FACE RECOGNITION USING HIDDEN MARKOV EIGENFACE MODELS

Yoshihiko Nankaku and Keiichi Tokuda

Department of Computer Science and Engineering
Nagoya Institute of Technology, Nagoya 466-8555, Japan
E-mail: {nankaku, tokuda}@ics.nitech.ac.jp

ABSTRACT

This paper proposes Hidden Markov Eigenface Models (HMEMs) in which the eigenfaces are integrated into Separable Lattice Hidden Markov Models (SL-HMMs). SL-HMMs have been proposed for modeling multi-dimensional data, e.g, images, image sequences, 3-D objects. In its application to face recognition, SL-HMMs can perform an elastic image matching in both horizontal and vertical directions. However, SL-HMMs still have a limitation that the observations are assumed to be generated independently from corresponding states; it is insufficient to represent variations in face images, e.g., lighting conditions, facial expressions, etc. To overcome this problem, the structure of Probabilistic Principal Component Analysis (PPCA) and Factor Analysis (FA) is used as a probabilistic representation of eigenfaces. The proposed model has good properties of both PPCA/FA and SL-HMMs: a linear feature extraction and invariances to size and location of images. In face recognition experiments on the XM2VTS database, the proposed model improved the performance significantly.

Index Terms— Face recognition, Eigenfaces, Hidden Markov models, Probabilistic principal component analysis, Factor analysis

1. INTRODUCTION

In face recognition, appearance-based approaches have been extensively investigated in which pixel values are directly used as a feature vector and applied statistical analysis to extract an efficient representation of images. Eigenface method [1] is one of the most popular methods belonging to this category. A linear feature extractor which is called "eigenfaces" is constructed by applying the Principal Component Analysis (PCA) to all training images of all classes. Face images are projected and classified on the subspace spanned by eigenfaces. Subspace method [2] is also a well known pattern recognition technique based on PCA and frequently applied to appearance based face recognition. This method assumes that class dependent linear subspace can represent variations of class images and the distance of an input image from the subspace is used as a similarity measure for the particular class. Although the classification measures are different among these methods, a linear feature extraction based on statistical analysis is an effective and principal technique for face recognition. However, if face images contain variations such as size and location (geometric variations), the recognition performance is significantly degraded, because it is inefficient to represent the change of size and location by a linear combination of eigenfaces. To avoid this problem, normalization processes for geometric variations are required prior to applying these methods.

Hidden Markov Model (HMM) based approaches are one of techniques which can deal with the 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 while retaining the good properties for modeling multi-dimensional data, Separable Lattice Hidden Markov Models (SL-HMMs) have been proposed [3]. The SL-HMMs have the composite structure of multiple hidden state sequences which interact to model the observation on a lattice. In case of 2-D lattices, the SL-HMM performs an elastic matching in both horizontal and vertical directions; this makes it possible to model not only invariances to the size and location of an object but also nonlinear warping in each dimension. However, SL-HMMs still have a limitation in its application of face recognition: the observations are assumed to be generated independently from corresponding states, it is insufficient to represent variations in face images, e.g., lighting condition, facial expression, etc.

In this paper, we propose Hidden Markov Eigenface Models (HMEMs) in which eigenfaces are integrated into SL-HMMs. In the proposed model, the eigenfaces are represented by probabilistic latent variable models, such as Probabilistic Principal Component Analysis (PPCA) and Factor Analysis (FA). The proposed model has good properties of both the PPCA/FA and SL-HMM: a linear feature extraction based on statistical analysis and size and location invariant image recognition. The parameters of HMEMs can be estimated via the expectation maximization (EM) algorithm for approximating the Maximum Likelihood (ML) estimate. However, similarly to the training of SL-HMMs, the exact expectation step is computationally intractable. To derive a feasible problem, we applied the variational EM algorithm to HMEMs. Variational methods approximate the posterior distribution over the hidden variables by a tractable distribution. A structure approximation is presented in which the factor variables and the hidden state sequences and are decoupled.

The rest of the paper is organized as follows. Section 2 describes probabilistic representations of eigenfaces using PPCA and FA, and section 3 explains SL-HMMs briefly. In section 4 and 5, we define the structure of HMEMs and derive its training algorithm, respectively. In Section 6, we describe face recognition experiments on the XM2VTS database and finally conclude in Section 7.

2. PROBABILISTIC EIGENFACE MODELS

Probabilistic Principal Component Analysis (PPCA) and Factor analysis (FA) are statistical methods for modeling the covariance structure with a small number of latent variables [4], [5]. A d-dimensional observation vector $\boldsymbol{o}$ is assumed to be generated from a q-dimensional factor vector $(q < d)$ and a d-dimensional noise vector:

$$\boldsymbol{o} = \boldsymbol{W}\boldsymbol{a} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{W}$ is a $d \times q$ matrix known as the factor loading matrix. The vector $\boldsymbol{a}$ is a latent variable assumed to be distributed according to a

Gaussian density $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and the noise vector $\boldsymbol{n}$ is distributed according to $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{\Sigma}$ is assumed to be a diagonal matrix, this model is called FA, and PPCA is a special case of FA in which the noise is isotropic, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. The output probability of the observation $\boldsymbol{o}$ given $\boldsymbol{a}$ can be written as

$$P(\boldsymbol{o} \mid \boldsymbol{a}, \Lambda) = \mathcal{N}(\boldsymbol{o} \mid \boldsymbol{W}\boldsymbol{a} + \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (2)$$

because the product $\boldsymbol{W}\boldsymbol{a}$ becomes a constant vector and added to the noise vector $\boldsymbol{n}$. Therefore, the marginal distribution of observation $\boldsymbol{o}$ is obtained by integrating out the latent variable $\boldsymbol{a}$:

$$\begin{aligned} P(\boldsymbol{o} \mid \Lambda) &= \int P(\boldsymbol{o} \mid \boldsymbol{a}, \Lambda) P(\boldsymbol{a} \mid \Lambda) d\boldsymbol{a} \\ &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{W}\boldsymbol{W}^\top + \boldsymbol{\Sigma}) \end{aligned} \quad (3)$$

From the above equation, it is obvious that PPCA/FA is a Gaussian distribution whose covariance matrix is constrained by the loading matrix and the noise covariance matrix. Therefore, PPCA/FA can capture the correlation structure among observations by a small number of parameters instead of using the full covariance matrix.

It is known that the maximum likelihood solution of PPCA find the subspace spanned by the principal eigenvectors of covariance matrix obtained by the conventional PCA [4]. In the proposed model, the eigenface method and the subspace method are performed by PPCA/FA instead of the conventional PCA. It should be noted that the subspace method using the conventional PCA is not exactly the same as the classifier using PPCA, because PPCA is a constrained Gaussian distribution and it takes into account of the probability of the factor vector. Taking the limit as $\sigma^2 \to 0$, PPCA is equivalent to the subspace method, though the density model becomes singular. The eigenface method using a single global projection can also be represented by the probabilistic form in which the parameters of the loading matrices are shared among all classes. Accordingly, although many structures can be considered in the presented framework, this paper focuses on the case that PPCA/FA are applied for modeling data of each classes individually.

3. SEPARABLE LATTICE HMMS

The separable lattice hidden Markov models (SL-HMMs) are defined for modeling multi-dimensional data. The observations of M-dimensional data, e.g., pixel values of an image and image sequence, are assumed to be given on a M-dimensional lattice:

$$\boldsymbol{O} = \{\boldsymbol{O}_{\boldsymbol{t}} \mid \boldsymbol{t} = (t^{(1)}, \ldots, t^{(m)}, \ldots, t^{(M)}) \in \boldsymbol{T}\} \quad (4)$$

where $\boldsymbol{t}$ denotes the coordinates of the lattice in M-dimensional space $\boldsymbol{T}$ and $t^{(m)} = 1, \ldots, T^{(m)}$ is the coordinate of the m-th dimension. The observation $\boldsymbol{O}_{\boldsymbol{t}}$ is emitted from corresponding state indicated by the hidden variable $\boldsymbol{S}_{\boldsymbol{t}} \in \boldsymbol{K}$. The hidden variable $\boldsymbol{S}_{\boldsymbol{t}} \in \boldsymbol{K}$ can take one of $K = \prod_m K^{(m)}$ states which assumed to be arranged on an M-dimensional state lattice $\boldsymbol{K} = \{1, \ldots, K\}$. Since the observation $\boldsymbol{O}_{\boldsymbol{t}}$ is dependent only on the state $\boldsymbol{S}_{\boldsymbol{t}}$ as in ordinary HMMs, dependencies between hidden variables determine the properties and the modeling ability of multi-dimensional HMMs.

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

$$\boldsymbol{S} = \{\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(m)}, \ldots, \boldsymbol{S}^{(M)}\} \quad (5)$$

$$\boldsymbol{S}^{(m)} = \{S_1^{(m)}, \ldots, S_{t^{(m)}}^{(m)}, \ldots, S_{T^{(m)}}^{(m)}\} \quad (6)$$

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

$$\boldsymbol{S}_{\boldsymbol{t}} = (S_{t^{(1)}}^{(1)}, S_{t^{(2)}}^{(2)}, \ldots, S_{t^{(M)}}^{(M)}) \quad (7)$$

This means that in the 2-D case, 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.

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

$$\begin{aligned} & P(\boldsymbol{O}, \boldsymbol{S} \mid \Lambda) \\ &= P(\boldsymbol{O} \mid \boldsymbol{S}, \Lambda) \prod_{m=1}^{M} P(\boldsymbol{S}^{(m)} \mid \Lambda) \quad (8) \\ &= \prod_{\boldsymbol{t}} P(\boldsymbol{O}_{\boldsymbol{t}} \mid \boldsymbol{S}_{\boldsymbol{t}}, \Lambda) \times \\ & \quad \prod_{m=1}^{M} \left[P(S_1^{(m)} \mid \Lambda) \prod_{t^{(m)}=2}^{T^{(m)}} P(S_{t^{(m)}}^{(m)} \mid S_{t^{(m)}-1}^{(m)}, \Lambda) \right] \quad (9) \end{aligned}$$

In the application of image modeling, SL-HMMs can perform an elastic matching in both horizontal and vertical directions by assuming the transition probabilities with left-to-right and top-to-bottom topologies. The training algorithm for the SL-HMMs using the variational EM algorithm are derived in [3].

4. HIDDEN MARKOV EIGENFACE MODELS

Hidden Markov Eigenface Models is defined as an integration of the PPCA/FA models and the SL-HMM. The basic idea of the proposed model is that the eigenfaces are generated from an SL-HMM. Figure 1 shows the graphical representation for HMEMs and the likelihood function is defined as follows:

$$P(\boldsymbol{O} \mid \Lambda) = \sum_{\boldsymbol{S}} \int P(\boldsymbol{O} \mid \boldsymbol{a}, \boldsymbol{S}, \Lambda) P(\boldsymbol{a} \mid \Lambda) P(\boldsymbol{S} \mid \Lambda) d\boldsymbol{a} \quad (10)$$

where $\boldsymbol{a}$ and $\boldsymbol{S}$ correspond to the factor vector of PPCA/FA and the state variables of SL-HMMs, respectively.

$$P(\boldsymbol{a} \mid \Lambda) = \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \quad (11)$$

$$P(\boldsymbol{S} \mid \Lambda) = \prod_{m=1}^{M} P(\boldsymbol{S}^{(m)} \mid \Lambda) \quad (12)$$

The output probabilities given $\boldsymbol{a}$ and $\boldsymbol{S}$ is defined as

$$\begin{aligned} P(\boldsymbol{O} \mid \boldsymbol{a}, \boldsymbol{S}, \Lambda) &= \prod_{\boldsymbol{t}} P(\boldsymbol{O}_{\boldsymbol{t}} \mid \boldsymbol{a}, \boldsymbol{S}_{\boldsymbol{t}}, \Lambda) \quad (13) \\ &= \prod_{\boldsymbol{t}} \mathcal{N}(\boldsymbol{O}_{\boldsymbol{t}} \mid \boldsymbol{W}_{\boldsymbol{S}_{\boldsymbol{t}}} \boldsymbol{a} + \boldsymbol{\mu}_{\boldsymbol{S}_{\boldsymbol{t}}}, \boldsymbol{\Sigma}_{\boldsymbol{S}_{\boldsymbol{t}}}) \quad (14) \end{aligned}$$

where $\boldsymbol{W}_{\boldsymbol{S}_{\boldsymbol{t}}}$ corresponds to the row elements of the loading matrix, and $\boldsymbol{\mu}_{\boldsymbol{S}_{\boldsymbol{t}}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{S}_{\boldsymbol{t}}}$ denote the mean vector and covariance matrix of noise vector, respectively. Incorporating the state transition structure into the loading matrix, eigenfaces can be transformed to match an input image and it performs size and location normalization. Once the state sequences are given, HMEMs are regarded as PPCA/FA models which are applied on the normalized data. Therefore, HMEMs overcome the limitation of SL-HMMs, i.e., the correlation among all observations can be modeled throughout the factor variables as well as the standard PPCA/FA. Thus, HMEMs have the

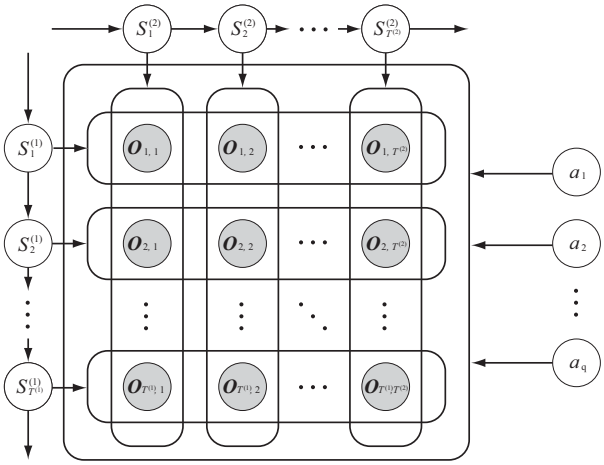

Fig. 1. Graphical representation for HMEMs.

both properties of PPCA/FA and SL-HMMs: a linear feature extraction based on statistical analysis and invariances to size and location of images. Moreover, the structure of HMEMs includes conventional PPCA/FA and SL-HMMs as special cases: HMEMs with zero factor become the conventional SL-HMMs, and HMEMs which have the same number of states as input images are equivalent to the conventional PPCA/FA.

5. TRAINING ALGORITHM FOR HMEMS

The parameters of HMEMs 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 $\mathcal{Q}$-function:

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

By maximizing the $\mathcal{Q}$-function with respect to model parameters Λ, the re-estimation formula in the M-step can be easily derived. However, the calculation of the posterior distribution $P(\boldsymbol{a}, \boldsymbol{S} \,|\, \boldsymbol{O}, \Lambda)$ 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 [6] to the training algorithm of HMEMs.

The variational methods approximate the posterior distribution over the hidden variables by a tractable distribution. Any distribution $Q(\boldsymbol{a}, \boldsymbol{S})$ over the hidden variables defines a lower bound on the log-likelihood:

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

where the Jensen's inequality has been applied. The difference between $\ln P(\boldsymbol{O} \,|\, \Lambda)$ and $\mathcal{F}$ is given by the Kullback-Leibler divergence between $Q(\boldsymbol{a}, \boldsymbol{S})$ and the posterior distribution $P(\boldsymbol{a}, \boldsymbol{S} \,|\, \boldsymbol{O}, \Lambda)$. Therefore maximizing the lower bound $\mathcal{F}$ with respect to $Q(\boldsymbol{a}, \boldsymbol{S})$ is equivalent to minimizing the KL divergence. In order to yield a tractable algorithm, it is necessary to consider a more restricted structure of $Q(\boldsymbol{a}, \boldsymbol{S})$ distributions. Given the structure, the parameters of $Q(\boldsymbol{a}, \boldsymbol{S})$ 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 Λ holding the other parameters fixed:

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

where C is the set of constrained distributions. In the M-step, the re-estimation formula in the standard EM algorithm can be used by calculating the expectations with respect to $Q(\boldsymbol{a}, \boldsymbol{S})$ instead of the true posterior distribution $P(\boldsymbol{a}, \boldsymbol{S} \,|\, \boldsymbol{O}, \Lambda)$. In this procedure, the lower bound $\mathcal{F}$ is guaranteed to increase instead of the value of the $\mathcal{Q}$-function.

The complexity and the approximation property of the variational EM algorithm are dependent on a constraint to the posterior distribution $Q(\boldsymbol{a}, \boldsymbol{S})$. Here we consider a constrained family of variational distributions for HMEMs by assuming that $Q(\boldsymbol{a}, \boldsymbol{S})$ factorizes over $\boldsymbol{a}$ and $\boldsymbol{S}^{(m)}$, so that

$$Q(\boldsymbol{a}, \boldsymbol{S}) = Q(\boldsymbol{a}) \prod_{m=1}^{M} Q(\boldsymbol{S}^{(m)}) \quad (17)$$

where $\int Q(a) d\boldsymbol{a} = 1$ and $\sum_{\boldsymbol{S}^{(m)}} Q(\boldsymbol{S}^{(m)}) = 1$, $m = 1, \ldots, M$. The optimal distributions of the subsets are obtained by maximizing $\mathcal{F}$ independently while keeping the other distributions fixed:

$$\begin{aligned} Q(\boldsymbol{a}) &\propto P(\boldsymbol{a} \,|\, \Lambda) \\ &\quad \times \exp\Big[\sum_{\boldsymbol{S}} \prod_{m} Q(\boldsymbol{S}^{(m)}) \ln P(\boldsymbol{O} \,|\, \boldsymbol{a}, \boldsymbol{S}, \Lambda) \Big] \end{aligned} \quad (18)$$

$$\begin{aligned} Q(\boldsymbol{S}^{(m)}) &\propto P(\boldsymbol{S}^{(m)} \,|\, \Lambda) \\ &\quad \times \exp\Big[\sum_{\boldsymbol{S}/\boldsymbol{S}^{(m)}} \int Q(\boldsymbol{a}) \prod_{n \neq m} Q(\boldsymbol{S}^{(n)}) \ln P(\boldsymbol{O} \,|\, \boldsymbol{a}, \boldsymbol{S}, \Lambda) d\boldsymbol{a} \Big] \end{aligned} \quad (19)$$

The E-step consists of the updates of $Q(\boldsymbol{a})$ and $Q(\boldsymbol{S}^{(m)})$ which interact through the expectations. By inspection, the distribution $Q(\boldsymbol{S}^{(m)})$ has the same structure as the posterior of standard HMMs; the forward-backward algorithm can be used to compute a new set of expectations. The expectations of $Q(\boldsymbol{a})$ can also be computed by the similar equations of the EM algorithm for PPCA/FA.

6. EXPERIMENTS

In order to demonstrate the modeling ability of HMEMs, face recognition experiments on the XM2VTS database [7] were conducted. We prepared eight images of 100 subjects; seven images were used for training and one image for testing. Face images of grayscale 64×64 pixels were extracted from the original images. In this process, two sets of data were prepared:

- "dataset1": the size- and location-normalized data (the original size and location in the database are used).
- "dataset2": 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 point and the range of size from 500×500 to 600×600 with fixed aspect.

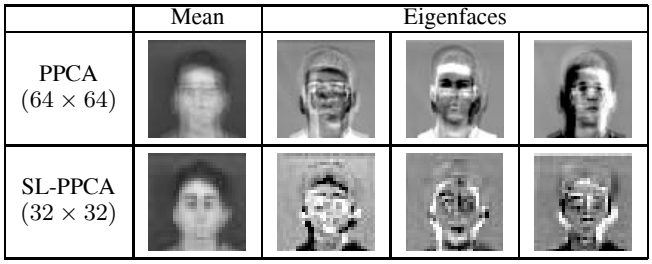

Fig. 2. Visualized mean vectors and eigenfaces on "dataset2."

To compare the recognition performances, the following models were constructed: the conventional PPCA and FA ("PPCA" and "FA"), the SL-HMM with single Gaussian distributions ("SL-HMM"), and the proposed HMEMs with the noise variances of PPCA and FA ("SL-PPCA" and "SL-FA"). The intensity values were used as a feature vector and modeled by SL-HMMs and HMEMs with 32×32 states. "PPCA" and "FA" were trained as "SL-PPCA" and "SL-FA" with 64×64 states, respectively. The number of factors was varied from one to six for PPCA/FA and HMEMs. The initial state transition was constructed by a linear segmentation and the loading matrices were initialized by Gaussian noise with zero mean and unit variance. The transition probabilities for state sequences were assumed to be a left-to-right (top-to-bottom) no skip topology.

Figure 2 shows the visualized mean vectors and eigenfaces obtained by "PPCA" and "SL-PPCA" using "dataset2." As seen in the figure, the mean vector of "PPCA" is blurred and the eigenfaces include variations in size and location. Although the spacial resolution of "SL-PPCA" are lower than "PPCA," the parts of face can be clearly identified in the mean vector of "SL-PPCA." It is also be considered that "SL-PPCA" eliminates the size and location variations from the eigenfaces by the state transitions of SL-HMMs.

Figure 3 and 4 show the recognition rates of "dataset1" and "dataset2," respectively. From the results, it can be seen that "SL-PPCA" and "SL-FA" achieved higher performance than "SL-HMM" in both datasets. This means that the eigenfaces of "SL-PPCA" and "SL-FA" appropriately modeled the correlation among the observations and the efficient features for face image modeling were extracted automatically. Although the recognition rate of "SL-FA" was decreased as increasing the number of factors, this is because the inaccurate estimation of noise variances due to insufficient amount of training data. In "dataset2," it is confirmed that the performances of "PPCA" and "FA" were significantly degraded by the size and location variations. However, the models including the structure of SL-HMMs ("SL-HMM," "SL-PPCA" and "SL-FA") could reduce the influence of the variations due to the ability to normalize the size and location of images.

7. CONCLUSION

We have proposed hidden Markov eigenface models which are defined as an integration of PPCA/FA and SL-HMMs, and derived a feasible training algorithm based on the variational EM algorithm. The proposed model has good properties of the both PPCA/FA and SL-HMMs: a linear feature extraction and invariances to size and location. In face recognition experiments on the XM2VTS database, the proposed model achieved better results than the conventional PPCA/FA and SL-HMMs. Investigation of optimal parameter sharing structures and an improvement of the training algorithm using the deterministic annealing EM algorithm will be future works.

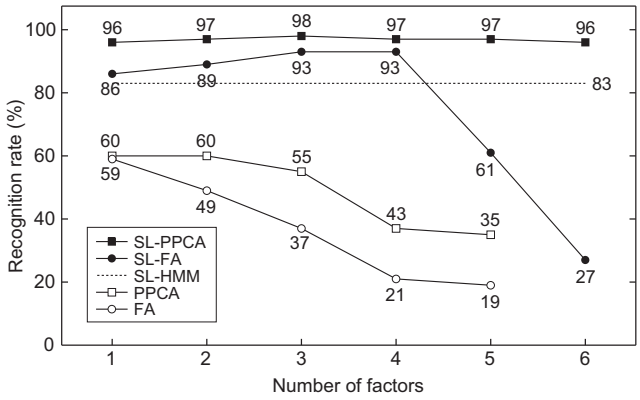

Fig. 3. Recognition rates on "dataset1" (without variations).

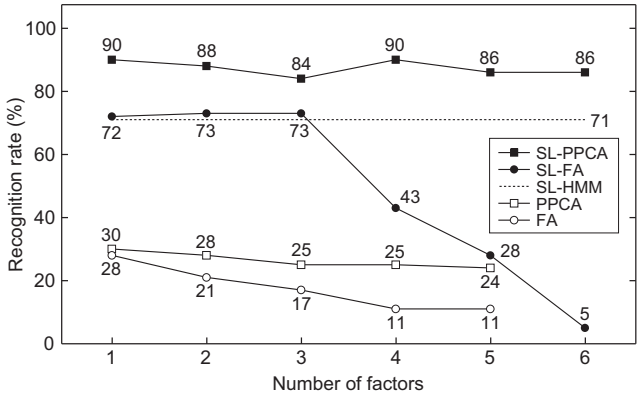

Fig. 4. Recognition rates on "dataset2" (with variations).

8. REFERENCES

[1] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," IEEE Computer Vision and Pattern Recognition, 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] 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.

[4] M. E. Tipping and C. M. Bishop, "Mixtures of Probabilistic Principal Component Analyzers," Neural Computation, 11(2), pp.443–482, 1999.

[5] D. Rubin and D. Thayer, "EM Algorithms for ML Factor Analysis," Psychometrika, vol. 47, pp. 69–76, 1982.

[6] M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, and L.K. Saul, "An introduction to Variational Methods for Graphical Models," Machine Learning, vol.37, pp.183–233, 1999.

[7] K. Messer, J. Mates, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," Audio- and Video-Based Biometric Person Authentication, pp.72–77, 1999.