HMM-Based Image Sequences Filtering

Hiroshi MATSUO and Hiroyasu KATO
Department of Electrical and Computer Engineering, Nagoya Institute of Technology
Gokiso, Showa, Nagoya, 4668555, JAPAN
matsuo@elcom.nitech.ac.jp

Abstract

It is possible to divide image sequences into some states, where each state has same motion phase. By applying such operation before recognition processing, accuracy of the recognition will be improved. New image sequence filtering by using HMM is proposed in this paper, which can divide image sequences into multiple states. Performance of the filtering is improved by doing re-learning of the observation symbol probability. In addition, effectiveness of the proposed method is shown by human identification using image sequences.

Keywords: Hidden Markov Models, Image sequences processing, Image sequences filtering, Human face identification, Image processing

1. Introduction

Recently, capturing of image sequences became possible with the improvement in the processing speed of computers, even if special hardware was not used. According to this improvement, various algorithms are proposed using image sequence recognition, such as optical flow calculation, tracking of moving object, gesture and face recognitions.

A lot of image sequence recognition algorithm using the Hidden Markov Models (HMM) are proposed[1],[2],[3]. In the recognition of image sequences such as gesture recognitions, whole image sequences itself contains information for recognition, and all image sequences are necessary for recognizing. However, image sequences can be divided into some phases which contains same meaning for recognition. And some phase which is not necessary for the recognition is included generally. Furthermore, these phases may give the adverse effect for recognition.

To overcome this problem, image sequences should be divided into multiple phases, and some phase are eliminated from image sequences. By applying these operation before recognition procedure, the improvement in recognition rate and robustness for recognition can be expected. Then, new image sequence filtering by using HMM which can eliminate some stumbling image sequences for recognition is proposed in this paper.

Proposed algorithm has the following features. (1) It is possible to obtain parameters such as transition and output probabilities using HMM which uses image sequence filtering by learning. (2) In conventional HMM recognition model, symbols generated from image sequences are used, then a lot of information for recognition will be lacked. Using proposed algorithm, raw image sequences and its current state in HMM can be used simultaneously after applying image sequences filtering. (3) The new algorithm in which the performance of the filtering can be improved by applying re-learning of output probability, is also proposed.

In addition, effectiveness of this algorithm is shown by human identification using image sequences with proposed filtering method.

2. Conventional recognition method using HMM

HMM

The HMM is a collection of states connected by transitions. Each transition has a pair of probability, a transition probability and an output probability. A formal characterization $\lambda = (\pi, \mathbf{A}, \mathbf{B})$ is as follows:

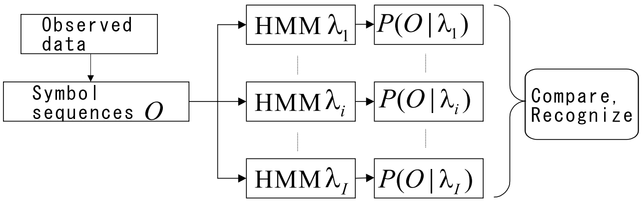


Figure 1. Conventional recognition model using HMM

$\pi = \{\pi_i\}$: The initial state distribution where π_i is the probability that the state S_i is the initial state. $\mathbf{A} = \{a_{ij}\}$: the state transition probability distributions where $\{a_{ij}\}$ is the probability of making a transition from state s_i to s_j. $\mathbf{B} = \{b_j(k)\}$: the observation symbol probability where $\{b_j(k)\}$ is the probability of emitting k at time t in state s_i. Let number of state be $(1 \leq i, j \leq N)$, number of symbol be $(1 \leq k \leq M)$.

Learning and recognition

Schematic diagrams of the general recognition method using HMM are shown in Fig.1. Each HMM parameter λ_i is learned using each symbol series O. The recognition is performed by choosing a HMM model with maximum probability $P(O|\lambda_i)$ in which observed symbol sequences $O = o_1, o_2, \cdots, o_T$ comes from phenomenon i. It also has elegant and efficient algorithms for learning and recognition, such as the Baum-Welch algorithm for learning, and Viterbi search algorithm[4] for recognition.

3. Image sequence filtering using HMM states and re-learning method of HMM parameters

Image sequences filtering using HMM states

In recognition using HMM, the information quantity will decrease because of converting observed phenomenon (image sequences) into the one dimensional observation symbol series. By using HMM for the phase division and recognition described as Fig.2, it becomes possible to compare observed image sequences and dictionaries made from learned image sequences. Furthermore raw image sequences can be used on recognition processes after applying image sequence filtering.

In generally forward algorithm for the recognition using HMM, but it is not possible to decide all transitions definitely, because all state transitions procedure is needed to obtain the generation probability. Therefore, it is also difficult to divide the image sequences into multiple phases with meanings.

However, by using the Viterbi algorithm which obtains the generation probability from the optimum condition series, it can make correspondence with one condition in every input image.

Each HMM parameter is correspondent to suitable phase with the meaning for image sequence filtering (we called this set of states “the goal states”) is described in the following.

Initial parameter of learning

The setting of the initial parameters is important for the Baum-Welch algorithm, because of this algorithm is based on obtaining local maximum value that maximizes generation probability of the input symbol series.

We use the following method for setting initial value in order to give the meaning to each HMM state. If the output from a state is suitable for the meaning of this state, symbol output probability $b_j(k)$ is set high. And if the output from this state is not suitable, output probability is set low. Furthermore state transition probability a_{ij} is also set according to transition probability of the goal state.

Filter design by re-learning HMM parameters

Even if the HMM parameter was learned by using Baum-Welch algorithm with the mentioned parameters, converged states do not always become the goal states. Then, new initial parameter setting algorithm is proposed for using re-learning procedure in order to bring closer to the goal states.

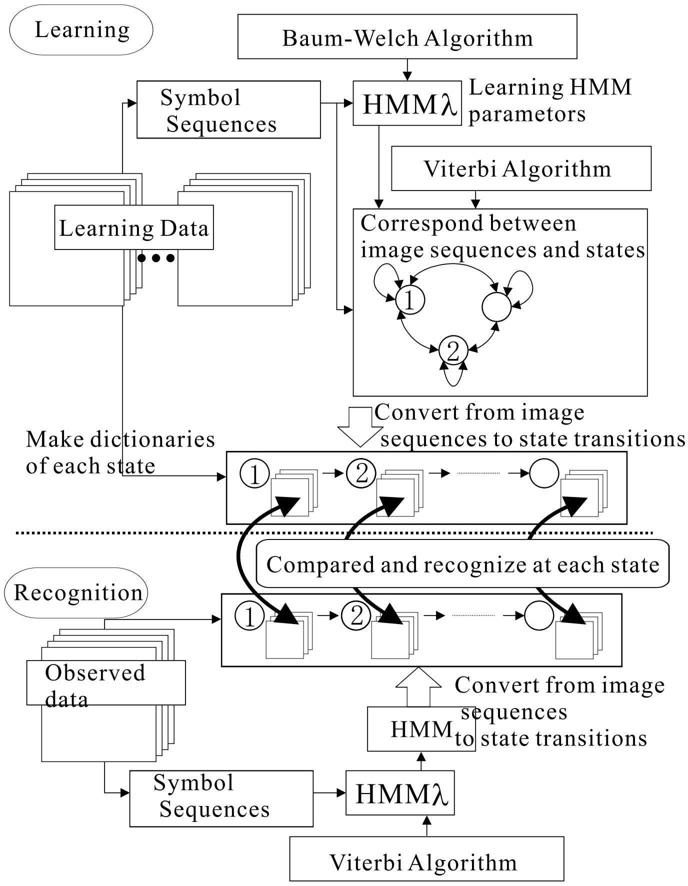


Figure 2. Image sequences filtering

First, initial transition probability on re-learning are set with constant value $\hat{a}_{ij} = \frac{1}{N} \quad (1 \leq i, j \leq N)$, in order to bring symbol output probability $b_j(k)$ close to the goal state. This is because to prevent that transition probabilities suppress the improvement of the output probability.

Output symbol probability $b_j(k)$ correspond to the goal state are marked $Label : High$, the others are marked $Label : Low$.

The initial parameter $\hat{b}_j(k)$ for re-learning algorithm from learned output probability $b_j(k)$ is described as follows.

(Step1) if $Label$ of $b_j(k)$ is $High$ goto Step2 else goto Step3.

(Step2) if $b_j(k) > th_{high}$ then $\hat{b}_j(k) = b_j(k)$ else $\hat{b}_j(k) = th_{high}$, and goto Step 4.

(Step3) if $b_j(k) < th_{low}$ then $\hat{b}_j(k) = b_j(k)$ else $\hat{b}_j(k) = th_{low}$, and goto 4.

(Step4) goto next symbol output probability.

Re-learning procedure is performed from these initial parameter $\hat{a}_{ij}, \hat{b}_j(k)$ using Baum-Welch algorithm, which make possible that HMM parameters approaches the goal state. Schematic diagram for re-learning procedure is shown in Fig.3.

4. Human face matching using image sequence filtering

In order to examine the effect of the proposed image sequences filter with HMM, the experiment of human identification is performed using image sequences by monitoring camera in the ceiling of the alley without exclusive lighting.

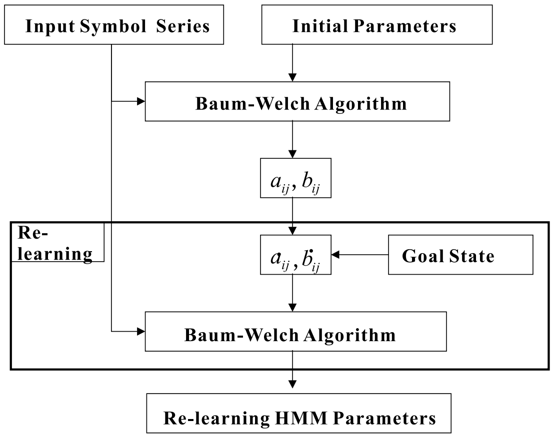


Figure 3. Re-learning

The human movement action described in Fig.4 was captured. Face region is trimmed manually and pixel size is normalized to 32×32 pixel[1]. Calculating the cross correlation between face image from 32 directions and input image, and finding the largest correlation image, input image sequences were converted into symbol sequences.

In the cognitive process, a direction dictionary that contains relation between face images and its direction is used. The input image sequences are converted into image sequences to calculate cross correlations between input face image and direction dictionary. Relation between symbols (labels of direction) and states are also shown in Fig.5. Six states (S_0 to S_5) are established, facing right, oblique right, front, oblique left, left and back. In Fig.6, a solid line shows the learning locus, and a dotted line shows the additional locus of input image sequence for filtering. Transition from S_0, S_1, S_2, S_3 to S_4 is a learned transition, then the aim of performing image sequence filter is (1) eliminate looped locus, (2) input image sequences is divided into learned transitions.

20 patterns (5 persons $\times$ 4 patterns) are used for learning. Each pattern has about 60 frames. To evaluate image sequence filtering, modified nearest neighbor method using state information from HMM is used. The category where the value of the following equation becomes a maximum was made to be the recognition result.

$$G(i) = \sum_{t, i \in D(l)} \max_{l=1,\cdots,L} f(B(t), D(l))$$

where f is cross correlation function, $B(t)$ is face image on time t, $D = \{D(l)\}$ is a set of dictionary data correspond to HMM state $B(t)$, i is a category.

Fig.7 shows recognition result in "no filtering", "filtering without re-learning" and "filtering with re-learning" The recognition using "no filtering" shows the result to compare all learning face data and input face image.

The vertical line shows the difference between maximum value G calculated from correct category and G calculated without correct categories. Hence, it becomes a positive value when recognized correctly, and an amount of this parameter shows the robustness of the recognition process.

Using simple nearest neighbor method, only person A and C would be able to recognize. Person B would be able to be recognized by applying the proposed filtering. Since most difference in max G without correct categories were increased in person A, C and E, improvement in the robustness can be confirmed.

In person B and D, the value calculated using re-learning method is less than the value of without re-learning. This is because the translation error from image sequences to symbol sequences.

Table 1 shows selected states and categories using each method. Cross correlations of no filtering method shows a maximum value with all data in the dictionary, selected category is the category in which the cross correlation becomes a maximum. Cross correlations of filtering with and without re-learning method shows a maximum value with same category in the dictionary. "—" shows deleted image using image sequence filtering.

[1] A lot of face region detection algorithms were proposed. But face detection is not a subject of this paper, then face region detection is performed manually.

Figure 4. Captured images

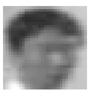
Performing image sequence filtering with HMM, correct categories are selected compared with no filtering method (simple nearest neighbor method). Furthermore by applying re-learning algorithm, the correspondence between image sequences and state of HMM becomes accurate.

5. Conclusion

New image sequence filtering algorithm using HMM is proposed. Image sequences can be divided into some arranged phase for reducing meaningless image sequences and for following recognition step. Re-learning algorithm for HMM is also proposed. Experiment of human identification using image sequence is also performed, and considerably good result has been achieved.

A simple modified nearest neighbor recognition algorithm using HMM state has been used in this experiment.

Table 1. Recognition result of person A

	Face Image						
No filtering	correlation (category)	0.993(A)	0.951(D)	0.953(C)	0.970(E)	0.972(B)	0.972(A)
HMM filtering	HMM state	S_0	S_5	S_4	S_2	S_5	S_3
	correlation (category)	0.993(A)	-	0.951(A)	0.960(A)	-	0.968(A)
HMM filtering with re-learning	HMM state	S_0	S_5	S_5	S_2	S_3	S_4
	correlation (category)	0.993(A)	-	-	0.960(A)	0.963(A)	0.969(A)

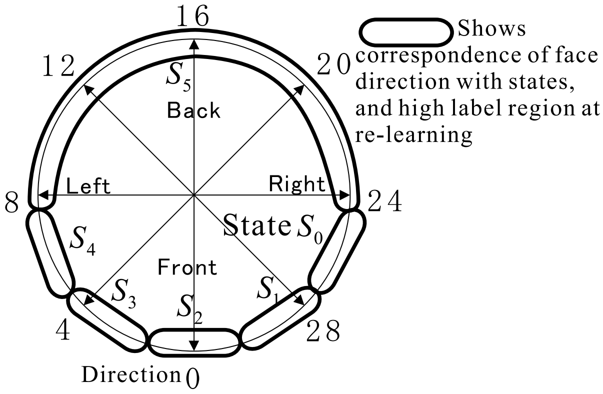


Figure 5. Relation between states and input symbols

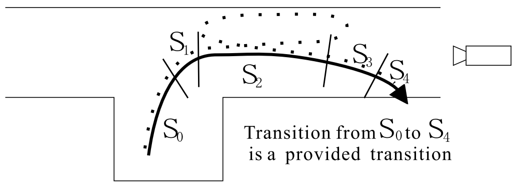


Figure 6. The human movement locus

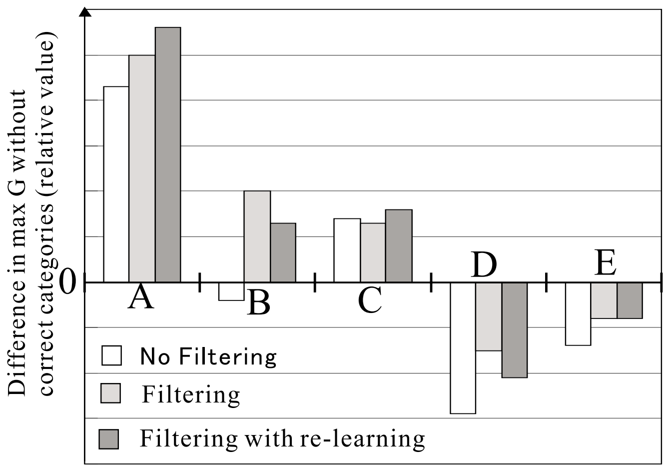


Figure 7. The experimental result

Future research will address the development of new recognition algorithm to use state of HMM more effectively.

References

[1] Hyeon-Kyu Lee and Jin H.kim, "Gesture spotting from continuous hand motion", Pattern Recognition Letters 19 pp.513-520, 1998.

[2] T. Starner and A.P. Pentland, "Visual recognition of American Sign Language using hidden Markov models", Int. Workshop on Automatic Face and Gesture Recognition, pp.189-194,1995

[3] H.I.Choi,P.K.Rhee, "Head Gesture recognition using HMMs","Pattern Recognition Letter 17,pp.213-221,1999

[4] A.J. Viterbi, "Error Bounds for Convolutions Codes and an Asymptotically Optimum Decoding Algorithm", IEEE Trans. Information Theory,Vol.13, pp.260-269, 1967