

## PAPER

# An Extension of Separable Lattice 2-D HMMs for Rotational Data Variations

Akira TAMAMORI<sup>†a)</sup>, *Nonmember*, Yoshihiko NANKAKU<sup>†b)</sup>, and Keiichi TOKUDA<sup>†c)</sup>, *Members*

**SUMMARY** This paper proposes a new generative model which can deal with rotational data variations by extending Separable Lattice 2-D HMMs (SL2D-HMMs). In image recognition, geometrical variations such as size, location and rotation degrade the performance. Therefore, the appropriate normalization processes for such variations are required. SL2D-HMMs can perform an elastic matching in both horizontal and vertical directions; this makes it possible to model invariance to size and location. To deal with rotational variations, we introduce additional HMM states which represent the shifts of the state alignments among the observation lines in a particular direction. Face recognition experiments show that the proposed method improves the performance significantly for rotational variation data.

**key words:** *image recognition, hidden Markov models, variational method*

## 1. Introduction

For many years, many researchers of pattern recognition have developed the field of image recognition as the main focus of pattern recognition and various techniques have been proposed. Especially, statistical approaches based on Principal Component Analysis (PCA) such as eigenface methods and subspace methods show good recognition performance in many applications [1], [2]. However, if images contain geometric variations such as size, location and rotation, the recognition performance is significantly degraded. Therefore, normalization processes for such geometric variations are required prior to applying these methods.

In many image recognition systems, the normalization process is included in the pre-process part of the classification, and heuristic normalization techniques are used. However, it is necessary to develop the normalization technique for each task, because such heuristic techniques usually use task dependent information. Furthermore, in image recognition, the final objective is not to accurately normalize images for human perception but to achieve a better recognition performance. Therefore, it is natural to use the same criterion for both training classifiers and normalization. This means that the normalization process should be integrated into classifiers.

Hidden Markov model (HMM) based techniques have been proposed as approaches for geometric variations. The

geometric matching between input images and model parameters is represented by discrete hidden variables and the normalization process is included in the calculation of probabilities. However, the extension of HMMs to multi-dimensions generally leads to an exponential increase in the amount of computation for its training algorithm. To reduce the computational complexity, the model structure needs to be constrained by limiting the number of possible alignments and/or assuming independence between hidden variables. Pseudo 2-D HMMs (or called embedded HMMs) have been proposed [3] and applied to many image recognition tasks. A pseudo 2-D HMM has a composite state structure for an efficient 2-D representation avoiding the complexity burden of a fully connected 2-D HMM. The states of a superior HMM in the horizontal direction are called super-states and each super-state has a one-dimensional HMM in the vertical direction instead of a probability density function. This assumption reduces the computational complexity and the maximum likelihood training algorithm has been derived [4]. However, the state alignments of consecutive observation lines in the vertical direction are calculated independently of each other and this hypothesis does not always hold true in practice.

Essentially, the studies of 2D dynamic programming (2D-DP) treat the same problem of the 2D-HMMs. The main difference between these studies is the definition of the cost function; The 2D-DP focuses on finding the mapping between two images with a pre-defined cost function, while the likelihood of 2D-HMMs is defined between an input image and the distribution which is estimated from multiple training images. Although some efficient approximation algorithms have been proposed for the 2D-DP problem [7]–[9], they still need high complicated costs and prior knowledge to determine the cost function is required for representing an accurate elastic matching dependently on image variations.

For another HMM based approach, Separable Lattice 2-D HMMs (SL2D-HMMs) have been proposed [5] to reduce the computational complexity while retaining the good properties for modeling multi-dimensional data. Furthermore, hidden Markov eigenface models (HMEMs) have been proposed [6] where the eigenface methods are integrated into SL2D-HMMs. The SL2D-HMMs have the composite structure of multiple hidden state sequences which interact to model the observation on a lattice. SL2D-HMMs perform an elastic matching in both horizontal and vertical directions; this makes it possible to model not only invari-

Manuscript received November 21, 2011.

Manuscript revised March 26, 2012.

<sup>†</sup>The authors are with the Department of Computer Science, Nagoya Institute of Technology, Nagoya-shi, 466–8555 Japan.

a) E-mail: matakai@sp.nitech.ac.jp

b) E-mail: nankaku@sp.nitech.ac.jp

c) E-mail: tokuda@sp.nitech.ac.jp

DOI: 10.1587/transinf.E95.D.2074

ance to the size and location of an object but also nonlinear warping in each dimension. However, SL2D-HMMs still cannot deal with rotational variations.

In this paper, we propose a new generative model which can deal with rotational data variations by extending SL2D-HMMs. To reduce the complexity, SL2D-HMMs have only one state sequence in each direction; this means that all horizontal/vertical lines of an observation lattice have the same state alignment for each direction. However, to represent the rotational variations, the models should have a different state alignment for each observation line and horizontal/vertical state alignments should be changed along with vertical/horizontal direction. Furthermore, it should take account of the dependency of the state alignments between consecutive observation lines to perform a continuous elastic matching. In this paper, we introduce additional HMM states which represent the shifts of the state alignments of the observation lines in a particular direction.

The parameters of the proposed model can be estimated via the Expectation Maximization (EM) algorithm for approximating the Maximum Likelihood (ML) estimate. However, similar to the training of SL2D-HMMs, the exact expectation step is computationally intractable. To derive a feasible algorithm, we applied the variational EM algorithm [10] to the our proposed model. The variational method approximates the posterior distribution over the hidden variables by a tractable distribution. The rest of the paper is organized as follows. Section 2 explains SL2D-HMMs briefly. Section 3 defines the structure of the model representing rotational variations and, the training algorithm for the proposed model is derived in Sect. 4. In Sect. 5, we describe face recognition experiments on the XM2VTS database and finally conclude in Sect. 6.

## 2. Separable Lattice 2-D HMMS

Separable lattice 2-D hidden Markov models are defined for modeling two-dimensional data. The observations of two-dimensional data, e.g., pixel values of an image and image sequence, are assumed to be given on a two-dimensional lattice:

$$\mathbf{O} = \{\mathbf{O}_t | t = (t^{(1)}, t^{(2)}) \in T\} \quad (1)$$

where  $t$  denotes the coordinates of the lattice in two-dimensional space  $T$  and  $t^{(m)} = 1, \dots, T^{(m)}$  is the coordinate of the  $m$ -th dimension. The observation  $\mathbf{O}_t$  is emitted from the state indicated by the hidden variable  $S_t \in K$ . The hidden variables  $S_t \in K$  can take one of  $K = K^{(1)}K^{(2)}$  states which assumed to be arranged on an two-dimensional state lattice  $K = \{(1, 1), (1, 2), \dots, (1, K^{(2)}), (2, 1), \dots, (K^{(1)}, K^{(2)})\}$ .

In other words, a set of hidden variables  $\{S_t | t \in T\}$  represents a segmentation of observations into the  $K$  states and each state corresponds to a segmented region in which the observation vectors are assumed to be generated from the same distribution. Since the observation  $\mathbf{O}_t$  is dependent only on the state  $S_t$  as in ordinary HMMs, dependen-

cies among hidden variables determine the properties and the modeling ability of two-dimensional HMMs.

To reduce the number of possible state sequences, we constrain the hidden variables to be composed of two Markov chains:

$$S = \{S^{(1)}, S^{(2)}\} \quad (2)$$

$$S^{(m)} = \{S_1^{(m)}, \dots, S_{t^{(m)}}^{(m)}, \dots, S_{T^{(m)}}^{(m)}\} \quad (3)$$

where  $S^{(m)}$  is the Markov chain along with the  $m$ -th coordinate and  $S_{t^{(m)}}^{(m)} \in \{1, \dots, K^{(m)}\}$ . In the separable lattice 2-D HMMS, the composite structure of hidden variables is defined as the product of hidden state sequences:

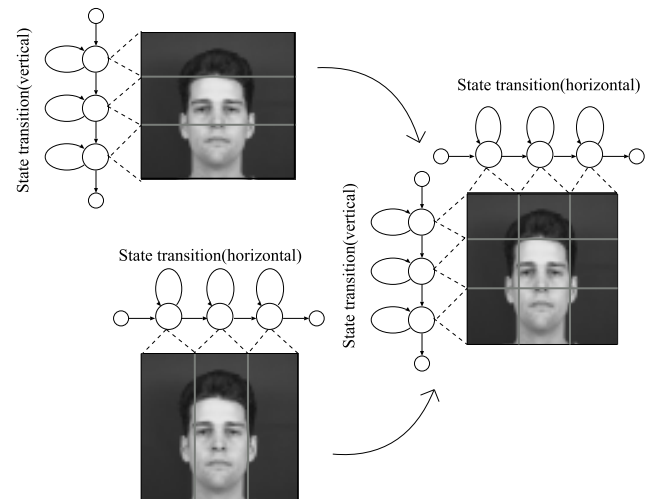
$$S_t = (S_{t^{(1)}}^{(1)}, S_{t^{(2)}}^{(2)}) \quad (4)$$

This means that the segmented regions of observations are constrained to be rectangles and this allows an observation lattice to be elastic in both vertical and horizontal directions. Using this structure, the number of possible state sequences can be reduced from  $\{\prod_m K^{(m)}\}^{\prod_m T^{(m)}}$  to  $\prod_m \{K^{(m)}\}^{T^{(m)}}$ . Figures 1 and 2 show the model structure of the separable lattice 2-D HMMS and its graphical representation, respectively.

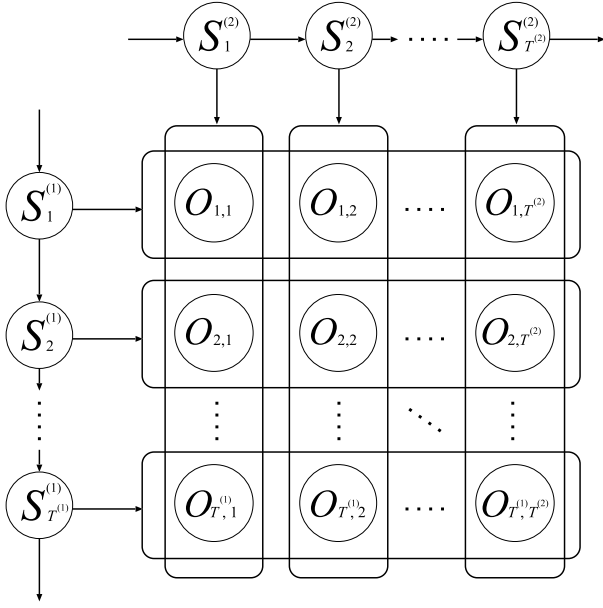
The joint probability of observation vectors  $\mathbf{O}$  and hidden variables  $S$  can be written as

$$\begin{aligned} P(\mathbf{O}, S | \Lambda) &= P(\mathbf{O} | S, \Lambda) \cdot \prod_{m=1,2} P(S^{(m)} | \Lambda) \\ &= \prod_t P(\mathbf{O}_t | S_t, \Lambda) \\ &\times \prod_{m=1,2} \left[ P(S_1^{(m)} | \Lambda) \prod_{t^{(m)}=2}^{T^{(m)}} P(S_{t^{(m)}}^{(m)} | S_{t^{(m)}-1}^{(m)}, \Lambda) \right] \end{aligned} \quad (5)$$

where  $\Lambda$  is a set of model parameters.



**Fig. 1** Model structure of the separable lattice 2-D HMMS: hidden state sequences are composed of independent two Markov chains.



**Fig. 2** Graphical representation of the separable lattice 2-D HMMs: the observation are emitted from the product of horizontal and vertical hidden state sequences.

### 3. Model Structure Representing Rotational Variations

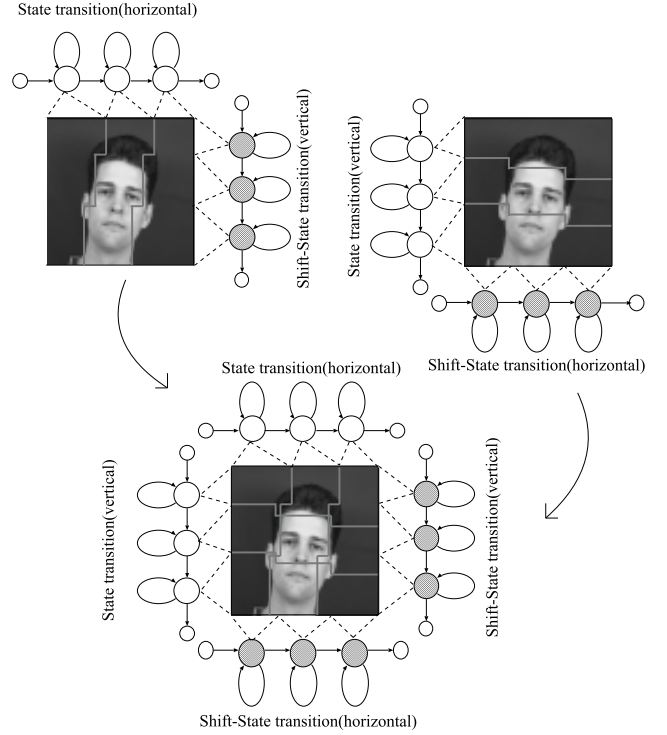
To represent rotational variations, the models should have a different state alignment for each observation line and horizontal/vertical state alignments should be changed along with vertical/horizontal direction. In this paper, we propose a new model structure with additional HMM states which represent the shifts of the state alignments of observation lines in a particular direction. Since the degree of the shift is controlled by the Markov chains, the proposed model can represent the dependency of the state alignments between consecutive observation lines. Therefore, the proposed model can perform a continuous elastic matching including rotational transformations. Figures 3 and 4 show the model structure of the proposed model and graphical representation for the proposed model, respectively.

The likelihood function of the proposed model is defined as follows:

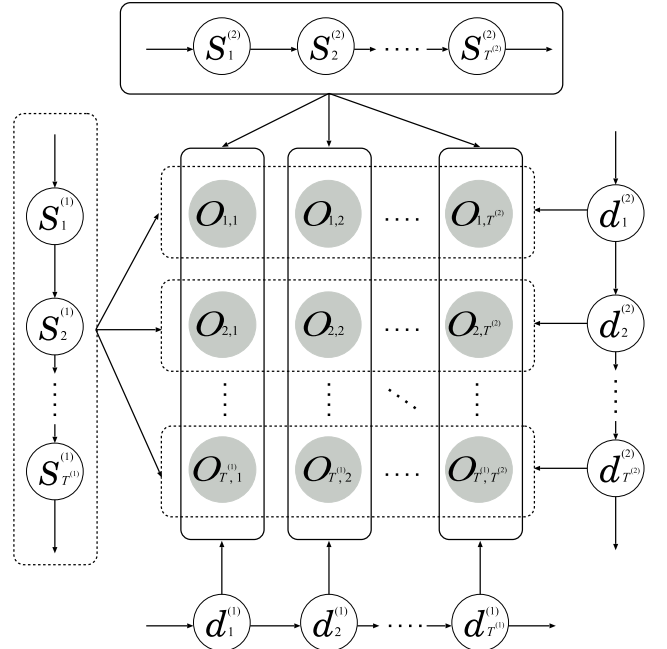
$$\begin{aligned}
 P(\mathbf{O}, \mathbf{S}, \mathbf{d} | \Lambda) &= P(\mathbf{O} | \mathbf{S}, \mathbf{d}, \Lambda) \cdot P(\mathbf{S} | \Lambda) \cdot P(\mathbf{d} | \Lambda) \\
 &= \prod_t P(\mathbf{O}_t | \mathbf{S}_t, \mathbf{d}_t, \Lambda) \\
 &\quad \times \prod_m P(\mathbf{S}^{(m)} | \Lambda) \cdot \prod_m P(\mathbf{d}^{(m)} | \Lambda)
 \end{aligned} \quad (6)$$

where  $\mathbf{S}$  represents the reference state sequences corresponding to the state sequences of SL2D-HMMs and  $\mathbf{d}$  represents the shift state sequences and consists of two Markov chains for each dimension:

$$\mathbf{d} = \{\mathbf{d}^{(1)}, \mathbf{d}^{(2)}\} \quad (7)$$



**Fig. 3** Model structure of the proposed model: The horizontal/vertical state alignments is changed along with vertical/horizontal state direction to represent the rotational variations.

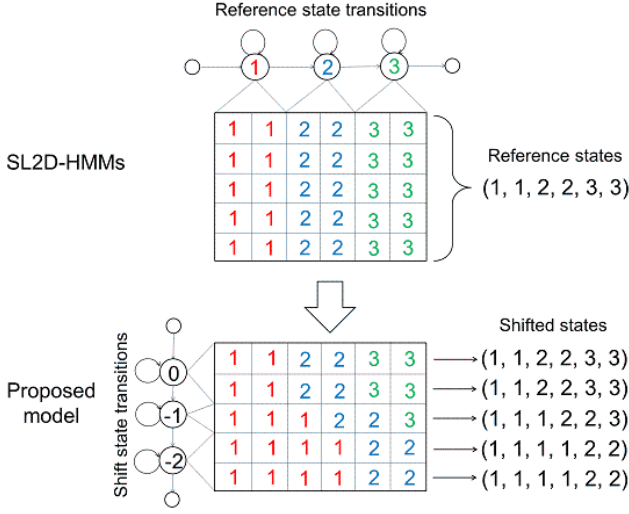


**Fig. 4** Graphical representation of the proposed model: The shift sequence affects the all data on the same observed line.

$$\mathbf{d}^{(m)} = \{d_1^{(m)}, d_2^{(m)}, \dots, d_{T^{(n)}}^{(m)}\} \quad (8)$$

$$d_{t^{(n)}}^{(m)} \in \{D_{min}^{(m)}, D_{min}^{(m)} + 1, \dots, D_{max}^{(m)}\}, n \neq m \quad (9)$$

where  $D_{min}^{(m)}$  and  $D_{max}^{(m)}$  represent the minimum and maximum



**Fig. 5** An example of state alignment of the proposed model for reference states and shifted states in horizontal direction: Without shift states (SL2D-HMMs), rectangle state alignments can be obtained while with shift states, monotonically shifted state alignments can be obtained in the proposed model.

shift of the  $m$ -th coordinate respectively, and  $S_{\vec{t}}$  is the shifted state defined as

$$S_{\vec{t}} = (S_{\vec{t}^{(1)}}^{(1)}, S_{\vec{t}^{(2)}}^{(2)}) = (S_{t^{(1)}+d_{t^{(1)}}}^{(1)}, S_{t^{(2)}+d_{t^{(2)}}}^{(2)}) \quad (10)$$

where the following boundary conditions are assumed:

$$S_{\vec{t}^{(m)}}^{(m)} = \begin{cases} 1 & (\vec{t}^{(m)} \leq 0) \\ K^{(m)} & (\vec{t}^{(m)} > T^{(m)}) \end{cases} \quad (11)$$

Figure 5 shows an example of the state alignment of the proposed model where monotonic alignment can be obtained by using shift states.

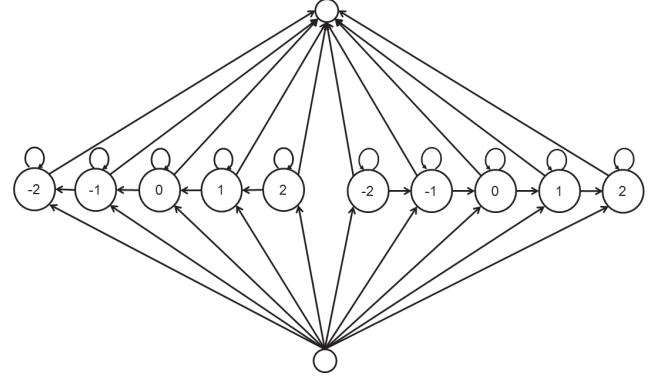
Model parameters of the proposed model are summarized as follows:

- **Parameters for state transition probability of reference states  $S$ :**

- 1)  $\Pi_S^{(m)} = \{\pi_{S,i}^{(m)} | 1 \leq i \leq K^{(m)}\}$ : the initial state probability distribution, where  $\pi_{S,i}^{(m)} = P(S_1^{(m)} = i | \Lambda)$  is the probability of state  $i$  at  $t^{(m)} = 1$  in the  $m$ -th state sequence  $S^{(m)}$ .
- 2)  $A_S^{(m)} = \{a_{S,ij}^{(m)} | 1 \leq i, j \leq K^{(m)}\}$ : the transition probability matrix, where  $a_{S,ij}^{(m)} = P(S_{t^{(m)}}^{(m)} = j | S_{t^{(m)}-1}^{(m)} = i, \Lambda)$  is the transition probability from state  $i$  to state  $j$  in the  $m$ -th state sequence  $S^{(m)}$ .

- **Parameters for state transition probability of shift states  $d$ :**

- 1)  $\Pi_d^{(m)} = \{\pi_{d,i}^{(m)} | 1 \leq i \leq K_d^{(m)}\}$ : the initial state probability distribution, where  $\pi_{d,i}^{(m)} = P(d_1^{(m)} = i | \Lambda)$  is the probability of state  $i$  at  $t^{(n)} = 1$  in the  $m$ -th state sequence  $d^{(m)}$ .



**Fig. 6** The example of topology of the transition probabilities of the  $m$ -th dimension shift states where  $D_{\min}^{(m)} = -2$  and  $D_{\max}^{(m)} = 2$ ; from this topology, monotonically increasing or decreasing sequence of the shift amount can be obtained and clockwise or counterclockwise rotational variations can be represented.

- 2)  $A_d^{(m)} = \{a_{d,ij}^{(m)} | D_{\min}^{(m)} \leq i, j \leq D_{\max}^{(m)}\}$ : the transition probability matrix, where  $a_{d,ij}^{(m)} = P(d_{t^{(n)}-1}^{(m)} = i, \Lambda)$  is the transition probability from state  $i$  to state  $j$  in the  $m$ -th state sequence  $d^{(m)}$ .

- **Parameters for output probability distribution:**

$B = \{b_k(\mathbf{O}_t) | k \in K\}$ : the output probability distributions, where  $b_k(\mathbf{O}_t)$  is the probability of observation vector  $\mathbf{O}_t$  at the state  $k$  on the state lattice  $K$  and assumed to be a single Gaussian distribution:  $P(\mathbf{O}_t | S_t = k) = \mathcal{N}(\mathbf{O}_t; \mu_k, \Sigma_k)$  where  $\mu_k$  and  $\Sigma_k$  are the mean vector and the covariance matrix, respectively.

Using the above shorthand notation, the proposed model is defined as

$$\Lambda = \{\Lambda_S^{(1)}, \Lambda_S^{(2)}, \Lambda_d^{(1)}, \Lambda_d^{(2)}, B\}, \quad (12)$$

$$\Lambda_S^{(m)} = \{\Pi_S^{(m)}, A_S^{(m)}\}, \quad (13)$$

$$\Lambda_d^{(m)} = \{\Pi_d^{(m)}, A_d^{(m)}\}. \quad (14)$$

The proposed model has potential to perform an continuous elastic matching beyond rotational variations. However, in this paper, the topology and the shift amounts are constrained to a special form which is expected to represent the continuous rotational variations. The example of the form for the  $m$ -th dimension where  $D_{\min}^{(m)} = -2$  and  $D_{\max}^{(m)} = 2$  is shown in Fig. 6.

## 4. Training Algorithm

### 4.1 Variational EM Algorithm

The parameters of the proposed model can be estimated via the Expectation Maximization (EM) algorithm which is an iterative procedure for approximating the Maximum Likelihood (ML) estimate. This procedure maximizes the expectation of the complete data log-likelihood so called Q-function:

$$Q(\Lambda, \Lambda') = \sum_{S, d} P(S, d | O, \Lambda) \ln P(O, S, d | \Lambda') \quad (15)$$

By maximizing the  $Q$ -function with respect to model parameters  $\Lambda$ , the re-estimation formula in the M-step can be easily derived. However, the calculation of the posterior distribution  $P(S, d | O, \Lambda)$  in the E-step is computationally intractable due to the combination of hidden variables. To derive a feasible problem, we applied the variational EM algorithm [10] to the training algorithm of the proposed model.

The variational methods approximate the posterior distribution over the hidden variables by a tractable distribution. Any distribution over the hidden variables defines a lower bound on the log-likelihood

$$\begin{aligned} \ln P(O | \Lambda) &= \ln \sum_S \sum_d Q(S, d) \frac{P(O, S, d | \Lambda)}{Q(S, d)} \\ &\geq \sum_S \sum_d Q(S, d) \ln \frac{P(O, S, d | \Lambda)}{Q(S, d)} \\ &= \mathcal{F}(Q, \Lambda) \end{aligned} \quad (16)$$

where Jensen's inequality has been applied. The difference between  $\ln P(O | \Lambda)$  and  $\mathcal{F}$  is given by the KL divergence between  $Q(S, d)$  and the posterior distribution of the hidden variables  $P(S, d | O, \Lambda)$ :

$$\begin{aligned} \mathcal{F}(Q, \Lambda) &= \sum_S \sum_d Q(S, d) \ln \frac{P(O, S, d | \Lambda)}{Q(S, d)} \\ &= \sum_S \sum_d Q(S, d) \ln P(O | \Lambda) \\ &\quad + \sum_S \sum_d Q(S, d) \ln \frac{P(S, d | O, \Lambda)}{Q(S, d)} \\ &= \ln P(O | \Lambda) - \text{KL}(Q \| P) \end{aligned} \quad (17)$$

Since the true log-likelihood  $\ln P(O | \Lambda)$  is independent of  $Q(S, d)$ , maximizing the lower bound  $\mathcal{F}$  is equivalent to minimizing the KL divergence. If we allow  $Q(S, d)$  to have complete flexibility then we see that the optimal  $Q(S, d)$  distribution is given by the true posterior  $P(S, d | O, \Lambda)$ , in the case where the KL divergence is zero and the bound becomes exact. In order to yield a tractable algorithm, it is necessary to consider a more restricted structure of  $Q(S, d)$  distributions. Given the structure, the parameters of  $Q(S, d)$  are varied so as to obtain the tightest possible bound, which maximizes  $\mathcal{F}$ .

The variational EM algorithm iteratively maximizes  $\mathcal{F}$  with respect to the  $Q$  and  $\Lambda$  holding the other parameters fixed:

$$\begin{aligned} \text{(E-step)} &: Q^{(k+1)} = \arg \max_{Q \in C} \mathcal{F}(Q, \Lambda^{(k)}) \\ \text{(M-step)} &: \Lambda^{(k+1)} = \arg \max_{\Lambda} \mathcal{F}(Q^{(k+1)}, \Lambda) \end{aligned}$$

where  $C$  is the set of constrained distributions. In this procedure, the lower bound  $\mathcal{F}$  is guaranteed to increase instead of the value of the  $Q$ -function.

The complexity and the approximation property of the

variational EM algorithm are dependent on a constraint to the posterior distribution  $Q(S, d)$  and it should be determined for each structure of graphical models. Here we consider a constrained family of variational distributions for the proposed model by assuming that  $Q(S, d)$  factorizes over subset  $S^{(m)}$  and  $d^{(m)}$  of the variables in  $S$  and  $d$ , so that

$$Q(S, d) = Q(S)Q(d) \quad (18)$$

$$= \prod_{m=1}^M Q(S^{(m)}) \prod_{m=1}^M Q(d^{(m)}) \quad (19)$$

where  $Q(S)$  and  $Q(d)$  are the posterior distribution over  $S$  and  $d$ , respectively. Also,  $\sum_{S^{(m)}} Q(S^{(m)}) = 1$  and  $\sum_{d^{(m)}} Q(d^{(m)}) = 1$ ,  $m = 1, \dots, M$ . The optimal distributions of the subsets are obtained by maximizing  $\mathcal{F}$  independently while keeping the other distributions fixed:

$$\begin{aligned} Q(S^{(m)}) &\propto P(S^{(m)} | \Lambda) \\ &\times \exp \left[ \sum_d Q(d) \sum_{S \setminus S^{(m)}} \prod_{n \neq m} Q(S^{(n)}) \ln P(O | S, d, \Lambda) \right] \end{aligned} \quad (20)$$

$$\begin{aligned} Q(d^{(m)}) &\propto P(d^{(m)} | \Lambda) \\ &\times \exp \left[ \sum_S Q(S) \sum_{d \setminus d^{(m)}} \prod_{n \neq m} Q(d^{(n)}) \ln P(O | S, d, \Lambda) \right] \end{aligned} \quad (21)$$

The E-step consists of the updates of  $Q(S^{(1)})$ ,  $Q(S^{(2)})$ ,  $Q(d^{(1)})$  and  $Q(d^{(2)})$ , which interact through the expectations. By inspection, the distribution  $Q(S^{(1)})$ ,  $Q(S^{(2)})$ ,  $Q(d^{(1)})$  and  $Q(d^{(2)})$  have the same structure as the posterior of standard HMMs. Therefore, the forward-backward algorithm can be used to compute the following expectations efficiently:

$$\langle \langle S_{t^{(m)}}, i \rangle \rangle = \sum_{S^{(m)}} Q(S^{(m)}) \delta(S_{t^{(m)}}^{(m)}, i) \quad (22)$$

$$\langle \langle S_{t^{(m)}-1}^{(m)}, i \rangle \langle S_{t^{(m)}}^{(m)}, j \rangle \rangle = \sum_{S^{(m)}} Q(S^{(m)}) \delta(S_{t^{(m)}-1}^{(m)}, i) \delta(S_{t^{(m)}}^{(m)}, j) \quad (23)$$

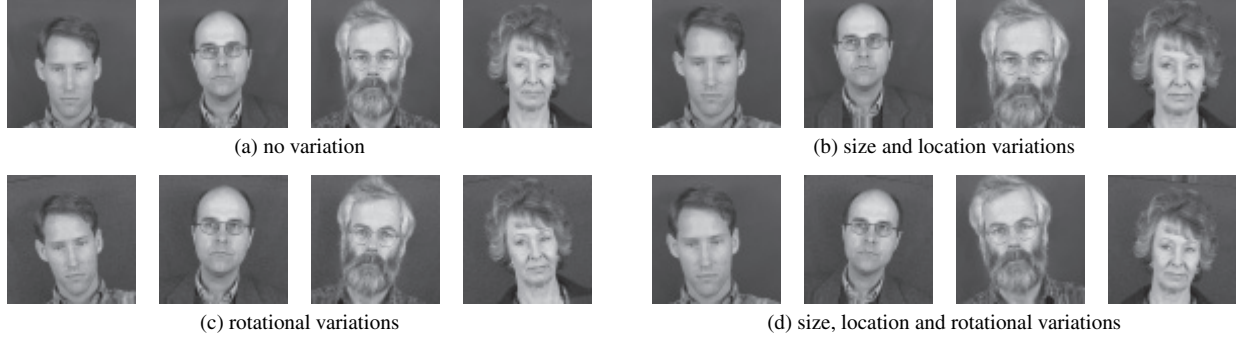
$$\langle \langle d_{t^{(n)}}^{(m)}, i \rangle \rangle = \sum_{d^{(m)}} Q(d^{(m)}) \delta(d_{t^{(n)}}^{(m)}, i) \quad (24)$$

$$\langle \langle d_{t^{(n)}-1}^{(m)}, i \rangle \langle d_{t^{(n)}}^{(m)}, j \rangle \rangle = \sum_{d^{(m)}} Q(d^{(m)}) \delta(d_{t^{(n)}-1}^{(m)}, i) \delta(d_{t^{(n)}}^{(m)}, j) \quad (25)$$

$$\begin{aligned} &\langle \langle S_{t^{(m)}+d_{t^{(n)}}}^{(m)}, k^{(m)} \rangle \langle d_{t^{(n)}}^{(m)}, l^{(m)} \rangle \rangle \\ &= \sum_{S^{(m)}} \sum_{d^{(m)}} Q(S^{(m)}) Q(d^{(m)}) \\ &\quad \times \delta(S_{t^{(m)}+d_{t^{(n)}}}^{(m)}, k^{(m)}) \delta(d_{t^{(n)}}^{(m)}, l^{(m)}) \end{aligned} \quad (26)$$

$$\langle \langle S_t, k \rangle \langle d_t, l \rangle \rangle = \prod_m \langle \langle S_{t^{(m)}+d_{t^{(n)}}}^{(m)}, k^{(m)} \rangle \langle d_{t^{(n)}}^{(m)}, l^{(m)} \rangle \rangle \quad (27)$$

where  $n \neq m$ . Using these expectations, the re-estimation



**Fig. 7** Examples of training data; with no variation (a) and with variations of size and location (b), with rotational variations (c) and with variations of size, location and rotations (d).

formula of the proposed model in the M-step are derived as follows.

$$\pi_{S,i}^{(m)} = \langle \langle S_1^{(m)}, i \rangle \rangle \quad (28)$$

$$\pi_{d,i}^{(m)} = \langle \langle d_1^{(m)}, i \rangle \rangle \quad (29)$$

$$a_{S,ij}^{(m)} = \frac{\sum_{t^{(m)}=2}^{T^{(m)}} \langle \langle S_{t^{(m)}-1}^{(m)}, i \rangle \langle S_{t^{(m)}}^{(m)}, j \rangle \rangle}{\sum_{t^{(m)}=1}^{T^{(m)}} \langle \langle S_{t^{(m)}}^{(m)}, i \rangle \rangle} \quad (30)$$

$$a_{d,ij}^{(m)} = \frac{\sum_{t^{(n)}=2}^{T^{(n)}} \langle \langle d_{t^{(n)}-1}^{(m)}, i \rangle \langle d_{t^{(n)}}^{(m)}, j \rangle \rangle}{\sum_{t^{(n)}=1}^{T^{(n)}} \langle \langle d_{t^{(n)}}^{(m)}, i \rangle \rangle} \quad (31)$$

$$\mu_k = \frac{\sum_t \sum_l \langle \langle S_t, k \rangle \langle d_t, l \rangle \rangle O_t}{\sum_t \sum_l \langle \langle S_t, k \rangle \langle d_t, l \rangle \rangle} \quad (32)$$

$$\Sigma_k = \frac{\sum_{t,l} \langle \langle S_t, k \rangle \langle d_t, l \rangle \rangle (O_t - \mu_k)(O_t - \mu_k)^T}{\sum_{t,l} \langle \langle S_t, k \rangle \langle d_t, l \rangle \rangle} \quad (33)$$

## 5. Experiments

### 5.1 Experimental Conditions

In order to demonstrate the modeling ability of the proposed model, face recognition experiments on the XM2VTS database [11] were conducted. We prepared eight images of 100 subjects; seven images are used for training and one image for testing. The face images were extracted from the original images (720×576 pixels and transformed into grayscale) and then sub-sampled to 64 × 64 pixels. In this process, we prepared four sets of data:

- “dataset 1”: the size- and location-normalized data.

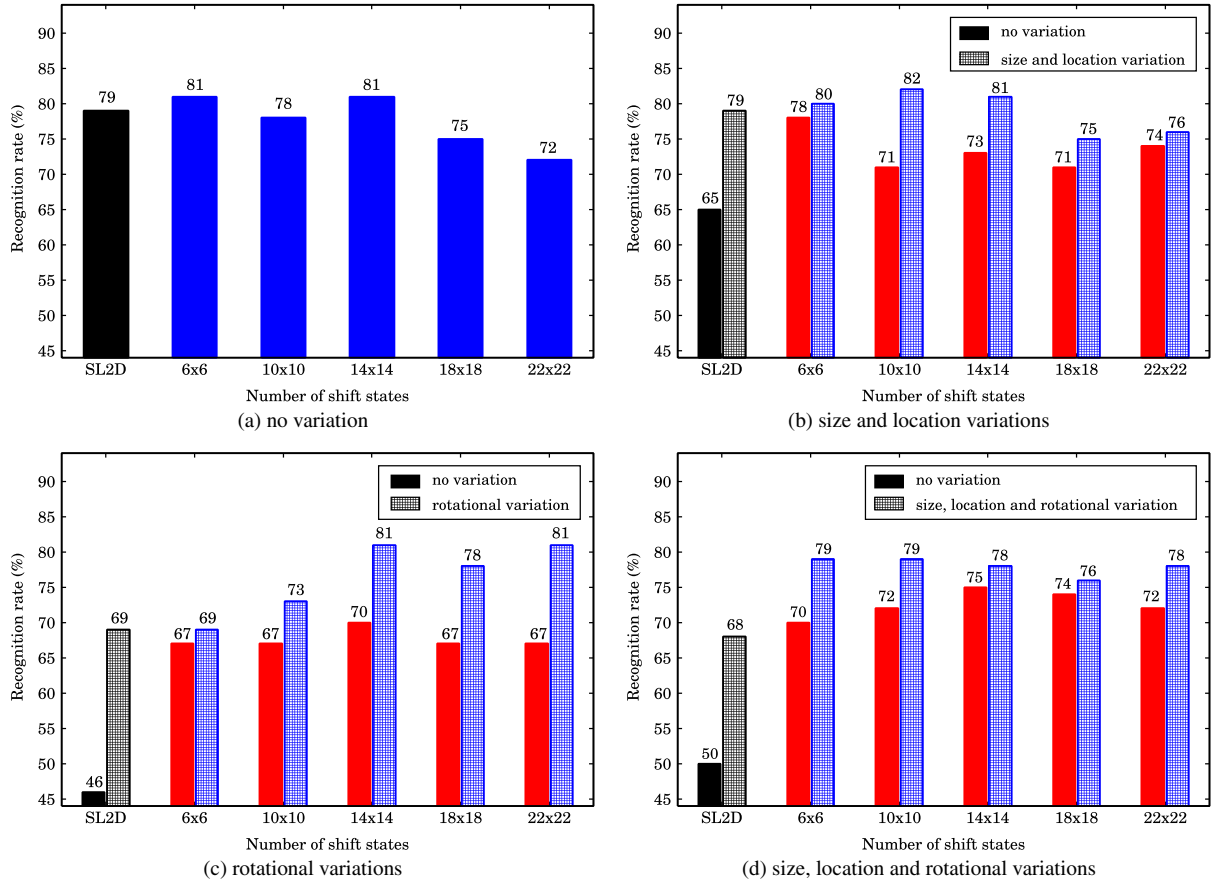
The original database does not include much variations of size and location, hence the center of the original images was used as the face location and the size was fixed to 550 × 550 pixels.

- “dataset 2”: the data with size and location variations. The sizes and locations were randomly generated by Gaussian distributions almost within the location shift of 40 × 20 pixels from the center and the range of size 500 × 500 ~ 600 × 600 with fixed aspect.
- “dataset 3”: the data with rotational variations. The rotation angles are randomly generated within −10 ~ 10 degrees from Gaussian distribution with 0.0 mean and 5.0 standard deviation.
- “dataset 4”: the data with size, location and rotational variations. The size and location variations were generated as well as “dataset 2” and the rotational variations were generated as well as “dataset 3”.

Figure 7 shows the examples of four datasets. Although it was already confirmed that the recognition performance was significantly improved with appropriate feature vectors such as 2-D discrete cosine transform coefficients or linear regression coefficients of images, the pixel intensity values were used as features in this paper. This is because the objective of this experiment was not to obtain the best performance of the proposed model but to demonstrate the property of the proposed model to normalize rotational variations. For the purpose of improving the recognition performance, the SL2D-HMMs were extended by integrating with a linear feature extraction such as probabilistic PCA or factor analyzers [6]. In the paper, it was confirmed that SL2D-HMMs and their extensions exceed the eigenface methods and subspace methods in face recognition experiments. The structure proposed in this paper can be easily integrated with a linear feature extraction as [6] for improving recognition performance.

The number of reference states was 24 × 24 and the number of shift states was varied among 6 × 6, 10 × 10, 14 × 14, 18 × 18 and 22 × 22, corresponding to the condi-





**Fig. 8** Recognition rates of the SL2D-HMMs and proposed model for each shift states tested on the dataset with no variation (a), with variations of size and location (b), with rotational variations (c) and with variations of size, location and rotations (d), respectively. In the figures, plain boxes and meshed ones represent the recognition rates of the models trained from the dataset with no variation and the same variation as the test dataset, respectively.

tions that  $-D_{min}^{(m)} = D_{max}^{(m)} = 1, 2, 3, 4$  and  $5$ , respectively. The number of reference states was previously optimized to give the best recognition performance on SL2D-HMMs. The transition probabilities for each sequence of reference states were assumed to be a left-to-right and top-to-bottom no skip topology and the transition probabilities for each sequence of shift states were assumed to be the topology as shown in Fig. 6.

## 5.2 Experimental Results

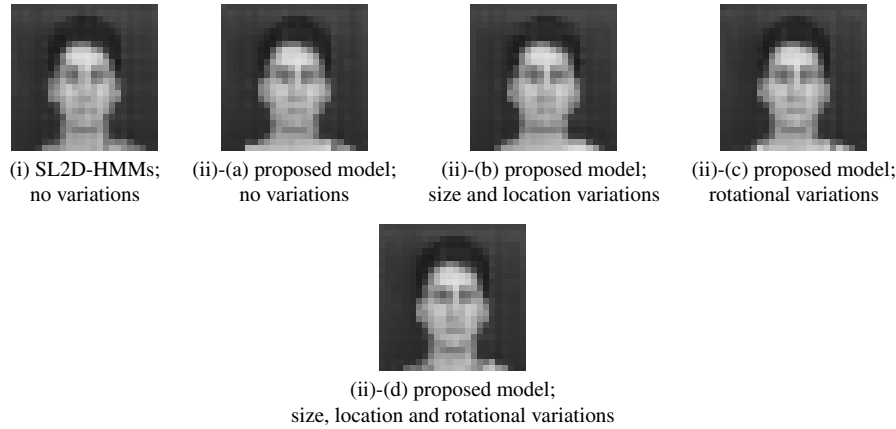
### 5.2.1 Recognition Performance

Figure 8(a), 8(b), 8(c) and 8(d) show the recognition rates of the test dataset with no variation (a), with variations of size and location (b), with rotational variations (c) and with variations of size, location and rotations (d), respectively. In the figures, plain boxes and meshed ones represent the recognition rates of the models trained from the dataset with no variation and the same variation as the test dataset, respectively.

From Fig. 8(b), it can be seen that the proposed

model possesses the comparable normalization ability to the SL2D-HMMs for size and location variations. Also, from Fig. 8(c), it can be seen that SL2D-HMMs degrade the recognition performance when they were trained and tested on “dataset3” where rotational variations were included, while the proposed model improves the recognition performance significantly compared with the SL2D-HMMs (meshed boxes). Especially, the highest recognition rate of 81% was obtained at  $14 \times 14$  and  $22 \times 22$  shift states, which is comparable to the recognition rate of SL2D-HMMs on “dataset 1.” This means that the proposed model can normalize rotational variations appropriately. It also can be seen that the proposed model improves the performance to rotational variations from Fig. 8(d) (meshed boxes). Particularly, the recognition rates of 79% at  $6 \times 6$ ,  $10 \times 10$ ,  $14 \times 14$  and  $22 \times 22$  shift states were obtained, which also indicates that the proposed model can normalize not only the size and location variations but also the rotational variations accurately.

Comparing the models trained from no variation datasets (plain boxes) and matched variation datasets (meshed boxes), the recognition rates of the matched varia-



**Fig. 9** Example of mean vectors: (i) is the mean vectors of the SL2D-HMMs. (ii) is the mean vectors of the proposed model. The number of shift state of (ii) is  $22 \times 22$ . They were estimated from the normalized data ("dataset 1").

tion were higher than those of the no variation datasets, even though no variation datasets were appropriately normalized. This is because the models over-fitted to the variation of the training datasets. However, from another point of the view, the proposed model can preserve the information of variation in the training data. It might be useful for some classification tasks, e.g., the model can use a kind of information that some target objects tend to rotate and the others are not for classification.

### 5.2.2 State Alignments

Figures 9 and 10 show the examples of mean vectors of SL2D-HMMs and the proposed model, and the visualized state alignments obtained by the Viterbi algorithm, respectively. In figure 9, the number of shift states of the proposed model is  $22 \times 22$ . The mean vectors were estimated from "dataset 1," "dataset 2," "dataset 3," and "dataset 4," respectively. The state alignments are represented by the mean vectors of the states corresponding to the observations of the test data. The values below the images represent the averaged log-likelihoods of the observation per pixel given the best alignments. When the visualized alignment is similar to the test data, it means that the model appropriately normalized the variations of the test data. The likelihood of the test data can also be regarded as an objective measure of the similarity; higher likelihood means that more preferable matching was obtained in terms of the maximum likelihood criterion.

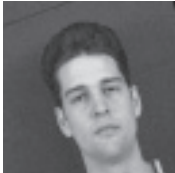




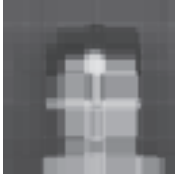
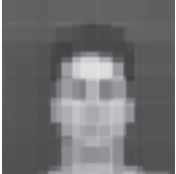

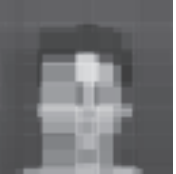





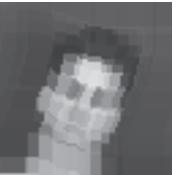
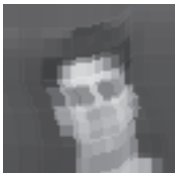

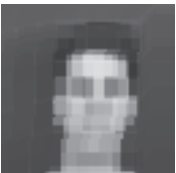


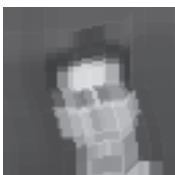
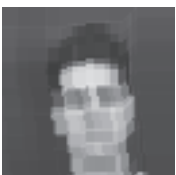

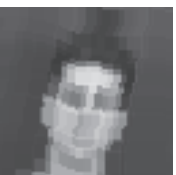

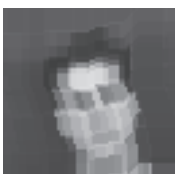
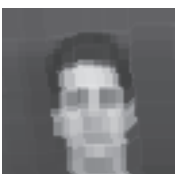
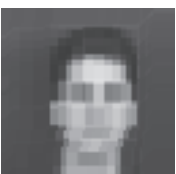
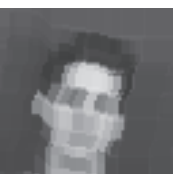
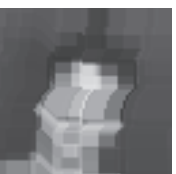
From the results, we can observe that SL2D-HMMs could not deal with the rotational variations due to the constraint of the model structure. The likelihood of the test data was also significantly decreased with increasing the rotational angle of the test data. Contrary to this, when the rotational angle of the test data was  $-10$ ,  $0$  or  $10$  degrees, the rotational variations of the data can be represented by the proposed model and the differences of the likelihood between  $0$  degree and  $10$ ,  $-10$  degrees were smaller than those of the SL2D-HMMs. It seemed that the maximum value of

the shift amount obtained by the proposed model was sufficient to represent the rotational angle  $\pm 10$  degrees. For the model (c) and (d), the maximum/minimum value of the rotational angle in the corresponding training dataset was between  $\pm 10$  degrees. This also led to the preferable results. On the other hand, when the rotational angle was larger, i.e.  $\pm 20$  degrees, the shift amount provided by the proposed model was not sufficient, so that the proposed model could not deal rotational variations compared to the results as the angle was  $\pm 10$ . Similarly, the proper state alignment of the reference state was not obtained. This is because, as shown in Eqs. (20) and (21), the reference state sequences and the shift state sequences are dependent on each other through the variational distributions. Therefore it was difficult to estimate the proper reference state sequences once the improper shift state sequences were estimated from the test data. From these results, it was suggested that the number of shift states need to be determined according to the degree of rotational variation.

## 6. Conclusion

We extended the model structure of separable lattice hidden Markov models for rotational variations. To represent rotational variations, the proposed model has additional HMM states which represent the shifts of the state alignments of observation lines in a particular direction. In face recognition experiments on the XM2VTS database, the proposed model achieved better results to the images than the conventional SL2D-HMMs. Moreover, the state alignments shows that the proposed model can normalize not only size and location variations but also rotational variations. The model parameter estimation from images with rotational variations and extensions to more flexible models will be future work.



	$\theta = 20^\circ$	$\theta = 10^\circ$	$\theta = 0^\circ$	$\theta = -10^\circ$	$\theta = -20^\circ$
test data					
SL2D-HMMs	 $\mathcal{F} = -4.56$	 $\mathcal{F} = -3.54$	 $\mathcal{F} = -3.13$	 $\mathcal{F} = -3.81$	 $\mathcal{F} = -4.60$
proposed model; (a) no variation	 $\mathcal{F} = -3.83$	 $\mathcal{F} = -3.32$	 $\mathcal{F} = -3.12$	 $\mathcal{F} = -3.29$	 $\mathcal{F} = -3.97$
proposed model; (b) size and location variations	 $\mathcal{F} = -4.17$	 $\mathcal{F} = -3.45$	 $\mathcal{F} = -3.11$	 $\mathcal{F} = -3.51$	 $\mathcal{F} = -4.14$
proposed model; (c) rotational variations	 $\mathcal{F} = -3.77$	 $\mathcal{F} = -3.27$	 $\mathcal{F} = -3.05$	 $\mathcal{F} = -3.38$	 $\mathcal{F} = -4.44$
proposed model; (d) size, location and rotational variations	 $\mathcal{F} = -3.68$	 $\mathcal{F} = -3.28$	 $\mathcal{F} = -3.12$	 $\mathcal{F} = -3.39$	 $\mathcal{F} = -4.19$

**Fig. 10** Examples of test data and the visualized state alignments: The  $\theta$  means the rotational angle for each test data. The  $\mathcal{F}$  means the estimated log-likelihood to test data.

## References

- [1] M. Turk and A. Pentland, "Face recognition using eigenface," *Proc. IEEE Comput. Vis. Pattern Recognit.*, pp.586–591, 1991.
- [2] S. Watanabe and N. Pakvasa, "Subspace method of pattern recognition," *1st International Joint Conference on Pattern Recognition*, pp.25–32, 1973.
- [3] S. Kuo and O.E. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.16, no.8, pp.842–848, 1994.
- [4] H. Othman and T. Aboulnasr, "A simplified second-order HMM with application to face recognition," *Proc. International Symposium on Circuits and Systems*, vol.2, pp.161–164, 2001.
- [5] D. Kurata, Y. Nankaku, K. Tokuda, T. Kitamura, and Z. Ghahramani, "Face recognition based on separable lattice HMMs," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp.737–740, 2006.
- [6] Y. Nankaku and K. Tokuda, "Face recognition based hidden Markov eigenface models," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp.469–472, 2007.
- [7] S. Uchida and H. Sakoe, "An approximation algorithm for two-dimensional warping," *IEICE Trans. Inf. & Syst.*, vol.E83-D, no.1, pp.109–111, Jan. 2000.
- [8] S. Uchida and H. Sakoe, "Piecewise linear two-dimensional warping," *Systems and Computers in Japan*, vol.32, no.12, pp.1–9, 2001.
- [9] N. Suto, T. Nishimura, R.H. Fujii, and R. Oka, "Spotting recognition of concave and convex reference image with pixel-wise correspondence using two-dimensional continuous dynamic programming," *IEICE Technical Report*, MVE2003-43, 2003.
- [10] M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, and L.K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol.37, pp.183–233, 1999.
- [11] K. Messer, J. Mates, J. Kitter, J. Luetttin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," *Audio and Video-Based Biometric Person Authentication*, pp.72–77, 1999.



**Yoshihiko Nankaku** received the B.E. degree in Computer Science, and the M.E. and Ph.D. degrees in the Department of Electrical and Electronic Engineering from the Nagoya Institute of Technology, Nagoya, Japan, in 1999, 2001, and 2004 respectively. After a year as a postdoctoral fellow at the Nagoya Institute of Technology, he is currently an Assistant Professor at the same Institute. His research interests include statistical machine learning, speech recognition, speech synthesis, image recognition, and multi-modal interface. He is a member of the Acoustical Society of Japan (ASJ).



**Keiichi Tokuda** received the B.E. degree in electrical and electronic engineering from Nagoya Institute of Technology, Nagoya, Japan, the M.E. and Dr.Eng. degrees in information processing from the Tokyo Institute of Technology, Tokyo, Japan, in 1984, 1986, and 1989, respectively. From 1989 to 1996 he was a Research Associate at the Department of Electronic and Electric Engineering, Tokyo Institute of Technology. From 1996 to 2004 he was a

Associate Professor at the Department of Computer Science, Nagoya Institute of Technology as Associate Professor, and now he is a Professor at the same institute. He is also an Invited Researcher at ATR Spoken Language Translation Research Laboratories, Japan and was a Visiting Researcher at Carnegie Mellon University from 2001 to 2002. He published over 60 journal papers and over 150 conference papers, and received 5 paper awards. He was a member of the Speech Technical Committee of the IEEE Signal Processing Society from 2000 to 2003. Currently he is a member of ISCA Advisory Council and an associate editor of IEEE Transactions on Audio, Speech & Language Processing, and acts as organizer and reviewer for many major speech conferences, workshops and journals. His research interests include speech coding, speech synthesis and recognition, and statistical machine learning.



**Akira Tamamori** received the B.E. degree in Computer Science, and the M.E. degrees in the Department of Scientific and Engineering Simulation from the Nagoya Institute of Technology, Nagoya, Japan, in 2008, 2010 respectively. He is currently a Ph.D. candidate at the same institute. His research interests include statistical machine learning, image recognition, speech recognition.