PARTLY-HIDDEN MARKOV MODEL AND ITS APPLICATION TO GESTURE RECOGNITION

Tetsunori Kobayashi, Satoshi Haruyama
Dept. EECE, Waseda University
Shijuku-ku, Tokyo 169, Japan

Abstract

A new pattern matching method, Partly-Hidden Markov model, is proposed for gesture recognition.

Hidden Markov Model, which is widely used for the time series pattern recognition, can deal with only piecewise stationary stochastic process. We solved this problem by introducing the modified second order Markov Model, in which the first state is hidden and the second one is observable.

As the results of 6 sign-language recognition test, the error rate was improved by 73% compared with normal HMM.

1 Introduction

Gesture recognition, the problem of recognizing human's body actions and hand movements through moving image processing, is one of attractive applications of time series pattern recognition [1],[2],[3]. This technique consists of two parts, one is the feature extraction rom moving images and the other is the pattern matching of time series of features. This paper focus on the pattern matching for the gesture recognition.

Other applications of time series pattern recognition involve speech recognition. In this field, many efforts have been made for a long time. DP matching and HMM are two major techniques for speech recognition. Recently, the superiority of HMM over DP matching become apparent. Now, almost of all systems adopt HMM [4], [5]. Therefore, it seems that HMM is promising in the field of gesture recognition also.

However, it is doubtful that HMM is effective enough for the gesture recognition because the gesture recognition problem has another difficulty which may not be dealt by HMM. That is the dynamical feature of time series parameters.

HMM consider that the output probability of feature vectors in each state is unique. That means HMM deal with only piecewise stationary process. This constraint is rather acceptable in speech recognition because speech can be considered as the set of stationary parts connected through rater short transient parts. In gesture, however, all parts are transient. Piecewise stationary procces is not adequate.

In this paper, we propose another technique for time series pattern matching which deal with transient process rather than piecewise stationary process.

2 Markov Model and Hidden Markov Model

In general, the output probability of feature vector x_t, $P_t(x_t)$, is given by the conditional probability of the all past observation $x_0 x_1 \cdots x_{t-1}$

$$P_t(x_t) = Pr(x_t | x_0 x_1 \cdots x_{t-1}). \quad (1)$$

In the Markov model, the number of the sequence which is used for the condition is truncated by fixed number K.

$$P_t(x_t) = Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1}). \quad (2)$$

Usually, we give a certain state, say S_i, to the sequence of $x_{t-K} x_{t-K+1} \cdots x_{t-1}$. The former equation become

$$P_t(x_t) = Pr(x_t | S_i). \quad (3)$$

A certain state, S_j, is also given to the sequence of $x_{t-K+1} \cdots x_{t-1} x_t$. In this case, if the output is x_t and the previous state is S_i then the next state should be S_j. That means state transition is observable. In the Markov model, very complicated models are required to deal with complicated phenomena.

In the Hidden Markov Model, the same representation is adopted for the output probability. However, in this case, the relation between the output sequence and state is not unique but probabilistic. The same state is given to the many varieties of output sequences. So simple model which deal with complicated phenomena can be realized. This is an advantage of the Hidden Markov Model. However, in HMM, since so many varieties of output sequence share the same state and the shared state is the only information which affect the output probability, the process dealt with by HMM is restricted to piecewise stationary process. This restriction is not suit for the gesture recognition.

3 Proposed Model: Partly-Hidden Markov Model

3.1 Outline and parameters of PHMM

In the proposed model, the output probability $Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1})$ is represented by second order model,

$$Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1}) = Pr(x_t | S_i^f, S_j^s). \quad (4)$$

Here, state S_i^f is given to the sequence of $x_{t-K}x_{t-K+1}\cdots x_{t-2}$ and state S_j^s is given to the out-out of x_{t-1}. We call S_i^f f-state (first state). And we call S_j^s s-state (second state). If both of these mappings are unique, this model is equivalent to the Markov Model. If both of them are probabilistic, it is equivalent to the Hidden Markov Model.

In the proposed model, mapping from the sequence of $x_{t-K}x_{t-K+1}\cdots x_{t-2}$ to state S_i^f is probabilistic and the mapping from the output x_{t-1} to the state S_j^s is unique. We call this model "Partly-Hidden Markov Model (PHMM)." Since the half of the conditional part of the output probability is shared in many varieties of output sequences, the number of the states S_i^f can be reduced and the complexity of the model can also be reduced. Since the output probability of x_t is conditioned by state S_j^s (that means it is conditioned by x_{t-1}), the model can deal with more complicated process than piecewise stationary.

In the proposed model, the probability that the output sequence $x_1x_2\cdots x_T$ (s-state transition is $x_0x_1x_2\cdots x_{T-1}$) comes from the model with the f-state transition $s_1s_2\cdots s_T$ is defined by following equation.

$$Ps = Pr(x_1x_2\cdots x_T s_1^f s_2^f\cdots s_T^f s_1^s s_2^s \cdots s_T^s) \quad (5)$$

Since $s_1^s s_2^s\cdots s_T^s = x_0x_1\cdots x_{T-1}$,

$$\begin{aligned} Ps &= Pr(x_1x_2\cdots x_T s_1^f s_2^f \cdots s_T^f x_0) \\ &= Pr(s_1^f s_1^s)Pr(x_1|s_1^f, s_1^s) \\ &\quad \cdot\prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f s_t^s)Pr(x_{t+1}|s_{t+1}^f s_{t+1}^s) \\ &= Pr(s_1^f, x_0)Pr(x_1|s_1^f x_0) \\ &\quad \cdot\prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f x_{t-1})Pr(x_{t+1}|s_{t+1}^f x_t) \quad (6)\end{aligned}$$

Also, since,

$$Pr(s_{t+1}^f \mid s_t^f x_{t-1}) = \frac{Pr(s_{t+1}^f|s_t^f)Pr(x_{t-1}|s_{t+1}^f s_t^f)}{Pr(x_{t-1}|s_t^f)} \quad (7)$$

$$Pr(x_{t+1}|s_{t+1}^f x_t) = \frac{Pr(x_{t+1}x_t|s_{t+1}^f)}{Pr(x_t|s_{t+1}^f)} \quad (8)$$

Eq. (6) becomes,

$$\begin{aligned} Ps &= Pr(s_1^f, x_0)Pr(x_1|s_1^f, x_0) \\ &\quad \cdot\prod_{t=1}^{T-1} \frac{Pr(s_{t+1}^f|s_t^f)Pr(x_{t-1}|s_{t+1}^f s_t^f)}{Pr(x_{t-1}|s_t^f)} \\ &\quad \cdot\frac{Pr(x_{t+1}x_t|s_{t+1}^f)}{Pr(x_t|s_{t+1}^f)}. \quad (9)\end{aligned}$$

$Pr(x_1x_2\cdots x_T)$ can be obtained by summing up Eq.(9) for all possible combination of F-state transition $s_1^f s_2^f\cdots s_T^f$.

From above discussion, it is found that PHMM can be expressed by following 5 parameters.

a_{ij} : the probability that the next f-state is S_j^f in case that the current f-state is S_i^f.

$b_i(x)$: the probability that the current S-state is x in case that the current f-state is S_i^f.

$c_{ij}(y)$: the probability that the current s-state (last output) is y in case that the current f-state is S_i^f and the next f-state is S_j^f.

$d_i(x,y)$: the probability that the current output is x and the current s-state (last output) is y in case that the current f-state is S_i^f.

$e_i(y)$: the probability that the initial s-state is y and the initial f-state is S_i^f.

3.2 Forward algorithm for PHMM

Using above parameters, we describe a fast algorithm to get $Pr(x_1x_2\cdots x_T)$.

The forward probability, $\alpha(j,t)$, is the probability that the F-sate arrive at S_j^f and generate x_0 through x_t at time t.

$$\alpha(j,t) = Pr(x_0, x_1, \cdots, x_t, s_t^f = S_j^f) \quad (10)$$

The $\alpha(j,t)$ can be represented following recursive form.

$$\alpha(j,t) = \begin{cases} e_j(x_0) & (t=1) \\ \sum_i \alpha(i,t-1)\frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})} \cdot\frac{d_j(x_t,x_{t-1})}{b_j(x_t)} & (t=2,\cdots,T)\end{cases} \quad (11)$$

After the iterative calculation of $\alpha(j,t)$ from $t=1$ to T, $Pr(x_1x_2\cdots x_T)$ can be obtained as the $\alpha(J,T)$.

3.3 Viterbi algorithm for PHMM

Since F-state is the hidden state here, Viterbi algorithm of PHMM is the algorithm which find the best F-state transition which maximize the output probability of $x_0x_1\cdots x_T$. Here, we define the optimal sequence of F-state by $\hat{S} = \hat{s}_1^f\hat{s}_2^f\cdots\hat{s}_T^f$, and the output probability of $x_1x_2\cdots x_T$ given by $\hat{S}$ as P_v.

$\beta(j,t)$ denote the optimal output probability of the sequence of x_0 - x_t given by the optimal F-state sequence in which F-state arrive at S_j^f at time t. $\delta(j,t)$

denote the pointer for the backtrack. Then,

$$\beta(j,t) = \begin{cases} e_j(x_0) & (t=1) \\ \max_i \beta(i,t-1)\frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})} & \\ \cdot\frac{d_j(x_t,x_{t-1})}{b_j(x_t)} & (t=2,\cdots,T) \end{cases} \quad (12)$$

$$\delta(j,t) = \begin{cases} 1 & (t=1) \\ \arg\max_i \beta(i,t-1)\frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})} & \\ \cdot\frac{d_j(x_t,x_{t-1})}{b_j(x_t)} & (t=2,\cdots,T) \end{cases} \quad (13)$$

P_v can be obtained by the following equation after getting β iteratively from $t = 1$ to T.

$$P_v = \beta(J,T) \quad (14)$$

P_v can be used the score in the Viterbi decoding.

$\hat{\boldsymbol{S}}$ can be obtained by the back-track using δ.

$$\hat{s}_t^f = \begin{cases} S_J^f & (t=T) \\ S^f_{\delta(f(\hat{s}^f_{t+1}),t+1)} & (t=T-1,\cdots,1) \end{cases} \quad (15)$$

Here, $f()$ denote the function which returns the subscript number of the given F-state. $\hat{\boldsymbol{S}}$ gives the optimal segmentation.

3.4 Training of PHMM

Training can be performed by the EM algorithm using Eq.(11) or segmental K-means algorithm based on the Eq.(12).

In the following experiments, we adopted the segmental K-means algorithm.

1. Define the initial values for PHMM parameters (a_{ij}, $b_i(x)$, $c_{ij}(x)$, $d_i(x,y)$, $e_i(x)$).
2. Repeat 2.1,2.2 till the terminate condition satisfied.
   - 2.1 Do Viterbi decoding using current model parameters and get optimal segment boundaries.
   - 2.2 Re-estimate PHMM parameters using given segment boundaries.

In the following experiment, terminate condition is to repeat fixed number of iterations. Gaussian distribution functions are used for the parameters $b_i(x)$, $c_{ij}(x)$, $d_i(x,y)$, $e_i(x)$.

4 Experiment

4.1 Experimental setup

(a) Categories for recognition

6 words of sign language, {"No", "Doubtful", "Goodbye", "Come here", "Speak", "Like it"} are selected. Only right hand is used to express these words. So, they are easy to be confused.

Figure 1 and 2 shows the examples of "Speak" and "Like it".

(b) Data

20 subjects played 6 words from Japanese sign language 4 times each.

(c) Feature parameter

We get 10 frames of image for every second. Then, we extracted head position (x_t^h, y_t^h), right hand position (x_t^r, y_t^r) and hand area(s_t) from each image. The position parameters are translated into relative parameter,

$$(x_t, y_t) = (x_t^r, y_t^r) - (x_t^h, y_t^h). \quad (16)$$

Then, the following vector, z_t, is used as feature parameter.

$$z_t = (x_t, y_t, \Delta x_t, \Delta y_t, \Delta s_t) \quad (17)$$

Here Δx_t denote difference parameter of x_t,

$$\Delta x_t = x_t - x_{t-1}. \quad (18)$$

(d) Comparative experimental condition

DP-D : Player-dependent test using DP matching. The same player's data are used as templates.

DP-I$_1$: Player-independent test using DP matching. The different player's templates are used.

DP-I$_4$: Player-independent test using multi- template DP matching. Four set of templates are selected by clustering method using different player's data.

HMM-I$_c$: Player-closed player-independent test using HMM. Training dataset include the same player's data.

HMM-I$_o$: Player-open player-independent test using HMM. Training dataset does not include the same player's data.

PHMM-I$_c$: Player-closed player-independent test using HMM. Training dataset include the same player's data.

PHMM-I$_o$: Player-open player-independent test using HMM. Training dataset does not include the same player's data.

For HMM and PHMM, we don't test player-dependent condition because the number of the samples in each subject is too small to train model.

4.2 Experimental results

Table 1 show the experimental results.

Table 1 Experimental results

	% Correct		
	-D	-I$_1$	-I$_4$
DP	98.1	72.1	79.2
		-I$_c$	-I$_o$
HMM		95.6	93.1
PHMM		98.8	93.5

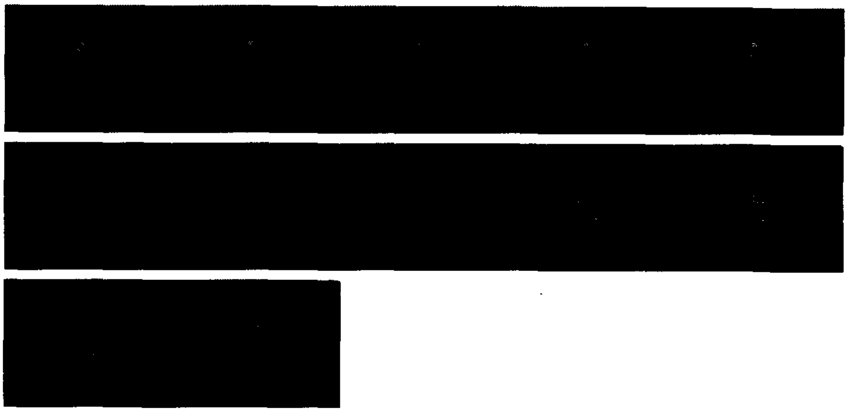

Figure 1: An example of captured image for gesture "Speak".

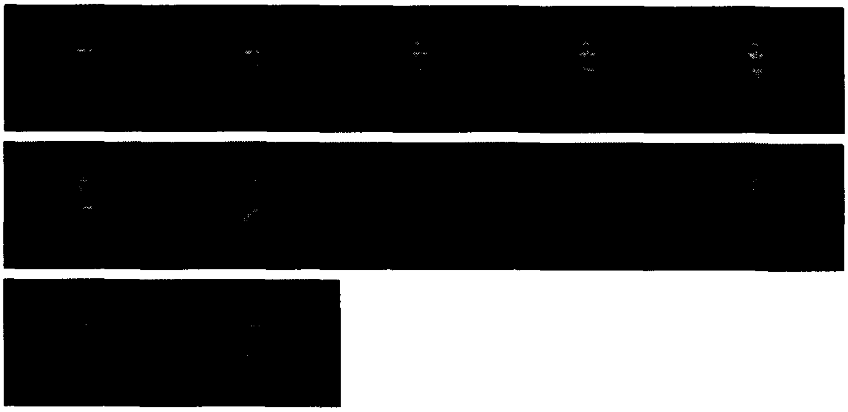

Figure 2: An example of captured image for gesture "Like it".

As for the player-closed test, score of PHMM is better than that of HMM by 3.2 point. This is 73% of error reduction. This score is better than player-dependent test using DP-matching.

As for the player-open test, performance of PHMM is slightly better than HMM. But this difference is not significant.

5 Conclusion

A new time series pattern matching method is proposed and it is applied to gesture recognition. As compared with HMM, PHMM improved the player-closed score by 73%. This results show the effectiveness of proposed method.

PHMM has more model parameters than HMM, so it is expected that the model need more training data. More study is needed to prove the effect of quantity of the training data.

References

[1] J.Yamato, J.Ohtani, K.Ishii, "Recognizing human action in time-sequential image using Hidden Markov Models", Trans. IEICE(D-II), **J76-D-II**, 12, pp.2556-2563, Dec. 1993.

[2] K.Takahashi, S.Seki, H.Kojima, R.Oka, "Spotting recognition of human gesture from time-varying images", Trans. IEICE(D-II),**J77-D-II**, 8, pp.1552-1561, Aug. 1994.

[3] T.Starner and A.Pentland, "Visual recognition of American sign language using Hidden Markov Models," Proc. Int'l Workshop on Automatic Face and Gesture Recognition 1995, June 1995.

[4] T.Kobayashi, "Speech recognition based on Hidden Markov Models," Trans. Institute of Electrical Engineers of Japan, 113-C, 5, pp.259-301 May 1993.

[5] X.D.Huang, Y.Ariki, M.A.Jack, "Hidden markov models for speech recognition", Edinburgh University Press, 1990.