r}_t and r_t^o are contour points of \hat{x}_t and x_t , respectively, and $\| \cdot \|^2$ denotes L^2 norm [2]. The first norm forces the current state to be positioned for smooth changes in a local interval while the second guarantees that the fitted curve remains unchanged. In figure 4 the smoothed sequence of shape vectors is shown as the white dots without any jump.

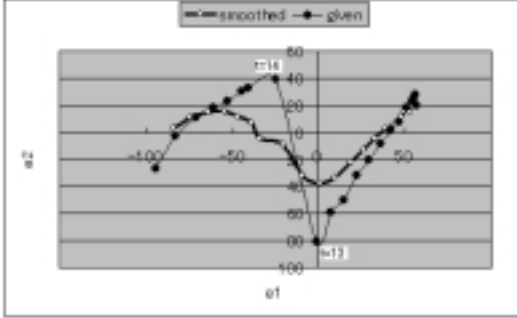


Figure 4. The result of the *regularized smoothing*: A given training sequence with an abrupt jump is smoothed after five iterations of the regularized smoothing.

4. Learning via EM

EM algorithm is a general iterative technique for finding maximum likelihood parameter estimates in problems where some variables are unobserved [5]. It is natural to use EM algorithm for our problem, in which unobserved variables are continuous state variables x_t and discrete state variables m_t . Assume that the probability density for observation sequence is parameterized using

$\lambda = \{F, D, Q, \pi, \Phi\}$, $p(O_T | \lambda)$. The log-likelihood is given by

$$\begin{aligned} \text{Likelihood}(\lambda | O_T) &= \log p(O_T | \lambda) \\ &= \log \sum_{M_T} \int_{X_T} p(M_T, X_T, O_T | \lambda) dX_T \end{aligned} \quad (13)$$

where M_T and X_T are sequences (of length T) of discrete states and continuous states, respectively. Neal and Hinton [14] showed that the auxiliary log-likelihood is given by

$$\begin{aligned} L &= \sum_{M_T} \int_{X_T} p(M_T, X_T | O_T, \bar{\lambda}) \cdot \log p(M_T, X_T, O_T | \lambda) dX_T \\ &= E_{\bar{p}}[\log p(M_T, X_T, O_T | \lambda)] \end{aligned} \quad (14)$$

where $\bar{p} = p(M_T, X_T | O_T, \bar{\lambda})$ and $\bar{\lambda}$ is the parameter set estimated previously. From figure 1 the joint probability for the sequences of states M_T, X_T and observations O_T can be factored as:

$$\begin{aligned} p((M_T, X_T, O_T | \lambda)) &= \prod_{j=1}^M [\pi_j p(x_1 | m_1) p(o_1 | x_1, m_1)]^{\psi_1(j)} \\ &\cdot \prod_{t=2}^T \prod_{j=1}^M [p(o_t | x_t, m_t)]^{\psi_t(j)} \\ &\cdot \prod_{t=2}^T \prod_{j=1}^M \prod_{i=1}^M \Phi_{i,j} p(x_t | x_{t-1}, m_{t-1}, m_t)^{\psi_{t-1}^{(i)} \psi_t(j)} \end{aligned} \quad (15)$$

where $\psi_t(k) = 1$ if $m_t = k$, otherwise 0. Based on the collapsing method in the presented switching linear model, then L can be approximately represented as the followings, up to constants:

$$\begin{aligned} L \approx \tilde{L} &= \sum_{t=2}^T \sum_{i,j=1}^M \left(p(m_{t-1} = i, m_t = j | O_T) \cdot \right. \\ &\left. \frac{1}{2} (\det(Q_j^{-1}) - \eta_{i|T}^{(i,j)} Q_j^{-1} \eta_{i|T}^{(i,j)}) \right) \\ &+ \sum_{t=2}^T \sum_{i,j=1}^M p(m_{t-1} = i, m_t = j | O_T) \log \Phi_{i,j} \\ &+ \sum_{i=1}^M p(m_i = i | O_T) \log \pi_i \end{aligned} \quad (16)$$

where $\eta_{i|T}^{(i,j)} = (x_i^{(i,j)} - F_j x_{i-1}^{(i)} - D_j)$ 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 $O_T, \bar{\lambda}$, we estimate continuous states $x_{i|T}^{(m_i)}$, joint-continuous states $x_{i|T}^{(m_{t-1}, m_t)}$, and probabilities of joint-discrete states and discrete states, $p(m_{t-1}, m_t | O_T)$ and $p(m_t | O_T)$, respectively. These estimations are performed through the forward and the backward processes described in section 2.1 and 2.2.

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

The above two steps are iterated until the likelihood value converges. The likelihood value can be computed by (17) given in the following section.



Figure 5. Tracked and recognized hand gestures: Under the complicated background three sets of sequence are given, *Scissors*, *Paper* and *Stone* sequences from the top. The black-colored contours are tracking results by the model with the smallest *AIC* value at each time while the gray-colored contours are by the others.

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 forward algorithm.

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

$$L(\lambda | O_\tau) = L_\tau = \prod_{t=1}^{\tau} p(o_t | O_{t-1}, \lambda) = \sum_{k_t=1}^r \log\left(\frac{1}{k_t}\right) \quad (17)$$

where k_t has been computed in the forward algorithm. At time τ , the goodness of the fit of a model out of the candidate models is evaluated by *AIC* criterion[18]:

$$AIC = -2L_\tau + 2n \quad (18)$$

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*.

6. Experimental Results

We have prepared three kinds of shape-changing hand gestures, which are called as *Paper*, *Scissors* and *Stone*, respectively. First of all, the *regularized smoothing* was applied to each training sequence. After preparing the three models by EM learning, we have performed experiments of tracking and recognition. Tracking is performed through the forward algorithm with respect to all models. At the same time, *AIC* for all models are computed by (18). Accordingly, an observed sequence can be recognized as the model with the smallest *AIC* at each time.

Figure 5 shows that outlines of the hand are well tracked under the complicated background. *AIC* values computed during tracking were plotted with respect to all three models in figure 6. We might reduce misjudgment in recognition through watching longer in that the differences in *AIC* between the correct model and the others increase as the frame number increases in figure 6. The three sets of hand gestures are modeled to have three discrete states. For example, three states of *Scissors* model correspond to *initial pause*, *moving* and *final pause*, respectively. Classification of the *Scissors* sequence into the discrete states is illustrated in figure 7.

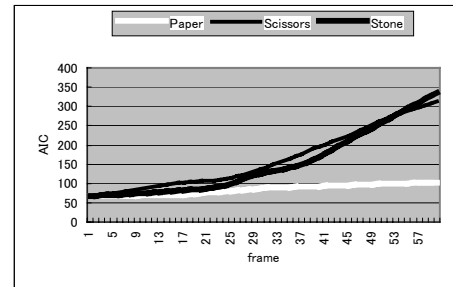


Figure 6. Graphs of *AIC* vs. frame: *AIC* values are computed and plotted with respect to three models in the case that the *Paper* sequence is given.

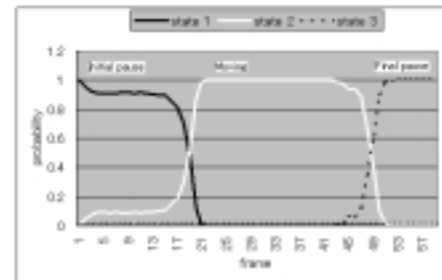


Figure 7. Classification of *Scissors* sequence: The black curve is for the probability of *initial pause* state, and the white and the dotted are for *moving* and *final pause*, respectively.

7. Conclusion and Discussion

We have presented a framework to track and recognize shape-changing gestures simultaneously. To model complex and rich dynamic behaviors of hands we introduced switching linear model in which shape vectors, which are parameterizations of hand contours by active contour model, are considered as state vectors. To overcome exponential complexity of exact inference in switching linear model an approximate inference was performed by a collapsing method in which some Gaussian distributions of state vectors are merged.

The parameters of the model are estimated via EM algorithm into which the collapsing method is incorporated. We also presented a smoothing method using regularization for smoothness in training state vectors. Through this process, we obtained state vectors shifted in the state space while real outlines of the hand remain invariant with allowable errors.

Recognition is performed by selection of a model out of some learned models through a criterion using log-likelihood values of each model. In some experiments we showed that shape-changing hand gestures are recognized and tracked simultaneously using the presented scheme.

However though we achieved satisfactory results in tracking under the complicated background, there still remains a problem that the allowable error bound in the approximate inference is not known, in other words, we have no information enough to make sure that the collapsing method can cope with all various backgrounds. As an alternative, there are Monte-Carlo-based methods in which a number of samples are used to represent the probability densities of state vectors[7][11][13]. Although this approach is plausible to complex-cluttered backgrounds, this is often considered not to be feasible for real time applications because the exponential number of samples is required for high dimensional space, which is general in shape-changing hand gestures.

Acknowledgements

This work was supported in part by the Ministry of Education, Science, Sports and Culture under the Grant-in-Aid for Scientific Research (12650249) and Ikeuchi CREST project.

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]Bobick, A. F., Wilson, A. D., "A State-based Approach to the

Representation and Recognition of Gesture", *IEEE Trans. PAMI*, Vol.19, pp. 1325-1337, 1997.
[4]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.
[5]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.
[6]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.
[7]Gordon, N. J., Salmond, D. J. and Smith, A., "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation", *IEE Proceedings-F*, Vol. 140, No. 2, pp. 107-113, 1993.
[8]Ghahramani, Z. and Hinton, G. E., "Variational Learning for Switching State-Space Models", *CRG-TR-96-3 of Toronto Univ.*, 1996.
[9]Harrison, P. J. and Stevens, C. F., "Bayesian Forecasting", *J. of the Royal Statistical Society*, Vol. B38, pp. 205-247, 1977.
[10]Isard, M. and Blake, A., "A mixed-state Condensation tracker with automatic model-switching", *Proc 6th Int. Conf. Computer Vision*, pp. 107-112, 1998.
[11]Isard, M. and Blake, A., "Condensation-Conditional Density Propagation for Visual Tracking", *Int. J. Computer Vision*, Vol. 29(1), pp. 5-28, 1998.
[12]Kim, C-J., "Dynamic Linear Models with Markov-Switching", *Journal of Econometrics*, Vol. 60, pp. 1-22, 1994.
[13]Kitagawa, G., "Mont Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models", *J. Computational and Graphical Statistics*, Vol. 5, No. 1, pp. 1-25, 1996.
[14]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.
[15]Ohno, H. and Yamamoto, M., "Gesture Recognition using Character Recognition Techniques on Two-dimensional Eigenspace", *Int. Con. Automatic Face and Gesture Recognition*, pp. 151-156, 1999.
[16]Pentland, A. and Liu, A., "Modeling and Prediction of Human Behavior", *Technical Reports 433*, MIT Media Lab., 1995.
[17]Pavlovic, V. and Rehg, J., "A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models", *Int. Con. Automatic Face and Gesture Recognition*, pp. 94-101, 1999.
[18]Sakamoto, Y., Ishiguro, M. and Kitagawa, G., *Akaike Information Criterion Statistics*, Tokyo:D.Reidel, 1986.
[19]Schlenzig, J., Hunter, E. and Jain, R., "Recursive Identification of Gesture Inputs using Hidden Markov Models", *Proc. Second Ann. Conf. on Appl. of Comp. Vision*, pp. 187-194, 1994.
[20]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.
[21]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.
[22]Wilson, A. D., Bobick, A. F., "Recognition and Interpretation of Parametric Gesture", *ICCV 98*, 1998.
[23]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.